Wolfgang Härdle

Nikolaus Hautsch

Ludger Overbeck

Editors

Applied Quantitative Finance

Second Edition

Applied Quantitative Finance

Second Edition

Wolfgang K. Härdle • Nikolaus Hautsch
Ludger Overbeck

Editors

Applied Quantitative Finance

Second Edition

Springer

Professor Dr. Wolfgang K. Härdle
Humboldt-Universität zu Berlin
CASE-Center for Applied
Statistics and Economics
Spandauer Straße 1
10178 Berlin
Germany
haerdle@wiwi.hu-berlin.de

Professor Dr. Nikolaus Hautsch
Humboldt-Universität zu Berlin
CASE-Center for Applied
Statistics and Economics
Spandauer Straße 1
10178 Berlin
Germany
nikolaus.hautsch@wiwi.hu-berlin.de

Professor Dr. Ludger Overbeck
Universität Gießen
Mathematical Finance and
Quantitative Risk Management
Arndtstraße 2
35392 Giessen
Germany
Ludger.Overbeck@math.uni-giessen.de

Cover design: WMX Design, Heidelberg

9 8 7 6 5 4 3 2 1

springer.com

Preface to the 2nd Edition

The second edition of this book widens the scope of presented methods and topics. We have introduced new chapters covering ongoing and currently dominant topics like Value-at-Risk, Credit Risk, the pricing of multivariate Bermudan options and Collaterized Debt Obligations, Realized Volatility and High-Frequency Econometrics. Since modern statistical methods, like e.g. copulae estimation, are increasingly important in quantitative finance such as in credit risk management, we have put more weight on the presentation of these themes. Moreover, we included more up-to-date data sets in our examples and applications. These modifications and give the text a higher degree of timeliness and strengthens the applicability. Accordingly, we the structure of the second edition is slightly changed.

The probably most important step towards readability and user friendliness of this book is that we have translated all numerical Quantlets into the R and Matlab language. The algorithms can be downloaded from the publisher's web sites. In the preparation of this 2nd edition, we received helpful input from Ying Chen and Song Song. We would like to thank them.

Wolfgang Karl Härdle, Nikolaus Hautsch and Ludger Overbeck

Berlin and London, August 2008

Preface to the 1st Edition

This book is designed for students and researchers who want to develop professional skill in modern quantitative applications in finance. The Center for Applied Statistics and Economics (CASE) course at Humboldt-Universität zu Berlin that forms the basis for this book is offered to interested students who have had some experience with probability, statistics and software applications but have not had advanced courses in mathematical finance. Although the course assumes only a modest background it moves quickly between different fields of applications and in the end, the reader can expect to have theoretical and computational tools that are deep enough and rich enough to be relied on throughout future professional careers.

The text is readable for the graduate student in financial engineering as well as for the inexperienced newcomer to quantitative finance who wants to get a grip on modern statistical tools in financial data analysis. The experienced reader with a bright knowledge of mathematical finance will probably skip some sections but will hopefully enjoy the various computational tools of the presented techniques. A graduate student might think that some of the econometric techniques are well known. The mathematics of risk management and volatility dynamics will certainly introduce him into the rich realm of quantitative financial data analysis.

The computer inexperienced user of this e-book is softly introduced into the interactive book concept and will certainly enjoy the various practical examples. The e-book is designed as an interactive document: a stream of text and information with various hints and links to additional tools and features. Our e-book design offers also a complete PDF and HTML file with links to world wide computing servers. The reader of this book may therefore without download or purchase of software use all the presented examples and methods via the enclosed license code number with a local XploRe Quantlet Server (XQS). Such XQ Servers may also be installed in a department or addressed freely on the web, click to www.xplore-stat.de and www.quantlet.com.

"Applied Quantitative Finance" consists of four main parts: Value at Risk, Credit Risk, Implied Volatility and Econometrics. In the first part Jaschke and Jiang treat the Approximation of the Value at Risk in conditional Gaussian Models and Rank and Siegl show how the VaR can be calculated using copulas.

The second part starts with an analysis of rating migration probabilities by Höse, Huschens and Wania. Frisch and Knöchlein quantify the risk of yield spread changes via historical simulations. This part is completed by an analysis of the sensitivity of risk measures to changes in the dependency structure between single positions of a portfolio by Kiesel and Kleinow.

The third part is devoted to the analysis of implied volatilities and their dynamics. Fengler, Härdle and Schmidt start with an analysis of the implied volatility surface and show how common PCA can be applied to model the dynamics of the surface. In the next two chapters the authors estimate the risk neutral state price density from observed option prices and the corresponding implied volatilities. While Härdle and Zheng apply implied binomial trees to estimate the SPD, the method by Huynh, Kervella and Zheng is based on a local polynomial estimation of the implied volatility and its derivatives. Blaskowitz and Schmidt use the proposed methods to develop trading strategies based on the comparison of the historical SPD and the one implied by option prices.

Recently developed econometric methods are presented in the last part of the book. Fengler and Herwartz introduce a multivariate volatility model and apply it to exchange rates. Methods used to monitor sequentially observed data are treated by Knoth. Chen, Härdle and Kleinow apply the empirical likelihood concept to develop a test about a parametric diffusion model. Schulz and Werwatz estimate a state space model of Berlin house prices that can be used to construct a time series of the price of a standard house. The influence of long memory effects on financial time series is analyzed by Blaskowitz and Schmidt. Mercurio proposes a methodology to identify time intervals of homogeneity for time series. The pricing of exotic options via a simulation approach is introduced by Lüssem and Schumacher The chapter by Franke, Holzberger and Müller is devoted to a nonparametric estimation approach of GARCH models. The book closes with a chapter of Aydınlı, who introduces a technology to connect standard software with the XploRe server in order to have access to quantlets developed in this book.

W. Härdle, T. Kleinow and G. Stahl

Berlin and Bonn, June 2002

Contents

2 Quantification of Spread Risk by Means of Historical Simulation 37
Christoph Frisch and Germar Knöchlein

3 A Copula-Based Model of the Term Structure of CDO Tranches 69
Umberto Cherubini, Sabrina Mulinacci and Silvia Romagnoli

4 VaR in High Dimensional Systems – a Conditional Correlation Approach 83
Helmut Herwartz and Bruno Pedrinha

II Credit Risk   103

5 Rating Migrations   105

Steffi Höse, Stefan Huschens and Robert Wania

6 Cross- and Autocorrelation in Multi-Period Credit Portfolio Models   125

Christoph K.J. Wagner

III Implied Volatility 191

9 Least Squares Kernel Smoothing of the Implied Volatility Smile 193
Matthias R. Fengler and Qihua Wang

10 Numerics of Implied Binomial Trees 209
Wolfgang Härdle and Alena Myšičková

11 Application of Extended Kalman Filter to SPD Estimation 233
Zdeněk Hlávka and Marek Svojik

12 Stochastic Volatility Estimation Using Markov Chain Simulation 249
Nikolaus Hautsch and Yangguoyi Ou

13 Measuring and Modeling Risk Using High-Frequency Data 275
Wolfgang Härdle, Nikolaus Hautsch and Uta Pigorsch

14 Valuation of Multidimensional Bermudan Options 295
Shih-Feng Huang and Meihui Guo

21 Canonical Dynamics Mechanism of Monetary Policy and Interest Rate   417

Jenher Jeng, Wei-Fang Niu, Nan-Jye Wang, and Shih-Shan Lin

Contributors

Denis Belomestny Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

Umberto Cherubini University of Bologna, Department of Mathematical Economics, Bologna, Italy

Mstislav Elagin Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

Matthias R. Fengler Sal. Oppenheim jr. & Cie. Kommanditgesellschaft auf Aktien, Frankfurt am Main

Christoph Frisch LRP Landesbank Rheinland-Pfalz, Germany

Vasyl Golosnoy Institut für Statistik und Ökonometrie, Christian-Albrechts-Universität Kiel, Germany

Meihui Guo Department of Applied Mathematics, National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C.

Nikolaus Hautsch Humboldt-Universität zu Berlin, CASE, Center for Applied Statistics and Economics, Germany

Wolfgang K. Härdle Humboldt-Universität zu Berlin, CASE, Center for Applied Statistics and Economics, Germany

Helmut Herwartz Institue for Statistics and Econometrics, University Kiel, Germany

Zdeněk Hlávka Univerzita Karlova v Praze, Department of Probability and Mathematical Statistics, Czech Republic

Steffi Höse Technische Universität Dresden, Germany

Shih-Feng Huang Department of Applied Mathematics, National Sun Yat-Sen University, Kaohsiung, Taiwan, R.O.C.

Stefan Huschens Technische Universität Dresden, Germany

Vahidin Jeleskovic Quantitative Products Laboratory, Deutsche Bank AG, Germany

Jenher Jeng   Graduate Institute of Finance and Institute of Mathematical Modelling and Scientific Computing, National Chiao Tung University; G5 Capital management, Taipei, R.O.C.

Ying Jiao   MIF ESILV and CMAP Ecole Polytechnique, France

Nicole El Karoui   CMAP Ecole Polytechnique, France

Torsten Kleinow   Department of Actuarial Mathematics and Statistics, Heriot-Watt University, United kingdom

Germar Knöchlein   LRP Landesbank Rheinland-Pfalz, Germany

David Kurtz   BlueCrest Capital Management Limited, London, United Kingdom

Shih-Shan Lin   G5 Capital management, Taipei, R.O.C.

Grigori N. Milstein   Institute of Physics and Applied Mathematics, Ural State University, Russia

Sabrina Mulinacci   University of Bologna, Department of Mathematical Economics, Bologna, Italy

Alena Myšičková   Humboldt-Universität zu Berlin, CASE, Center for Applied Statistics and Economics, Germany

Wei-Fang Niu   Department of Quantitative Finance, National Tsing Hua University; G5 Capital management, Taipei. R.O.C.

Ostap Okhrin   Humboldt-Universität zu Berlin, CASE, Center for Applied Statistics and Economics, Germany

Yarema Okhrin   Lehrstuhl für Quantitative Methoden, insb. Statistik, Europa-Universität Viadrina, Frankfurt (Oder), Germany

Yangguoyi Ou   Humboldt-Universität zu Berlin, CASE, Center for Applied Statistics and Economics, Germany

Ludger Overbeck   Institut für Mathematik, Universität Gießen, Germany

Bruno Pedrinha   Institue of Economics, University Kiel, Germany

Uta Pigorsch   University of Mannheim, Department of Economics, Germany

Silvia Romagnoli   University of Bologna, Department of Mathematical Economics, Bologna, Italy

Wolfgang Schmid   European University, Department of Statistics, Frankfurt (Oder), Germany

Rainer Schulz University of Aberdeen Business School, Aberdeen, United Kingdom

Maria Sokolova Moscow State University, Russia

Vladimir Spokoiny Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

Markus Staiber Allianz Lebensversicherungs-AG Leipzig, Germany

Marek Svojik Humboldt-Universität zu Berlin, CASE, Center for Applied Statistics and Economics, Germany

Christoph K.J. Wagner Allianz Risk Transfer, Zürich

Nan-Jye Wang G5 Capital management, Taipei, R.O.C.

Qihua Wang The University of Hong Kong, Department of Statistics and Actuarial Science, HK, China

Robert Wania Technische Universität Dresden, Dresden

Martin Wersing Humboldt Universität zu Berlin, Germany

Axel Werwatz Technical Universität Berlin, Germany

Frequently Used Notation

$x \stackrel{\text{def}}{=} \ldots$ x is defined as ...
$\mathbb{R}$ real numbers
$\Re$ real part
$\overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{\infty, \infty\}$
$A^{\top}$ transpose of matrix A
$X \sim D$ the random variable X has distribution D
$\mathsf{E}[X]$ expected value of random variable X
$\mathsf{Var}(X)$ variance of random variable X
$\mathrm{Std}(X)$ standard deviation of random variable X
$\mathrm{Cov}(X, Y)$ covariance of two random variables X and Y
$\mathrm{N}(\mu, \Sigma)$ normal distribution with expectation μ and covariance matrix Σ, a
similar notation is used if Σ is the correlation matrix
cdf denotes the cumulative distribution function
pdf denotes the probability density function
$\stackrel{\mathcal{L}}{\to}$ convergence in distribution
Φ standard normal cummulative distribution function
φ standard normal density function
χ_p^2 chi-squared distribution with p degrees of freedom
t_p t-distribution (Student's) with p degrees of freedom
W_t Wiener process
$\mathrm{vech}(B)$ half-vectorization operator stacking the elements of a $(m \times m)$ matrix
B from the main diagonal downwards in a $m(m+1)/2$ dimensional column
vector
$\mathrm{P}[A]$ or $\mathrm{P}(A)$ probability of a set A
$\mathbf{1}$ indicator function
$(F \circ G)(x) \stackrel{\text{def}}{=} F\{G(x)\}$ for functions F and G
$x \approx y$ x is approximately equal to y
$\alpha_n = \mathcal{O}(\beta_n)$ iff $\frac{\alpha_n}{\beta_n} \longrightarrow$ constant, as $n \longrightarrow \infty$
$\alpha_n = o(\beta_n)$ iff $\frac{\alpha_n}{\beta_n} \longrightarrow 0$, as $n \longrightarrow \infty$
$\mathcal{F}_t$ is the information set generated by all information available at time t
Let A_n and B_n be sequences of random variables.
$A_n = \mathcal{O}_p(B_n)$ iff $\forall \varepsilon > 0 \; \exists M, \; \exists N$ such that $\mathrm{P}[|A_n/B_n| > M] < \varepsilon, \; \forall n > N$.
$A_n = o_p(B_n)$ iff $\forall \varepsilon > 0 \; : \; \lim_{n \to \infty} \mathrm{P}[|A_n/B_n| > \varepsilon] = 0$.

$'\odot'$ matrix multiplication by element

vechl(B) an operator stacking the elements below the diagonal of a symmetric $(m \times m)$ matrix B in a $\{m(m-1)/2\}$ dimensional column vector

$\otimes$ Kronecker product

Part I

Value at Risk

1 Modeling Dependencies with Copulae

Wolfgang Härdle, Ostap Okhrin and Yarema Okhrin

1.1 Introduction

The modeling and estimation of multivariate distributions is one of the most critical issues in financial and economic applications. The distributions are usually restricted to the class of multivariate elliptical distributions. This limits the analysis to a very narrow class of candidate distribution and requires the estimation of a large number of parameters. Two further problems are illustrated in Figure 1.1. The scatter plot in the first figure shows realizations of two Gaussian random variables, the points are symmetric and no extreme outliers can be observed. In contrast, the second picture exhibits numerous outliers. The outliers in the first and third quadrants show that extreme values often occur simultaneously for both variables. Such behavior is observed in crisis periods, when strong negative movements on financial markets occur simultaneously. In the third figure we observe that the dependency between negative values is different compared to positive values. This type of non-symmetric dependency cannot be modeled by elliptical distributions, because they impose a very specific radially symmetric dependency structure. Both types of dependencies are often observed in financial applications. The assumption of Gaussian distribution is therefore rarely consistent with the empirical evidence and possibly leads to incorrect inferences from financial models. Moreover, the correlation coefficient is equal for all three samples, despite clear differences in the dependencies. This questions the suitability of the correlation coefficient as the key measure of dependence for financial data.

The seminal result of Sklar (1959) provides a partial solution to these problems. It allows the separation of marginal distributions from the dependency structure between the random variables. Since the theory on modeling and estimation of univariate distributions is well established compared to the multivariate case, the initial problem reduces to modeling the dependency by copulae. This approach has several important advantages. Firstly, it dramatically widens the class of candidate distribution. Secondly, it allows a simple
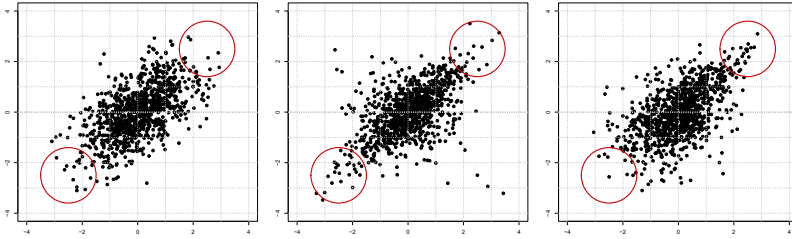
Figure 1.1. Scatter plots of bivariate samples with different
dependency structures

construction of distributions with less parameters than imposed by elliptical
models. Thirdly, the copula-based models reflect the real-world relationships
on financial markets better.

The purpose of this chapter is twofold. Firstly, to provide the theoretical
background, dealing with the estimation, simulation and testing of copula-
based models, and secondly, to discuss several important applications of cop-
ulae to financial problems. The chapter is structured as follows. The next
section provides a review of bivariate copulae. Here we also consider differ-
ent copula families and dependency measures. The third section extends the
discussion to a multivariate framework. The fourth and fifth sections provide
estimation and simulation techniques. The final section illustrates the use
of copulae in financial problems. We omit all proofs and follow the notation
used in Joe (1997).

1.2 Bivariate Copulae

Modeling and measuring the dependency between two random variables using
copulae is the subject of this section. There are several equivalent definitions
of the copula function. We define it as a bivariate distribution function with
both marginal distributions being uniform on $[0, 1]$.

DEFINITION 1.1 *The bivariate copula is a function C: $[0, 1]^2 \rightarrow [0, 1]$
with the following properties:*

1. *For every $u_1, u_2 \in [0, 1]$ $C(u_1, 0) = 0 = C(0, u_2)$.*

2. *For every $u_1, u_2 \in [0, 1]$ $C(u_1, 1) = u_1$ and $C(1, u_2) = u_2$.*

3. *For every (u_1, u_2), $(u'_1, u'_2) \in [0, 1]^2$ such that $u_1 \leq u_2$ and $u'_1 \leq u'_2$*

$$C(u_2, u'_2) - C(u_2, u'_1) - C(u_1, u'_2) + C(u_1, u'_1) \geq 0.$$

Copulae gained their popularity due to a seminal paper by Sklar (1959), where this term was first coined. The separation of the bivariate distribution function into the copula function and margins is formalized in the next theorem (Nelsen (2006), Theorem 2.3.3).

PROPOSITION 1.1 *Let F be a bivariate distribution function with margins F_1 and F_2, then there exists a copula C such that*

$$F(x_1, x_2) = C\{F_1(x_1), F_2(x_2)\}, \quad x_1, x_2 \in \overline{\mathbb{R}}. \tag{1.1}$$

If F_1 and F_2 are continuous then C is unique. Otherwise C is uniquely determined on $F_1(\overline{\mathbb{R}}) \times F_2(\overline{\mathbb{R}})$.

Conversely, if C is a copula and F_1 and F_2 are univariate distribution functions, then function F in (1.1) is a bivariate distribution function with margins F_1 and F_2.

The theorem allows us to depart an arbitrary continuous bivariate distribution into its marginal distributions and the dependency structure. The latter is defined by the copula function.

The representation (1.1) also shows how new bivariate distributions can be constructed. We can extend the class of standard elliptical distributions by keeping the same elliptical copula function and varying the marginal distributions or vice versa. Going further we can take elliptical margins and impose some non-symmetric form of dependency by considering non-elliptical copulae. This shows that copulae substantially widen the family of elliptical distributions. To determine the copula function of a given bivariate distribution we use the transformation

$$C(u_1, u_2) = F\{F_1^{-1}(u_1), F_2^{-1}(u_2)\}, \quad u_1, u_2 \in [0, 1], \tag{1.2}$$

where F_i^{-1}, $i = 1, 2$ are generalized inverses of the marginal distribution functions.

Since the copula function is a bivariate distribution with uniform margins, it follows that the copula density could be determined in the usual way

$$c(u_1, u_2) = \frac{\partial^2 C(u_1, u_2)}{\partial u_1 \partial u_2}, \quad u_1, u_2 \in [0, 1]. \tag{1.3}$$

Being armed with the Theorem 1.1 and (1.3) we could write density function $f(\cdot)$ of the bivariate distribution F in terms of copula as follows

$$f(x_1, x_2) = c\{F_1(x_1), F_2(x_2)\}f_1(x_1)f_2(x_2), \quad x_1, x_2 \in \overline{\mathbb{R}}.$$

A very important property of copulae is given in Theorem 2.4.3 in Nelsen (2006), in it, it is shown that copula is invariant under strictly monotone transformations. This implies that the copulae capture only those features of the joint distribution, which are invariant under increasing transformations.

1.2.1 Copula Families

Naturally there is an infinite number of different copula functions satisfying the assumptions of Definition 1. In this section we discuss in details three important sub-classes of simple, elliptical and Archimedean copulae.

Simplest Copulae

We are often interested in some extreme, special cases, like independence and perfect positive or negative dependence. If two random variables X_1 and X_2 are stochastically independent, from the Theorem 1.1 the structure of such a relationship is given by the product (independence) copula defined as

$$\Pi(u_1, u_2) = u_1 u_2, \quad u_1, u_2 \in [0, 1].$$

The contour diagrams of the bivariate density function with product copula and either Gaussian or t-distributed margins are given in Figure 1.2.

Another two extremes are the lower and upper Fréchet-Hoeffding bounds. They represent the perfect negative and positive dependences respectively

$$W(u_1, u_2) = \max(0, u_1 + u_2 - 1) \text{ and } M(u_1, u_2) = \min(u_1, u_2), \quad u_1, u_2 \in [0, 1].$$

If $C = W$ and $(X_1, X_2) \sim C(F_1, F_2)$ then X_2 is a decreasing function of X_1. Similarly, if $C = M$, then X_2 is an increasing function of X_1. In general we can argue that an arbitrary copula which represents some dependency structure lies between these two bounds, i.e.

$$W(u_1, u_2) \le C(u_1, u_2) \le M(u_1, u_2), \quad u_1, u_2 \in [0, 1].$$

The bounds serve as benchmarks for the evaluation of the dependency magnitude.

Elliptical Family

Due to the popularity of Gaussian and t-distributions in financial applications, the elliptical copulae also play an important role. The construction of

this type of copulae is based directly on the Theorem 1.1 and (1.2). By the Theorem 2.3.7 of Nelsen (2006) bivariate copula is elliptical (has reflection symmetry) if and only if

$$C(u_1, u_2, \theta) = u_1 + u_2 - 1 + C(1 - u_1, 1 - u_2, \theta), \quad u_1, u_2 \in [0, 1].$$

From (1.2) the Gaussian copula and its copula density are given by

$$
\begin{aligned}
C_N(u_1, u_2, \delta) &= \boldsymbol{\Phi}_\delta\{\Phi^{-1}(u_1), \Phi^{-1}(u_2)\}, \\
c_N(u_1, u_2, \delta) &= (1 - \delta^2)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(1 - \delta^2)^{-1}(u_1^2 + u_2^2 - 2\delta u_1 u_2) \right\} \\
&\quad \times \exp\left\{ \frac{1}{2}(u_1^2 + u_2^2) \right\}, \quad \text{for all } u_1, u_2 \in [0, 1], \ \delta \in [-1, 1]
\end{aligned}
$$

where Φ is the distribution function of $N(0, 1)$, Φ^{-1} is the functional inverse of Φ and $\boldsymbol{\Phi}_\delta$ denotes the bivariate standard normal distribution function with the correlation coefficient δ. The level plots of the respective density are given in Figure 1.2. The t-distributed margins lead to more mass and variability in the tails of the distribution. However, the curves are symmetric, which reflects the ellipticity of the underlying copula.

In the bivariate case the t-copula and its density are given by

$$
\begin{aligned}
C_t(u_1, u_2, \nu, \delta) &= \int_{-\infty}^{t_\nu^{-1}(u_1)} \int_{-\infty}^{t_\nu^{-1}(u_2)} \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\pi\nu\sqrt{(1 - \delta^2)}} \\
&\quad \times \left(1 + \frac{x_1^2 - 2\delta x_1 x_2 + x_2^2}{(1 - \delta^2)\nu} \right)^{-\frac{\nu}{2} - 1} dx_1 dx_2, \\
c_t(u_1, u_2, \nu, \delta) &= \frac{f_{\nu\delta}\{t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)\}}{f_\nu\{t^{-1}(u_1)\}f_\nu\{t^{-1}(u_2)\}}, \quad u_1, u_2, \delta \in [0, 1],
\end{aligned}
$$

where δ denotes the correlation coefficient, ν is the number of degrees of freedom. $f_{\nu\delta}$ and f_ν are joint and marginal t-distributions respectively, while t_ν^{-1} denotes the quantile function of the t_ν distribution. In-depth analysis of the t-copula is done in Demarta and McNeil (2004).

Using (1.2) we can derive the copula function for an arbitrary elliptical distribution. The problem is, however, that such copulae depend on the inverse distribution functions and these are rarely available in an explicit form. Therefore, the next class of copulae and their generalizations provide an important flexible and rich family of alternatives to the elliptical copulae.

Archimedean Family

Opposite to elliptical copulae, the Archimedean copulae are not constructed using (1.2), but are related to Laplace transforms of bivariate distribution

functions (see Section 1.6.2). Let $\mathbb{L}$ denote the class of Laplace transforms which consists of strictly decreasing differentiable functions Joe (1997), i.e.

$$\mathbb{L} = \{\phi : [0; \infty) \to [0, 1] \,|\, \phi(0) = 1, \ \phi(\infty) = 0; \ (-1)^j \phi^{(j)} \geq 0; \ j = 1, \ldots, \infty\}.$$

The function $C : [0, 1]^2 \to [0, 1]$ defined as

$$C(u_1, u_2) = \phi\{\phi^{-1}(u_1) + \phi^{-1}(u_2)\}, \quad u_1, u_2 \in [0, 1]$$

is a 2-dimensional Archimedean copula, where $\phi \in \mathbb{L}$ and is called the generator of the copula. It is straightforward to show that $C(u_1, u_2)$ satisfies the conditions of Definition 1. The generator usually depends on some parameters, however, generators with a single parameter θ are mainly considered. Joe (1997) and Nelsen (2006) provide a thoroughly classified list of popular generators for Archimedean copulae and discuss their properties. The most useful in financial applications (see Patton (2004)) appears to be the Gumbel copula with the generator function

$$\phi(x, \theta) = \exp\left(-x^{1/\theta}\right), \quad 1 \leq \theta < \infty, \ x \in [0, \infty].$$

It leads to the copula function

$$\begin{aligned} C(u_1, u_2, \theta) &= \exp\left[-\left\{(-\log u_1)^\theta + (-\log u_2)^\theta\right\}^{1/\theta}\right], \\ & \quad 1 \leq \theta < \infty, \ u_1, u_2 \in [0, 1]. \end{aligned}$$

Consider a bivariate distribution based on the Gumbel copula with univariate extreme valued marginal distributions. Genest and Rivest (1989) show that this distribution is the only bivariate extreme value distribution based on an Archimedean copula. Moreover, all distributions based on Archimedean copulae belong to its domain of attraction under common regularity conditions.

In contrary to the elliptical copulae, the Gumbel copula leads to asymmetric contour diagrams in Figure 1.2. The Gumbel copula shows stronger linkage between positive values, however, more variability and more mass in the negative tail. The opposite is observed for the Clayton copula with the generator and copula functions

$$\begin{aligned} \phi(x, \theta) &= (\theta x + 1)^{-\frac{1}{\theta}}, \quad 1 \leq \theta < \infty, \ \theta \neq 0, \ x \in [0, \infty], \\ C(u_1, u_2, \theta) &= (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}, \quad 1 \leq \theta < \infty, \ \theta \neq 0, \ u_1, u_2 \in [0, 1]. \end{aligned}$$

Another popular copula generator is the Frank generator given by

$$\phi(x, \theta) = \theta^{-1} \log\{1 - (1 - e^{-\theta})e^{-x}\}, \quad 0 \leq \theta < \infty, \ x \in [0, \infty].$$

The respective Frank copula is the only elliptical Archimedean copula, with the copula function

$$C(u_1, u_2, \theta) = -\theta^{-1} \log \left\{ \frac{1 - e^{-\theta} - (1 - e^{-\theta u_1})(1 - e^{-\theta u_2})}{1 - e^{-\theta}} \right\},$$
$$0 \leq \theta < \infty, \ u_1, u_2 \in [0, 1].$$

1.2.2 Dependence Measures

Since copulae define the dependency structure between random variables, there is a relationship between the copulae and different dependency measures. The classical measures for continuous random variables are Kendall's τ and Spearman's ρ. Similarly as copula functions, these measures are invariant under strictly increasing transformations. They are equal to 1 or -1 under perfect positive or negative dependence respectively. In contrast to τ and ρ, the Pearson correlation coefficient measures the linear dependence and, therefore, is unsuitable for measuring nonlinear relationships. Next we discuss the relationship between τ, ρ and the underlying copula function.

DEFINITION 1.2 *Let F be a continuous bivariate cumulative distribution function with the copula C. Moreover, let $(X_1, X_2) \sim F$ and $(X_1', X_2') \sim F$ be independent random pairs. Then Kendall's τ_2 is given by*

$$\tau_2 = \mathrm{P}\{(X_1 - X_1')(X_2 - X_2') > 0\} - \mathrm{P}\{(X_1 - X_1')(X_2 - X_2') < 0\}$$
$$= 2\mathrm{P}\{(X_1 - X_1')(X_2 - X_2') > 0\} - 1 = 4 \iint\limits_{[0,1]^2} C(u_1, u_2) \, dC(u_1, u_2) - 1.$$

Kendall's τ represents the difference between the probability of two random concordant pairs and the probability of two random discordant pairs.

For most copula functions with a single parameter θ there is a one-to-one relationship between θ and the Kendall's τ_2. For example, it holds that

$$\tau_2(\text{Gaussian and } t) = \frac{2}{\pi} \arcsin \delta,$$
$$\tau_2(\text{Archimedean}) = 4 \int_0^1 \frac{\phi^{-1}(t)}{(\phi^{-1}(t))'} \, dt + 1, \quad (\text{Genest and MacKay (1986)}),$$
$$\tau_2(\Pi) = 0, \quad \tau_2(W) = 1, \quad \tau_2(M) = -1.$$

This implies, that for Gaussian, t and an arbitrary Archimedean copula we can estimate the unknown copula parameter θ using a type of method of

Gaussian margins t **margins**
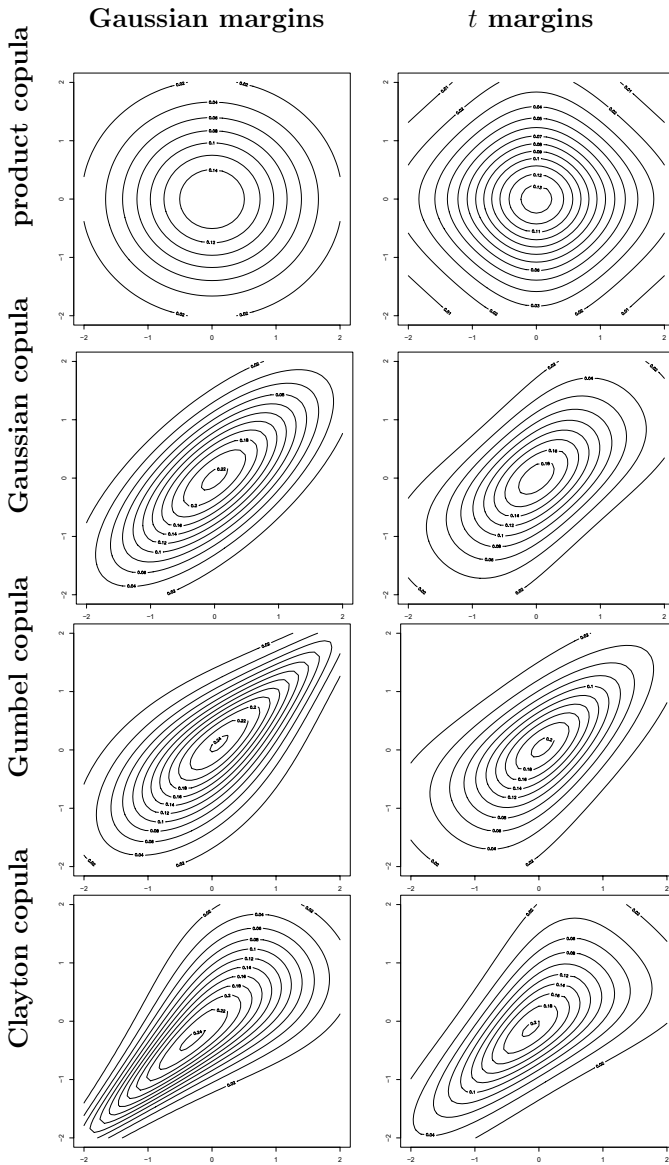
Figure 1.2. Contour diagrams for product, Gaussian, Gumbel and Clayton copulae with Gaussian (left column) and t_3 distributed (right column) margins.

moments procedure with a single moment condition. This requires, however, an estimator of τ_2. Naturally (Kendall (1970)) it is computed by

$$\hat{\tau}_{2n} = \frac{4}{n(n-1)} P_n - 1,$$

where n stands for the sample size and P_n denotes the number of concordant pairs, e.g. such pairs (X_1, X_2) and (X_1', X_2') that $(X_1 - X_1')(X_2 - X_2') > 0$. However, as argued by Genest, Ghoudi and Rivest (1995) the MM estimator of copula parameters is highly inefficient (see Section 1.4). Next we provide the definition and similar results for the Spearman's ρ.

DEFINITION 1.3 *Let F be a continuous bivariate distribution function with the copula C and the univariate margins F_1 and F_2 respectively. Assume that $(X_1, X_2) \sim F$. Then the Spearman's ρ is given by*

$$\rho_2 = 12 \iint\limits_{\mathbb{R}^2} F_1(x_1) F_2(x_2) \, dF(x_1, x_2) - 3 = 12 \iint\limits_{[0,1]^2} u_1 u_2 \, dC(u_1, u_2) - 3.$$

Similarly as for Kendall's τ, we provide the relationship between Spearman's ρ and copulae.

$$\rho_2(\text{Gaussian and } t) = \frac{6}{\pi} \arcsin \frac{\delta}{2},$$
$$\rho_2(\Pi) = 0, \qquad \rho_2(W) = 1, \qquad \rho_2(M) = -1.$$

Unfortunately there is no explicit representation of Spearman's ρ_2 for Archimedean in terms of generator functions as by Kendall's τ. The estimator of ρ is easily computed using

$$\hat{\rho}_{2n} = \frac{12}{n(n+1)(n-1)} \sum_{i=1}^{n} R_i S_i - 3\frac{n+1}{n-1},$$

where R_i and S_i denote the ranks of two samples. For a detailed discussion and relationship between these two measures we refer to Fredricks and Nelsen (2004), Chen (2004), etc.

1.3 Multivariate Copulae

In this section we generalize the above theory to the multivariate case. First we define the copula function and state Sklar's theorem.

DEFINITION 1.4 *A d-dimensional copula is a function C: $[0,1]^d \to [0,1]$ with the following properties:*

1. $C(u_1, \ldots, u_d)$ *is increasing in each component $u_i \in [0,1]$, $i = 1, \ldots, d$.*
2. $C(1, \ldots, 1, u_i, 1, \ldots, 1) = u_i$ *for all $u_i \in [0,1]$, $i = 1, \ldots, d$.*
3. *For all $(u_1, \ldots, u_d)$, $(u'_1, \ldots, u'_d) \in [0,1]^d$ with $u_i < u'_i$ we have*

$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1 + \cdots + i_d} C(v_{j1}, \ldots, v_{jd}) \geq 0,$$

where $v_{j1} = u_j$ and $v_{j2} = u'_j$, for all $j = 1, \ldots, d$.

Thus a d-dimensional copula is the distribution function on $[0,1]^d$ where all marginal distributions are uniform on [0,1]. In the Sklar's theorem the very importance of copulae in the area of multivariate distributions has been recapitulated in an exquisite way.

PROPOSITION 1.2 (Sklar (1959)) *Let F be a multivariate distribution function with margins $F_1, \ldots, F_d$, then there exists the copula C such that*

$$F(x_1, \ldots, x_d) = C\{F_1(x_1), \ldots, F_d(x_d)\}, \quad x_1, \ldots, x_d \in \overline{\mathbb{R}}.$$

If F_i are continuous for $i = 1, \ldots, d$ then C is unique. Otherwise C is uniquely determined on $F_1(\overline{\mathbb{R}}) \times \cdots \times F_d(\overline{\mathbb{R}})$.

Conversely, if C is a copula and $F_1, \ldots, F_d$ are univariate distribution functions, then function F defined above is a multivariate distribution function with margins $F_1, \ldots, F_d$.

The representation in Sklar's Theorem can be used to construct new multivariate distributions by changing either the copula function or the marginal distributions. For an arbitrary continuous multivariate distribution we can determine its copula from the transformation

$$C(u_1, \ldots, u_d) = F\{F_1^{-1}(u_1), \ldots, F_d^{-1}(u_d)\}, \quad u_1, \ldots, u_d \in [0,1], \qquad (1.4)$$

where F_i^{-1} are inverse marginal distribution functions. Copula density and density of the multivariate distribution with respect to copula are

$$c(u_1, \ldots, u_d) = \frac{\partial^d C(u_1, \ldots, u_d)}{\partial u_1 \ldots \partial u_d}, \quad u_1, \ldots, u_d \in [0,1],$$

$$f(x_1, \ldots, x_d) = c\{F_1(x_1), \ldots, F_d(x_d)\} \prod_{i=1}^{d} f_i(x_i), \quad x_1, \ldots, x_d \in \overline{\mathbb{R}}.$$

For the multivariate case as well as for the bivariate case copula function is invariant under monotone transformations.

1.3.1 Copula Families

It is straightforward to generalize the independence copula and the upper and lower Fréchet-Hoeffdings bounds to the multivariate case. The independence copula is defined by the product

$$\Pi(u_1, \ldots, u_d) = \prod_{i=1}^{d} u_i.$$

The upper and lower Fréchet-Hoeffdings bounds are given by

$$W(u_1, \ldots, u_d) = \max\Big(0, \sum_{i=1}^{d} u_i + 1 - d\Big),$$
$$M(u_1, \ldots, u_d) = \min(u_1, \ldots, u_d), \quad u_1, \ldots, u_d \in [0, 1].$$

respectively. An arbitrary copula $C(u_1, \ldots, u_d)$ lies between the upper and lower Fréchet-Hoeffdings bounds

$$W(u_1, \ldots, u_d) \leq C(u_1, \ldots, u_d) \leq M(u_1, \ldots, u_d).$$

Note, however, that the lower Fréchet-Hoeffding bound is not a copula function for $d > 2$.

The generalization of elliptical copulae to $d > 2$ is straightforward. In the Gaussian case we have:

$$C_N(u_1, \ldots, u_d, \boldsymbol{\Sigma}) = \boldsymbol{\Phi_\Sigma}\{\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)\},$$
$$c_N(u_1, \ldots, u_d, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} \times$$
$$\exp\left\{-\frac{[\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_d)](\boldsymbol{\Sigma}^{-1} - \boldsymbol{I})[\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_k)]}{2}\right\}^{\top},$$
$$\text{for all } u_1, \ldots, u_d \in [0, 1],$$

where $\boldsymbol{\Phi_\Sigma}$ is a d-dimensional normal distribution with zero mean and the correlation matrix $\boldsymbol{\Sigma}$. The variances of the variables are imposed by the marginal distributions. Note, that in the multivariate case the implementation of elliptical copulae is very involved due to technical difficulties with multivariate cdf's.

Archimedean and Hierarchical Archimedean copulae

In contrast to the bivariate case, the multivariate setting allows construction methods for copulae. The simplest multivariate generalization of the

Archimedean copulae is $C : [0,1]^d \rightarrow [0,1]$ and is defined as

$$C(u_1, \ldots, u_d) = \phi\{\phi^{-1}(u_1) + \cdots + \phi^{-1}(u_d)\}, \quad u_1, \ldots, u_d \in [0,1], \quad (1.5)$$

where $\phi \in \mathbb{L}$. This definition provides a simple, but rather limited technique for the construction of multivariate copulae. The whole complex multivariate dependency structure is determined by a single copula parameter. Furthermore, the multivariate Archimedean copulae imply that the variables are exchangeable. This means, that the distribution of $(u_1, \ldots, u_d)$ is the same as of $(u_{j_1}, \ldots, u_{j_d})$ for all $j_\ell \neq j_v$. This is certainly not an acceptable assumption in practical applications.

A much more flexible method is provided by hierarchical Archimedean copulae (HAC), discussed by Joe (1997), Whelan (2004), Savu and Trede (2006), Embrechts, Lindskog and McNeil (2003), Okhrin, Okhrin and Schmid (2007). In the most general case of fully nested copulae, the copula function is given by

$$\begin{aligned} C(u_1, \ldots, u_d) &= \phi_{d-1}\{\phi_{d-1}^{-1} \circ \phi_{d-2}\big(\ldots [\phi_2^{-1} \circ \phi_1\{\phi_1^{-1}(u_1) + \phi_1^{-1}(u_2)\} \quad (1.6) \\ &+ \phi_2^{-1}(u_3)] + \cdots + \phi_{d-2}^{-1}(u_{d-1})\big) + \phi_{d-1}^{-1}(u_d)\} \\ &= \phi_{d-1}[\phi_{d-1}^{-1} \circ C(\{\phi_1, \ldots, \phi_{d-2}\})(u_1, \ldots, u_{d-1}) + \phi_{d-1}^{-1}(u_d)] \end{aligned}$$

for $\phi_{d-i}^{-1} \circ \phi_{d-j} \in \mathbb{L}^*$, $i < j$, where

$$\begin{aligned} \mathbb{L}^* = \{\omega : [0;\infty) &\rightarrow [0,\infty) \,|\, \omega(0) = 0, \\ \omega(\infty) &= \infty; \, (-1)^{j-1}\omega^{(j)} \geq 0; \, j = 1, \ldots, \infty\}, \end{aligned}$$

and "$\circ$" is the composition operator. In contrast to the usual Archimedean copula (1.5), the HAC defines the whole dependency structure in a recursive way. At the lowest level, the dependency between the first two variables is modeled by a copula function with the generator ϕ_1, i.e. $z_1 = C(u_1, u_2) = \phi_1\{\phi_1^{-1}(u_1) + \phi_1^{-1}(u_2)\}$. At the second level another copula function is used to model the dependency between z_1 and u_3, etc. Note, that the generators ϕ_i can come from the same family and differ only through the parameter or, to introduce more flexibility, come from different generator families. As an alternative to the fully nested model, we can consider copula functions, with arbitrary chosen combinations at each copula level, so-called partially nested copulae. For example the following 4-dimensional copula, where the first and the last two variables are joined by individual copulae with generators ϕ_{12} and ϕ_{34}. Further, the resulted copulae are combined by a copula with the generator ϕ.

$$\begin{aligned} C(u_1, u_2, u_3, u_4) &= \phi\big(\phi^{-1}[\phi_{12}\{\phi_{12}^{-1}(u_1) + \phi_{12}^{-1}(u_2)\}] \quad (1.7) \\ &+ \phi^{-1}[\phi_{34}\{\phi_{34}^{-1}(u_3) + \phi_{34}^{-1}(u_4)\}]\big). \end{aligned}$$

Whelan (2004) and McNeil (2007) provide tools for generating samples from Archimedean copulae, Savu and Trede (2006) derived the density of such copulae and Joe (1997) proves their positive quadrant dependence (see Theorem 4.4). Okhrin et al. (2007) considered methods for determining the optimal structure of the HAC and provided asymptotic theory for the estimated parameters.

1.3.2 Dependence Measures

MMeasuring dependence in a multivariate framework is a tedious task. This is due to the fact that, the generalizations of bivariate measures are not unique. One of the multivariate extensions of the Kendall's τ and its estimator is proposed in Barbe, Genest, Ghoudi and Rémillard (1996)

$$\tau_d = \frac{2^d}{2^{d-1}-1}\, \mathsf{E}(V) - 1 = \frac{2^d}{2^{d-1}-1}\int t\, dK(t) - 1, \tag{1.8}$$

$$\hat{\tau}_{dn} = \frac{2^d}{2^{d-1}-1}\cdot\frac{1}{n}\sum_{i=1}^{n} V_{in} - 1 = \frac{2^d}{2^{d-1}-1}\int t\, dK_n(t) - 1, \tag{1.9}$$

where $V_{in} = \frac{1}{n-1}\sum_{m=1}^{n}\prod_{j=1}^{d}\mathbf{1}(x_{jm}\le x_{im})$ and $V = C\{F_1(X_1),\ldots,F_d(X_d)\}\in$ $[0,1]$. $K_n(t)$ and $K(t)$ are distribution functions of V_{in} and V respectively. The expression in (1.8) implies that τ_d is an affine transformation of the expectation of the value of the copula. Genest and Rivest (1993) and Barbe et al. (1996) provide in-depth investigation and derivation of the distribution K.

A multivariate extension of Spearman's ρ based on multivariate copula was introduced in Wolff (1980):

$$\rho_d = \frac{d+1}{2^d-(d+1)}\left\{2^d\int\cdots\int_{[0,1]^d} C(u_1,\ldots,u_d)\, du_1\ldots du_d - 1\right\}.$$

Schmid and Schmidt (2006a) and Schmid and Schmidt (2006b) discuss its properties and provide a detailed analysis of its estimator given by

$$\hat{\rho}_{dn} = \frac{d+1}{2^d-d-1}\left[\frac{2^d}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\{1-\hat{F}(x_{ij})\} - 1\right].$$

A version of the pairwise Spearman's ρ was introduced in Kendall (1970)

$$\rho_r = 2^2\sum_{m<l}\binom{d}{2}^{-1}\iint_{[0,1]^2} C_{ml}(u,v)\, du\, dv - 1,$$

where C_{ml} denotes the bivariate copula of the variables m and l.

Generalizations

There are numerous techniques which allow for the construction of new types of copulae from simple, elliptical or Archimedean copulae. For example, copula families B11 and B12 (Joe (1997)) arise as a combination of the upper Fréchet-Hoeffding bound and the product copula

$$
\begin{aligned}
C_{B11}(u_1, u_2, \theta) &= \theta M(u_1, u_3) + (1 - \theta)\Pi(u_1, u_2) \\
&= \theta \min\{u_1, u_2\} + (1 - \theta)u_1 u_2, \\
C_{B12}(u_1, u_2, \theta) &= M^\theta(u_1, u_2)\Pi^{1-\theta}(u_1, u_2) \\
&= (\min\{u_1, u_2\})^\theta (u_1 u_2)^{1-\theta}, \quad u_1, u_2, \theta \in [0, 1].
\end{aligned}
$$

For the family B11 we used the property, that every convex combination of copulae is a copula too. Family B12 is also known as Spearman or Cuadras-Augé copula, which is a weighted geometric mean of the upper Fréchet-Hoeffding bound and the product copula. Further generalization is done by using power mean over the upper Fréchet-Hoeffding bound and the product copula

$$
\begin{aligned}
C_p(u_1, u_2, \theta_1, \theta_2) &= \{\theta_1 M^{\theta_2}(u_1, u_2) + (1 - \theta_1)\Pi^{\theta_2}(u_1, u_2)\}^{1/\theta_2} \\
&= \{\theta_1 \min(u_1, u_2)^{\theta_2} + (1 - \theta_1)(u_1 u_2)^{\theta_2}\}^{1/\theta_2}, \\
&\quad \theta_1 \in [0, 1], \theta_2 \in \mathbb{R}.
\end{aligned}
$$

Nelsen (2006), Chapter 3 provides further methods of constructing multivariate copulae, one of them is based on the Archimedean n-copulae. This family of copulae arises from simple multivariate Archimedean copula from reparametrization $\lambda = e^{-\phi t}$. We get

$$
C(u_1, \ldots, u_d) = \lambda^{-1}\{\lambda(u_1) \ldots \lambda(u_d)\} = \lambda^{-1}[\Pi\{\lambda(u_1), \ldots, \lambda(u_d)\}].
$$

The function λ is known as a *multiplicative generator* of C. Replacing product copula Π with an arbitrary copula C_1 of dimension d we get a new copula family, investigated in Morillas (2005).

Another popular approach to modeling multivariate distributions is based on *vines*. This class was introduced in Joe (1996) and then discussed by Bedford and Cooke (2001), Bedford and Cooke (2002), Kurowicka and Cooke (2006), Aas, Czado, Frignessi and Bakken (2006) and Berg and Aas (2007). The idea is based on the decomposition of a multivariate density into $d(d-1)/2$ bivariate densities. In the literature we have only come accross two types of

such structures D-vines and canonical vines. For the D-vine the density is

$$f(x_1, \ldots, x_d) =$$

$$= \prod_{m=1}^{d} f(x_m) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{ji}\{F(x_i|x_{i+1}, \ldots, x_{i+j-1}), F(x_{i+j}|x_{i+1}, \ldots, x_{i+j-1})\},$$

where the conditional distribution is computed as a derivative with respect to known arguments (in details in Sections 2.5 and 2.6.1). To get the copula function we integrate the density over the d-dimensional hyper cube. As noted by Berg and Aas (2007) there are only $d(d-1)/2$ possible copulae to be described using vines. For $d = 10$ it is only 45 different models, while using HAC as in Okhrin et al. (2007) more than 300 million copulae are available. However the estimation of the parameters of the model and simulation from the copula are faster when vines are used.

1.4 Estimation Methods

The estimation of a copula-based multivariate distribution involves both the estimation of the copula parameters $\boldsymbol{\theta}$ and the estimation of the margins F_j, $j = 1, \ldots, d$. The properties and quality of the estimator of $\boldsymbol{\theta}$ heavily depend on the estimators of F_j, $j = 1, \ldots, d$. We distinguish between a parametric and a nonparametric specification of the margins. If we are interested only in the dependency structure, the estimator of $\boldsymbol{\theta}$ should be independent of any parametric models for the margins. In practical applications, however, we are interested in a complete distribution model and, therefore, parametric models for margins are preferred (see Joe (1997)).

In the bivariate case a standard method of estimating the univariate parameter θ is based on Kendall's τ statistic by Genest and Rivest (1993). The estimator of τ complemented by the method of moments allows for the estimation of the parameters. However, as shown in Genest et al. (1995) the maximum-likelihood method lead to substantially more efficient and general estimators. For non-parametrically estimated margins, Genest et al. (1995) show the consistency and asymptotic normality of ML estimators and derive the moments of the asymptotic distribution. The maximum-likelihood estimation can be performed simultaneously for the parameters of the margins and of the copula function. Alternatively, a two-stage procedure can be applied, where we estimate the parameters of margins at the first stage and the copula parameters at the second stage (see Joe (1997), Joe (2005)). Fermanian and Scaillet (2003), Chen, Fan and Patton (2004) and Chen, Fan and Tsyrennikov (2006) analyze the case of nonparametrically estimated margins.

Chen and Huang (2007) considered a fully nonparametric estimation of the copula. Next we provide details on both approaches.

Parametric margins

Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1^\top, \ldots, \boldsymbol{\alpha}_d^\top)^\top$ denote the vector of parameters of marginal distributions and $\boldsymbol{\theta}$ parameters of the copula. The classical full ML estimator $\hat{\boldsymbol{\eta}}$ of $\boldsymbol{\eta} = (\boldsymbol{\alpha}^\top, \boldsymbol{\theta}^\top)^\top$ solves the system

$$\frac{\partial \mathcal{L}(\boldsymbol{\eta}, \mathbf{X})}{\partial \boldsymbol{\eta}^\top} = \mathbf{0},$$

$$\text{where } \mathcal{L}(\boldsymbol{\eta}, \mathbf{X}) = \sum_{i=1}^n \log\left[c\{F_1(x_{1i}, \boldsymbol{\alpha}_1), \ldots, F_d(x_{di}, \boldsymbol{\alpha}_d), \boldsymbol{\theta}\} \prod_{j=1}^d f_j(x_{ji}, \boldsymbol{\alpha}_j) \right]$$

$$= \sum_{i=1}^n \Big[\log c\{F_1(x_{1i}, \boldsymbol{\alpha}_1), \ldots, F_d(x_{di}, \boldsymbol{\alpha}_d), \boldsymbol{\theta}\}$$

$$+ \sum_{j=1}^d \log f_j(x_{ji}, \boldsymbol{\alpha}_j) \Big].$$

Following the standard theory on ML estimation, the estimator is efficient and asymptotically normal, however, it is often computationally demanding to solve the system simultaneously. Alternatively the multistage optimization proposed in Joe (1997), Chapter 10 also known as *inference of margins*, can be applied. First, we estimate separately the parameters of the margins and then use them in the estimation of the copula parameters as known quantities. The above optimization problem is then replaced by

$$\left(\frac{\partial \mathcal{L}_1}{\partial \boldsymbol{\alpha}_1^\top}, \ldots, \frac{\partial \mathcal{L}_d}{\partial \boldsymbol{\alpha}_d^\top}, \frac{\partial \mathcal{L}_{d+1}}{\partial \boldsymbol{\theta}^\top} \right)^\top = \mathbf{0}, \tag{1.10}$$

$$\text{where } \quad \mathcal{L}_j = \sum_{i=1}^n l_j(\mathbf{X}_i), \text{ for } j = 1, \ldots, d+1,$$

$$l_j(\mathbf{X}_i) = \log f_j(x_{ji}, \boldsymbol{\alpha}_j), \text{ for } j = 1, \ldots, d, i = 1, \ldots, n,$$

$$l_{d+1}(\mathbf{X}_i) = \log\left[c\{F_1(x_{1i}, \boldsymbol{\alpha}_1), \ldots, F_d(x_{di}, \boldsymbol{\alpha}_d)\} \right], \text{ for } i = 1, \ldots, n.$$

The first d components in (1.10) correspond to the usual ML estimation of the parameters of the marginal distributions. The last component reflects the estimation of the copula parameters. Detailed discussion on this method could be found in Joe and Xu (1996). Note, that this procedure does not lead to efficient estimators; however, as argued by Joe (1997) the loss in the

efficiency is modest. The advantage of the two-stage procedure lies in the dramatic reduction of the numerical complexity. This is especially pronounced in the case of hierarchical Archimedean copulae (see Okhrin et al. (2007)). This method is a special case of the generalized method of moments with an identity weighting matrix (see Cherubini, Luciano and Vecchiato (2004), Section 4.5).

Canonical Maximum Likelihood

In this section we consider a nonparametric estimation of the marginal distributions. The asymptotic properties of the multistage estimators of $\boldsymbol{\theta}$ do not depend explicitly on the type of the nonparametric estimator, but on its convergence properties. Here we use the rectangular kernel (histogram). The estimator is given by

$$\hat{F}_j(x) = \frac{1}{n+1} \sum_{i=1}^{n} \mathbf{1}(x_{ji} \leq x), \quad j = 1, \ldots, d.$$

The factor $n/(n+1)$ is used to bound the cdf from one. Let $\hat{F}_1, \ldots, \hat{F}_d$ denote the nonparametric estimators of $F_1, \ldots, F_d$. The canonical ML estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ solves the system by maximizing the pseudo log-likelihood with estimated margins $\hat{F}_1, \ldots, \hat{F}_d$, i.e.

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^\top} = \mathbf{0},$$

$$\text{where} \quad \mathcal{L} = \sum_{i=1}^{n} l(\mathbf{X}_i),$$

$$l(\mathbf{X}_i) = log\left[c\{\hat{F}_1(x_{1i}), \ldots, \hat{F}_d(x_{di})\}\right], \text{ for } i = 1, \ldots, n.$$

As in the parametric case, the semiparametric estimator $\hat{\boldsymbol{\theta}}$ is asymptotically normal under suitable regularity conditions. This method was first used in Oakes (2005) and then investigated by Genest et al. (1995) and Shih and Louis (1995). For properties we refer to these papers.

1.5 Goodness-of-Fit Tests for Copulae

In this section we review the goodness-of-fit (GOF) tests for copulae. With the GOF tests we test whether the underlying copula is equal to some target

copula function or belongs to some copula family. The test problem could be written as a composite or a simple null hypothesis

$$H_0: \quad C \in \mathcal{C}_0, \quad \text{against} \quad H_1: \quad C \notin \mathcal{C}_0,$$

$$H_0: \quad C = C_0, \quad \text{against} \quad H_1: \quad C \neq C_0,$$

where $\mathcal{C}_0$ is some known parametric family of copulae, C_0 is some known target copula and C is the underlying true copula. The test problem is in general equivalent to the GOF tests for multivariate distributions. However, since the margins are estimated we cannot apply the standard test procedures directly.

Several related tests have been introduced into the literature. As a simple generalization of the standard χ^2 an adopted χ^2-test is proposed in Fermanian (2005) (Section 2) which is based directly on the distance between C and C_0. Genest and Rivest (1993) consider, in a bivariate setup, a test based on the true and empirical distributions of the pseudo-variable $Z = C_0(X, Y)$. As a measure they use the L_2 norm. This approach is extended to the multivariate case and other measures of proximity by Barbe et al. (1996), Wang and Wells (2000), Genest, Quessy and Rémillard (2006). Wang and Wells (2000) propose to compute a Crámer-von-Mises statistic of the form

$$S_{n\xi} = \int_{\xi}^{1} \{K_n(w) - K(w)\}^2 \, dw, \quad \xi \in (0, 1),$$

where $K_n(w)$ and $K(w)$ are empirical and theoretical K-distributions from Section 2.3.2. However exact p-values for this statistic cannot be computed explicitly. Savu and Trede (2004) propose a χ^2-test based on the K-distribution. Tests of LR type were proposed in Chen and Fan (2005). Unfortunately in most cases the distribution of the test statistic does not follow a standard distribution and either bootstrap or other computationally intensive methods should be used.

An alternative approach is based on the probability integral transform introduced in Rosenblatt (1952) and applied in Breymann, Dias and Embrechts (2003), Chen et al. (2004). The idea of the transformation is to construct the variables

$$\begin{aligned} Y_1 &= F_1(X_1), \\ Y_j &= C_0\{F_j(X_j)|F_1(X_1), \ldots, F_{j-1}(X_{j-1})\}, \quad \text{for} \quad j = 2, \ldots, d, \end{aligned}$$

where the conditional copula is defined as

$$C_0(u_j|u_1, \ldots, u_{j-1}) = \frac{\frac{\partial^{j-1}}{\partial u_1 \ldots \partial u_{j-1}} C_0(u_1, \ldots, u_j, 1, \ldots, 1)}{\frac{\partial^{j-1}}{\partial u_1 \ldots \partial u_{j-1}} C_0(u_1, \ldots, u_{j-1}, 1, \ldots, 1)}.$$

Under H_0 the variables Y_i, for $i = 1, \ldots, d$ are independently and uniformly distributed on $[0, 1]$. Since the variables Y_i are not directly observable, we compute the pseudo variables $\hat{Y}_{ji}$ defined by

$$
\begin{aligned}
\hat{Y}_{1i} &= \hat{F}_1(X_{1i}), \\
\hat{Y}_{ji} &= C\{\hat{F}_j(X_{ji})|\hat{F}_1(X_{1i}), \ldots, \hat{F}_{j-1}(X_{j-1,i})\},
\end{aligned}
\tag{1.11}
$$

for $j = 2, \ldots, d$, $i = 1, \ldots, n$. Chen et al. (2004) proposed two tests based on $\hat{Y}_{ji}$. Both can be used for our purposes, however here we discuss the second test. Consider the variable $W = \sum_{j=1}^{d}[\Phi^{-1}(Y_j)]^2$. Under H_0 it holds that $W \sim \chi_d^2$. Similarly as Y_j's, W is not observed and its pseudo-observations are computed as $\hat{W}_i = \sum_{j=1}^{d}[\Phi^{-1}(Y_{ji})]^2$. Breymann et al. (2003) assume that estimating margins and copula parameters does not significantly affect the distribution of $\hat{W}_i$ and apply a standard χ^2 test directly to the pseudo-observations.

Chen et al. (2004) develop a kernel-based test for the distribution of W and, thus, account for estimation errors. Let $\hat{g}_W(w)$ denote the kernel estimator of the density of W, i.e. $\hat{g}_W(w) = \frac{1}{nh}\sum_{i=1}^{n} K_h\{w, F_{\chi_d^2}(\hat{W}_i)\}$, where K_h is the univariate boundary kernel with the second order kernel function $k(\cdot)$. Under H_0 the density $g_W(w)$ is equal to one. As a measure of divergency we use $\hat{J}_n = \int_0^1 \{\hat{g}_W(w) - 1\}^2 dw$. Assuming non-parametric estimator of the marginal distributions Chen et al. (2004) prove under regularity conditions that

$$
T_n = (n\sqrt{h}\hat{J}_n - c_n)/\sigma \to \mathrm{N}(0, 1),
$$

where the parameters are defined in Chen et al. (2004). The proof of this statement does not depend explicitly on the type of the non-parametric estimator of the marginals F_i, but uses the order of $\hat{F}_j(X_{ji}) - F_j(X_{ji})$ as a function of n. It can be shown that if the parametric families of marginal distributions are correctly specified and their parameters are consistently estimated, then the statement holds also if we use parametric estimators for marginal distributions. Since the test is distribution-free it is convenient to use it as a GOF measure for different copulae in different dimensions. Moreover as argued by Chen et al. (2004), the power and size of the test are comparable with other more sophisticated tests.

1.6 Simulation Methods

Monte-Carlo simulations are often a single reliable solution method in many financial problems. Within the simulation study the random variables are

generated from some prescribed distributions. There are numerous methods of simulating from copula-based distributions (see Frees and Valdez (1998), Whelan (2004), Marshall and Olkin (1988), McNeil (2007), Embrechts, McNeil and Straumann (1999), Frey and McNeil (2003), Devroye (1986), etc.). Here we focus on two of them, on the conditional inversion method and on the method proposed by Marshall and Olkin (1988) for Archimedean copulae with generalizations to hierarchical Archimedean copulae by McNeil (2007).

1.6.1 Conditional Inverse Method

The conditional inverse method is a general approach aimed of simulating random variables from an arbitrary multivariate distribution. Here we sketch this method on the example of simulating from copulae. The idea is to generate random variables recursively from the conditional distributions. Let $u_1, \ldots, u_d$ be the sample we generate and let $v_1, \ldots, v_d \sim U(0, 1)$ be a uniformly distributed random sample. We set $u_1 = v_1$. The rest of the variables we generate using the recursion $u_i = C_i^{-1}(v_i | u_1, \ldots, u_{i-1})$ for $i = 2, \ldots, d$, where $C_i = C(u_1, \ldots, u_i, 1, \ldots, 1)$ and the conditional distribution of U_i is given by

$$C_i(u_i | u_1, \ldots, u_{i-1}) = \mathrm{P}(U_i \leq u_i | U_1 = u_1 \ldots U_{i-1} = u_{i-1}) = \frac{\frac{\partial^{i-1} C_i(u_1, \ldots, u_i)}{\partial u_1 \ldots \partial u_{i-1}}}{\frac{\partial^{i-1} C_{i-1}(u_1, \ldots, u_{i-1})}{\partial u_1 \ldots \partial u_{i-1}}}.$$

The method is numerically expensive, since it depends on higher order derivatives of C and the inverse of the conditional distribution function.

1.6.2 Marshal-Olkin Method

For simulating from Archimedean copulae a simpler method is introduced in Marshall and Olkin (1988). The idea of the method is based on the fact that the Archimedean copulae are derived from Laplace transforms. Let M be a univariate cumulative distribution function of a positive random variable (so that $M(0) = 0$) and ϕ is the Laplace transform of M, i.e.

$$\phi(s) = \int_0^\infty \exp\{-sw\} \, dM(w), \ s \geq 0.$$

For any univariate distribution function F, a unique distribution G exists such that

$$F(x) = \int_0^\infty G^\alpha(x) \, dM(\alpha) = \phi\{-\log G(x)\}.$$

Considering d different univariate distributions $F_1, \ldots, F_d$, we obtain that

$$C(u_1, \ldots, u_d) = \int_0^\infty \prod_{i=1}^d G_i^\alpha \, dM(\alpha) = \phi \left[\sum_{i=1}^d \phi^{-1}\{F_i(u_i)\} \right]$$

is a multivariate distribution function. To add even more generality we replace the product of univariate distributions G_i with an arbitrary copula function R

$$C(u_1, \ldots, u_d) = \int_0^\infty \ldots \int_0^\infty R(G_1^\alpha, \ldots, G_d^\alpha) \, dM(\alpha).$$

Note that for the classical Archimedean copula R is equal to a product copula. Following the paper of Marshall and Olkin (1988) we proceed with the following three steps to make a draw from a distribution described by an Archimedean copula:

1. generate an observation u from M;
2. generate observations $(v_1, \ldots, v_d)$ from R;
3. the generated vector is computed by $\mathbf{x} = \{G_1^{-1}(v_1^{1/u}), \ldots, G_d^{-1}(v_d^{1/u})\}$.

This method works much faster than the classical conditional inverse technique. The drawback is that the distribution M can be determined explicitly only for a few generator functions ϕ. This can be done, for example, for Frank, Gumbel and Clayton families (see McNeil (2007), Marshall and Olkin (1988)). The same problem arises in the case of hierarchical copulae, where $\phi_i \circ \phi_{i+1}^{-1}$ should satisfy the properties of generator functions. A slightly modified but more simple procedure for simulating from hierarchical Archimedean copulae is considered in McNeil (2007).

1.7 Applications to Finance

The dependency plays a key role in many financial applications. Elliptical distributions, with the correlation coefficient as the main measure of dependency, constitute a well established class of dependency models commonly used in finance. However, the symmetry assumption and the imposed tail behavior do not reflect the empirical evidence on financial time series. This leads to numerous extensions of Gaussian models to copula-based distributions. In this section we discuss three such extensions. Firstly, we consider the asset allocation problem with non-Gaussian asset returns. Secondly, we discuss the peculiarities of the Value-at-Risk estimation in the non-elliptical framework. Thirdly, we consider the time series models with the residuals following a copula-based distribution.

1.7.1 Asset Allocation

In this section we illustrate the extension of the classical asset allocation problem to copula-based models following Patton (2004). Further discussion and application of the impact of copula-based distribution on portfolio selection procedures can be found in Longin and Solnik (2001) and Hennessy and Lapan (2002).

We consider an investor with a CRRA utility function $U(x) = (1 - \gamma)^{-1} x^{1-\gamma}$ willing to allocate his wealth to d risky assets. We denote the d-dimensional vector of continuously compounded asset returns at time $t + 1$ by $\mathbf{r}_{t+1} = (r_{1,t+1}, \ldots, r_{d,t+1})^\top$ and the vector of portfolio weights by $\mathbf{w} = (w_1, \ldots, w_d)^\top$. Let F_{t+1} be the d-dimensional distribution function of $\mathbf{r}_{t+1}$ with the mean μ_{t+1} and covariance matrix $\mathbf{\Sigma}_{t+1}$. The aim is to forecast F_{t+1} for the time period $t + 1$ using the data up to time t. The estimator is denoted by $\hat{F}_{t+1}$ with the mean $\hat{\mu}_{t+1}$, the covariance matrix $\hat{\mathbf{\Sigma}}_{t+1}$ and the density $\hat{f}_{t+1}$. The objective of the investor is to maximize the expected utility at the time point $t + 1$. This leads to the optimization problem

$$\max_{\mathbf{w} \in \mathcal{W}} \mathsf{E}\, U(1 + \mathbf{w}^\top \mathbf{r}_{t+1}). \tag{1.12}$$

In the case of no-short-sales constraint we set $\mathcal{W} = \{\mathbf{w} \in [0, 1]^d : \mathbf{w}^\top \mathbf{1} = 1\}$ else we set $\mathcal{W} = \{\mathbf{w} \in \mathbb{R}^d : \mathbf{w}^\top \mathbf{1} = 1\}$. The conditional expectation in (1.12) implies that we integrate the utility with respect to the forecasted distribution $\hat{F}_{t+1}$. This reduces the problem (1.12) to the problem

$$\max_{\mathbf{w} \in \mathcal{W}} \int \cdots \int U(1 + \mathbf{w}^\top \mathbf{r}_{t+1}) \hat{f}_{t+1}(\mathbf{r}_{t+1}) d\mathbf{r}_{t+1}.$$

There are several alternative parametric approaches to modeling F_{t+1}. Let $\mathbf{\Sigma}_{d,t+1}$ denote the diagonal matrix containing only the main diagonal of $\mathbf{\Sigma}_{t+1}$. Then $\mathbf{\Sigma}_{t+1} = \mathbf{\Sigma}_{d,t+1}^{1/2} \mathbf{R}_{t+1} \mathbf{\Sigma}_{d,t+1}^{1/2}$, where $\mathbf{R}_{t+1}$ denotes the correlation matrix. A standard approach is to define the model of the asset returns in the form

$$\mathbf{\Sigma}_{d,t}^{-1/2}(\mathbf{r}_t - \mu_t) \sim \mathrm{N}_d(\mathbf{0}, \mathbf{R}_t), \tag{1.13}$$

where the conditional moments μ_t and $\mathbf{\Sigma}_t$ are modeled by a GARCH-in-mean type of process (Franke, Härdle and Hafner (2008)). As a simpler alternative we can consider a Bayesian framework where F_{t+1} denotes the predictive distribution of the asset returns as in Barberis (2000). The unknown parameters of the conditional moments are usually estimated numerically using the ML methodology.

To introduce a copula-based distribution into the asset allocation we deviate from the normality assumption and, following the Sklar's theorem, assume

that $F = C(F_1, \ldots, F_d)$. Thus the model (1.14) is replaced with the model

$$\Sigma_{d,t}^{-1/2}(\mathbf{r}_t - \mu_t) \sim C(F_1, \ldots, F_d) \tag{1.14}$$

with some given functional forms of the copula and the marginal distributions. Similarly as above, the parameters of the conditional moments, of the copula and of the marginal distributions are estimated using the ML method.

In Patton (2004) the investor allocates his wealth between small cap and large cap stocks (i.e. $d = 2$). The conditional mean is defined as linear function of the lagged asset returns and additional explanatory variables. The conditional variance is stated in the TARCH(1,1) form. The rotated Gumbel copula with skewed t margins are used to construct the bivariate distribution of the residuals. This model reveals the highest likelihood function and the lowest AIC and BIC criterion. It is concluded that unconstrained portfolios derived from the normality assumption performed worse in 9 of 10 different trading strategies compared to the Gumbel model.

1.7.2 Value-at-Risk

One of the main advantages of copulae is the fact that they allow for flexible modeling of the tail behavior of multivariate distributions. Since the tail behavior explains the simultaneous outliers of asset returns, it is of special interest in risk management. Therefore, in this section we illustrate the use of copulae for computing the Value-at-Risk (VaR) of portfolios following Embrechts et al. (1999) and Junker and May (2005). The VaR of a portfolio at level α is defined as the lower α-quantile of the distribution of the portfolio return $r_p = \mathbf{w}^\top \mathbf{r}$, i.e.

$$VaR(\alpha) = F_{r_p}^{-1}(\alpha).$$

The VaR is a reasonable measure of risk if we assume that the returns are elliptically distributed. This follows from the fact that VaR is a coherent risk measure (see Embrechts et al. (1999)). Moreover, the assumption of ellipticity implies that minimizing the variance in the Markowitz problem also minimizes the VaR, the expected shortfall and any other coherent measure of risk. However, this statement is false in the non-elliptical case. Moreover, regarding the effect of diversification the variance is the smallest (highest) for perfect negative (positive) correlation of the assets. This also holds for the VaR in the elliptical case, however, not for the non-elliptical distributions (see Embrechts et al. (1999), Theorem 5). This implies that for copula-based distribution the VaR should be used with caution and its computation should be awarded more attention.

Consider the probability that the portfolio return r_p does not exceed some predetermined value ξ, i.e. $\mathrm{P}(r_p \leq \xi)$. Our aim is to determine the lower α-quantile of the distribution of r_p, or, equivalently, to determine such ξ that $\mathrm{P}(r_p \leq \xi) = \alpha$. Note that

$$r_p = \mathbf{w}^\top \mathbf{r} = \sum_{i=1}^{d} w_i r_i = \sum_{i=1}^{d} w_i F_i^{-1}(u_i),$$

where F_i denote the marginal distributions of individual asset returns, $u_i = F_i(r_i) \sim U[0,1]$ for all $i = 1, \ldots, d$ and $u_1, \ldots, u_d \sim C$. The copula C defines the dependency structure between the asset returns. This implies that

$$\mathrm{P}(r_p \leq \xi) = \int_{\mathcal{U}} c(u_1, \ldots, u_d) du_1 \ldots du_d, \qquad (1.15)$$

with

$$\mathcal{U} = \{[0,1]^{d-1} \times [0, u_d(\xi)]\}, \quad u_d(\xi) = F_d\Big[\xi/w_d - \sum_{i=1}^{d-1} w_i F_i^{-1}(u_i)/w_d\Big].$$

For fixed α, the VaR is determined by solving (1.15) numerically for ξ. Direct multidimensional numerical integration is a tedious task which can be substantially simplified by using the Monte-Carlo integration. For this purpose we have to generate random samples from C using methods described in Section 1.6.

Junker and May (2005) apply the above methodology to a portfolio consisting of two assets, Hoechts and Volkswagen shares. The returns are standardized by the sample mean and the conditional volatility from the GARCH(1,1) process. The copula function is defined as a convex linear combination of the Frank copula and its survival copula. It is concluded that empirical or t-margins and asymmetric copula-based dependency structures provide the best fit in terms of χ^2 goodness-of-fit test of Diebold, Gunther and Tay (1998). Moreover, the VaR estimator from this model well approximates the empirical estimator. The assumption of Gaussian GARCH(1,1) standardized returns renders the worst results.

1.7.3 Time Series Modeling

Time series models constitute one of the most important tools in dealing with financial data. However, multivariate modeling used up to now does not properly describe financial and economic time series. The reason is that

these models are mainly based on the Gaussian or on elliptical distributions. Nowadays there are numerous papers extending the classical time series models to model copula distributed residuals. First we consider the semiparametric copula-based multivariate dynamic model (SCOMDY) of Chen and Fan (2006). Let $\{\mathbf{y}_t^\top, \mathbf{x}_t^\top\}_{t=1}^n$ be stochastic processes, where $\mathbf{x}_t$ contains the exogenous variables and the d-dimensional vector $\mathbf{y}_t$ contains the variables of interest. Let $\mathcal{F}_{t-1}$ denote the information up to the time point t. They specified the model in the following way

$$\mathbf{y}_t = \mu_t(\boldsymbol{\theta}_1) + \sqrt{\mathbf{H}_t(\boldsymbol{\theta})}\boldsymbol{\varepsilon}_t,$$

where

$$
\begin{aligned}
\boldsymbol{\theta} &= \{\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top\}^\top \\
\mu_t(\boldsymbol{\theta}_1) &= (\mu_{1,t}(\boldsymbol{\theta}_1), \ldots, \mu_{d,t}(\boldsymbol{\theta}_1))^\top \\
&= \mathsf{E}\{\mathbf{y}_t | \mathcal{F}_{t-1}\} \\
\mathbf{H}_t(\boldsymbol{\theta}) &= \mathrm{diag}\{h_{1,t}(\boldsymbol{\theta}), \ldots, h_{d,t}(\boldsymbol{\theta})\} \\
&= \mathrm{diag}\{h_{1,t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2), \ldots, h_{d,t}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)\} \\
&= \mathrm{diag}\left(\mathsf{E}[\{\mathbf{y}_{1t} - \mu_t(\boldsymbol{\theta}_1)\}^2 | \mathcal{F}_{t-1}], \ldots, \mathsf{E}[\{\mathbf{y}_{dt} - \mu_t(\boldsymbol{\theta}_1)\}^2 | \mathcal{F}_{t-1}]\right).
\end{aligned}
$$

$\mu_t(\cdot)$ is the true conditional mean of the $\mathbf{y}_t$ given $\mathcal{F}_{t-1}$ and $h_{jt}(\cdot)$ is the true conditional variance of the $\mathbf{y}_{jt}$ given $\mathcal{F}_{t-1}$. The residuals are assumed to be serially independent with zero mean and unit variances, i.e. $\mathsf{E}[\varepsilon_{jt}] = 0$ and $\mathsf{E}[\varepsilon_{jt}^2] = 1$ for $j = 1, \ldots, d$. The joint distribution of $\boldsymbol{\varepsilon}$ is assumed to be given by $C\{F_1(\varepsilon_1), \ldots, F_d(\varepsilon_d)\}$, where the margins and the copula function are unknown. This general specification includes the standard processes GARCH, ARCH, VAR as special cases, however, it allows for much more flexibility in the choice of the dependency structure of the residuals. For example considering $\boldsymbol{\theta}_1 = (\boldsymbol{\delta}_1^\top, \ldots, \boldsymbol{\delta}_d^\top)^\top$, $\boldsymbol{\theta}_2 = (\kappa_1, \ldots, \kappa_d; \beta_1, \ldots, \beta_d; \gamma_1, \ldots, \gamma_d)^\top$, $\mu_t = (\mathbf{x}_{1t}^\top \boldsymbol{\delta}_1, \ldots, \mathbf{x}_{dt}^\top \boldsymbol{\delta}_d)^\top$, $\mathbf{H}_t = \mathrm{diag}\{h_{1t}, \ldots, h_{dt}\}$ and the copula for $\boldsymbol{\varepsilon}$ is assumed to be the Gaussian copula, where $\kappa_j > 0$, $\beta_j \geq 0$, $\gamma_j \geq 0$ and $\beta_j + \gamma_j < 1$ for $j = 1, \ldots, d$ we get GARCH(1,1) model with normal innovations

$$
\begin{aligned}
y_{jt} &= \mathbf{x}_{jt}^\top \boldsymbol{\delta}_j + \sqrt{h_{jt}}\varepsilon_{jt} \\
h_{jt} &= \kappa_j + \beta_j h_{j,t+1} + \gamma_j (y_{j,t-1} - \mathbf{x}_{j,t-1}^\top \boldsymbol{\delta}_j)^2, \quad j = 1, \ldots, d.
\end{aligned}
$$

Chen and Fan (2006) consider maximum likelihood estimators of the parameters in these models and establish large sample properties when the copula is mis-specified. For the choice between two SCOMDY models they introduce a pseudo likelihood ratio test and provide the limiting distribution of the test statistic.

In contrast to the paper by Chen and Fan (2006), Fermanian and Scaillet (2003) consider a nonparametric estimation of copulae for time series and

derived asymptotic properties of the kernel estimator of copulae. Further generalization is discussed in Giacomini, Härdle and Spokoiny (2008), where the parameter of the copula is assumed to be time dependent. The aim is to determine the periods with constant dependency structure.

1.8 Simulation Study and Empirical Results

In this section we illustrate the considered algorithms on simulated and real-world data. The next sub-section contains the simulation study, where we show that the aggregated binary structures outperform the alternative strategies of artificial data. In Section 1.8.2 we used Bayes and Akaike information criterion to compare the performance of the HAC-based model with the classical Gaussian and t-models on real data.

1.8.1 Simulation Study

Setup of the study

The aim of this simulation study is the comparison of grouping methods on the example of simulated data. We consider two different true structures $s = (123)(45)$ and $s = (12(34))5$ with the Gumbel generator function given by

$$\phi^{-1} = \{-\log(u)\}^\theta, \qquad \phi = \exp(-u^{\frac{1}{\theta}}).$$

This naturally corresponds to the Gumbel copula. The parameters are set for the first structure equal to $\theta_{123} = 4$, $\theta_{45} = 3$ and $\theta_{(123)(45)} = 2$ and for the second structure to $\theta_{34} = 4$, $\theta_{12(34)} = 3$ and $\theta_{(12(34))5} = 2$. Without loss of generality the marginal distributions are taken as uniform on $[0, 1]$. We simulate a sample of 1000 observations. The procedure is repeated 101 times. This number is selected to simplify the interpretation and computation of median structures.

For the simulation we use the conditional inversion method. This method is also used by Frees and Valdez (1998) and Whelan (2004) and we discussed it in Section 1.6. The copula parameters are estimated using the multistage ML method with the nonparametric estimation of margins based on the Epanechnikov kernel. The vector of bandwidths $\mathbf{h} = \{h_i\}_{i=1,\dots,d}$ in the estimation of the density and in the estimation of the distribution function is based on the Silverman's rule of thumb.

Discussion of the results

The results of the simulation study are summarized in Table 1.1 for the first structure and in Table 1.2 for the second structure. For each simulated data set and for each structure we compare the fit and the structure obtained from the grouping procedures. We consider the simple Archimedean copula (sAC) and groupings based on the Chen et al. (2004) test statistics (Chen), on the θ (θ), binary copulae (θ_{binary}) and aggregated binary copulae ($\theta_{binary\ aggr.}$). As benchmark models we consider the 5-dimensional multivariate normal distribution (N) with $\hat{\Sigma}$ and $\hat{\mu}$ estimated from the data; the multivariate Gaussian copula with nonparametric margins ($N_{nonparam.}$); the multivariate t-distribution with eight degrees of freedom and with $\hat{\Sigma}$ and $\hat{\mu}$ estimated from the data set (t_8); the multivariate t-copula with eight degrees of freedom and nonparametric margins ($t_{nonparam.}$).

For each grouping method and each benchmark we compute the Kullback-Leibler divergence from the empirical distribution function as in Giacomini et al. (2008) and the test statistic of Chen et al. (2004). The Kulback-Leibler functional for the distribution functions estimated using two different methods is

$$\mathcal{K}(\hat{F}_{method\ 1}, \hat{F}_{method\ 2}) = \frac{1}{n} \sum_{i=1}^{n} \log \left\{ \frac{\widehat{f}_{method\ 1}(x_{1i}, \ldots, x_{di})}{\widehat{f}_{method\ 2}(x_{1i}, \ldots, x_{di})} \right\}$$

The Kullback-Leibler divergence for the multivariate distribution which is based on copula, can be regarded as a distance between two copula densities.

The first blocks of Table 1.1 and Table 1.2 contain the results for groupings based on the Kullback-Leibler divergence. The columns "$\mathcal{K}$" contain the value of the Kullback-Leibler divergence which is the closest to the median divergence given in parenthesis. The corresponding structure and the test statistic of Chen et al. (2004) are given in the columns "copula structure" and "Chen." respectively. The variance of the Kullback-Leibler divergence is given in the last column. The same holds for the lower blocks of both tables, however, here we find the structure which has the test statistics of Chen et al. (2004) which is the closest to the median of the test statistics. Note that we provide the results for the median performance measures and not for the best replications of the simulation study. This makes the conclusions more robust.

The results show that the grouping method based on the aggregated binary structure is dominant. It provides for both structures the smallest Kullback-Leibler divergence as well as the lowest test statistics. The simple binary copula also provides good results, however, we see that some of the parameters are very close. This indicates that the variables can be joined together

into an aggregated copula. Although, the method based on θ's performs better than the benchmark strategies, it leads, however, to incorrect structures and much higher goodness-of-fit measures compared to the binary copula. The grouping based on the test statistic of Chen et al. (2004) provides very poor results, which indicates a low power of the test against similar structures. Similarly, the ignorance of the hierarchical structure of the distribution imposed by the simple Archimedean copulae leads to the worst results among copula-based methods. The comparison with normal and t-distributions is possible only on the basis of the Kullback-Leibler divergence. We see that, despite of the substantially larger number of parameters, the normal and t-distributions cannot outperform the θ-based grouping methods. Thus we conclude that the proposed grouping methodology based on the aggregated binary structure provides robust and precise results. Moreover, the method is computationally more efficient than the considered alternatives.

method	copula structure	Chen.	$\mathcal{K}$ $(\hat{\mu}_{\mathcal{K}})$	$\hat{\sigma}^2_{\mathcal{K}}(10^{-3})$
N			1.074 (1.074)	4.0
$N_{nonparam.}$			0.282 (0.283)	1.0
t_8			1.104 (1.104)	3.0
$t_{nonparam.}$			0.199 (0.199)	0.0
sAC	$(1.2.3.4.5)$	85.535463	0.811 (0.809)	3.0
CHEN	$((1.3)_{4.517}.(2.4.5)_{2.341})_{2.34}$	31.432	0.611 (0.613)	78.0
θ	$((1.3.4.5)_{2.288}.2)_{2.286}$	82.510	0.697 (0.560)	142.0
θ_{binary}	$(((1.(2.3)_{4.39})_{4.282}.5)_{2.078}.4)_{2.077}$	3.929	0.132 (0.133)	0.4
$\theta_{binary\ aggr.}$	$(((1.3)_{4.26}.2)_{3.868}.(4.5)_{3.093})_{2.259}$	2.737	0.022 (0.021)	0.0
method	copula structure	Chen. $(\hat{\mu}_{Chen})$	$\mathcal{K}$	$\hat{\sigma}^2_{Chen\,stat}$
sAC	$(1.2.3.4.5)$	88.842 (88.850)	0.704	68.127
CHEN	$((1.2)_{4.316}.(3.4.5)_{2.256})_{2.255}$	31.558 (32.419)	0.585	490.059
θ	$((1.2.4.5)_{2.376}.3)_{2.375}$	56.077 (56.910)	0.769	1407.632
θ_{binary}	$((((1.2)_{4.487}.3)_{4.469}.5)_{2.247}.4)_{2.246}$	4.789 (4.827)	0.112	4.388
$\theta_{binary\ aggr.}$	$(((1.3)_{4.228}.2)_{3.68}.(4.5)_{3.369})_{2.333}$	2.253 (2.248)	0.021	1.914

Table 1.1. Model fit for the true structure (123)(45): Averages of the Kullback-Leibler Divergence and Averages of the Chen Statistics separately

1.8.2 Empirical Example

In this subsection we apply the proposed estimation techniques to financial data. We consider the daily returns of four companies listed in DAX index: Commerzbank (CBK), Merck (MRK), Thyssenkrupp (TKA) and Volkswagen (VOW). The sample period covers more than 2300 observations from

method	copula structure	Chen.	$\mathcal{K}$ $(\hat{\mu}_{\mathcal{K}})$	$\hat{\sigma}^2_{\mathcal{K}}(10^{-3})$
N			1.088 (1.089)	4
N$_{nonparam.}$			0.289 (0.289)	1
t_8			1.113 (1.114)	3
$t_{nonparam.}$			0.202 (0.202)	1
sAC	$(1.2.3.4.5)$	78.604	0.502 (0.502)	2
CHEN	$(((1.2)_{3.22}.3)_{3.177}.(4.5)_{2.116})_{2.114}$	8.544	0.305 (0.304)	23
θ	$(((1.2.3)_{3.207}.4)_{3.205}.5)_{2.15}$	5.741	0.079 (0.079)	0
θ_{binary}	$(((1.(3.4)_{4.157})_{3.099}.2)_{3.012}.5)_{2.028}$	2.293	0.003 (0.003)	0
$\theta_{binary\ aggr.}$	$(((3.4)_{4.32}.1.2)_{3.268}.5)_{1.83}$	1.220	0.019 (0.019)	0
method	**copula structure**	**Chen.** $(\hat{\mu}_{Chen.})$	$\mathcal{K}$	$\hat{\sigma}^2_{Chen\,stat}$
sAC	$(1.2.3.4.5)$	86.245 (86.278)	0.480	142.714
CHEN	$(((1.3)_{2.835}.5)_{1.987}.(2.4)_{2.898})_{1.986}$	16.263 (16.512)	0.453	281.615
θ	$(((1.2.4)_{3.009}.3)_{3.007}.5)_{1.973}$	4.235 (4.222)	0.083	6.229
θ_{binary}	$(((1.(3.4)_{4.122})_{3.155}.2)_{3.07}.5)_{2.027}$	1.934 (1.955)	0.000	1.520
$\theta_{binary\ aggr.}$	$(((3.4)_{4.195}.1.2)_{3.305}.5)_{1.724}$	2.561 (2.526)	0.014	3.287

Table 1.2. Model fit for the true structure (12(34))5: Averages of the Kullback-Leibler Divergence and Averages of the Chen Statistics separately

13.11.1998 to 18.10.2007. Margins are estimated nonparametrically with Epanechnikov kernel, normal and t-distributed with three degrees of freedom. The results are given in Tables 1.3, 1.4 and 1.5 respectively. For goodness-of-fit measures we choose BIC (Bayes or Schwarz Information Criteria) and AIC (Akaike Information Criteria) and provide the value of the likelihood as intermediate results.

We fit the following multivariate copula functions to the data: HAC with binary and binary aggregated structure, simple Archimedean copula. For comparison purposes we also provide the results for the multivariate normal distribution and multivariate t-distribution with eight degrees of freedom in each table. We also provide the optimal binary and aggregated binary HACs and the simple Archimedean copula for all types of the margins.

We calculate the maximum likelihood value as described in Section 1.4. For the copula-base distributions we use

$$ML = \sum_{i=1}^{n} \log\{c(u_1, \dots, u_d, \boldsymbol{\theta}) f_1(u_1) \dots f_d(u_d)\},$$

where c is the copula density and f_i for $i = 1, \dots, d$ are marginal densities. For the multivariate normal and t-distribution, we computed the likelihood

as

$$ML = \sum_{i=1}^{n} \log\{f(u_1, \ldots, u_d, \boldsymbol{\theta})\},$$

where f denotes the joint multivariate density function and $\boldsymbol{\theta}$ is the set of parameters. To penalize the likelihood for large number of parameters we consider the AIC and BIC criterion computed as

$$AIC = -2ML + 2m, \qquad BIC = -2ML + 2\log(m),$$

where m is the number of the parameters to be estimated. The values of ML for the best structure should the the highest, while AIC and BIC should be as small as possible.

We emphasize with bold font the best strategy in each column and with italic the worst strategies. We can conclude that the multivariate t distribution outperforms all other methods and shows the best results for all types of the margins. Nevertheless, note that with the properly selected marginal distributions and copula function, the HAC outperforms the normal distribution. Moreover, note that we considered only HACs based on the Gumbel generator functions. Alternative generator specifications and HACs dependent on several different generators may outperform the t-distribution as well.

	ML	AIC	BIC
HAC	28319.3701	-56632.7402	-56614.9003
HAC$_{binary}$	28319.3701	-56632.7402	-56614.9003
AC	*28028.5201*	-56055.0403	-56049.0937
N	28027.4098	*-56026.8195*	*-55943.5669*
t_8	**28726.8637**	**-57425.7273**	**-57342.4747**

Table 1.3. Information Criteria: Nonparametric Margins

$$\begin{aligned}
\text{Optimal binary structure} \;&=\; (((\text{CBK VOW})_{1.5631}\ \text{TKA})_{1.4855}\ \text{MRK})_{1.1437} \\
\text{Optimal structure} \;&=\; (((\text{CBK VOW})_{1.5631}\ \text{TKA})_{1.4855}\ \text{MRK})_{1.1437} \\
\text{Simple Archimedean Copula} \;&=\; (\text{CBK MRK TKA VOW})_{1.4116}
\end{aligned}$$

	ML	AIC	BIC
HAC	27961.0997	-55918.1995	-55906.3062
HAC_binary	27961.2399	-55916.4799	-55898.6400
AC	*27737.7392*	*-55473.4784*	*-55467.5317*
N	28027.4098	-56026.8195	-55943.5669
t_8	**28726.8637**	**-57425.7273**	**-57342.4747**

Table 1.4. Information Criteria: Normal margins

$$
\begin{aligned}
\text{Optimal binary structure} \;=\;& (((\text{CBK VOW})_{1.3756}\ \text{TKA})_{1.3571}\ \text{MRK})_{1.1071} \\
\text{Optimal structure} \;=\;& ((\text{CBK TKA VOW})_{1.3756}\ \text{MRK})_{1.1071} \\
\text{Simple Archimedean Copula} \;=\;& (\text{CBK MRK TKA VOW})_{1.1944}
\end{aligned}
$$

	ML	AIC	BIC
HAC	28613.9640	-57223.9280	-57212.0347
HAC_binary	28612.2069	-57218.4138	-57200.5740
AC	28404.8899	-56807.7798	-56801.0347
N	*28027.4098*	*-56026.8195*	*-55943.5669*
t_8	**28726.8637**	**-57425.7273**	**-57342.4747**

Table 1.5. Information Criteria: t margins

$$
\begin{aligned}
\text{Optimal binary structure} \;=\;& (((\text{CBK VOW})_{1.3416}\ \text{TKA})_{1.3285}\ \text{MRK})_{1.1007} \\
\text{Optimal structure} \;=\;& ((\text{CBK TKA VOW})_{1.3416}\ \text{MRK})_{1.1007} \\
\text{Simple Archimedean Copula} \;=\;& (\text{CBK MRK TKA VOW})_{1.1987}
\end{aligned}
$$

1.9 Summary

In this chapter we provide a detailed review of the copula models in discrete time. We review the construction and simulation of bivariate and multivariate copula models. For practical applications we discuss the alternative estimation procedures and goodness-of-fit tests. Special attention is paid to the hierarchical Archimedean copulae. The chapter is complemented with an extensive simulation study and an application to financial data.

Bibliography

Aas, K., Czado, C., Frignessi, A. and Bakken, H. (2006). Pair-copula constructions of multiple dependence, *Ins.: Mathematics Econ.* . forthcoming.

Barbe, P., Genest, C., Ghoudi, K. and Rémillard, B. (1996). On Kendalls's process, *J. Multivariate Anal.* **58**: 197–229.

Barberis, N. (2000). Investing for the long run when returns are predictable, *J. Finance* **55**(1): 225–264.

Bedford, T. and Cooke, R. M. (2001). Probability density decomposition for conditionally dependent random variables modeled by vines, *Annals of Mathematical and Artificial Intelligence* **32**: 245–268.

Bedford, T. and Cooke, R. M. (2002). Vines – a new graphical model for dependent random variables, *Ann. Statist.* **30**(4): 1031–1068.

Berg, D. and Aas, K. (2007). Models for construction of multivariate dependence, *Technical Report SAMBA/23/07*, Norwegian Computing Center.

Breymann, W., Dias, A. and Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance, *Quant. Finance* **1**: 1–14.

Chen, S. X. and Huang, T. (2007). Nonparametric estimation of copula functions for dependence modeling, *The Canadian Journal of Statistics* . forthcoming.

Chen, X. and Fan, Y. (2005). Pseudo-likelihood ratio tests for model selection in semiparametric multivariate copula models, *The Canadian Journal of Statistics* **33**(2): 389–414.

Chen, X. and Fan, Y. (2006). Estimation and model selection of semiparametric copula-based multivariate dynamic models under copula misspesification, *J. Econometrics* **135**: 125–154.

Chen, X., Fan, Y. and Patton, A. (2004). Simple tests for models of dependence between multiple financial time series, with applications to U.S. equity returns and exchange rates, *Discussion paper 483*, Financial Markets Group, London School of Economics.

Chen, X., Fan, Y. and Tsyrennikov, V. (2006). Efficient estimation of semiparametric multivariate copula models, *J. Amer. Statist. Assoc.* **101**(475): 1228–1240.

Chen, Y.-P. (2004). A note on the relationship between Spearman's ρ and Kendall's τ for extreme order statistics, *Journal of Statistical Planning and Inference* **137**(7): 2165–2171.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*, John Wiley & Sons, New York.

Demarta, S. and McNeil, A. J. (2004). The *t*-copula and related copulas, *International Statistical Review* **73**(1): 111–129.

Devroye, L. (1986). *Non-uniform Random Variate Generation*, Springer Verlag, New York.

Diebold, F. X., Gunther, T. and Tay, A. S. (1998). Evaluating density forecasts, with applications to financial risk management, *Int. Econ. Rev.* **39**: 863–883.

Embrechts, P., Lindskog, F. and McNeil, A. J. (2003). Modeling dependence with copulas and applications to risk management, *in* S. T. Rachev (ed.), *Handbook of Heavy Tailed Distributions in Finance*, Elsevier, North-Holland.

Embrechts, P., McNeil, A. J. and Straumann, D. (1999). Correlation and dependence in risk management: Properties and pitfalls, *RISK* pp. 69–71.

Fermanian, J.-D. (2005). Goodness-of-fit tests for copulas, *J. Multivariate Anal.* **95**(1): 119–152.

Fermanian, J.-D. and Scaillet, O. (2003). Nonparametric estimation of copulas for time series, *J. Risk* **5**: 25–54.

Franke, J., Härdle, W. and Hafner, C. (2008). *Statistics of Financial Market, 2nd edition*, Springer Verlag, Berlin.

Fredricks, G. A. and Nelsen, R. B. (2004). On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variables, *Journal of Statistical Planning and Inference* **137**(7): 2143–2150.

Frees, E. and Valdez, E. (1998). Understanding relationships using copulas, *N. Amer. Actuarial J.* **2**: 1–125.

Frey, R. and McNeil, A. J. (2003). Dependent defaults in models of portfolio credit risk, *J. Risk* **6**(1): 59–92.

Genest, C., Ghoudi, K. and Rivest, L.-P. (1995). A semi-parametric estimation procedure of dependence parameters in multivariate families of distributions, *Biometrika* **82**: 543–552.

Genest, C., Quessy, J.-F. and Rémillard, B. (2006). Goodness-of-fit procedures for copula models based on the probability integral transformation, *Scandinavian Journal of Statistics* **33**: 337–366.

Genest, C. and Rivest, L.-P. (1989). A characterization of Gumbel family of extreme value distributions, *Statist. Probab. Letters* **8**: 207–211.

Genest, C. and Rivest, L.-P. (1993). Statistical inference procedures for bivariate Archimedean copulas, *J. Amer. Statist. Assoc.* **88**: 1034–1043.

Giacomini, E., Härdle, W. K. and Spokoiny, V. (2008). Inhomogeneous dependency modelling with time varying copulae, *J. Bus. Econ. Statist.* . forthcoming.

Hennessy, D. A. and Lapan, H. E. (2002). The use of Archimedean copulas to model portfolio allocations, *Math. Finance* **12**: 143–154.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters, *in* L. Rüschendorf, B. Schweizer and M. Taylor (eds), *Distribution with fixed marginals and related topics*, IMS Lecture Notes – Monograph Series, Institute of Mathematical Statistics.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*, Chapman & Hall, London.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models, *J. Multivariate Anal.* **94**: 401–419.

Joe, H. and Xu, J. J. (1996). The estimation method of inference functions for margins for multivariate models, *Technical Report 166*, Department of Statistics, University of British Columbia.

Junker, M. and May, A. (2005). Measurement of aggregate risk with copulas, *Econometrics J.* **8**: 428–454.

Kendall, M. (1970). *Rank Correlation Methods*, Griffin, London.

Kurowicka, M. and Cooke, R. M. (2006). *Uncertainty Analysis with High Dimensional Dependence Modelling*, John Wiley & Sons, New York.

Longin, F. and Solnik, B. (2001). Extreme correlation of international equity markets, *J. Finance* **56**: 649–676.

Marshall, A. W. and Olkin, J. (1988). Families of multivariate distributions, *J. Amer. Statist. Assoc.* **83**: 834–841.

McNeil, A. J. (2007). Sampling nested Archimedean copulas, *Journal of Statistical Computation and Simulation* . forthcoming.

Morillas, P. M. (2005). A method to obtain new copulas from a given one, *Metrika* **61**: 169–184.

Nelsen, R. B. (2006). *An Introduction to Copulas*, Springer Verlag, New York.

Oakes, D. (2005). Multivariate survival distributions, *Journal of Nonparametric Statistics* **3**(3-4): 343–354.

Okhrin, O., Okhrin, Y. and Schmid, W. (2007). On the structure and estimation of hierarchical Archimedean copulas, *SFB 649 discussion series*, Humbodt-Universität zu Berlin.

Patton, A. J. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation, *Journal of Financial Econometrics* **2**: 130–168.

Rosenblatt, M. (1952). Remarks on a multivariate transformation, *Ann. Math. Statist.* **23**: 470–472.

Savu, C. and Trede, M. (2004). Goodness-of-fit tests for parametric families of Archimedean copulas, *Discussion paper*, University of Muenster.

Savu, C. and Trede, M. (2006). Hierarchical Archimedean copulas, *Discussion paper*, University of Muenster.

Schmid, F. and Schmidt, R. (2006a). Bootstrapping Spearman's multivariate rho, *in* A. Rizzi and M. Vichi (eds), *COMPSTAT, Proceedings in Computational Statistics*, pp. 759–766.

Schmid, F. and Schmidt, R. (2006b). Multivariate extensions of Spearman's rho and related statistics, *Statist. Probab. Letters* **77**(4): 407–416.

Shih, J. H. and Louis, T. A. (1995). Inferences on the association parameter in copula models for bivariate survival data, *Biometrics* **51**: 1384–1399.

Sklar, A. (1959). Fonctions dé repartition á n dimension et leurs marges, *Publ. Inst. Stat. Univ. Paris* **8**: 299–231.

Wang, W. and Wells, M. (2000). Model selection and semiparametric inference for bivariate failure-time data, *J. Amer. Statist. Assoc.* **95**: 62–76.

Whelan, N. (2004). Sampling from Archimedean copulas, *Quant. Finance* **4**: 339–352.

Wolff, E. (1980). N-dimensional measures of dependence, *Stochastica* **4**(3): 175–188.

2 Quantification of Spread Risk by Means of Historical Simulation

Christoph Frisch and Germar Knöchlein

2.1 Introduction

Modeling spread risk for interest rate products, i.e., changes of the yield difference between a yield curve characterizing a class of equally risky assets and a riskless benchmark curve, is a challenge for any financial institution seeking to estimate the amount of economic capital utilized by trading and treasury activities. With the help of standard tools this contribution investigates some of the characteristic features of yield spread time series available from commercial data providers. From the properties of these time series it becomes obvious that the application of the parametric variance-covariance-approach for estimating idiosyncratic interest rate risk should be called into question. Instead we apply the non-parametric technique of historical simulation to synthetic zero-bonds of different riskiness, in order to quantify general market risk and spread risk of the bond. The quality of value-at-risk predictions is checked by a backtesting procedure based on a mark-to-model profit/loss calculation for the zero-bond market values. From the backtesting results we derive conclusions for the implementation of internal risk models within financial institutions.

2.2 Risk Categories – a Definition of Terms

For the analysis of obligor-specific and market-sector-specific influence on bond price risk we make use of the following subdivision of "price risk", Gaumert (1999), Bundesaufsichtsamt für das Kreditwesen (2001).

1. General market risk: This risk category comprises price changes of a financial instrument, which are caused by changes of the general

market situation. General market conditions in the interest rate sector are characterized by the shape and the moves of benchmark yield curves, which are usually constructed from several benchmark instruments. The benchmark instruments are chosen in such a way so that they allow for a representative view on present market conditions in a particular market sector.

2. Residual risk: Residual risk characterizes the fact that the actual price of a given financial instrument can change in a way different from the changes of the market benchmark (however, abrupt changes which are caused by events in the sphere of the obligor are excluded from this risk category). These price changes cannot be accounted for by the volatility of the market benchmark. Residual risk is contained in the day-to-day price variation of a given instrument relative to the market benchmark and, thus, can be observed continuously in time. Residual risk is also called *idiosyncratic risk*.

3. Event risk: Abrupt price changes of a given financial instrument relative to the benchmark, which significantly exceed the continuously observable price changes due to the latter two risk categories, are called event risk. Such price jumps are usually caused by events in the sphere of the obligor. They are observed infrequently and irregularly.

Residual risk and event risk form the two components of so-called specific price risk or *specific risk* — a term used in documents on banking regulation, Bank for International Settlements (1998a), Bank for International Settlements (1998b) — and characterize the contribution of the individual risk of a given financial instrument to its overall risk.

The distinction between general market risk and residual risk is not unique but depends on the choice of the benchmark curve, which is used in the analysis of general market risk: The market for interest rate products in a given currency has a substructure (market-sectors), which is reflected by product-specific (swaps, bonds, etc.), industry-specific (bank, financial institution, retail company, etc.) and rating-specific (AAA, AA, A, BBB, etc.) yield curves. For the most liquid markets (USD, EUR, JPY), data for these sub-markets is available from commercial data providers like Bloomberg. Moreover, there are additional influencing factors like collateral, financial restrictions etc., which give rise to further variants of the yield curves mentioned above. Presently, however, hardly any standardized data on these factors is available from data providers.

The larger the universe of benchmark curves a bank uses for modeling its interest risk, the smaller is the residual risk. A bank, which e.g. only uses product-specific yield curves but neglects the influence of industry- and rating-

specific effects in modelling its general market risk, can expect specific price risk to be significantly larger than in a bank which includes these influences in modeling general market risk. The difference is due to the consideration of product-, industry- and rating-specific spreads over the benchmark curve for (almost) riskless government bonds. This leads to the question, whether the risk of a spread change, the *spread risk*, should be interpreted as part of the general market risk or as part of the specific risk. The uncertainty is due to the fact that it is hard to define what a market-sector is. The definition of benchmark curves for the analysis of general market risk depends, however, critically on the market sectors identified.

We will not further pursue this question in the following but will instead investigate some properties of this spread risk and draw conclusions for modeling spread risk within internal risk models. We restrict ourselves to the continuous changes of the yield curves and the spreads, respectively, and do not discuss event risk. In this contribution different methods for the quantification of the risk of a fictive USD zero bond are analyzed. Our investigation is based on time series of daily market yields of US treasury bonds and US bonds (banks and industry) of different credit quality (rating) and time to maturity.

2.3 Yield Spread Time Series

Before we start modeling the interest rate and spread risk we will investigate some of the descriptive statistics of the spread time series. Our investigations are based on commercially available yield curve histories. The Bloomberg dataset we use in this investigation consists of daily yield data for US treasury bonds as well as for bonds issued by banks and financial institutions with ratings AAA, AA+/AA, A+, A, A− (we use the Standard & Poor's naming convention) and for corporate/industry bonds with ratings AAA, AA, AA−, A+, A, A−, BBB+, BBB, BBB−, BB+, BB, BB−, B+, B, B−. The data we use for the industry sector covers the time interval from March 09 1992 to June 08 2000 and corresponds to 2147 observations. The data for banks/financial institutions covers the interval from March 09 1992 to September 14 1999 and corresponds to 1955 observations. We use yields for 3 and 6 month (3M, 6M) as well as 1, 2, 3, 4, 5, 7, and 10 year maturities (1Y, 2Y, 3Y, 4Y, 5Y, 7Y, 10Y). Each yield curve is based on information on the prices of a set of representative bonds with different maturities. The yield curve, of course, depends on the choice of bonds. Yields are option-adjusted but not corrected for coupon payments. The yields for the chosen maturities are constructed by Bloomberg's interpolation algorithm for yield curves. We use the USD

treasury curve as a benchmark for riskless rates and calculate yield spreads relative to the benchmark curve for the different rating categories and the two industries. We correct the data history for obvious flaws using complementary information from other data sources. Some parts of our analysis in this section can be compared with the results given in Kiesel, Perraudin and Taylor (1999).

2.3.1 Data Analysis

We store the time series of the different yield curves in individual files. The file names, the corresponding industries and ratings and the names of the matrices used in the XploRe code are listed in Table 2.2. Each file contains data for the maturities 3M to 10Y in columns 4 to 12. XploRe creates matrices from the data listed in column 4 of Table 2.2 and produces summary statistics for the different yield curves. As example files the data sets for US treasury and industry bonds with rating AAA are provided. The output of the **Q**summarize command for the INAAA curve is given in Table 2.1.

	Minimum	Maximum	Mean	Median	Std.Error
3M	3.13	6.93	5.0952	5.44	0.95896
6M	3.28	7.16	5.2646	5.58	0.98476
1Y	3.59	7.79	5.5148	5.75	0.95457
2Y	4.03	8.05	5.8175	5.95	0.86897
3Y	4.4	8.14	6.0431	6.1	0.79523
4Y	4.65	8.21	6.2141	6.23	0.74613
5Y	4.61	8.26	6.3466	6.36	0.72282
7Y	4.75	8.3	6.5246	6.52	0.69877
10Y	4.87	8.36	6.6962	6.7	0.69854

Table 2.1. Output of **Q**summarize for the INAAA curve.
QXFGsummary

The long term means are of particular interest. Therefore, we summarize them in Table 2.3. In order to get an impression of the development of the treasury yields in time, we plot the time series for the USTF 3M, 1Y, 2Y, 5Y, and 10Y yields. The results are displayed in Figure 2.1, **Q**XFGtreasury . The averaged yields within the observation period are displayed in Figure 2.2 for USTF, INAAA, INBBB2, INBB2 and INB2, **Q**XFGyields .

In the next step we calculate spreads relative to the treasury curve by subtracting the treasury curve from the rating-specific yield curves and store

Industry	Rating	File Name	Matrix Name
Government	riskless	USTF	USTF
Industry	AAA	INAAA	INAAA
Industry	AA	INAA2.DAT	INAA2
Industry	AA-	INAA3.DAT	INAA3
Industry	A+	INA1.DAT	INA1
Industry	A	INA2.DAT	INA2
Industry	A-	INA3.DAT	INA3
Industry	BBB+	INBBB1.DAT	INBBB1
Industry	BBB	INBBB2.DAT	INBBB2
Industry	BBB-	INBBB3.DAT	INBBB3
Industry	BB+	INBB1.DAT	INBB1
Industry	BB	INBB2.DAT	INBB2
Industry	BB-	INBB3.DAT	INBB3
Industry	B+	INB1.DAT	INB1
Industry	B	INB2.DAT	INB2
Industry	B-	INB3.DAT	INB3
Bank	AAA	BNAAA.DAT	BNAAA
Bank	AA+/AA	BNAA12.DAT	BNAA12
Bank	A+	BNA1.DAT	BNA1
Bank	A	BNA2.DAT	BNA2
Bank	A-	BNA3.DAT	BNA3

Table 2.2. Data variables

them to variables SINAAA, SINAA2, etc. For illustrative purposes we display time series of the 1Y, 2Y, 3Y, 5Y, 7Y, and 10Y spreads for the curves INAAA, INA2, INBBB2, INBB2, INB2 in Figure 2.3, **Q**XFGseries .

We run the summary statistics to obtain information on the mean spreads. Our results, which can also be obtained with the **Q**mean command, are collected in Table 2.4, **Q**XFGmeans .

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	7Y	10Y
USTF	4.73	4.92	5.16	5.50	5.71	5.89	6.00	6.19	6.33
INAAA	5.10	5.26	5.51	5.82	6.04	6.21	6.35	6.52	6.70
INAA2	5.19	5.37	5.59	5.87	6.08	6.26	6.39	6.59	6.76
INAA3	5.25	-	5.64	5.92	6.13	6.30	6.43	6.63	6.81
INA1	5.32	5.50	5.71	5.99	6.20	6.38	6.51	6.73	6.90
INA2	5.37	5.55	5.76	6.03	6.27	6.47	6.61	6.83	7.00
INA3	-	-	5.84	6.12	6.34	6.54	6.69	6.91	7.09
INBBB1	5.54	5.73	5.94	6.21	6.44	6.63	6.78	7.02	7.19
INBBB2	5.65	5.83	6.03	6.31	6.54	6.72	6.86	7.10	7.27
INBBB3	5.83	5.98	6.19	6.45	6.69	6.88	7.03	7.29	7.52
INBB1	6.33	6.48	6.67	6.92	7.13	7.29	7.44	7.71	7.97
INBB2	6.56	6.74	6.95	7.24	7.50	7.74	7.97	8.34	8.69
INBB3	6.98	7.17	7.41	7.71	7.99	8.23	8.46	8.79	9.06
INB1	7.32	7.53	7.79	8.09	8.35	8.61	8.82	9.13	9.39
INB2	7.80	7.96	8.21	8.54	8.83	9.12	9.37	9.68	9.96
INB3	8.47	8.69	8.97	9.33	9.60	9.89	10.13	10.45	10.74
BNAAA	5.05	5.22	5.45	5.76	5.99	6.20	6.36	6.60	6.79
BNAA12	5.14	5.30	5.52	5.83	6.06	6.27	6.45	6.68	6.87
BNA1	5.22	5.41	5.63	5.94	6.19	6.39	6.55	6.80	7.00
BNA2	5.28	5.47	5.68	5.99	6.24	6.45	6.61	6.88	7.07
BNA3	5.36	5.54	5.76	6.07	6.32	6.52	6.68	6.94	7.13

Table 2.3. Long term mean for different USD yield curves
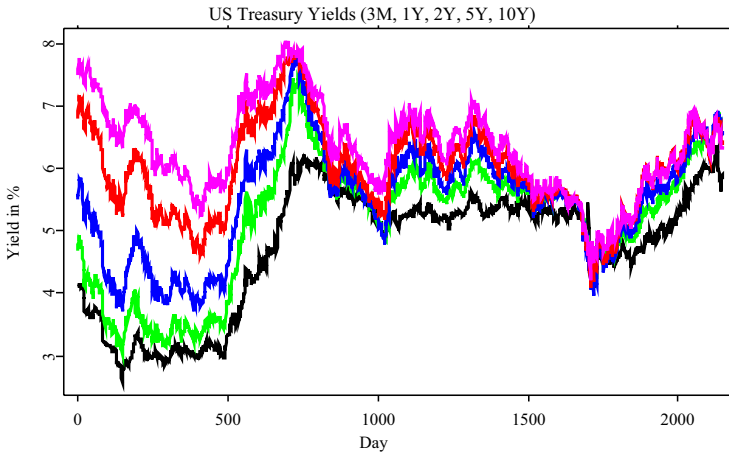
Figure 2.1. US Treasury Yields. ⚲XFGtreasury

Now we calculate the 1-day spread changes from the observed yields and store them to variables DASIN01AAA, etc. We run the ⚲descriptive routine to
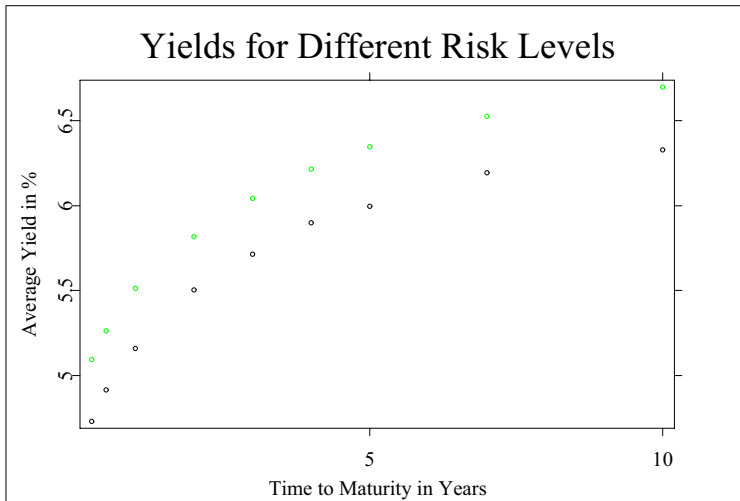
Figure 2.2. Averaged Yields. ◘XFGyields
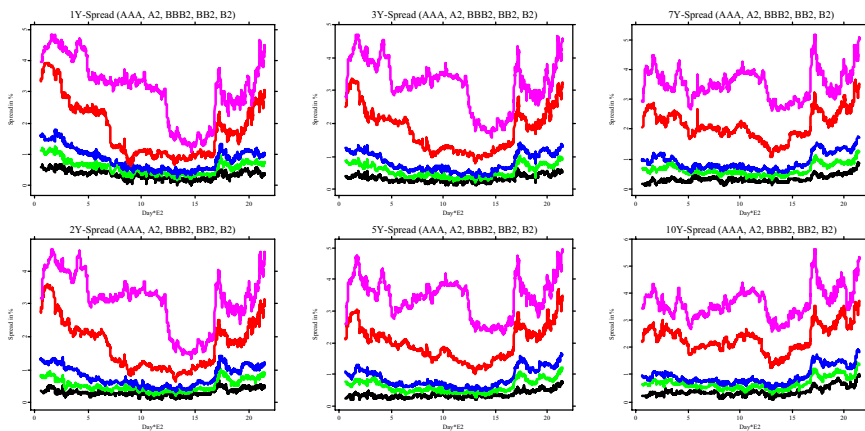
Figure 2.3. Credit Spreads. ◘XFGseries

calculate the first four moments of the distribution of absolute spread changes. Volatility as well as skewness and kurtosis for selected curves are displayed in Tables 2.5, 2.6 and 2.7.

◘XFGchange

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	7Y	10Y
INAAA	36	35	35	31	33	31	35	33	37
INAA2	45	45	43	37	37	36	40	39	44
INAA3	52	-	48	42	42	40	44	44	49
INA1	58	58	55	49	49	49	52	53	57
INA2	63	63	60	53	56	57	62	64	68
INA3	-	-	68	62	63	64	69	72	76
INBBB1	81	82	78	71	72	74	79	83	86
INBBB2	91	91	87	80	82	82	87	90	94
INBBB3	110	106	103	95	98	98	104	110	119
INBB1	160	156	151	142	141	140	145	151	164
INBB2	183	182	179	173	179	185	197	215	236
INBB3	225	225	225	221	228	233	247	259	273
INB1	259	261	263	259	264	271	282	294	306
INB2	306	304	305	304	311	322	336	348	363
INB3	373	377	380	382	389	400	413	425	441
BNAAA	41	39	38	33	35	35	41	43	47
BNAA12	50	47	45	40	42	42	49	52	56
BNA1	57	59	57	52	54	54	59	64	68
BNA2	64	65	62	57	59	60	65	71	75
BNA3	72	72	70	65	67	67	72	76	81

Table 2.4. Mean spread in basis points p.a.

For the variable `DASIN01AAA[,12]` (the 10 year AAA spreads) we demonstrate the output of the ⬛`descriptive` command in Table 2.8.

Finally we calculate 1-day relative spread changes and run the ⬛`descriptive` command. The results for the estimates of volatility, skewness and kurtosis are summarized in Tables 2.9, 2.10 and 2.11.   ⬛`XFGrelchange`

2.3.2 Discussion of Results

Time Development of Yields and Spreads: The time development of US treasury yields displayed in Figure 2.1 indicates that the yield curve was steeper at the beginning of the observation period and flattened in the second half. However, an inverse shape of the yield curve occurred hardly ever. The long term average of the US treasury yield curve, the lowest curve in Figure 2.2, also has an upward sloping shape.

The time development of the spreads over US treasury yields displayed in Figure 2.3 is different for different credit qualities. While there is a large variation of spreads for the speculative grades, the variation in the investment grade sector is much smaller. A remarkable feature is the significant spread

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	7Y	10Y
INAAA	4.1	3.5	3.3	2.3	2.4	2.2	2.1	2.2	2.5
INAA2	4.0	3.5	3.3	2.3	2.4	2.2	2.2	2.2	2.5
INAA3	4.0	-	3.3	2.2	2.3	2.2	2.2	2.2	2.5
INA1	4.0	3.7	3.3	2.3	2.4	2.2	2.2	2.2	2.6
INA2	4.1	3.7	3.3	2.4	2.4	2.1	2.2	2.3	2.5
INA3	-	-	3.4	2.4	2.4	2.2	2.2	2.3	2.6
INBBB1	4.2	3.6	3.2	2.3	2.3	2.2	2.1	2.3	2.6
INBBB2	4.0	3.5	3.4	2.3	2.4	2.1	2.2	2.3	2.6
INBBB3	4.2	3.6	3.5	2.4	2.5	2.2	2.3	2.5	2.9
INBB1	4.8	4.4	4.1	3.3	3.3	3.1	3.1	3.9	3.4
INBB2	4.9	4.6	4.5	3.8	3.8	3.8	3.7	4.3	4.0
INBB3	5.5	5.1	4.9	4.3	4.4	4.2	4.1	4.7	4.3
INB1	6.0	5.2	4.9	4.5	4.5	4.4	4.4	4.9	4.6
INB2	5.6	5.2	5.2	4.8	4.9	4.8	4.8	5.3	4.9
INB3	5.8	6.1	6.4	5.1	5.2	5.1	5.1	5.7	5.3
BNAAA	3.9	3.5	3.3	2.5	2.5	2.3	2.2	2.3	2.6
BNAA12	5.4	3.6	3.3	2.4	2.3	2.2	2.1	2.3	2.6
BNA1	4.1	3.7	3.2	2.1	2.2	2.1	2.0	2.2	2.6
BNA2	3.8	3.5	3.1	2.3	2.2	2.0	2.1	2.2	2.5
BNA3	3.8	3.5	3.2	2.2	2.2	2.1	2.1	2.2	2.5

Table 2.5.   volatility for absolute spread changes in basis points p.a.

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	10Y
INAAA	0.1	0.0	-0.1	0.6	0.5	0.0	-0.5	0.6
INAA2	0.0	-0.2	0.0	0.4	0.5	-0.1	-0.2	0.3
INA2	0.0	-0.3	0.1	0.2	0.4	0.1	-0.1	0.4
INBBB2	0.2	0.0	0.2	1.0	1.1	0.5	0.5	0.9
INBB2	-0.2	-0.5	-0.4	-0.3	0.3	0.5	0.4	-0.3

Table 2.6.   Skewness for absolute 1-day spread changes (in σ^3).

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	10Y
INAAA	12.7	6.0	8.1	10.1	16.8	9.1	11.2	12.8
INAA2	10.5	6.4	7.8	10.1	15.8	7.8	9.5	10.0
INA2	13.5	8.5	9.2	12.3	18.2	8.2	9.4	9.8
INBBB2	13.7	7.0	9.9	14.5	21.8	10.5	13.9	14.7
INBB2	11.2	13.0	11.0	15.8	12.3	13.2	11.0	11.3

Table 2.7.   Kurtosis for absolute spread changes (in σ^4).

increase for all credit qualities in the last quarter of the observation period which coincides with the emerging market crises in the late 90s. The term

```
===========================================================
Variable 10Y
===========================================================

Mean              0.000354147
Std.Error         0.0253712      Variance      0.000643697

Minimum             -0.18        Maximum             0.2
Range                0.38

Lowest cases                     Highest cases
      1284:         -0.18              1246:         0.14
      1572:         -0.14              1283:         0.14
      1241:         -0.13              2110:         0.19
      1857:         -0.11              1062:         0.19
       598:          -0.1              2056:          0.2

Median                0
25% Quartile        -0.01        75% Quartile        0.01

Skewness          0.609321       Kurtosis         9.83974

Observations                     2146
Distinct observations              75

Total number of {-Inf,Inf,NaN}    0
===========================================================
```

Table 2.8. Output of ◖descriptive for the 10 years AAA spread.

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	7Y	10Y
INAAA	36.0	19.2	15.5	8.9	8.4	8.0	6.4	7.8	10.4
INAA2	23.5	13.1	11.2	7.2	7.4	6.4	5.8	6.2	7.6
INAA3	13.4	-	9.0	5.8	6.2	5.3	5.0	5.8	6.4
INA1	13.9	9.2	7.7	5.7	5.6	4.7	4.5	4.6	5.7
INA2	11.5	8.1	7.1	5.1	4.9	4.3	4.0	4.0	4.5
INA3	-	-	6.4	4.6	4.3	3.8	3.5	3.5	4.1
INBBB1	8.1	6.0	5.4	3.9	3.7	3.3	3.0	3.2	3.8
INBBB2	7.0	5.3	5.0	3.3	3.3	2.9	2.8	2.9	3.3
INBBB3	5.7	4.7	4.4	3.2	3.0	2.7	2.5	2.6	2.9
INBB1	4.3	3.8	3.4	2.5	2.4	2.2	2.1	2.5	2.2
INBB2	3.7	3.3	3.0	2.2	2.1	2.0	1.8	2.0	1.7
INBB3	3.2	2.8	2.5	2.0	1.9	1.8	1.6	1.8	1.5
INB1	3.0	2.4	2.1	1.7	1.7	1.6	1.5	1.6	1.5
INB2	2.3	2.1	1.9	1.6	1.6	1.5	1.4	1.5	1.3
INB3	1.8	2.2	2.3	1.3	1.3	1.2	1.2	1.3	1.1
BNAAA	37.0	36.6	16.9	9.8	9.0	8.2	6.1	5.9	6.5
BNAA12	22.8	9.7	8.3	7.0	6.3	5.8	4.6	4.8	5.5
BNA1	36.6	10.1	7.9	5.6	4.8	4.4	3.8	3.9	4.4
BNA2	17.8	8.0	6.6	4.5	4.1	3.6	3.4	3.3	3.7
BNA3	9.9	6.9	5.6	3.7	3.6	3.3	3.1	3.1	3.4

Table 2.9. Volatility for relative spread changes in %

structure of the long term averages of the rating-specific yield curves is also normal. The spreads over the benchmark curve increase with decreasing credit quality.

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	10Y
INAAA	2.3	4.6	4.3	2.2	2.3	2.1	0.6	4.6
INAA2	5.4	2.6	3.7	1.6	2.0	0.6	0.8	1.8
INA2	7.6	1.5	1.2	0.9	1.6	0.8	0.9	0.8
INBBB2	5.5	0.7	0.8	0.8	1.4	0.8	0.7	0.8
INBB2	0.8	0.4	0.6	0.3	0.4	0.5	0.3	-0.2

Table 2.10. Skewness for relative spread changes (in σ^3).

Curve	3M	6M	1Y	2Y	3Y	4Y	5Y	10Y
INAAA	200.7	54.1	60.1	27.8	28.3	33.9	16.8	69.3
INAA2	185.3	29.5	60.5	22.1	27.4	11.0	17.5	23.0
INA2	131.1	22.1	18.0	13.9	26.5	16.4	18.5	13.9
INBBB2	107.1	13.9	16.9	12.0	20.0	14.0	16.6	16.7
INBB2	16.3	11.9	12.9	12.4	11.0	10.1	10.2	12.0

Table 2.11. Kurtosis for relative spread changes (in σ^4).

Mean Spread: The term structure of the long term averages of the rating-specific yield curves, which is displayed in Figure 2.3, is normal (see also Table 2.4). The spreads over the benchmark curve increase with decreasing credit quality. For long maturities the mean spreads are larger than for intermediate maturities as expected. However, for short maturities the mean spreads are larger compared with intermediate maturities.

Volatility: The results for the volatility for absolute 1-day spread changes in basis points p.a. are listed in Table 2.5. From short to intermediate maturities the volatilities decrease. For long maturities a slight volatility increase can be observed compared to intermediate maturities. For equal maturities volatility is constant over the investment grade ratings, while for worse credit qualities a significant increase in absolute volatility can be observed. Volatility for relative spread changes is much larger for short maturities than for intermediate and long maturities. As in the case of absolute spread changes, a slight volatility increase exists for the transition from intermediate to long maturities. Since absolute spreads increase more strongly with decreasing credit quality than absolute spread volatility, relative spread volatility decreases with decreasing credit quality (see Table 2.9).

Skewness: The results for absolute 1-day changes (see Table 2.6) are all close to zero, which indicates that the distribution of changes is almost symmetric. The corresponding distribution of relative changes should have a positive skewness, which is indeed the conclusion from the results in Table 2.10.

Kurtosis: The absolute 1-day changes lead to a kurtosis, which is significantly larger than 3 (see Table 2.6). Thus, the distribution of absolute changes is leptokurtic. There is no significant dependence on credit quality or maturity. The distribution of relative 1-day changes is also leptokurtic (see Table 2.10). The deviation from normality increases with decreasing credit quality and decreasing maturity.

We visualize symmetry and leptokursis of the distribution of absolute spread changes for the `INAAA 10Y` data in Figure 2.4, where we plot the empirical distribution of absolute spreads around the mean spread in an averaged shifted histogram and the normal distribution with the variance estimated from historical data.

<div align="right">◘ XFGdist</div>
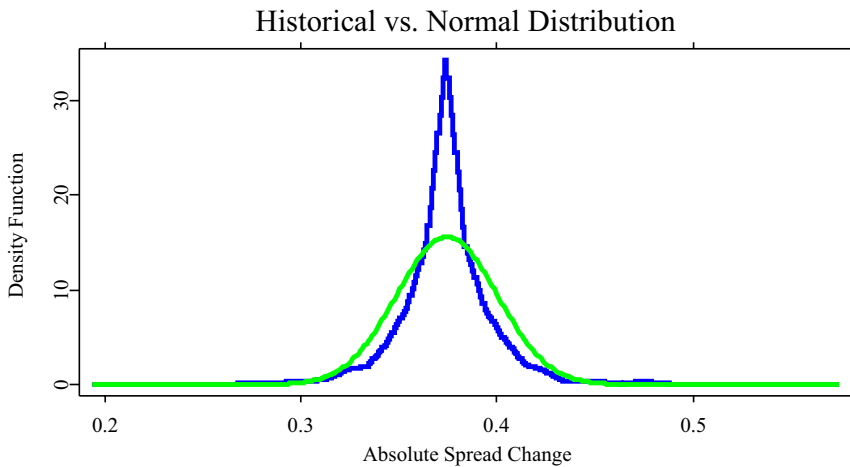
Figure 2.4. Historical distribution and estimated normal distribution. ◘ XFGdist

We note that by construction the area below both curves is normalized to one. We calculate the 1%, 10%, 90% and 99% quantiles of the spread distribution with the ◘ `quantile` command. Those quantiles are popular in market risk management. For the data used to generate Figure 2.4 the results are 0.30%, 0.35%, 0.40%, and 0.45%, respectively. The corresponding quantiles of the plotted normal distribution are 0.31%, 0.34%, 0.41%, 0.43%. The differences are less obvious than the difference in the shape of the distributions. However, in a portfolio with different financial instruments, which is exposed to different risk factors with different correlations, the difference

in the shape of the distribution can play an important role. That is why a simple variance-covariance approach, J.P. Morgan (1996) and Kiesel et al. (1999), seems not adequate to capture spread risk.

2.4 Historical Simulation and Value at Risk

We investigate the behavior of a fictive zero-bond of a given credit quality with principal 1 USD, which matures after T years. In all simulations $t = 0$ denotes the beginning and $t = T$ the end of the lifetime of the zero-bond. The starting point of the simulation is denoted by t_0, the end by t_1. The observation period, i.e., the time window investigated, consists of $N \geq 1$ trading days and the holding period of $h \geq 1$ trading days. The confidence level for the VaR is $\alpha \in [0, 1]$. At each point in time $0 \leq t \leq t_1$ the risky yields $R_i(t)$ (full yield curve) and the riskless treasury yields $B_i(t)$ (benchmark curve) for any time to maturity $0 < T_1 < \cdots < T_n$ are contained in our data set for $1 \leq i \leq n$, where n is the number of different maturities. The corresponding spreads are defined by $S_i(t) = R_i(t) - B_i(t)$ for $1 \leq i \leq n$.

In the following subsections 2.4.1 to 2.4.5 we specify different variants of the historical simulation method which we use for estimating the distribution of losses from the zero-bond position. The estimate for the distribution of losses can then be used to calculate the quantile-based risk measure Value-at-Risk. The variants differ in the choice of risk factors, i.e., in our case the components of the historical yield time series. In Section 2.6 we describe how the VaR estimation is carried out with XploRe commands provided that the loss distribution has been estimated by means of one of the methods introduced and can be used as an input variable.

2.4.1 Risk Factor: Full Yield

1. Basic Historical Simulation:

We consider a historical simulation, where the risk factors are given by the full yield curve, $R_i(t)$ for $i = 1, \ldots, n$. The yield $R(t, T - t)$ at time $t_0 \leq t \leq t_1$ for the remaining time to maturity $T - t$ is determined by means of linear interpolation from the adjacent values $R_i(t) = R(t, T_i)$ and $R_{i+1}(t) = R(t, T_{i+1})$ with $T_i \leq T - t < T_{i+1}$ (for reasons of simplicity we do not consider remaining times to maturity $T - t < T_1$ and $T - t > T_n$):

$$R(t, T - t) = \frac{[T_{i+1} - (T - t)]R_i(t) + [(T - t) - T_i]R_{i+1}(t)}{T_{i+1} - T_i} . \qquad (2.1)$$

The present value of the bond $PV(t)$ at time t can be obtained by discounting,

$$PV(t) = \frac{1}{\left[1 + R(t, T - t)\right]^{T-t}}, \qquad t_0 \leq t \leq t_1. \tag{2.2}$$

In the historical simulation the relative risk factor changes

$$\Delta_i^{(k)}(t) = \frac{R_i\left(t - k/N\right) - R_i\left(t - (k+h)/N\right)}{R_i\left(t - (k+h)/N\right)}, \qquad 0 \leq k \leq N - 1, \tag{2.3}$$

are calculated for $t_0 \leq t \leq t_1$ and each $1 \leq i \leq n$. Thus, for each scenario k we obtain a new fictive yield curve at time $t + h$, which can be determined from the observed yields and the risk factor changes,

$$R_i^{(k)}(t + h) = R_i(t)\left[1 + \Delta_i^{(k)}(t)\right], \qquad 1 \leq i \leq n, \tag{2.4}$$

by means of linear interpolation. This procedure implies that the distribution of risk factor changes is stationary between $t - (N - 1 + h)/N$ and t. Each scenario corresponds to a drawing from an identical and independent distribution, which can be related to an i.i.d. random variable $\varepsilon_i(t)$ with variance one via

$$\Delta_i(t) = \sigma_i \varepsilon_i(t). \tag{2.5}$$

This assumption implies homoscedasticity of the volatility of the risk factors, i.e., a constant volatility level within the observation period. If this were not the case, different drawings would originate from different underlying distributions. Consequently, a sequence of historically observed risk factor changes could not be used for estimating the future loss distribution.

In analogy to (2.1) for time $t + h$ and remaining time to maturity $T - t$ one obtains

$$R^{(k)}(t + h, T - t) = \frac{[T_{i+1} - (T - t)]R_i^{(k)}(t) + [(T - t) - T_i]R_{i+1}^{(k)}(t)}{T_{i+1} - T_i}$$

for the yield. With (2.2) we obtain a new fictive present value at time $t + h$:

$$PV^{(k)}(t + h) = \frac{1}{\left[1 + R^{(k)}(t + h, T - t)\right]^{T-t}}. \tag{2.6}$$

In this equation we neglected the effect of the shortening of the time to maturity in the transition from t to $t + h$ on the present value. Such an approximation should be refined for financial instruments whose time to maturity/time to expiration is of the order of h, which is not relevant for the constellations investigated in the following.

Now the fictive present value $PV^{(k)}(t+h)$ is compared with the present value for unchanged yield $R(t+h, T-t) = R(t, T-t)$ for each scenario k (here the remaining time to maturity is not changed, either).

$$PV(t+h) = \frac{1}{\{1 + R(t+h, T-t)\}^{T-t}}. \tag{2.7}$$

The loss occurring is

$$L^{(k)}(t+h) = PV(t+h) - PV^{(k)}(t+h) \qquad 0 \le k \le N-1, \tag{2.8}$$

i.e., losses in the economic sense are positive while profits are negative. The VaR is the loss which is not exceeded with a probability α and is estimated as the $[(1-\alpha)N + 1]$-th-largest value in the set

$$\{L^{(k)}(t+h) \mid 0 \le k \le N-1\}.$$

This is the $(1-\alpha)$-quantile of the corresponding empirical distribution.

2. Mean Adjustment:

A refined historical simulation includes an adjustment for the average of those relative changes in the observation period which are used for generating the scenarios according to (2.3). If for fixed $1 \le i \le n$ the average of relative changes $\Delta_i^{(k)}(t)$ is different from 0, a trend is projected from the past to the future in the generation of fictive yields in (2.4). Thus the relative changes are corrected for the mean by replacing the relative change $\Delta_i^{(k)}(t)$ with $\Delta_i^{(k)}(t) - \overline{\Delta}_i(t)$ for $1 \le i \le n$ in (2.4):

$$\overline{\Delta}_i(t) = \frac{1}{N} \sum_{k=0}^{N-1} \Delta_i^{(k)}(t), \tag{2.9}$$

This mean correction is presented in Hull (1998).

3. Volatility Updating:

An important variant of historical simulation uses volatility updating Hull (1998). At each point in time t the exponentially weighted volatility of relative historical changes is estimated for $t_0 \le t \le t_1$ by

$$\sigma_i^2(t) = (1-\gamma) \sum_{k=0}^{N-1} \gamma^k \{\Delta_i^{(k)}(t)\}^2, \qquad 1 \le i \le n. \tag{2.10}$$

The parameter $\gamma \in [0, 1]$ is a decay factor, which must be calibrated to generate a best fit to empirical data. The recursion formula

$$\sigma_i^2(t) = (1-\gamma)\sigma_i^2(t - 1/N) + \gamma\{\Delta_i^{(0)}(t)\}^2, \qquad 1 \le i \le n, \tag{2.11}$$

is valid for $t_0 \leq t \leq t_1$. The idea of volatility updating consists in adjusting the historical risk factor changes to the present volatility level. This is achieved by a renormalization of the relative risk factor changes from (2.3) with the corresponding estimation of volatility for the observation day and a multiplication with the estimate for the volatility valid at time t. Thus, we calculate the quantity

$$\delta_i^{(k)}(t) = \sigma_i(t) \cdot \frac{\Delta_i^{(k)}(t)}{\sigma_i(t - (k+h)/N)}, \qquad 0 \leq k \leq N - 1. \tag{2.12}$$

In a situation, where risk factor volatility is heteroscedastic and, thus, the process of risk factor changes is not stationary, volatility updating cures this violation of the assumptions made in basic historical simulation, because the process of re-scaled risk factor changes $\Delta_i(t)/\sigma_i(t))$ is stationary. For each k these renormalized relative changes are used in analogy to (2.4) for the determination of fictive scenarios:

$$R_i^{(k)}(t+h) = R_i(t)\big\{1 + \delta_i^{(k)}(t)\big\}, \qquad 1 \leq i \leq n, \tag{2.13}$$

The other considerations concerning the VaR calculation in historical simulation remain unchanged.

4. Volatility Updating and Mean Adjustment:

Within the volatility updating framework, we can also apply a correction for the average change according to 2.4.1(2). For this purpose, we calculate the average

$$\overline{\delta}_i(t) = \frac{1}{N} \sum_{k=0}^{N-1} \delta_i^{(k)}(t), \tag{2.14}$$

and use the adjusted relative risk factor change $\delta_i^{(k)}(t) - \overline{\delta}_i(t)$ instead of $\delta_i^{(k)}(t)$ in (2.13).

2.4.2 Risk Factor: Benchmark

In this subsection the risk factors are relative changes of the benchmark curve instead of the full yield curve. This restriction is adequate for quantifying general market risk, when there is no need to include spread risk. The risk factors are the yields $B_i(t)$ for $i = 1, \ldots, n$. The yield $B(t, T - t)$ at time t for remaining time to maturity $T - t$ is calculated similarly to (2.1) from

adjacent values by linear interpolation,

$$B(t, T-t) = \frac{\{T_{i+1} - (T-t)\}B_i(t) + \{(T-t) - T_i\}B_{i+1}(t)}{T_{i+1} - T_i}. \qquad (2.15)$$

The generation of scenarios and the interpolation of the fictive benchmark curve is carried out in analogy to the procedure for the full yield curve. We use

$$\Delta_i^{(k)}(t) = \frac{B_i\big(t - k/N\big) - B_i\big(t - (k+h)/N\big)}{B_i\big(t - (k+h)/N\big)}, \quad 0 \le k \le N-1, \qquad (2.16)$$

and

$$B_i^{(k)}(t+h) = B_i(t)\big[1 + \Delta_i^{(k)}(t)\big], \qquad 1 \le i \le n. \qquad (2.17)$$

Linear interpolation yields

$$B^{(k)}(t+h, T-t) = \frac{\{T_{i+1} - (T-t)\}B_i^{(k)}(t) + \{(T-t) - T_i\}B_{i+1}^{(k)}(t)}{T_{i+1} - T_i}.$$

In the determination of the fictive full yield we now assume that the spread remains unchanged within the holding period. Thus, for the k-th scenario we obtain the representation

$$R^{(k)}(t+h, T-t) = B^{(k)}(t+h, T-t) + S(t, T-t), \qquad (2.18)$$

which is used for the calculation of a new fictive present value and the corresponding loss. With this choice of risk factors we can introduce an adjustment for the average relative changes or/and volatility updating in complete analogy to the four variants described in the preceding subsection.

2.4.3 Risk Factor: Spread over Benchmark Yield

When we take the view that risk is only caused by spread changes but not by changes of the benchmark curve, we investigate the behavior of the spread risk factors $S_i(t)$ for $i = 1, \ldots, n$. The spread $S(t, T-t)$ at time t for time to maturity $T-t$ is again obtained by linear interpolation. We now use

$$\Delta_i^{(k)}(t) = \frac{S_i\big(t - k/N\big) - S_i\big(t - (k+h)/N\big)}{S_i\big(t - (k+h)/N\big)}, \quad 0 \le k \le N-1, \qquad (2.19)$$

and

$$S_i^{(k)}(t+h) = S_i(t)\big\{1 + \Delta_i^{(k)}(t)\big\}, \qquad 1 \le i \le n. \qquad (2.20)$$

Here, linear interpolation yields

$$S^{(k)}(t+h, T-t) = \frac{\{T_{i+1} - (T-t)\}S_i^{(k)}(t) + \{(T-t) - T_i\}S_{i+1}^{(k)}(t)}{T_{i+1} - T_i}.$$

Thus, in the determination of the fictive full yield the benchmark curve is considered deterministic and the spread stochastic. This constellation is the opposite of the constellation in the preceding subsection. For the k-th scenario one obtains

$$R^{(k)}(t+h, T-t) = B(t, T-t) + S^{(k)}(t+h, T-t). \qquad (2.21)$$

In this context we can also work with adjustment for average relative spread changes and volatility updating.

2.4.4 Conservative Approach

In the conservative approach we assume full correlation between risk from the benchmark curve and risk from the spread changes. In this worst case scenario we add (ordered) losses, which are calculated as in the two preceding sections from each scenario. From this loss distribution the VaR is determined.

2.4.5 Simultaneous Simulation

Finally, we consider simultaneous relative changes of the benchmark curve and the spreads. For this purpose (2.18) and (2.21) are replaced with

$$R^{(k)}(t+h, T-t) = B^{(k)}(t+h, T-t) + S^{(k)}(t, T-t), \qquad (2.22)$$

where, again, corrections for average risk factor changes or/and volatility updating can be added. We note that the use of relative risk factor changes is the reason for different results of the variants in subsection 2.4.1 and this subsection.

2.5 Mark-to-Model Backtesting

A backtesting procedure compares the VaR prediction with the observed loss. In a mark-to-model backtesting the observed loss is determined by calculation of the present value before and after consideration of the actually observed risk factor changes. For $t_0 \leq t \leq t_1$ the present value at time $t+h$ is

calculated with the yield $R(t + h, T - t)$, which is obtained from observed data for $R_i(t + h)$ by linear interpolation, according to

$$PV(t) = \frac{1}{\left\{1 + R(t + h, T - t)\right\}^{T-t}}. \tag{2.23}$$

This corresponds to a loss $L(t) = PV(t) - PV(t + h)$, where, again, the shortening of the time to maturity is not taken into account.

The different frameworks for the VaR estimation can easily be integrated into the backtesting procedure. When we, e.g., only consider changes of the benchmark curve, $R(t + h, T - t)$ in (2.23) is replaced with $B(t + h, T - t) + S(t, T - t)$. On an average $(1 - \alpha) \cdot 100$ per cent of the observed losses in a given time interval should exceed the corresponding VaR (outliers). Thus, the percentage of observed losses is a measure for the predictive power of historical simulation.

2.6 VaR Estimation and Backtesting

In this section we explain, how a VaR can be calculated and a backtesting can be implemented with the help of XploRe routines. We present numerical results for the different yield curves. The VaR estimation is carried out with the help of the ◘`VaRest` command. The ◘`VaRest` command calculates a VaR for historical simulation, if one specifies the method parameter as "EDF" (empirical distribution function). However, one has to be careful when specifying the sequence of asset returns which are used as input for the estimation procedure. If one calculates zero-bond returns from relative risk factor changes (interest rates or spreads) the complete empirical distribution of the profits and losses must be estimated anew for each day from the N relative risk factor changes, because the profit/loss observations are not identical with the risk factor changes.

For each day the N profit/loss observations generated with one of the methods described in subsections 2.4.1 to 2.4.5 are stored to a new row in an array `PL`. The actual profit and loss data from a mark-to-model calculation for holding period h are stored to a one-column-vector `MMPL`. It is not possible to use a continuous sequence of profit/loss data with overlapping time windows for the VaR estimation. Instead the ◘`VaRest` command must be called separately for each day. The consequence is that the data the ◘`VaRest` command operates on consists of a row of $N + 1$ numbers: N profit/loss values contained in the vector `(PL[t,])'`, which has one column and N rows followed by the actual mark-to-model profit or loss `MMPL[t,1]` within holding

period h in the last row. The procedure is implemented in the quantlet `XFGpl` which can be downloaded from quantlet download page of this book.
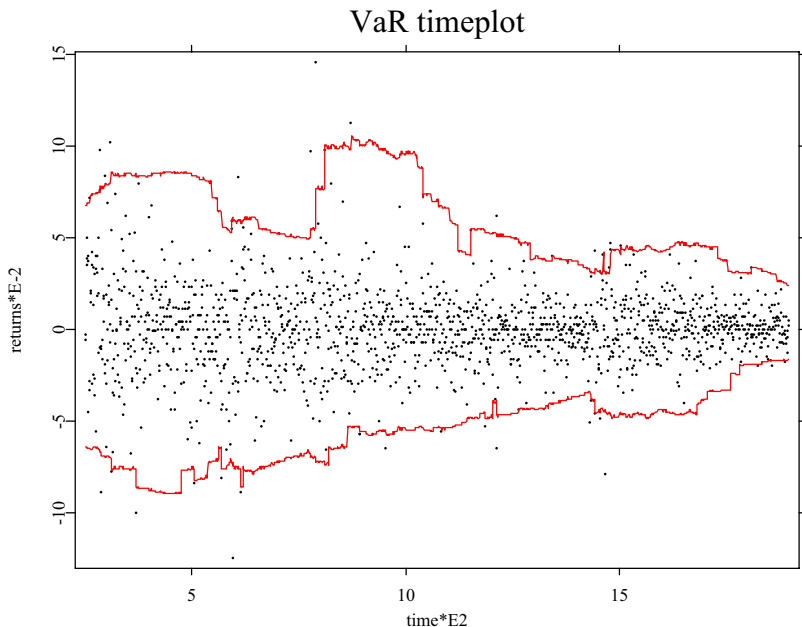
Figure 2.5.   VaR time plot basic historical simulation.
🔵 `XFGtimeseries`

The result is displayed for the `INAAA` curve in Figures. 2.5 (basic historical simulation) and 2.6 (historical simulation with volatility updating).   The time plots allow for a quick detection of violations of the VaR prediction. A striking feature in the basic historical simulation with the full yield curve as risk factor is the platform-shaped VaR prediction, while with volatility updating the VaR prediction decays exponentially after the occurrence of peak events in the market data. This is a consequence of the exponentially weighted historical volatility in the scenarios. The peak VaR values are much larger for volatility updating than for the basic historical simulation.

In order to find out, which framework for VaR estimation has the best predictive power, we count the number of violations of the VaR prediction and divide it by the number of actually observed losses. We use the 99% quantile, for which we would expect an violation rate of 1% for an optimal VaR estimator. The history used for the drawings of the scenarios consists of $N = 250$ days, and the holding period is $h = 1$ day. For the volatility updating we use
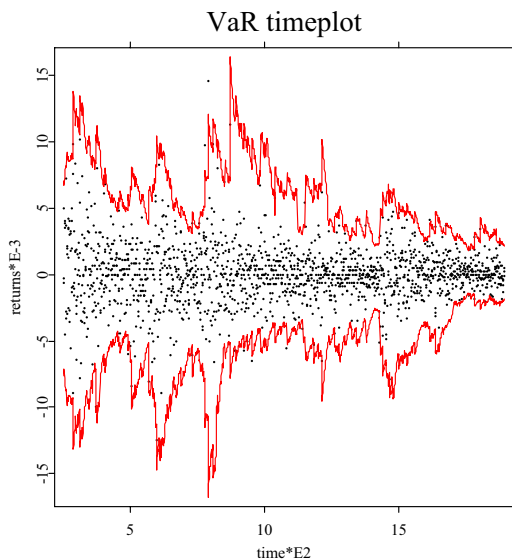
VaR timeplot

Figure 2.6. VaR time plot historical simulation with volatility updating. ⬛Q XFGtimeseries2

a decay factor of $\gamma = 0.94$, J.P. Morgan (1996). For the simulation we assume that the synthetic zero-bond has a remaining time to maturity of 10 years at the beginning of the simulations. For the calculation of the first scenario of a basic historical simulation $N + h - 1$ observations are required. A historical simulation with volatility updating requires $2(N + h - 1)$ observations preceding the trading day the first scenario refers to. In order to allow for a comparison between different methods for the VaR calculation, the beginning of the simulations is $t_0 = [2(N + h - 1)/N]$. With these simulation parameters we obtain 1646 observations for a zero-bond in the industry sector and 1454 observations for a zero-bond in the banking sector.

In Tables 2.12 to 2.14 we list the percentage of violations for all yield curves and the four variants of historical simulation V1 to V4 (V1 = Basic Historical Simulation; V2 = Basic Historical Simulation with Mean Adjustment; V3 = Historical Simulation with Mean Adjustment; V4 = Historical Simulation with Volatility Updating and Mean Adjustment). In the last row we display the average of the violations of all curves. Table 2.12 contains the results for the simulation with relative changes of the full yield curves and of the yield spreads over the benchmark curve as risk factors. In Table 2.13 the risk factors are changes of the benchmark curves. The violations in the

conservative approach and in the simultaneous simulation of relative spread and benchmark changes are listed in Table 2.14.

⊙ XFGexc

Curve	Full yield				Spread curve			
	V1	**V2**	**V3**	**V4**	**V1**	**V2**	**V3**	**V4**
INAAA	1,34	1,34	1,09	1,28	1,34	1,34	1,34	1,34
INAA2	1,34	1,22	1,22	1,22	1,46	1,52	1,22	1,22
INAA3	1,15	1,22	1,15	1,15	1,09	1,09	0,85	0,91
INA1	1,09	1,09	1,46	1,52	1,40	1,46	1,03	1,09
INA2	1,28	1,28	1,28	1,28	1,15	1,15	0,91	0,91
INA3	1,22	1,22	1,15	1,22	1,15	1,22	1,09	1,15
INBBB1	1,28	1,22	1,09	1,15	1,46	1,46	1,40	1,40
INBBB2	1,09	1,15	0,91	0,91	1,28	1,28	0,91	0,91
INBBB3	1,15	1,15	1,09	1,09	1,34	1,34	1,46	1,52
INBB1	1,34	1,28	1,03	1,03	1,28	1,28	0,97	0,97
INBB2	1,22	1,22	1,22	1,34	1,22	1,22	1,09	1,09
INBB3	1,34	1,28	1,28	1,22	1,09	1,28	1,09	1,09
INB1	1,40	1,40	1,34	1,34	1,52	1,46	1,09	1,03
INB2	1,52	1,46	1,28	1,28	1,34	1,40	1,15	1,15
INB3	1,40	1,40	1,15	1,15	1,46	1,34	1,09	1,15
BNAAA	1,24	1,38	1,10	1,10	0,89	0,89	1,03	1,31
BNAA1/2	1,38	1,24	1,31	1,31	1,03	1,10	1,38	1,38
BNA1	1,03	1,03	1,10	1,17	1,03	1,10	1,24	1,24
BNA2	1,24	1,31	1,24	1,17	0,76	0,83	1,03	1,03
BNA3	1,31	1,24	1,17	1,10	1,03	1,10	1,24	1,17
Average	**1,27**	**1,25**	**1,18**	**1,20**	**1,22**	**1,24**	**1,13**	**1,15**

Table 2.12. Violations full yield and spread curve (in %)

Curve	V1	V2	V3	V4
INAAA, INAA2, INAA3, INA1, INA2, INA3, INBBB1, INBBB2, INBBB3, INBB1, INBB2, INBB3, INB1, INB2, INB3	1,52	1,28	1,22	1,15
BNAAA, BNAA1/2, BNA1, BNA2, BNA3	1,72	1,44	1,17	1,10
Average	**1,57**	**1,32**	**1,20**	**1,14**

Table 2.13. Violations benchmark curve (in %)

	conservative approach				simultaneous simulation			
Curve	**V1**	**V2**	**V3**	**V4**	**V1**	**V2**	**V3**	**V4**
INAAA	0,24	0,24	0,30	0,30	1,22	1,28	0,97	1,03
INAA2	0,24	0,30	0,36	0,30	1,22	1,28	1,03	1,15
INAA3	0,43	0,36	0,30	0,30	1,22	1,15	1,09	1,09
INA1	0,36	0,43	0,55	0,55	1,03	1,03	1,03	1,09
INA2	0,49	0,43	0,49	0,49	1,34	1,28	0,97	0,97
INA3	0,30	0,36	0,30	0,30	1,22	1,15	1,09	1,09
INBBB1	0,43	0,49	0,36	0,36	1,09	1,09	1,03	1,03
INBBB2	0,49	0,49	0,30	0,30	1,03	1,03	0,85	0,79
INBBB3	0,30	0,30	0,36	0,36	1,15	1,22	1,03	1,03
INBB1	0,36	0,30	0,43	0,43	1,34	1,34	1,03	0,97
INBB2	0,43	0,36	0,43	0,43	1,40	1,34	1,15	1,09
INBB3	0,30	0,30	0,36	0,36	1,15	1,15	0,91	0,91
INB1	0,43	0,43	0,43	0,43	1,34	1,34	0,91	0,97
INB2	0,30	0,30	0,30	0,30	1,34	1,34	0,97	1,03
INB3	0,30	0,30	0,36	0,30	1,46	1,40	1,22	1,22
BNAAA	0,62	0,62	0,48	0,48	1,31	1,31	1,10	1,03
BNAA1/2	0,55	0,55	0,55	0,48	1,24	1,31	1,10	1,17
BNA1	0,62	0,62	0,55	0,55	0,96	1,03	1,10	1,17
BNA2	0,55	0,62	0,69	0,69	0,89	1,96	1,03	1,03
BNA3	0,55	0,55	0,28	0,28	1,38	1,31	1,03	1,10
Average	**0,41**	**0,42**	**0,41**	**0,40**	**1,22**	**1,22**	**1,03**	**1,05**

Table 2.14.   Violations in the conservative approach and simultaneous simulation(in %)

2.7 P-P Plots

The evaluation of the predictive power across all possible confidence levels $\alpha \in [0, 1]$ can be carried out with the help of a transformation of the empirical distribution $\{L^{(k)} \mid 0 \leq k \leq N - 1\}$. If F is the true distribution function of the loss L within the holding period h, then the random quantity $F(L)$ is (approximately) uniformly distributed on $[0, 1]$. Therefore we check the values $F_e\big[L(t)\big]$ for $t_0 \leq t \leq t_1$, where F_e is the empirical distribution. If the prediction quality of the model is adequate, these values should not differ significantly from a sample with size $250\,(t_1 - t_0 + 1)$ from a uniform distribution on $[0, 1]$.

The P-P plot of the transformed distribution against the uniform distribution (which represents the distribution function of the transformed empirical distribution) should therefore be located as closely to the main diagonal as possible. The mean squared deviation from the uniform distribution (MSD) summed over all quantile levels can serve as an indicator of the predictive

power of a quantile-based risk measure like VaR. The ⬛XFGpp quantlet creates a P-P plot and calculates the MSD indicator.

2.8 Q-Q Plots

With a quantile plot (Q-Q plot) it is possible to visualize whether an ordered sample is distributed according to a given distribution function. If, e.g., a sample is normally distributed, the plot of the empirical quantiles vs. the quantiles of a normal distribution should result in an approximately linear plot. Q-Q plots vs. a normal distribution can be generated with the following command:

⬛ VaRqqplot (matrix(N,1)|MMPL,VaR,opt)

2.9 Discussion of Simulation Results

In Figure 2.7 the P-P plots for the historical simulation with the full yield curve (INAAA) as risk factor are displayed for the different variants of the simulation. From the P-P plots it is apparent that mean adjustment significantly improves the predictive power in particular for intermediate confidence levels (i.e., for small risk factor changes).

Figure 2.8 displays the P-P plots for the same data set and the basic historical simulation with different choices of risk factors. A striking feature is the poor predictive power for a model with the spread as risk factor. Moreover, the over-estimation of the risk in the conservative approach is clearly reflected by a sine-shaped function, which is superposed on the ideal diagonal function.

In Figs. 2.9 and 2.10 we show the Q-Q plots for basic historic simulation and volatility updating using the INAAA data set and the full yield curve as risk factors. A striking feature of all Q-Q plots is the deviation from linearity (and, thus, normality) for extreme quantiles. This observation corresponds to the leptokurtic distributions of time series of market data changes (e.g. spread changes as discussed in section 2.3.2).

2.9.1 Risk Factor: Full Yield

The results in Table 2.12 indicate a small under-estimation of the actually observed losses. While volatility updating leads to a reduction of violations,
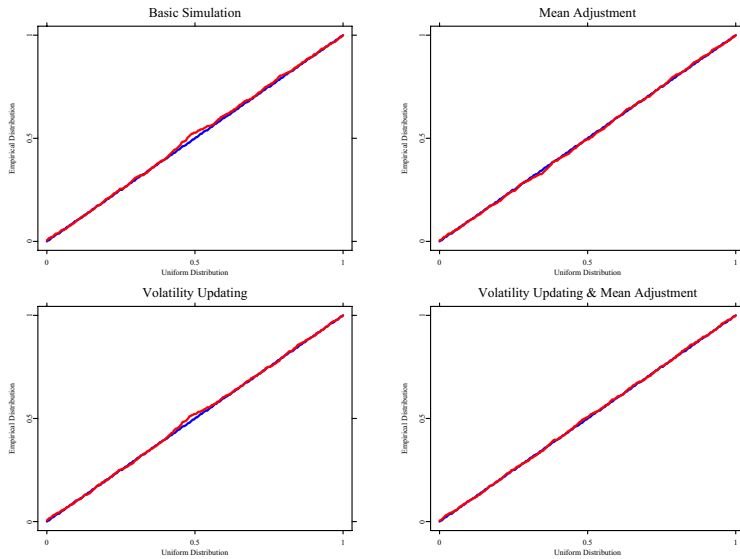
Figure 2.7. P-P Plots variants of the simulation. ⓠXFGpp

this effect is not clearly recognizable for the mean adjustment. The positive results for volatility updating are also reflected in the corresponding mean squared deviations in Table 2.15. Compared with the basic simulation, the model quality can be improved. There is also a positive effect of the mean adjustment.

2.9.2 Risk Factor: Benchmark

The results for the number of violations in Table 2.13 and the mean squared deviations in Table 2.16 are comparable to the analysis, where risk factors are changes of the full yield. Since the same relative changes are applied for all yield curves, the results are the same for all yield curves. Again, the application of volatility updating improves the predictive power and mean adjustment also has a positive effect.

2.9.3 Risk Factor: Spread over Benchmark Yield

The number of violations (see Table 2.12) is comparable to the latter two variants. Volatility updating leads to better results, while the effect of mean
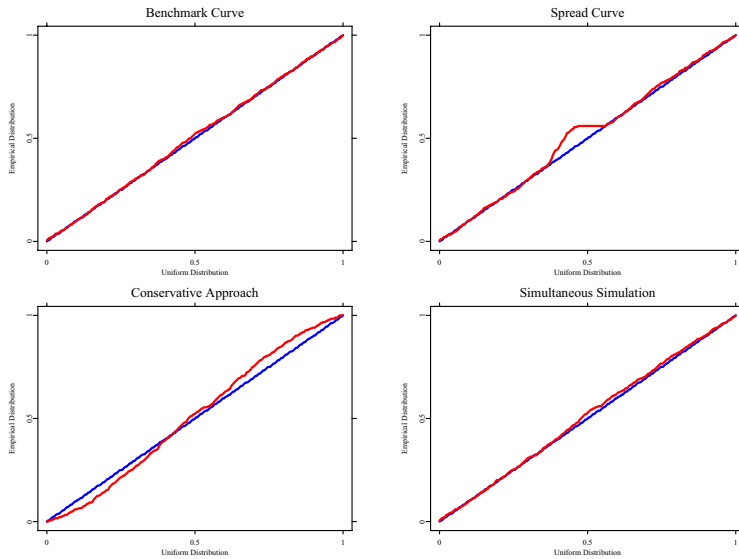
Figure 2.8. P-P Plots choice of risk factors. XFGpp

adjustment is only marginal. However, the mean squared deviations (see Table 2.15) in the P-P plots are significantly larger than in the case, where the risk factors are contained in the benchmark curve. This can be traced back to a partly poor predictive power for intermediate confidence levels (see Figure 2.8). Mean adjustment leads to larger errors in the P-P plots.

2.9.4 Conservative Approach

From Table 2.14 the conclusion can be drawn, that the conservative approach significantly over-estimates the risk for all credit qualities. Table 2.17 indicates the poor predictive power of the conservative approach over the full range of confidence levels. The mean squared deviations are the worst of all approaches. Volatility updating and/or mean adjustment does not lead to any significant improvements.

2.9.5 Simultaneous Simulation

From Tables 2.14 and 2.17 it is apparent that simultaneous simulation leads to much better results than the model with risk factors from the full yield curve,
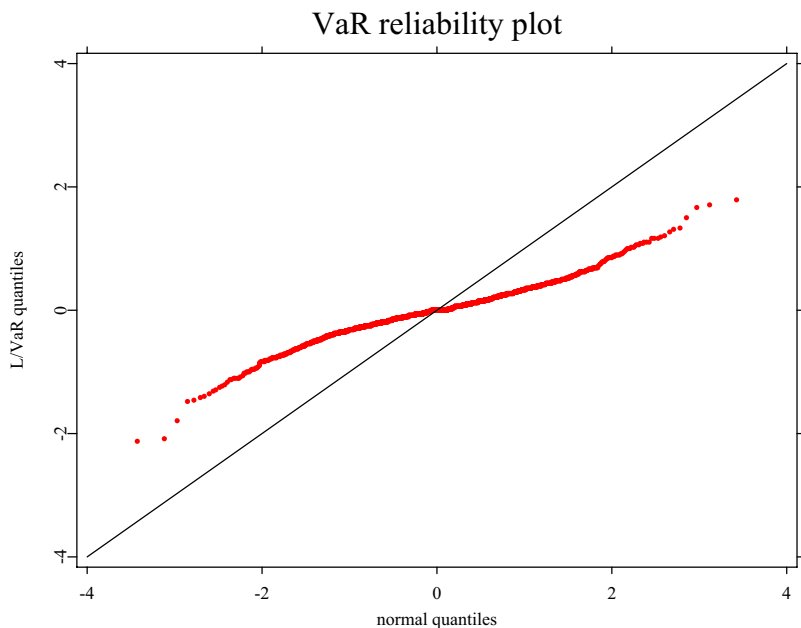
Figure 2.9. Q-Q Plot for basic historical simulation.

when volatility updating is included. Again, the effect of mean adjustment does not in general lead to a significant improvement. These results lead to the conclusion that general market risk and spread risk should be modeled independently, i.e., that the yield curve of an instrument exposed to credit risk should be modeled with two risk factors: benchmark changes and spread changes.

2.10 Internal Risk Models

In this contribution it is demonstrated that XploRe can be used as a tool in the analysis of time series of market data and empirical loss distributions. The focus of this contribution is on the analysis of spread risk. Yield spreads are an indicator of an obligor's credit risk. The distributions of spread changes are leptokurtic with typical fat tails, which makes the application of conventional variance-covariance risk models problematic. That is why in this contribution we prefer the analysis of spread risk by means of historical simulation. Since it is not a priori clear, how spread risk should be inte-

VaR reliability plot
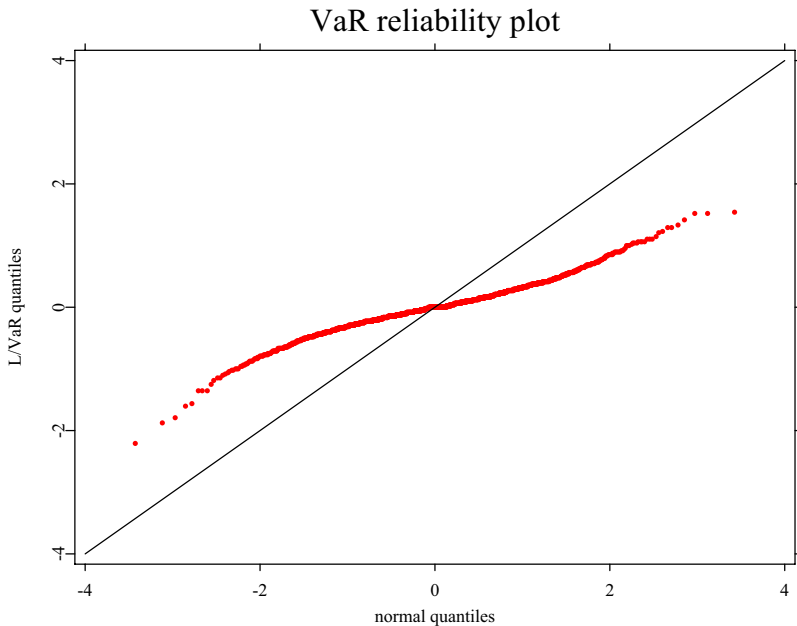
Figure 2.10. Q-Q plot for volatility updating.

grated in a risk model for interest rate products and how it can be separated from general market risk, we investigate several possibilities, which include modelling the full yield curve (i.e., consideration of only one risk factor category, which covers both benchmark and spread risk) as well as separately modelling spread risk and benchmark risk. The aggregation of both risk categories is carried out in a conservative way (addition of the risk measure for both risk categories) as well as coherently (simultaneous simulation of spread and benchmark risk). Moreover, in addition to the basic historical simulation method we add additional features like mean adjustment and volatility updating. Risk is quantified by means of a quantile-based risk measure in this contribution - the VaR. We demonstrate the differences between the different methods by calculating the VaR for a fictive zero-bond.

The numerical results indicate, that the conservative approach over-estimates the risk of our fictive position, while the simulation results for the full yield as single risk factor are quite convincing. The best result, however, is delivered by a combination of simultaneous simulation of spread and benchmark risk and volatility updating, which compensates for non-stationarity in the risk factor time series. The conclusion from this contribution for model-builders

	full yield				spread curve			
Curve	**V1**	**V2**	**V3**	**V4**	**V1**	**V2**	**V3**	**V4**
INAAA	0,87	0,28	0,50	0,14	8,13	22,19	8,14	16,15
INAA2	0,45	0,36	0,32	0,16	6,96	21,41	7,25	15,62
INAA3	0,54	0,41	0,43	0,23	7,91	21,98	7,97	15,89
INA1	0,71	0,27	0,41	0,13	7,90	15,32	8,10	8,39
INA2	0,50	0,39	0,42	0,17	9,16	15,15	9,51	6,19
INA3	0,81	0,24	0,58	0,24	9,53	12,96	9,61	7,09
INBBB1	0,71	0,29	0,54	0,13	9,59	15,71	9,65	11,13
INBBB2	0,33	0,34	0,26	0,12	11,82	14,58	11,59	10,72
INBBB3	0,35	0,59	0,40	0,34	7,52	11,49	7,78	6,32
INBB1	0,31	0,95	0,26	0,28	4,14	4,57	3,90	1,61
INBB2	0,52	0,49	0,36	0,19	6,03	3,63	5,89	2,12
INBB3	0,53	0,41	0,36	0,17	3,11	3,65	3,09	1,67
INB1	0,51	0,29	0,38	0,15	3,59	1,92	2,85	1,16
INB2	0,51	0,48	0,31	0,22	4,29	2,31	3,41	1,42
INB3	0,72	0,38	0,32	0,16	3,70	2,10	2,99	3,02
BNAAA	0,59	0,19	0,48	0,56	10,13	17,64	9,74	11,10
BNAA1/2	0,54	0,21	0,45	0,46	5,43	13,40	5,73	7,50
BNA1	0,31	0,12	0,29	0,25	8,65	17,19	8,09	8,21
BNA2	0,65	0,19	0,57	0,59	6,52	12,52	6,95	6,45
BNA3	0,31	0,19	0,32	0,29	6,62	9,62	6,59	3,80
Average	**0,54**	**0,35**	**0,40**	**0,25**	**7,04**	**11,97**	**6,94**	**7,28**

Table 2.15. MSD P-P Plot for the full yield and the spread curve($\times 10\,000$)

Curve	**V1**	**V2**	**V3**	**V4**
INAAA, INAA2, INAA3	0,49	0,23	0,26	0,12
INA1	0,48	0,23	0,26	0,12
INA2, INA3, INBBB1, INBBB2, INBBB3, INBB1, INBB2	0,49	0,23	0,26	0,12
INBB3	0,47	0,23	0,25	0,12
INB1	0,49	0,23	0,26	0,12
INB2	0,47	0,23	0,25	0,12
INB3	0,48	0,23	0,26	0,12
BNAAA, BNAA1/2	0,42	0,18	0,25	0,33
BNA1	0,41	0,18	0,23	0,33
BNA2	0,42	0,18	0,25	0,33
BNA3	0,41	0,18	0,24	0,33
Average	**0,47**	**0,22**	**0,25**	**0,17**

Table 2.16. MSD P-P-Plot benchmark curve ($\times 10\,000$)

in the banking community is, that it should be checked, whether the full yield curve or the simultaneous simulation with volatility updating yield satisfactory results for the portfolio considered.

	conservative approach				simultaneous simulation			
Curve	V1	V2	V3	V4	V1	V2	V3	V4
INAAA	14,94	14,56	14,00	13,88	1,52	0,64	0,75	0,40
INAA2	13,65	13,51	14,29	14,31	0,79	0,38	0,40	0,23
INAA3	14,34	13,99	13,66	13,44	0,79	0,32	0,49	0,27
INA1	15,39	15,60	15,60	15,60	0,95	0,40	0,52	0,29
INA2	13,95	14,20	14,32	14,10	0,71	0,55	0,50	0,39
INA3	14,73	14,95	14,45	14,53	0,94	0,30	0,59	0,35
INBBB1	13,94	14,59	14,05	14,10	1,00	0,33	0,43	0,17
INBBB2	13,74	13,91	13,67	13,73	0,64	0,52	0,45	0,29
INBBB3	13,68	14,24	14,10	14,09	0,36	0,78	0,31	0,31
INBB1	19,19	20,68	18,93	19,40	0,73	1,37	0,52	0,70
INBB2	13,21	14,17	14,79	15,15	0,30	0,82	0,35	0,51
INBB3	15,19	16,47	15,40	15,67	0,55	0,65	0,15	0,21
INB1	15,47	15,64	15,29	15,51	0,53	0,44	0,19	0,26
INB2	14,47	14,93	15,46	15,77	0,24	0,55	0,24	0,24
INB3	14,78	14,67	16,77	17,03	0,38	0,44	0,27	0,22
BNAAA	14,80	15,30	16,30	16,64	1,13	0,33	0,99	0,96
BNAA1/2	13,06	13,45	14,97	15,43	0,73	0,16	0,57	0,50
BNA1	11,95	11,83	12,84	13,08	0,52	0,26	0,44	0,41
BNA2	13,04	12,58	14,31	14,56	0,78	0,13	0,51	0,58
BNA3	12,99	12,70	15,19	15,42	0,34	0,18	0,58	0,70
Average	**14,33**	**14,60**	**14,92**	**15,07**	**0,70**	**0,48**	**0,46**	**0,40**

Table 2.17. MSD P-P Plot for the conservative approach and the simultaneous simulation($\times 10\,000$)

Bibliography

Bank for International Settlements (1998a). Amendment to the Capital Accord to incorporate market risks, www.bis.org. (January 1996, updated to April 1998).

Bank for International Settlements (1998b). Overview of the Amendment to the Capital Accord to incorporate market risk, www.bis.org. (January 1996, updated to April 1998).

Bundesaufsichtsamt für das Kreditwesen (2001). Grundsatz I/Modellierung des besonderen Kursrisikos, Rundschreiben 1/2001, www.bakred.de.

Gaumert, U. (1999). *Zur Diskussion um die Modellierung besonderer Kursrisiken in VaR-Modellen, Handbuch Bankenaufsicht und Interne Risikosteuerungsmodelle*, Schäffer-Poeschel.

Hull, J. C. (1998). Integrating Volatility Updating into the Historical Simulation Method for Value at Risk, *Journal of Risk* .

J.P. Morgan (1996). RiskMetrics, *Technical report*, J.P. Morgan, New York.

Kiesel, R., Perraudin, W. and Taylor, A. (1999). The Structure of Credit Risk. Working Paper, London School of Economics.

3 A Copula-Based Model of the Term Structure of CDO Tranches

Umberto Cherubini, Sabrina Mulinacci and Silvia Romagnoli

3.1 Introduction

A large literature has been devoted to the evaluation of CDO tranches in a cross-section setting. The main idea is that the cross-section dependence of the times to default of the assets or names of the securitization deal is specified, the dynamics of the losses in the pool of credits is simulated accordingly and the value of tranches is recovered. The dependence structure is usually represented in terms of copula functions, which provides flexibility and allows to separate the specification of marginal default probability distributions and dependence. The application of copula was first proposed by Li (2000) and nowadays is a common practice in the market of basket credit derivatives, particularly in the version of factor copulas (see Gregory and Laurent (2005) for a review and Burtshell, Gregory and Laurent (2005) for a comparison of the approach). This approach is obviously feasible for CDOs with a limited number of assets. Even for those, however, there is an issue that remains to be solved, and it has to do with the temporal consistency of prices.

This has become more and more evident with the growth of the so called *standardized CDO* markets. CDX or iTraxx tranches are in fact traded for the same degrees of protection (attachment/detachment) and on the same set of names for different time horizons. This has naturally raised the question whether the premium quoted for protection on a 5 year horizon for a CDX tranche was consistent with the observed price of protection for 10 years on the same tranche. For example, take the $(0, 3\%)$ equity tranche of a CDX. On June 15 2005 the CDX equity tranche $(0, 3\%)$ quoted a price of 42.375% and the 10 price for the same tranche on a 10 year horizon was 63% (remember that it is market convention to price equity tranche premia upfront plus 500 bp running premium). It is natural to assess that buying protection on a 10 year tranche should be weakly more expensive that buying the same

protection for a 5 year horizon. The question is how huge this difference in price ought to be.

A major flaw of copula functions applications to CDO pricing is that it has not provided an answer to this question. This is due to the static nature of copula functions, or at least to the fact that they have been used as static objects in these pricing problems. An alternative approach would be to directly specify the stochastic process of losses, for example using a general Lévy process. This choice comes at the cost of losing almost all of the flexibility provided by the copula approach. The infinite divisibility property of Lévy processes, for example, would imply that the probability distribution of losses on every time interval of the same length would have the same distribution. We could of course handle this by resorting to a suitable time change Carr and Wu (2004). Even after this, a crucial assumption would be retained, which is that losses in the second 5 year period should be independent from those in the first 5 years. Of course everyone who has been at least once exposed to studies on business cycles would hardly subscribe this statement. For this reason, it is possible that Lévy processes be too restrictive to explain the term structure behavior of premia in the market of basket credit derivatives.

The excessive flexibility of the copula tool on one side versus the excessive restriction of Lévy processes specification on the other raises the issue of an approach bridging the two together, for example using copulas to represent temporal dependence of losses, and encompassing the Lévy representation of them as a special case. This is the main question addressed in this paper, and the solution is handled by retaining the copula pricing methodology. To make the problem more formal, denote X_s the amount of losses cumulated up to time s on a basket of assets or names (say it is $s = 5$) and X_t the amount of losses ut to time $t = 10$. Denote $Y_{t-s} = X_t - X_s$ the amount of losses in the period between 5 and 10 years. Temporal consistency requires to define the dependency structure between X_s, X_t and Y_{t-s}. The current approach is to specify different copula functions for X_s and X_t. The Lévy process approach is to define a specific process for X_s, assuming Y_s to be iid. Our approach is to select a dependence structure between X_s and Y_{t-s} and the corresponding marginal distributions to recover the distribution of X_t. Of course, the process could be in turn iterated to recover the distribution of X_{t+s} based on the dependence structure of X_t and Y_s and their marginal distributions, and so on. In other words, by tuning the distributions of the losses over a set of time intervals and the dependence structure between each of them and the corresponding amount of cumulated losses at the beginning of the period, we can recursively recover a set of distributions of cumulated losses. From this, we can recover the term structure of the premia for protection on such losses. This preserves the flexibility of copula functions and encom-

passes Lévy processes as a special case in which the distribution of losses is assumed to be the same across all the intervals of time, and independent of the cumulated losses at the beginning of the period. The approach applies copula functions to represent the temporal dynamics in first order Markov process (see the seminal paper by Darsow, Nguyen and Olsen (1992) and the recent applications in econometrics, Ibragimov (2005), Gagliardini and Gourieroux (2005)). The technical contribution to this stream of literature is the proposal of a way to construct a class of such copulas starting from the dependence structure between the levels of a variable and its increments.

The plan of the paper is as follows. In section 3.2 we briefly describe a copula-based dynamics of the losses in a basket. In section 3.3 the show how to recover the distribution of the cumulated losses at the end of a period given the distributions of the cumulated losses at the beginning and the increment of losses in the period in a fully general setting. In section 3.4 we apply the theoretical analysis to recover the algorithm for the cumulated losses distribution. In section 3.5 we present a sensitivity analysis of the term structure of CDX premia to changes in the temporal correlation. Section 3.6 concludes.

3.2 A Copula-Based Model of Basket Credit Losses Dynamics

We assume a filtered probability space $\{\mathbf{\Omega}, \Im_t, \mathrm{P}\}$ satisfying the usual conditions with P the risk-neutral measure. We consider a set of losses $\{Y_1, Y_2, ..., Y_n\}$ in a set of periods limited by a dates $\{t_0, t_1, t_2, ..., t_n\}$. The set of cumulated losses are every time is given by $X_i = X_{i-1} + Y_i$. We assume that the derivative contract starts at time t_0 so that $X_0 = 0$. We assume that each loss Y_i is endowed with a probability distribution F_{Y_i}, $i = 1, ...n$. We want to recover the set of distributions F_{X_i}. Of course, we have $F_{Y_1} = F_{X_1}$. We assume a set of copula functions C_{X_{i-1}, Y_i} representing the dependence structure between the losses in a period and the cumulated losses at the beginning of that period. Our task is to jointly determine: i) the probability distribution of each X_i and ii) the temporal dependence structure between X_{i-1} and X_i. It is possible to show that this will generate a first order Markov process in the representation of Darsow, Nguyen and Olsen (1992).

3.3 Stochastic Processes with Dependent Increments

In this paragraph we focus on a single date t_i, $i = 2, ..., n$ so that we drop reference to time for simplicity. Let X, Y be two random variables with continuous c.d.f. F_X and F_Y, representing losses at the beginning of the period and losses during the period respectively. Let $C_{X,Y}(w, \lambda)$ be the copula function that describes their mutual dependence.

We begin by reminding a standard result in the copula functions literature (see Cherubini, Luciano and Vecchiato (2004) and Nelsen (2006)) that will be heavily used throughout the paper, stating that the partial derivative of a copula function corresponds to the conditional probability distribution. Formally, we have

LEMMA 3.1 *Let X and Y be two real-valued random variables on the same probability space $(\Omega, \mathcal{A}, \mathrm{P})$ with corresponding copula $C(w, \lambda)$ and continuous marginals F_X and F_Y. Then for every $x, y \in \mathbb{R}$, setting $\mathcal{F}_X = w$ and $F_Y = \lambda$, we have that*

$$D_1 C \{F_X(x), F_Y(y)\} \ \text{is a version of } \mathrm{P}\left(Y \leq y | X\right) \ \text{and}$$

$$D_1 C \{F_X(x), F_Y(Y)\} \ \text{is a version of } \mathrm{P}\left(X \leq x | Y\right).$$

For every continuous c.d.f. F we define the generalized inverse $F^{-1} : (0, 1) \rightarrow \mathbb{R}$ as

$$F^{-1}(w) = \inf \left\{l \in \mathbb{R} : F(l) \geq w\right\}. \tag{3.1}$$

We have

Proposition 3.1 *Let X e Y be two real-valued random variables on the same probability space $(\Omega, \mathcal{A}, \mathrm{P})$ with corresponding copula $C_{X,Y}(w, \lambda)$ and continuous marginals F_X and F_Y. Then,*

$$C_{X,X+Y}(u, v) = \int_0^u D_1 C_{X,Y} \left[w, F_Y\{F_{X+Y}^{-1}(v) - F_X^{-1}(w)\}\right] dw, \ \ \forall u, v \in [0, 1],$$

$$\tag{3.2}$$

and

$$F_{X+Y}(t) = \int_0^1 D_1 C_{X,Y} \left[w, F_Y\{t - F_X^{-1}(w)\}\right] dw. \tag{3.3}$$

Proof:

$$F_{X,X+Y}(s,t) = \mathrm{P}\left(X \le s, X + Y \le t\right) =$$

$$= \int_{-\infty}^{s} \mathrm{P}\left(X + Y \le t | X = x\right) dF_X(x) =$$

$$= \int_{-\infty}^{s} \mathrm{P}\left(Y \le t - x | X = x\right) dF_X(x) =$$

$$= \int_{-\infty}^{s} D_1 C_{X,Y}\left\{F_X(x), F_Y(t - x)\right\} dF_X(x) =$$

$$= \int_{0}^{F_X(s)} D_1 C_{X,Y}\left[w, F_Y\{t - F_X^{-1}(w)\}\right] dw$$

where we made the substitution $w = F_X(x) \in (0, 1)$.

Then, the copula function of X and $X + Y$ is

$$C_{X,X+Y}(u, v) = F_{X,X+Y}\left\{F_X^{-1}(u), F_{X+Y}^{-1}(v)\right\} =$$

$$= \int_{0}^{u} D_1 C_{X,Y}\left[w, F_Y\{F_{X+Y}^{-1}(v) - F_X^{-1}(w)\}\right] dw.$$

Moreover

$$F_{X+Y}(t) = \lim_{s \to +\infty} F_{X,X+Y}(s, t) = \int_{0}^{1} D_1 C_{X,Y}\left[w, F_Y\{t - F_X^{-1}(w)\}\right] dw.$$

$\square$

DEFINITION 3.1 *Let F, G be two continuous c.d.f. and C a copula function. We define the* **C-convolution** *of F and G the c.d.f.*

$$F \overset{C}{*} G(t) = \int_{0}^{1} D_1 C\left[w, G\{t - F^{-1}(w)\}\right] dw$$

Proposition 3.2 *Let F, G, H be three continuous c.d.f., $C(w, \lambda)$ a copula function and*

$$\widehat{C}(u, v) = \int_{0}^{u} D_1 C\left[w, F\{G^{-1}(v) - H^{-1}(w)\}\right] dw.$$

$\widehat{C}(u, v)$ is a copula function iff

$$G = H \overset{C}{*} F. \tag{3.4}$$

Proof:

Let us assume (3.4) to hold. Since there exist a probability space and two random variables X and Y with joint distribution function $F(x, y) = C(F_X(x), F_Y(y))$, thanks to Proposition 3.1, $\widehat{C}$ is a copula function.

Viceversa. Let $\widehat{C}$ be a copula function. Necessarily $\widehat{C}(1, v) = v$ holds. But

$$\widehat{C}(1, v) = \int_0^1 D_1 C \left[w, F \left\{ G^{-1}(v) - H^{-1}(w) \right\} \right] dw =$$

$$= H \overset{C}{*} F \left\{ G^{-1}(v) \right\}$$

and

$$H \overset{C}{*} F \left\{ G^{-1}(v) \right\} = v$$

for all $v \in (0, 1)$ if and only if $G = H \overset{C}{*} F$. $\square$

DEFINITION 3.2 *Let F and H be two continuous c.d.f. and C a copula function. We define the copula function*

$$\widehat{C}(u, v) = \int_0^u D_1 C \left[w, F\{ (H \overset{C}{*} F)^{-1}(v) - H^{-1}(w) \} \right] dw. \qquad (3.5)$$

REMARK 3.1 Independence

If C is the product copula, the C-convolution of H and F coincides with the convolution of H and F, while the copula $\widehat{C}$ defined through (3.5) takes the form

$$\widehat{C}(u, v) = \int_0^u F\{ (H * F)^{-1}(v) - H^{-1}(w) \} dw. \qquad (3.6)$$

REMARK 3.2 The co-monotonic case *In the case $C(w, \lambda) = w \wedge \lambda$, it is easy to verify*

$$H \overset{C}{*} F(t) = \int_0^1 \mathbf{1}\{ (0, F\{ t - H^{-1}(w) \}) \}(w) dw =$$

$$= \int_0^1 \mathbf{1}\{ w : F^{-1}(w) + H^{-1}(w) < t \}(w) dw$$

$$= \sup \left\{ w \in (0, 1) : H^{-1}(w) + F^{-1}(w) < t \right\}$$

and

$$\widehat{C}(u,v) = \int_0^u \mathbf{1}\{(0, F[\{H \overset{C}{*} F\}^{-1}(v) - H^{-1}(w)])\}(w)dw =$$

$$= \int_0^u \mathbf{1}\{w : H^{-1}(w) + F^{-1}(w) < (H \overset{C}{*} F)^{-1}(v)\}(w)dw \qquad (3.7)$$

$$= u \wedge \sup\left\{w \in (0,1) : F^{-1}(w) + H^{-1}(w) < (H \overset{C}{*} F)^{-1}(v)\right\}$$

$$= u \wedge v.$$

3.4 An Algorithm for the Propagation of Losses

In this Section we use the analysis above to recover a recursive algorithm for the computation of the distribution of losses over an increasing set of dates. We simply consider the same setting and the same symbology of section 3.2.

Thanks to (3.3) it is possible to compute the distribution of every X_i through the following iterated formula

$$F_{X_i}(t) = F_{X_{i-1}+Y_i}(t) =$$

$$= \int_0^1 D_1 C_{X_{i-1},Y_i}\left[w, F_{Y_i}\{t - F_{X_{i-1}}^{-1}(w)\}\right]dw$$

with $F_{X_1} = F_{Y_1}$.

From the distribution of losses it would be easy to integrate the value of protection of equity tranches with increasing detachment points.

Here below we describe the algorithm for the propagation of the probability distribution of losses based on the previous results. The algorithm consists of the following steps:

1. Start with the distribution of losses in the first period and set $X_1 = Y_1$

2. Numerically compute the integral in equation 3.4 yielding F_{X_2}

3. Go back to step 2 and use F_{X_2} and F_{Y_2} to compute F_{X_3}

4. Iterate until $i < n$

The algorithm is very general and may be applied to any problem of propagation of a distribution of increments. It may be thought of as an empirical way to generalize the scaling law of temporal aggregation of measures.

3.5 Empirical Analysis

Here we apply the model above to market data with the goal to gauge the sensitivity of the prices of tranches to changes in temporal dependence, and how these effects compare with those due to changes in cross-section dependence. Even though the structure of the algorithm is very flexible and general, and may allow for complex cross-section and temporal dependence structures, we keep the model simple. We assume that the joint distribution of losses in every period is given by the Vasicek (1991) formula

$$\Pr(Z < z) = \Phi \left\{ \frac{\sqrt{1-\rho^2}\,\Phi^{-1}(z) - \Phi^{-1}(p)}{\rho} \right\} \tag{3.8}$$

where Z represents the fraction of losses, ρ represent asset correlation of the names in the basket and p denotes their default probability. Notice that the model is homogeneous, meaning that all firms are assumed to have the same default probability and same asset correlation. As for temporal dependence, we have used the main Archimedean copulas, namely the Frank, Gumbel and Clayton. In order to save space, we report the results for the Frank, which for its symmetry properties is the member of the class that corresponds to the Gaussian distribution in the elliptical class.

The details of the analysis are as follows. We used the CDS quotes of the names included in the CDX, series number 8. For each of the 125 names, we bootstrapped the probability of default in each year, up 10 years. The data refers to the series number 8, and was collected on July 3rd 2007. The joint distribution of losses is computed using the Vasicek formula above, using the average default probability of the names and a dependence figure. Then, the joint distribution of losses is propagated forward using the algorithm described in the previous section. Finally, the prices of the tranches are computed using the standard market conventions.

In the base scenario, we assume that cross-section correlation is 40% in the Vasicek formula and the losses in each period are assumed to be independent. Then we change cross-section correlation by decreasing it to 20% and we compare the effects with changes in temporal dependence. We tried scenarios with both positive and negative dependence corresponding to a Kendall τ figures of 20% and 40%.

In Figure 3.1 we report the distribution of losses in the base scenario and in the hypothesis of a decrease of correlation from 40% to 20%. The curve describing the percentiles is tilted, and the percentile for a low percentage of losses is higher, while the percentile of a high percentage of losses is lower.
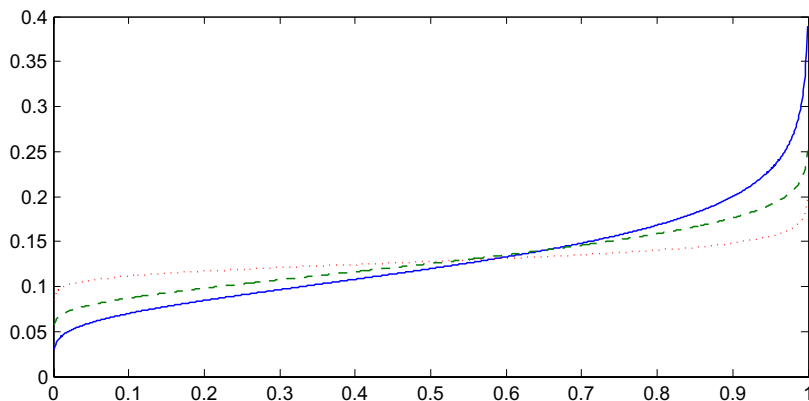
Figure 3.1. Probability Distribution of Losses. Different Cross-Section and Temporal Dependence. Blue line represents the base scenario, green sketched line assume cross section rho=0.2 and red dotted line a temporal dependence Kendall $\tau = -0.2$. ◘ XFGProbability

In financial terms, this means that risk is more idiosyncratic than in the base scenario. The third curve relaxes the assumption of temporal independence of losses. If the amount of losses in a period is negatively associated to cumulated losses at the beginning of the period the increase in idiosyncratic risk is even higher. In Figure 3.2 we report the effect of changes in temporal dependence on the distribution of losses. As temporal dependence decreases, the distribution is twisted upward.

The results on the distribution of losses foreshadow that equity tranches should be short on both cross-section and temporal correlation. Figure 3.3 reports the term structure of the equity tranche for different maturities. Contrary to the market practice, the price is reported in terms of running basis premium instead of upfront, but this does not make any difference for the discussion. As we move from the base scenario to a decrease in cross-section dependence the value of equity tranche increases. As expected, if we also allow for negative dependence between period losses and cumulated losses the impact on equity tranches is even higher. Figure 3.4 reports the effect on the senior tranche 15-30%. As expected, the impact is of opposite sign. The decrease in both cross-section and temporal dependence brings about a decrease in value of the 15-30% tranche. Figures 3.5 and 3.6 reports instead the effect of changes of temporal dependence on the base scenario. Notice that in this case, temporal correlation has a relevant impact only on the 10 year maturity, which is typically the most difficult to calibrate.
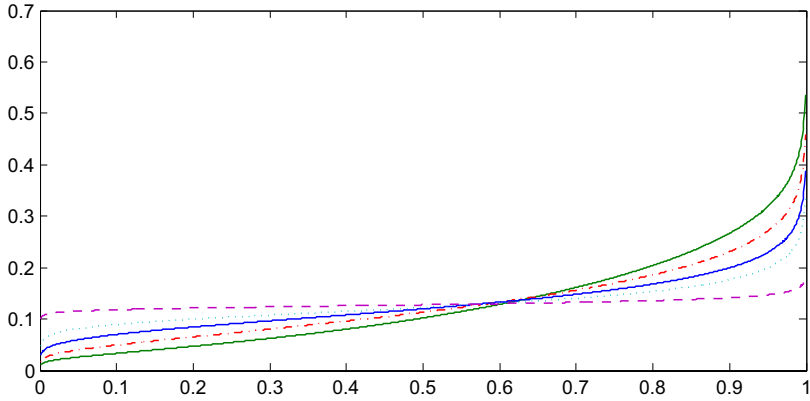
Figure 3.2.   Probability Distribution of Losses: 10 Year
Horizon. Different Temporal Dependence. Blue line repre-
sents the base scenario, green line the Frank copula for a
temporal dependence Kendall $\tau = 0.4$, red sketched-dotted
line the Frank copula for a temporal dependence Kendall
$\tau = 0.2$, light blue dotted line the Frank copula for a tem-
poral dependence Kendall $\tau = -0.2$ and purple sketched
line the Frank copula for a temporal dependence Kendall
$\tau = -0.4$. ◙ XFGProbability
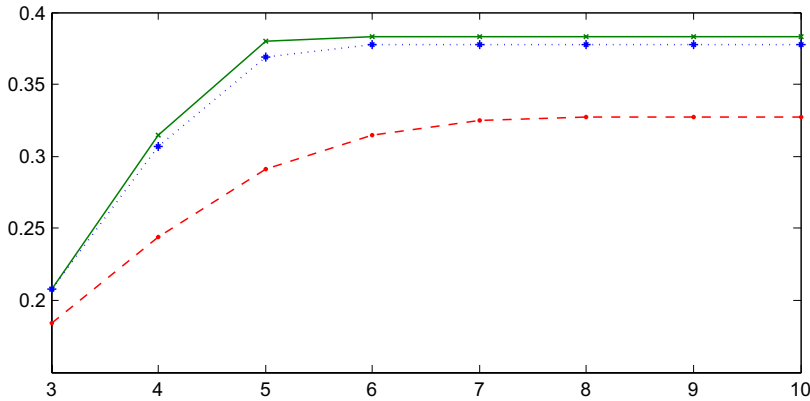
Figure 3.3.   Term Structure of Equity Tranche.   Different
Cross-Section and Temporal Dependence. Red sketched line
represents the base scenario, blue dotted line assume cross sec-
tion rho=0.2 and green line a temporal dependence Kendall
$\tau = -0.2$. ◙ XFGUpFront
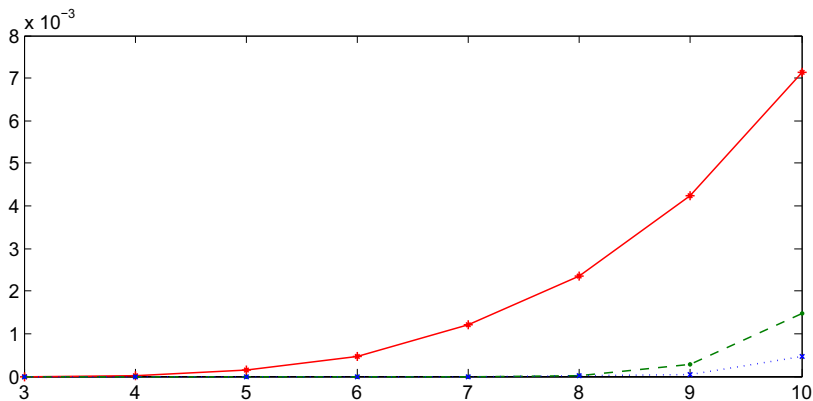
Figure 3.4. Term Structure of the Senior 15-30% Tranche. Different Cross-Section and Temporal Dependence. Red line represents the base scenario, green sketched line assume cross section rho=0.2 and blue dotted line a temporal dependence Kendall $\tau = -0.2$. **Q XFGRunningBasis**
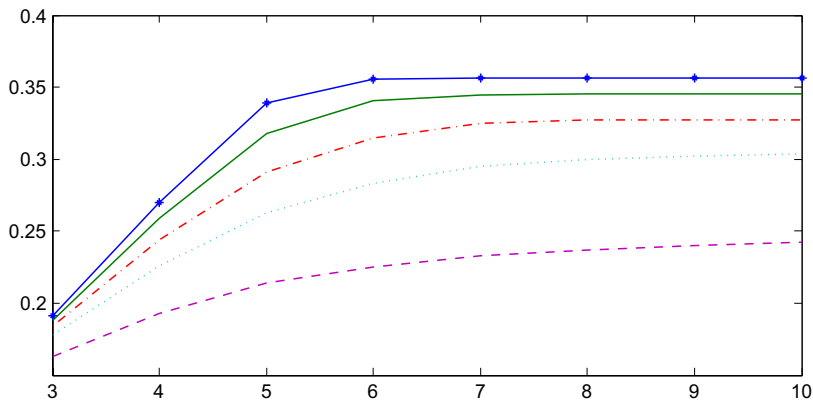
Figure 3.5. Term Structure of the Equity Tranche. Different Temporal Dependence. Red sketched-dotted line represents the base scenario, light blue dotted line for a temporal dependence Kendall $\tau = 0.2$, purple sketched line for a temporal dependence Kendall $\tau = 0.4$, green line for a temporal dependence Kendall $\tau = -0.2$ and blue marked line for a temporal dependence Kendall $\tau = -0.4$. **Q XFGUpFront**
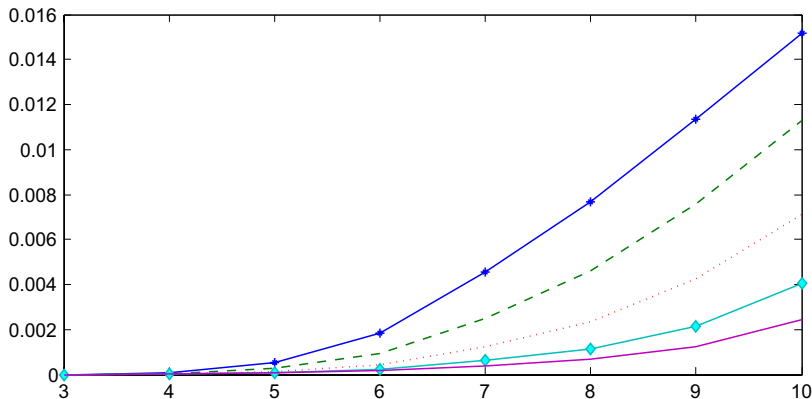
Figure 3.6. Term Structure of the Senior 15-30% Tranche.
Different Temporal Dependence. Red dotted line represents
the base scenario, green sketched line for a temporal depen-
dence Kendall $\tau = 0.2$, blue marked line for a temporal depen-
dence Kendall $\tau = 0.4$, light blue marked line for a temporal
dependence Kendall $\tau = -0.2$ and purple line for a temporal
dependence Kendall $\tau = -0.4$. ⊡ XFGRunningBasis

3.6 Concluding Remarks

In this paper we have proposed a copula based model to jointly model cross-
section and temporal dependence. Cross section dependence of losses is spec-
ified for every period, while cumulated losses are propagated from a period to
the next one by integrating a conditional distribution of the losses in a period
given the cumulated losses at the beginning of the period. The conditional
distribution is modelled using copula functions. We apply our strategy to
market data showing the sensitivity of the distribution of cumulated losses to
changes in temporal dependence. Changes in temporal dependence turn out
to have the same sign of impact on the value tranches as changes in cross-
section dependence. Namely, equity tranches are made more valuable by a
decrease in temporal dependence while senior tranches are made cheaper.

Using copulas to model both cross-section and temporal dependence in mul-
tivariate credit products open a new research program in the application of
these tools to these products. It is the ideal approach to try to jointly model
the entire term structure of the different tranches, which is one of the open
issues in the pricing of standard CDO (iTraxx and CDX). Future work will be
devoted to the calibration of the model with the goal to identify and separate
the cross-section and temporal correlation implied by the market prices.

Bibliography

Burtshell, X., Gregory, J. and Laurent, J.P. (2005). A comparative analysis of CDO pricing models, *working paper,BNP Paribas*.

Carr, P.and Wu, L. (2004). Time Changed Lévy Processes and Option Pricing,*Journal of Financial Economics*, **17**: 113–141.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). Copula Methods in Finance, *John Wiley Finance Series, Chichester, U.K.*

Cherubini, U.and Romagnoli, S. (2007). The Dependence Structure of Running Maxima and Minima: Theoretical Results and Option Pricing Application, *Working Paper, University of Bologna*.

Darsow, W.F.,Nguyen, B. and Olsen, E.T. (1992). Copulas and Markov Processes, *Illinois Journal of Mathematics*, **36** : 600–642.

Gagliardini, P., and Gourieroux, C. (2005). An Efficient Non Parametric Estimator For Models With Non Linear Dependence, *Journal of Econometrics*, **137** : 189–229.

Gregory, J. and Laurent, J.P. (2005). Basket default swaps, CDOs and factor copulas, *Journal of Risk*, **7**(4) : 103–122.

Ibragimov, R. (2005). Copula based characterization and modeling for time series, *Harvard Institute of Economic Research, Discussion paper n. 2094*.

Li, D. (2000). On default correlation: a copula function approach, *Journal of Fixed Income*, **9**: 43–54.

Nelsen, R. (2006), *Introduction to Copulas, Springer Verlag*.

Schmitz, V. (2003). Copulas and Stochastic Processes, *Ph.D. dissertation, Aachen University*.

Vasicek, O.A. (1991). Limiting Loan Loss Distribution, *KMV Corporation working paper*.

4 VaR in High Dimensional Systems – a Conditional Correlation Approach

Helmut Herwartz and Bruno Pedrinha

In empirical finance multivariate volatility models are widely used to capture both volatility clustering and contemporaneous correlation of asset return vectors. In higher dimensional systems, parametric specifications often become intractable for empirical analysis owing to large parameter spaces. On the contrary, feasible specifications impose strong restrictions that may not be met by financial data as, for instance, constant conditional correlation (CCC). Recently, dynamic conditional correlation (DCC) models have been introduced as a means to solve the trade off between model feasibility and flexibility. Here we employ alternatively the CCC and the DCC modeling framework to evaluate the Value-at-Risk associated with portfolios comprising major U.S. stocks. In addition, we compare their performance with corresponding results obtained from modeling portfolio returns directly via univariate volatility models.

4.1 Introduction

Volatility clustering, i.e. positive correlation of price variations observed on speculative markets, motivated the introduction of autoregressive conditionally heteroscedastic (ARCH) processes by Engle (1982) and its popular generalizations by Bollerslev (1986) (Generalized ARCH, GARCH) and Nelson (1991) (Exponential GARCH). Being univariate in nature, however, these models neglect a further stylized feature of empirical price variations, namely contemporaneous correlation over a cross section of assets, stock or foreign exchange markets (Engle, Ito and Lin, 1990; Hamao, Masulis and Ng, 1990; Hafner and Herwartz, 1998; Engle and Sheppard, 2001; Lee and Long, 2008).

The covariance between asset returns is of essential importance in finance. Effectively, many problems in financial theory and practice like asset allocation,

hedging strategies or Value-at-Risk (VaR) evaluation require some formalization not merely of univariate risk measures but rather of the entire covariance matrix (Bollerslev et al., 1988; Cecchetti, Cumby and Figlewski, 1988). Similarly, pricing of options with more than one underlying asset will require some (dynamic) forecasting scheme for time varying variances and covariances as well (Duan, 1995).

When modeling time dependent second order moments, a multivariate model is a natural framework to take cross sectional information into account. Over recent years, multivariate volatility models have been attracting high interest in econometric research and practice. Popular examples of multivariate volatility models comprise the GARCH model class recently reviewed by Bauwens, Laurent and Rombouts (2006). Numerous versions of the multivariate GARCH (MGARCH) model suffer from huge parameter spaces. Thus, their scope in empirical finance is limited since the dimension of vector valued systems of asset returns should not exceed five (Ding and Engle, 1994). Factor structures (Engle, Ng and Rothschild, 1990) and so-called correlation models (Bollerslev, 1990) have been introduced to cope with the curse of dimensionality in higher dimensional systems. The latter start from univariate GARCH specifications to describe volatility patterns and formalize in a second step the conditional covariances implicitly via some model for the systems' conditional correlations. Recently, dynamic conditional correlation models have been put forth by Engle (2002), Engle and Sheppard (2001) and Tse and Tsui (2002) that overcome the restrictive CCC pattern (Bollerslev, 1990) while retaining its computational feasibility.

Here, we will briefly review two competing classes of MGARCH models, namely the half-vec model family and correlation models. The latter will be applied to evaluate the VaR associated with portfolios comprised by stocks listed in the Dow Jones Industrial Average (DJIA) index. We compare the performance of models building on constant and dynamic conditional correlation. Moreover, it is illustrated how a univariate volatility model performs in comparison with both correlation models.

The remainder of this paper is organized as follows. The next section introduces the MGARCH model and briefly mentions some specifications that fall within the class of so-called half-vec MGARCH models. Correlation models are the focus of Section 4.3 where issues like estimation or inference within this model family are discussed in some detail. An empirical application of basic correlation models to evaluate the VaR for portfolios comprising U.S. stocks is provided in Section 4.5.

4.2 Half-Vec Multivariate GARCH Models

Let $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})^\top$ denote an N-dimensional vector of serially uncorrelated components with mean zero. The latter could be directly observed or estimated from a multivariate regression model. The process ε_t follows a multivariate GARCH process if it has the representation

$$\varepsilon_t | \mathcal{F}_{t-1} \sim N(0, \Sigma_t), \Sigma_t = [\sigma_{ij,t}], \qquad (4.1)$$

where Σ_t is measurable with respect to information generated up to time $t - 1$, formalized by means of the filtration $\mathcal{F}_{t-1}$. The $N \times N$ conditional covariance matrix, $\Sigma_t = \mathsf{E}[\varepsilon_t \varepsilon_t^\top | \mathcal{F}_{t-1}]$, has typical elements $\sigma_{ij,t}$ with $i = j$ $(i \neq j)$ indexing conditional variances (covariances). In a multivariate setting potential dependencies of the second order moments in Σ_t on $\mathcal{F}_{t-1}$ become easily intractable for practical purposes.

The assumption of conditional normality in (4.1) allows to specify the likelihood function for observed processes ε_t, $t = 1, 2, \ldots, T$. In empirical applications of GARCH models, it turned out that conditional normality of speculative returns is more an exception than the rule. Maximizing the misspecified Gaussian log-likelihood function is justified by quasi maximum likelihood (QML) theory. Asymptotic theory on properties of the QML estimator in univariate GARCH models is well developed (Bollerslev and Wooldridge, 1992; Lee and Hansen, 1994; Lumsdaine, 1996). Recently, a few results on consistency (Jeantheau, 1998) and asymptotic normality (Comte and Lieberman, 2003; Ling and McAleer, 2003) have been derived for multivariate processes.

The so-called half-vec specification encompasses all MGARCH variants that are linear in (lagged) second order moments or squares and cross products of elements in (lagged) ε_t. Let vech(B) denote the half-vectorization operator stacking the elements of a $(m \times m)$ matrix B from the main diagonal downwards in a $m(m+1)/2$ dimensional column vector. We concentrate the formalization of MGARCH models on the MGARCH(1,1) case which is, by far, the dominating model order used in the empirical literature (Bollerslev et al., 1994). Within the so-called half-vec representation of the GARCH(1, 1) model Σ_t is specified as follows:

$$\mathrm{vech}(\Sigma_t) = c + A \, \mathrm{vech}(\varepsilon_{t-1} \varepsilon_{t-1}^\top) + G \, \mathrm{vech}(\Sigma_{t-1}). \qquad (4.2)$$

In (4.2) the matrices A and G each contain $\{N(N+1)/2\}^2$ elements. Deterministic covariance components are collected in c, a column vector of dimension $N(N+1)/2$. On the one hand, the half-vec model in (4.2) allows a very general dynamic structure of the multivariate volatility process. On the

other hand this specification suffers from huge dimensionality of the relevant parameter space which is of order $\mathcal{O}(N^4)$. In addition, it might be cumbersome or even impossible in applied work to restrict the admissible parameter space such that the time path of implied matrices Σ_t is positive definite.

To reduce the dimensionality of MGARCH models, numerous avenues have been followed that can be nested in the general class of half-vec models. Prominent examples in this vein of research are the Diagonal model (Bollerslev et al., 1988), the BEKK model (Baba, Engle, Kraft and Kroner, 1990; Engle and Kroner, 1995), the Factor GARCH (Engle, Ng and Rothschild, 1990), the orthogonal GARCH (OGARCH) (Alexander, 1998; Alexander, 2001) or the generalized OGARCH model put forth by van der Weide (2002). Evaluating the merits of these proposals requires to weight model parsimony and computational issues against the implied loss of generality. For instance, the BEKK model is convenient to allow for cross sectional dynamics of conditional covariances, and weak restrictions have been formalized keeping Σ_t positive definite over time (Engle and Kroner, 1995). Implementing the model will, however, involve simultaneous estimation of $\mathcal{O}(N^2)$ parameters such that the BEKK model has been rarely applied in higher dimensional systems ($N > 4$). Factor models build upon univariate factors, such as, an observed stock market index (Engle, Ng and Rothschild, 1990) or underlying principal components (Alexander, 1998; Alexander, 2001). The latter are assumed to exhibit volatility dynamics which are suitably modeled by univariate GARCH-type models. Thereby, factor models drastically reduce the number of model parameters undergoing simultaneous estimation. Model feasibility is, however, paid with restrictive correlation dynamics implied by the (time invariant) loading coefficients. Moreover, it is worthwhile mentioning that in case of factor specifications still $\mathcal{O}(N)$ parameters have to be estimated jointly when maximizing the Gaussian (quasi) likelihood function.

4.3 Correlation Models

4.3.1 Motivation

Correlation models comprise a class of multivariate volatility models that is not nested within the half-vec specification. Similar to factor models correlation models circumvent the curse of dimensionality by separating the empirical analysis in two steps. First, univariate volatility models are employed to estimate volatility dynamics of each asset specific return process ε_{it}, $i = 1, \ldots, N$. In a second step Σ_t is obtained imposing some parsimo-

nious structure on the correlation matrix (Bollerslev, 1990). Thus, in the framework of correlation models we have

$$\Sigma_t = V_t(\theta) R_t(\phi) V_t(\theta), \tag{4.3}$$

where $V_t = \text{diag}(\sqrt{\sigma_{11,t}}, \ldots, \sqrt{\sigma_{NN,t}})$ is a diagonal matrix having as typical elements the square roots of the conditional variances estimates $\sigma_{ii,t}$. The latter could be obtained from some univariate volatility model specified with parameter vectors θ_i stacked in $\theta = (\theta_1^\top, \ldots, \theta_N^\top)^\top$. If univariate GARCH(1,1) models are used for the conditional volatilities $\sigma_{ii,t}$, θ_i will contain 3 parameters such that θ is of length $3N$. Owing to its interpretation of a correlation matrix the diagonal elements in $R(\phi)$ are unity ($r_{ii} = 1, i = 1, \ldots, N$). From the general representation in (4.3) it is apparent that alternative correlation models particularly differ with regard to the formalization of the correlation matrix $R_t(\phi)$ specified with parameter vector ϕ.

In this section, we will highlight a few aspects of correlation models. First, a log-likelihood decomposition is given that motivates the stepwise empirical analysis. Then, two major variants of correlation models are outlined, the early CCC model (Bollerslev, 1990) and the DCC approach introduced by Engle (2002) and Engle and Sheppard (2001). Tools for inference in correlation models that have been applied in the empirical part of the paper are collected in an own subsection. Also, a few remarks on recent generalizations of the basic DCC specification are provided.

4.3.2 Log-Likelihood Decomposition

The adopted separation of volatility and correlation analysis is motivated by a decomposition of the Gaussian log-likelihood function (Engle, 2002) applying to the model in (4.1) and (4.3):

$$
\begin{aligned}
l(\theta, \phi) &= -\frac{1}{2} \left\{ \sum_{t=1}^{T} N \log(2\pi) + \log(|\Sigma_t|) + \varepsilon_t^\top \Sigma_t^{-1} \varepsilon_t \right\} \tag{4.4} \\
&= -\frac{1}{2} \left\{ \sum_{t=1}^{T} N \log(2\pi) + 2\log(|V_t|) + \log(|R_t|) + \varepsilon_t^\top \Sigma_t^{-1} \varepsilon_t \right\} \\
&= \sum_{t=1}^{T} l_t(\theta, \varphi),
\end{aligned}
$$

$$\tag{4.5}$$

$$l_t(\theta, \phi) = l_t^V(\theta) + l_t^C(\theta, \phi), \tag{4.6}$$

$$l_t^V(\theta) = -\frac{1}{2}\left\{N \log 2\pi + 2\log(|V_t(\theta)|) + \varepsilon_t^\top V_t(\theta)^{-2}\varepsilon_t\right\} \tag{4.7}$$

$$l_t^C(\theta, \phi) = -\frac{1}{2}\left(\log|R_t(\phi)| + v_t^\top R_t(\phi)^{-1}v_t - v_t^\top v_t\right). \tag{4.8}$$

According to (4.7) and (4.8), the maximization of the log likelihood function may proceed in two steps. First, univariate volatility models are used to maximize the volatility component, $l_t^V(\theta)$, and conditional on first step estimates $\widehat{\theta}$, the correlation part $l_t^C(\theta, \phi)$ is maximized in a second step. To perform a sequential estimation procedure efficiently, it is required that the volatility and correlation parameters are variation free (Engle, Hendry and Richard, 1983) meaning that there are no cross relationships linking single parameters in θ and ϕ when maximizing the Gaussian log-likelihood function. In the present case, the parameters in θ will impact on $v_t = V_t^{-1}\varepsilon_t$, $v_t = (v_{1t}, v_{2t}, \ldots, v_{Nt})^\top$, and, thus, the condition necessary to have full information and limited information estimation equivalent is violated. Note, however, that univariate GARCH estimates $(\widehat{\theta})$ will be consistent. Thus, owing to the huge number of available observations which is typical for empirical analyses of financial data, the efficiency loss involved with a sequential procedure is likely to be smaller in comparison with the gain in estimation feasibility.

4.3.3 Constant Conditional Correlation Model

Bollerslev (1990) proposes a constant conditional correlation (CCC) model

$$\sigma_{ij,t} = r_{ij}\sqrt{\sigma_{ii,t}\sigma_{jj,t}},\ i, j = 1, \ldots, N,\ i \neq j. \tag{4.9}$$

Given positive time paths of the systems' volatilities, positive definiteness of Σ_t is easily guaranteed for the CCC model ($|r_{ij}| < 1, i \neq j$). As an additional objective of this specification, it is important to notice that the estimation of the correlation pattern may avoid iterative QML estimation of the $\{N(N-1)/2\}$ correlation parameters r_{ij} comprising $R_t(\phi) = R$. Instead, one may generalize the idea of variance targeting (Engle and Mezrich, 1996) towards the case of correlation targeting. Then, $D = \mathsf{E}[v_t v_t^\top]$ is estimated as the unconditional covariance matrix of standardized returns, $v_t = V_t^{-1}\varepsilon_t$, and R is the correlation matrix implied by D. With $'\odot'$ denoting matrix multiplication by element we have formally

$$\widehat{R} = \widehat{D}^{*-1/2}\widehat{D}\widehat{D}^{*-1/2},\ \widehat{D} = \frac{1}{T}\sum_{t=1}^{T} v_t v_t^\top,\ \widehat{D}^* = \widehat{D} \odot I_N. \tag{4.10}$$

The price paid for the feasibility of CCC is, however, the assumption of a rather restrictive conditional correlation pattern which is likely at odds with empirical systems of speculative returns. Applying this model in practice therefore requires at least some pretest for constant correlation (Tse, 2000; Engle, 2002).

4.3.4 Dynamic Conditional Correlation Model

The dynamic conditional correlation model introduced by Engle (2002) and Engle and Sheppard (2001) preserves the analytic separability of the models' volatilities and correlations, but allows a richer dynamic structure for the latter. For convenience, we focus the representation of the DCC model again on the DCC(1,1) case formalizing the conditional correlation matrix $R_t(\phi)$ as follows:

$$R_t(\phi) = \{Q_t^*(\phi)\}^{-1/2} Q_t(\phi) \{Q_t^*(\phi)\}^{-1/2}, \; Q_t^*(\phi) = Q_t(\phi) \odot I_N, \qquad (4.11)$$

with

$$Q_t(\phi) = R(1 - \alpha - \beta) + \alpha v_{t-1} v_{t-1}^\top + \beta Q_{t-1}(\phi) \qquad (4.12)$$

and R is a positive definite (unconditional) correlation matrix of v_t.

Sufficient conditions guaranteeing positive definiteness of the time path of conditional covariance matrices Σ_t implied by (4.3), (4.11) and (4.12) are given in Engle and Sheppard (2001). Apart from well known positivity constraints to hold for the univariate GARCH components, the DCC(1,1) model will deliver positive definite covariances if $\alpha > 0$, $\beta > 0$ while $\alpha + \beta < 1$ and λ_{min}, the smallest eigenvalue of R, is strictly positive, i.e. $\lambda_{min} > \delta > 0$. It is worthwhile to point out that the DCC framework not only preserves the separability of volatility and correlation estimation, but also allows to estimate the nontrivial parameters in R via correlation targeting described in (4.10).

Given consistent estimates of unconditional correlations $r_{ij}, i \neq j$, the remaining parameters describing the correlation dynamics are collected in the two-dimensional vector $\varphi = (\alpha, \beta)^\top$. Note that, making use of correlation targeting the number of parameters undergoing nonlinear iterative estimation in the DCC model is constant $(= 2)$, and, thus, avoids the curse of dimensionality even in case of very large systems of asset returns.

Instead of estimating the model in three steps one could alternatively estimate the unconditional correlation parameters in R and the coefficients in φ

jointly. Note that the number of unknown parameters in R is $\mathcal{O}(N^2)$. Formal representations of first and second order derivatives to implement two step estimation and inference can be found in Hafner and Herwartz (2007). We prefer the three step approach here, since it avoids iterative estimation procedures in large parameter spaces.

4.3.5 Inference in the Correlation Models

QML-inference on significance of univariate GARCH parameter estimates is discussed in Bollerslev and Wooldridge (1992). Analytical expressions necessary to evaluate the asymptotic covariance matrix are given in Bollerslev (1986). In the empirical part of the chapter we will not provide univariate GARCH parameter estimates at all to economize on space. Two issues of evaluating parameter significance remain, inference for correlation estimates given in (4.10) and for estimated DCC parameters $\widehat{\varphi}$. We consider these two issues in turn:

1. Inference for unconditional correlations

   Conditional on estimates $\widehat{\theta}$, we estimate R from standardized univariate GARCH residuals as formalized in (4.10). The elements in $\widehat{R}$ are obtained as a nonlinear and continuous transformation of the elements in $\widehat{D}$, i.e. $\widehat{R} = \widehat{D}^{*-1/2}\widehat{D}\widehat{D}^{*-1/2}$. Denote with vechl($B$) an operator stacking the elements below the diagonal of a symmetric $(m \times m)$ matrix B in a $\{m(m-1)/2\}$ dimensional column vector $b_l = \text{vechl}(B)$. Thus, $\widehat{r}_l = \text{vechl}(\widehat{R})$ collects the nontrivial elements in $\widehat{R}$. Standard errors for the estimates in $\widehat{r}_l$ can be obtained from a robust estimator of the covariance of the (nontrivial) elements in $\widehat{D}$, $\widehat{d} = \text{vech}(\widehat{D})$, via the delta method. To be precise, we estimate the covariance of $\widehat{r}_l$ by means of the following result (Ruud, 2000):

   $$\sqrt{T}(\widehat{r}_l - r_l) \xrightarrow{\mathcal{L}} \text{N}\left(0, \mathcal{H}(\widehat{r})\mathcal{G}\mathcal{H}(\widehat{r})^{\top}\right), \qquad (4.13)$$

   where $\mathcal{G}$ is an estimate of the covariance matrix of the elements in d, $\mathcal{G} = \widehat{\text{Cov}}(\widehat{d})$, and $\mathcal{H}(\widehat{r})$ is a $\{N(N-1)/2 \times (N(N+1)/2)\}$ dimensional matrix collecting the first order derivatives $\partial r_l/\partial d^{\top}$ evaluated at $\widehat{d}$. We determine $\mathcal{G}$ by means of the covariance estimator

   $$\mathcal{G} = \frac{1}{T}\sum_{t=1}^{T}(vv)_t(vv)_t^{\top}, \ (vv)_t = \text{vech}(v_t v_t^{\top}) - \widehat{d}. \qquad (4.14)$$

The derivatives in $\mathcal{H}(r)$ are derived from a result in Hafner and Herwartz (2007) as

$$\frac{\partial r_l}{\partial d^\top} = P_{N,-}^\top (D^* \otimes D^*) P_N + P_{N,-}^\top (DD^* \otimes I_N + I_N \otimes DD^*) P_N \frac{\partial \mathrm{vech}(D^*)}{\partial \mathrm{vech}(D)^\top}$$

and

$$\frac{\partial \mathrm{vech}(D^*)}{\partial \mathrm{vech}(D)^\top} = -\frac{1}{2} \mathrm{diag}\left[\mathrm{vech}\left\{ (I_N \odot D)^{-3/2} \right\} \right],$$

where the matrices $P_{N,-}$ and P_N serve as duplication matrices (Lütkepohl, 1996) such that $\mathrm{vec}(B) = P_{N-}\mathrm{vechl}(B)$ and $\mathrm{vec}(B) = P_N \mathrm{vech}(B)$.

2. Inference for correlation parameters

The correlation parameters are estimated by maximizing the correlation part, $l^C(\theta, \phi)$, of the Gaussian (quasi) log-likelihood function. When evaluating the estimation uncertainty associated with $\widehat{\varphi} = (\widehat{\alpha}, \widehat{\beta})^\top$, the sequential character of the estimation procedure has to be taken into account. To provide standard errors for QML estimates $\widehat{\varphi}$, we follow a GMM approach introduced in Newey and McFadden (1994), which works in case of sequential GMM estimation under typical regularity conditions. In particular, it is assumed that all steps of a sequential estimation procedure are consistent. The following result on the asymptotic behavior of $\widehat{\gamma} = (\widehat{\theta}^\top, \widehat{\varphi}^\top)^\top$ applies:

$$\sqrt{T}(\widehat{\gamma} - \gamma) \sim N(0, \mathcal{N}^{-1}\mathcal{M}(\mathcal{N}^{-1})^\top). \tag{4.15}$$

In (4.15) $\mathcal{M}$ is the (estimated) expectation of the outer product of the scores of the log-likelihood function evaluated at $\widehat{\gamma}$,

$$\mathcal{M} = \frac{1}{T} \sum_{t=1}^T \left(\frac{\partial l_t}{\partial \gamma}\right) \left(\frac{\partial l_t}{\partial \gamma}\right)^\top, \quad \frac{\partial l_t}{\partial \gamma} = \left(\frac{\partial l_t^V}{\partial \theta^\top}, \frac{\partial l_t^C}{\partial \varphi^\top}\right)^\top. \tag{4.16}$$

Compact formal representations for the derivatives in (4.16) can be found in Hafner and Herwartz (2007) and Bollerslev (1986). The matrix $\mathcal{N}$ in (4.15) has a lower blockdiagonal structure containing (estimates) of expected second order derivatives, i.e.

$$\mathcal{N} = \begin{pmatrix} \mathcal{N}_{11} & 0 \\ \mathcal{N}_{21} & \mathcal{N}_{22} \end{pmatrix},$$

with

$$\mathcal{N}_{11} = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t^V}{\partial \theta \partial \theta^\top}, \ \mathcal{N}_{21} = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t^C}{\partial \varphi \partial \theta^\top}, \ \mathcal{N}_{22} = \frac{1}{T} \sum_{t=1}^T \frac{\partial^2 l_t^C}{\partial \varphi \partial \varphi^\top}.$$

Formal representations of the latter second order quantities are provided in Hafner and Herwartz (2007).

4.3.6 Generalizations of the DCC Model

Generalizing the basic DCC(1,1) model in (4.11) and (4.12) towards higher model orders is straightforward and in analogy to the common GARCH volatility model. In fact, it turns out that the DCC(1,1) model is often sufficient to capture empirical correlation dynamics (Engle and Sheppard, 2001). Tse and Tsui (2002) propose a direct formalization of the dynamic correlation matrix R_t as a weighted average of unconditional correlation, lagged correlation and a local correlation matrix estimated over a time window comprising the M most recent GARCH innovation vectors ξ_{t-i}, $i = 1, ..., M$, $M \geq N$. As discussed so far, dynamic correlation models are restrictive in the sense that asset specific dynamics are excluded. Hafner and Franses (2003) discuss a generalized DCC model where the parameters α and β in (4.12) are replaced by outer products of N-dimensional vectors, e.g. $\widetilde{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_N)^\top$, obtaining

$$Q_t = R(1 - \widetilde{\alpha}\widetilde{\alpha}^\top - \widetilde{\beta}\widetilde{\beta}^\top) + \widetilde{\alpha}\widetilde{\alpha}^\top \odot v_{t-1}v_{t-1}^\top + \widetilde{\beta}\widetilde{\beta}^\top \odot Q_{t-1}. \tag{4.17}$$

From (4.17) it is apparent that implied time paths of conditional correlations show asset specific characteristics. Similar to the generalization of the basic GARCH volatility model towards threshold specifications (Glosten, Jagannathan and Runkle, 1993) one may also introduce asymmetric dependencies of Q_t on $\mathrm{vech}(v_t v_t^\top)$ as in Cappiello, Engle and Sheppard (2003). A semiparametric conditional correlation model is provided by Hafner, van Dijk and Franses (2006). In this model the elements in Q_t are determined via local averaging where the weights entering the nonparametric estimates depend on a univariate factor as, for instance, market volatility or market returns.

4.4 Value-at-Risk

Financial institutions and corporations can suffer financial losses in their portfolios or treasury department due to unpredictable and sometimes extreme movements in the financial markets. The recent increase in volatility in financial markets and the surge in corporate failures are driving investors, management and regulators to search for ways to quantify and measure risk exposure. One answer came in the form of Value-at-Risk (VaR) being the

minimum loss a portfolio can have with a given probability over a specific time horizon (Jorion, 2001; Christoffersen, Hahn and Inoue, 2001).

The VaR of some portfolio (.) may be defined as a one-sided confidence interval of expected h-periods ahead losses:

$$\text{VaR}_{t+h,\zeta}^{(.)} = \Xi_t^{(.)}(1 + \bar{\xi}_{t+h,\zeta}), \tag{4.18}$$

where $\Xi_t^{(.)}$ is the value of a portfolio in time t and $\bar{\xi}_{t+h,\zeta}$ is a time dependent quantile of the conditional distribution of portfolio returns $\xi_{t+h}^{(.)}$ such that

$$\text{P}[\xi_{t+h}^{(.)} < \bar{\xi}_{t+h,\zeta}] = \zeta, \ \bar{\xi}_{t+h,\zeta} = \sigma_{t+h} z_\zeta, \tag{4.19}$$

and z_ζ is a quantile from an unconditional distribution with unit variance. In the light of the assumption of conditional normality in (4.1), we will take the quantiles z_ζ from the Gaussian distribution. As outlined in (4.18) and (4.19) the quantities $\bar{\xi}_{t+h,\zeta}$ and σ_{t+h} generally depend on the portfolio composition. For convenience our notation, however, does not indicate this relationship. Depending on the risk averseness of the agent the parameter ζ is typically chosen as some small probability, for instance, $\zeta = 0.005, 0.01, 0.05$.

4.5 An Empirical Illustration

4.5.1 Equal and Value Weighted Portfolios

We analyze portfolios comprised by all 30 stocks listed in the Dow Jones Industrial Average (DJIA) over the period Jan, 2nd, 1990 to Jan, 31st, 2005. Measured at the daily frequency, 3803 observations are used for the empirical analysis. Two alternative portfolio compositions are considered. In the first place we analyze a portfolio weighting each asset equally. Returns of this equal weight portfolio (EWP) are obtained from asset specific returns (ε_{it}, $i = 1, \ldots, N$) as

$$\xi_t^{(e)} = \sum_{i=1}^{N} w_{it}^{(e)} \varepsilon_{it}, \quad w_{it}^{(e)} = N^{-1}.$$

Secondly, we consider value weighted portfolios (VWP) determined as:

$$\xi_t^{(v)} = \sum_{i=1}^{N} w_{it}^{(v)} \varepsilon_{it}, \ w_{it}^{(v)} = w_{it-1}(1 + \varepsilon_{it-1})/w_t^{(v)}, \ w_t^{(v)} = \sum_i w_{it-1}(1 + \varepsilon_{it-1}).$$

Complementary to an analysis of EWP and VWP, dynamics of minimum variance portfolios (MVP) could also be of interest. The MVP, however, will

typically depend on some measure of the assets' volatilities and covariances. The latter, in turn, depend on the particular volatility model used for the analysis. Since the comparison of alternative measures of volatility in determining VaR is a key issue of this investigation we will not consider MVP to immunize our empirical results from impacts of volatility specific portfolio compositions.

Our empirical comparison of alternative approaches to implement VaR concentrates on the relative performance of one step ahead ex-ante evaluations of VaR ($h = 1$). Note that the (M)GARCH model specifies covariance matrices Σ_t or univariate volatilities σ_t^2 conditional on $\mathcal{F}_{t-1}$. Therefore, we practically consider the issue of two step ahead forecasting when specifying

$$\text{VaR}_{t+1,\zeta}^{(.)}|\mathcal{F}_{t-1} = \text{VaR}^{(.)}(\widehat{\sigma}_{t+1}^2), \ \widehat{\sigma}_{t+1}^2|\mathcal{F}_{t-1} = E[(\xi_{t+1}^{(.)})^2|\mathcal{F}_{t-1}].$$

The performance of alternative approaches to forecast VaR is assessed by means of the relative frequency of actual hits observed over the entire sample period, i.e.

$$\text{hf}_\zeta^{(.)} = \frac{1}{3800} \sum_{t=3}^{3802} \mathbf{1}(\xi_t^{(.)} < \bar{\xi}_{t,\zeta}), \tag{4.20}$$

where $\mathbf{1}(.)$ is an indicator function. To determine the forecasted conditional standard deviation entering the VaR we adopt three alternative strategies. As a benchmark, we consider standard deviation forecasts obtained from univariate GARCH processes fitted directly to the series of portfolio returns $\xi_t^{(.)}$. For the two remaining strategies, we exploit forecasts of the covariance matrix, $\widehat{\Sigma}_{t+1} = E[\varepsilon_{t+1}\varepsilon_{t+1}^\top|\mathcal{F}_{t-1}]$, to determine VaR. Note that given portfolio weights $w_t = (w_{1t}, w_{2t}, \ldots, w_{Nt})^\top$ the expected conditional variance of the portfolio is $\widehat{\sigma}_{t+1}^2 = w^\top\widehat{\Sigma}_{t+1}w$. Feasible estimates for the expected covariance matrix are determined alternatively by means of the CCC and DCC model.

The empirical exercises first cover a joint analysis of all assets comprising the DJIA. Moreover, we consider 1000 portfolios composed of 5 securities randomly drawn from all assets listed in the DJIA. Implementing the volatility parts of both the CCC and the DCC model, we employ alternatively the symmetric GARCH(1,1) and the threshold GARCH(1,1) model as introduced by Glosten, Jagannathan and Runkle (1993). Opposite to the symmetric GARCH model, the latter accounts for a potential leverage effect (Black, 1976) stating that volatility is larger in the sequel of bad news (negative returns) in comparison with good news (positive returns).

		G	TG	G		TG	
		N=30		N=5			
		Loss frequencies					
	$\zeta e + 03$	hf	hf	hf	s(hf)	hf	s(hf)
		EWP					
D	5.00	8.15	7.36	7.56	.033	7.13	.034
	10.0	13.2	12.4	11.7	.041	11.2	.042
	50.0	41.6	41.8	40.4	.075	40.3	.078
C	5.00	10.8	9.73	7.78	.034	7.36	.035
	10.0	14.2	14.2	11.9	.040	11.5	.042
	50.0	42.6	41.8	40.8	.074	40.7	.077
U	5.00	11.6	11.6	8.70	.036	8.36	.037
	10.0	14.7	14.7	13.2	.045	12.9	.045
	50.0	47.3	47.3	43.5	.076	44.0	.077
		VWP					
D	5.00	6.58	7.10	7.86	.033	7.55	.033
	10.0	12.9	11.8	11.9	.043	11.6	.041
	50.0	41.6	40.5	40.3	.076	40.4	.078
C	5.00	9.21	9.21	8.18	.036	7.90	.035
	10.0	14.5	13.4	12.3	.043	12.1	.043
	50.0	42.6	41.8	41.1	.072	41.3	.071
U	5.00	9.99	9.99	8.71	.037	8.62	.035
	10.0	15.5	15.5	13.0	.048	12.9	.048
	50.0	43.7	43.7	42.6	.095	43.2	.098
		Estimation results					
D	$\widehat{\alpha}$	2.8e-03	2.8e-03	6.6e-03	4.5e-05	6.7e-03	4.8e-05
	t_α	17.5	17.3				
	$\widehat{\beta}$	.992	.992	.989	8.3e-05	.989	9.5e-05
	t_β	1.8e+03	1.8e+03				

Table 4.1. Estimation results and performance of VaR estimates. G and TG are short for GARCH(1,1) and TGARCH(1,1) models for asset specific volatilities, respectively. D,C and U indicate empirical results obtained from DCC, CCC and univariate GARCH(1,1) models applied to evaluate forecasts of conditional variances of equal weight (EWP) and value weighted portfolios (VWP). Entries in hf and s(hf) are relative frequencies of extreme losses and corresponding standard errors, respectively.
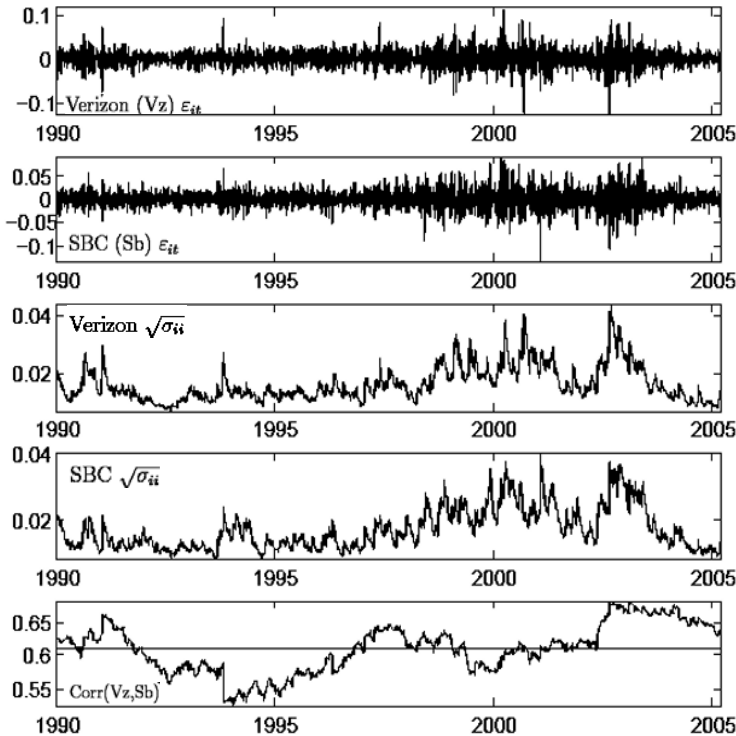
Figure 4.1. Returns, conditional volatilities and correlations for Verizon and SBC communications.

4.5.2 Estimation Results

A few selected estimation results are given in Table 4.1. Since we investigate 30 assets or 1000 random portfolios each containing $N = 5$ securities we refrain from providing detailed results on univariate GARCH(1,1) or TGARCH(1,1) estimates. Moreover, we leave estimates of the unconditional correlation matrix R undocumented since the number of possible correlations in our sample is $N(N-1)/2 = 435$.

The lower left part of Table 4.1 provides estimates of the DCC parameters α and β and corresponding t-ratios for the analysis of all assets comprising the DJIA. Although the estimated α parameter governing the impact of lagged GARCH innovations on the conditional correlation matrix is very small (around $2.8 \cdot 10^{-3}$ for both implementations of the DCC model) it is significant at any reasonable significance level. The relative performance of
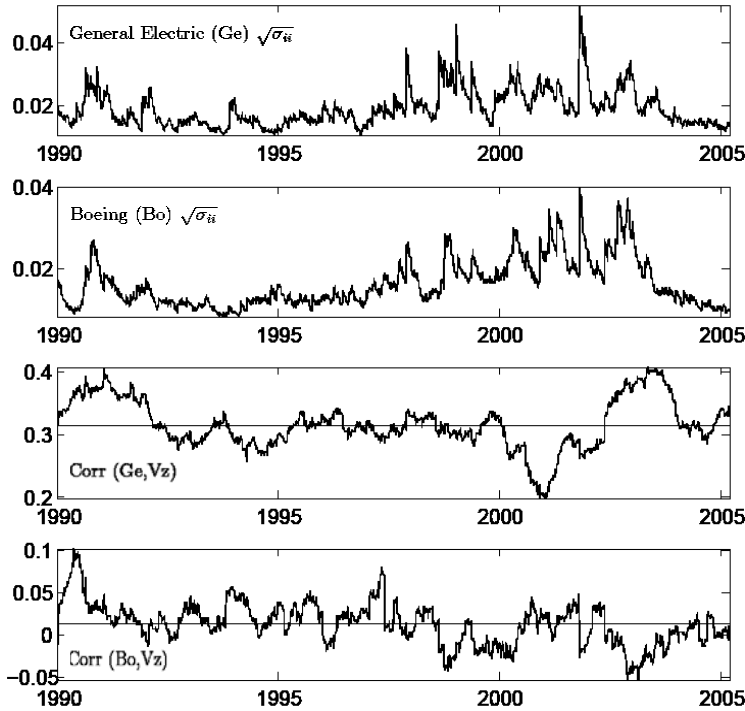
Figure 4.2. Conditional volatilities for General Electric and Boeing and conditional correlations with Verizon.

the CCC and DCC model may also be evaluated in terms of the models' log likelihood difference. Using symmetric and asymmetric volatility models for the diagonal elements of Σ_t the log likelihood difference between DCC and CCC is 645.66 and 622.00, respectively. Since the DCC specification has only two additional parameters, it apparently provides a substantial improvement of fitting multivariate returns. It is also instructive to compare, for the DCC case say, the log likelihood improvement achieved when employing univariate TGARCH instead of a symmetric GARCH. Interestingly, implementing the DCC model with asymmetric GARCH the improvement of the log likelihood is 'only' 236.27, which is to be related to the number of $N = 30$ additional model parameters. Reviewing the latter two results one may conclude that dynamic correlation is a more striking feature of US stock market returns than leverage.

The sum of both DCC parameter estimates, $\widehat{\alpha} + \widehat{\beta}$, is slightly below unity and, thus, the estimated model of dynamic covariances is stationary. The
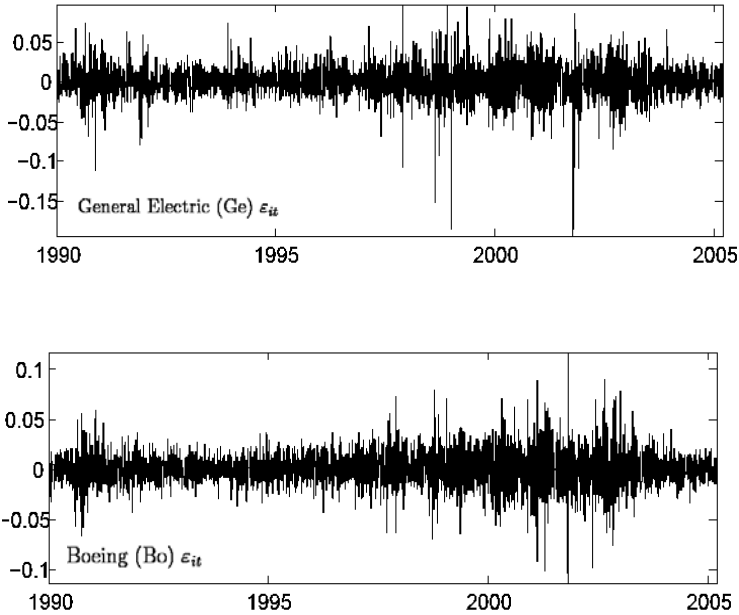
Figure 4.3. Returns for General Electric and Boeing

lower right part of Table 4.1 gives average estimates obtained for the DCC parameters when modeling 1000 portfolios randomly composed of five securities contained in the DJIA. We also provide an estimator of the empirical standard error associated with the latter average. Irrespective of using a symmetric or asymmetric specification of univariate volatility models, estimates for α are small throughout. According to the reported standard error estimates, however, the true α parameter is apparently different from zero at any reasonable significance level.

The maximum over all 435 unconditional correlations is obtained for two firms operating on the telecommunication market, namely Verizon Communications and SBC Communications. To illustrate the performance of the DCC model and compare it with the more restrictive CCC counterpart, Figure 4.1 provides the return processes for these two assets, the corresponding time paths of conditional standard deviations as implied by TGARCH(1,1) models and the estimated time paths of conditional correlations implied by the DCC model fitted over all assets contained in the DJIA. Facilitating the interpretation of the results, we also give the level of unconditional correlation.

Apparently, the univariate volatility models provide accurate descriptions of the return variability for both assets. Not surprisingly, estimated volatility turns out to be larger over the last third of the sample period in comparison with the first half. Although conditional correlation estimates vary around their unconditional level, the time path of correlation estimates exhibits only rather slow mean reversion. Interestingly, over the last part of the sample period the conditional correlation measured between Verizon and SBC increases with the volatilities of both securities.

As mentioned, Verizon and SBC provide the largest measure of unconditional correlation within the DJIA over the considered sample period. To illustrate that time varying conditional correlation with slow mean reversion is also an issue for bivariate returns exhibiting medium or small correlation we provide the conditional correlation estimates for Verizon and General Electric (medium unconditional correlation) and Verizon and Boeing (small unconditional correlation) in Figure 4.2. For completeness Figure 4.3 provides empirical return processes for General Electric and Boeing.

The upper part of Table 4.1 shows relative frequencies of realized losses exceeding the one step ahead ex-ante VaR forecasts. We provide average relative frequencies when summarizing the outcome for 1000 portfolios with random composition. To facilitate the discussion of the latter results all frequencies given are multiplied with a factor of 1000.

The relative frequency of empirical hits of dynamic VaR estimates at the 5% level is uniformly below the nominal probability, indicating that dynamic VaR estimates are too conservative on average. For the remaining probability levels $\zeta = 0.5\%$ and $\zeta = 1\%$ the empirical frequencies of hitting the VaR exceed the nominal probability. We concentrate the discussion of empirical results on the latter cases. With regard to the performance of alternative implementations of VaR it is worthwhile to mention that the basic results are qualitatively similar for EWP in comparison with VWP. Similarly, employing an asymmetric GARCH model instead of symmetric GARCH has only minor impacts on the model comparison between the univariate benchmark and the CCC and DCC model, respectively. For the latter reason, we focus our discussion of the relative model performance on VaR modeling for EWP with symmetric GARCH(1,1) applied to estimate conditional variances.

Regarding portfolios composed of 30 securities, it turns out that for both probability levels $\zeta = 1\%$ and $\zeta = 0.5\%$ the empirical frequencies of hitting the dynamic VaR estimates are closest to the nominal level for the DCC model and worst for modeling portfolio returns directly via univariate GARCH. Although it provides the best empirical frequencies of hitting the VaR, the DCC model still underestimates (in absolute value) on average the true quantile.

For instance, the 0.5% VaR shows an empirical hit frequency of 0.82% (EWP) and 0.66% (VWP), respectively. Drawing randomly 5 out of 30 assets to form portfolios, and regarding the average empirical frequencies of hitting the VaR estimates, we obtain almost analogous results in comparison with the case $N = 30$. The reported standard errors of average frequencies, however, indicate that the discussed differences of nominal and empirical probabilities are significant at a 5% significance level since the difference between both exceeds twice the standard error estimates.

In summary, using the CCC and DCC model and, alternatively, univariate GARCH specifications to determine VaR, it turns out that the former outperform the univariate GARCH as empirical loss frequencies are closer to the nominal VaR coverage. DCC based VaR estimates in turn outperform corresponding quantities derived under the CCC assumption. Empirical frequencies of large losses, however, exceed the corresponding nominal levels if the latter are rather small, i.e. 0.5% and 1%. This might indicate that the DCC framework is likely to restrictive to hold homogeneously over a sample period of the length (more than 15 years) considered in this work. More general versions of dynamic correlation models are available but allowance of asset specific dynamics requires simultaneous estimation of $O(N)$ parameters.

Bibliography

Alexander, C.O. (1998). Volatility and Correlation: Methods, Models and Applications, in Alexander (ed.), *Risk Management and Analysis: Measuring and Modelling Financial Risk*, New York: John Wiley.

Alexander, C.O. (2001). *Orthogonal GARCH*, in Alexander (ed.), *Mastering Risk*, Volume II, 21–38, Prentice Hall.

Baba, Y., Engle, R.F., Kraft, D.F. and Kroner, K.F. (1990). Multivariate simultaneous generalized ARCH, *mimeo, Department of Economics, University of California, San Diego*.

Bauwens, L., Laurent, S. and Rombouts, J.V.K. (2006). Multivariate GARCH models: A survey, *Journal of Applied Econometrics* **31**: 79–109.

Black, F. (1976). Studies of Stock Market Volatility Changes; *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177–181.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* **31**: 307-327.

Bollerslev, T. (1990). Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach, *Review of Economics and Statistics* **72**: 498–505.

Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). GARCH Models, in: Engle, R.F., and McFadden, D.L. (eds.) Handbook of Econometrics, Vol. 4, Elsevier, Amsterdam, 2961–3038.

Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988). A Capital Asset Pricing Model with Time-Varying Covariances, *Journal of Political Economy* **96**: 116–131.

Bollerslev, T. and Wooldridge, J.M. (1992). Quasi–Maximum Likelihood Estimation and Inference in Dynamic Models with Time–Varying Covariances, *Econometric Reviews* **11**: 143–172.

Cappiello, L., Engle, R.F. and Sheppard, K. (2003). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns, European Central Bank Working paper No. 204.

Cecchetti, S.G., Cumby, R.E. and Figlewski, S. (1988). Estimation of the Optimal Futures Hedge, *Review of Economics and Statistics* **70**: 623-630.

Christoffersen, P., Hahn, J. and Inoue, A. (2001). Testing, Comparing and Combining Value at Risk Measures, *Journal of Empirical Finance* **8**, 325–342.

Comte, F. and Lieberman, O. (2003). Asymptotic Theory for Multivariate GARCH Processes, *Journal of Multivariate Analysis* **84**, 61–84.

Ding, Z. and Engle, R.F. (1994). Large Scale Conditional Covariance Matrix Modeling, Estimation and Testing, *mimeo*, UCSD.

Duan, J.C. (1995). The GARCH Option Pricing Model, *Mathematical Finance* **5**: 13–32.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation, *Econometrica* **50**: 987–1008.

Engle, R.F. (2002). Dynamic Conditional Correlation - A Simple Class of Multivariate GARCH Models, *Journal of Business and Economic Statistics* **22**: 367–381.

Engle, R.F. and Kroner, K.F. (1995). Multivariate Simultaneous Generalized ARCH, *Econometric Theory* **11**: 122–150.

Engle, R.F. and Mezrich, J. (1996). GARCH for Groups, *Risk* **9**: 36–40.

Engle, R.F. and Sheppard, K. (2001). Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH, UCSD DP 2001-15.

Engle, R.F., Hendry, D.F. and Richard, J.F. (1983). Exogeneity, *Econometrica* **51**: 277–304.

Engle, R.F., Ito, T. and Lin, W.L. (1990). Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market, *Econometrica* **58**: 525–542.

Engle, R.F., Ng, V.M. and Rothschild, M. (1990). Asset Pricing with a Factor-ARCH Structure: Empirical Estimates for Treasury Bills, *Journal of Econometrics* **45**: 213–237.

Glosten, L.R., Jagannathan, R. and Runkle, D.E. (1993). On the Relation between the Expected value and the Volatility of the Normal Excess Return on Stocks, *Journal of Finance* **48**: 1779–1801.

Hafner, C.M. and Franses, P.H. (2003). A Generalized Dynamic Conditional Correlation Model for Many Assets, *Econometric Institute Report* **18**, Erasmus University Rotterdam.

Hafner, C.M. and Herwartz, H. (1998). Structural Analysis of Portfolio Risk using Beta Impulse Response Functions, *Statistica Neerlandica* **52**: 336-355.

Hafner, C.M. and Herwartz, H. (2007). Analytical Quasi Maximum Likelihood Inference in Multivariate Volatility Models, *Metrika*, forthcoming.

Hafner, C.M., van Dijk, D. and Franses, P.H. (2006). Semi-Parametric Modelling of Correlation Dynamics, in Fomby and Hill (eds.), *Advances in Econometrics* (Vol. 20) Part A, 51–103, New York: Elsevier Science.

Hamao, Y., Masulis, R.W. and Ng, V.K. (1990). Correlations in Price Changes and Volatility across International Stock Markets, *Review of Financial Studies* **3**: 281–307.

Jeantheau, T. (1998). Strong Consistency of Estimators for Multivariate ARCH Models, *Econometric Theory* **14**: 70-86.

Jorion, P. (2001) Value at Risk: the New Benchmark for Managing Financial Risk, 2nd ed, New York, McGraw Hill.

Lee, S.W. and Hansen, B.E. (1994). Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator, *Econometric Theory* **10**: 29–52.

Lee, T.H. and Long, X. (2008). Copula Based Multivariate GARCH Model with Uncorrelated Dependent Errors, *Journal of Econometrics* forthcoming.

Ling, S. and McAleer, M. (2003). Asymptotic Theory for a Vector ARMA-GARCH Model, *Econometric Theory* **19**: 280–309.

Lütkepohl, H. (1996). *Handbook of Matrices*, Wiley, Chichester.

Lumsdaine, R. (1996). Consistency and Asymptotic Normality of the Quasi-Maximum Likelihood Estimator in IGARCH(1,1) and Covariance Stationary GARCH(1,1) Models, *Econometrica* **64**: 575–596.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* **59**: 347–370.

Newey, W. and McFadden, D. (1994). Large Sample Estimation and Hypothesis Testing, in Engle and McFadden (eds.), *Handbook of Econometrics* (Vol. 4), New York: Elsevier Science, 2113-2245.

Ruud, P.A. (2000). An Introduction to Classical Econometric Theory, University of California, Berkeley.

Tse, Y.K. (2000). A Test for Constant Correlations in a Multivariate GARCH Model, *Journal of Econometrics* **98**: 107–127.

Tse, Y.K. and Tsui, A.K.C. (2002). A Multivariate GARCH Model with Time-Varying Correlations, *Journal of Business and Economic Statistics* **20**: 351–362.

van der Weide, R. (2002). GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model, *Journal of Applied Econometrics* **17**: 549–564.

Part II

Credit Risk

5 Rating Migrations

Steffi Höse, Stefan Huschens and Robert Wania

The bond rating is one of the most important indicators of a corporation's credit quality and therefore its default probability. It was first developed by Moody's in 1914 and by Poor's Corporation in 1922 and it is generally assigned by external agencies to publicly traded debts (Altman and Kao, 1992). Apart from the external ratings by independent rating agencies, there are internal ratings by banks and other financial institutions (Basel Committee on Banking Supervision, 2006). External rating data by agencies are available for many years, in contrast to internal ratings. Their short history in most cases does not exceed 5–10 years. Both types of ratings are usually recorded on an ordinal scale and labeled alphabetically or numerically. For the construction of a rating system see Crouhy, Galai and Mark (2001).

A change in a rating reflects the assessment that the company's credit quality has improved (upgrade) or deteriorated (downgrade). Analyzing these rating migrations including default is one of the preliminaries for credit risk models in order to measure future credit loss. In such models, the matrix of rating transition probabilities, the so called transition matrix, plays a crucial role. It allows the calculation of the joint distribution of future ratings for borrowers in a portfolio (Gupton, Finger and Bhatia, 1997). An element of a transition matrix gives the probability that an obligor with a certain initial rating migrates to another rating by the risk horizon. For the econometric analysis of transition data see Lancaster (1990).

In a study by Jarrow, Lando and Turnbull (1997) rating transitions were modeled as a time-homogeneous Markov chain, so future rating changes are not affected by the rating history (Markov property). The probability of changing from one rating to another is constant over time (homogeneous), which is assumed solely for simplicity of estimation. Empirical evidence indicates that transition probabilities are time-varying. Nickell, Perraudin and Varotto (2000) show that different transition matrices are identified across various factors such as the obligor's domicile and industry and the stage of business cycle. The latter has also been studied by Lando and Skødeberg (2002), Bangia, Diebold, Kronimus, Schagen and Schuermann (2002) and Krüger, Stötzel and Trück (2005).

Rating migrations are reviewed from a statistical point of view throughout this chapter. The way from the observed data to the estimated one-year transition probabilities is shown and estimates for the standard deviations of the transition rates are given. In further extension, dependent rating migrations are discussed. In particular, the modeling by a threshold normal model is presented.

Time stability of transition matrices is one of the major issues for credit risk estimation. Therefore, a chi-square test of homogeneity for the estimated rating transition probabilities is applied. The test is illustrated by an example and is compared to a simpler approach using standard errors. Further, assuming time stability, multi-period rating transitions are discussed. An estimator for multi-period transition matrices is given and its distribution is approximated by bootstrapping. Finally, the change of the composition of a credit portfolio caused by rating migrations is considered. The expected composition and its variance is calculated for independent migrations.

To demonstrate the computational solution of the described problems, program code for some examples is provided. The statistical software GAUSS is used, see www.aptech.com. Example programs and data can be downloaded from www.tu-dresden.de/wwqvs/f-publ.htm.

This chapter is an updated version of Höse, Huschens and Wania (2002) with minor changes and corrections. Additional literature has been included.

5.1 Rating Transition Probabilities

In this section, the way from raw data to estimated rating transition probabilities is described. First, migration events of the same kind are counted. The resulting migration counts are transformed into migration rates, which are used as estimates for the unknown transition probabilities. These estimates are complemented with estimated standard errors for two cases, for independence and for a special correlation structure.

5.1.1 From Credit Events to Migration Counts

It is assumed that credits or credit obligors are rated in d categories ranging from 1, the best rating category, to the category d containing defaulted credits. The raw data consist of a collection of *migration events*. The n observed

migration events form a $n \times 2$ matrix with rows

$$(e_{i1}, e_{i2}) \in \{1, \dots, d-1\} \times \{1, \dots, d\}, \quad i = 1, \dots, n. \tag{5.1}$$

Thereby, e_{i1} characterizes the rating of i-th credit at the beginning and e_{i2} the rating at the end of the risk horizon, which is usually one year. Subsequently, migration events of the same kind are aggregated in a $(d-1) \times d$ matrix $\mathbf{C}$ of *migration counts*, where the generic element

$$c_{jk} \stackrel{\text{def}}{=} \sum_{i=1}^{n} \mathbf{1}\{(e_{i1}, e_{i2}) = (j, k)\} \tag{5.2}$$

is the number of migration events from j to k. Thereby, $\mathbf{1}\{\cdot\}$ denotes the indicator function, which is one if the logical expression in brackets is true and otherwise zero. Clearly, their total sum is

$$\sum_{j=1}^{d-1} \sum_{k=1}^{d} c_{jk} = n.$$

5.1.2 Estimating Rating Transition Probabilities

It is assumed that the observations e_{i1} and e_{i2} are realizations of the random variables $\widetilde{e}_{i1}$ and $\widetilde{e}_{i2}$. In the following, only the conditional probability distribution

$$p_{jk} \stackrel{\text{def}}{=} \mathrm{P}(\widetilde{e}_{i2} = k | \widetilde{e}_{i1} = j), \quad \sum_{k=1}^{d} p_{jk} = 1$$

is of interest, where p_{jk} is the probability that a credit migrates from an initial rating $j = 1, \dots, d-1$ to rating $k = 1, \dots, d$. These probabilities are the so called *rating transition (or migration) probabilities*. Note that the distribution of the indicator variable $\mathbf{1}\{\widetilde{e}_{i2} = k\}$, conditional on $\widetilde{e}_{i1} = j$, is a Bernoulli distribution with success parameter p_{jk},

$$\mathbf{1}\{\widetilde{e}_{i2} = k\} \,|\, \widetilde{e}_{i1} = j \ \sim \ \mathrm{Ber}(p_{jk}). \tag{5.3}$$

In order to estimate these rating transition probabilities the number of migrations starting from rating j are defined as

$$n_j = \sum_{k=1}^{d} c_{jk}, \quad j = 1, \dots, d-1. \tag{5.4}$$

In the following it is assumed that a fixed vector of initial ratings $(e_{11}, \dots, e_{n1})$ is given with $n_j > 0$ for $j = 1, \dots, d-1$. Thus, all following probabilities

are conditional probabilities given $(\widetilde{e}_{11}, \ldots, \widetilde{e}_{n1}) = (e_{11}, \ldots, e_{n1})$. The composition of the portfolio at the beginning of the period is then given by $(n_1, \ldots, n_{d-1})$ and

$$\left(\sum_{j=1}^{d-1} c_{j1}, \ldots, \sum_{j=1}^{d-1} c_{jd} \right) \tag{5.5}$$

is the composition of the portfolio at the end of the period, where the last element is the *number of defaulted credits*. The observed *migration rate* from j to k,

$$\widehat{p}_{jk} \stackrel{\text{def}}{=} \frac{c_{jk}}{n_j}, \tag{5.6}$$

is the natural estimate of the unknown transition probability p_{jk}.

If the migration events are independent, i. e., the variables $\widetilde{e}_{12}, \ldots, \widetilde{e}_{n2}$ are stochastically independent, c_{jk} is the observed value of the binomially distributed random variable

$$\widetilde{c}_{jk} \sim \mathrm{B}(n_j, p_{jk}),$$

and therefore the standard deviation of $\widehat{p}_{jk}$ is

$$\sigma_{jk} = \sqrt{\frac{p_{jk}(1 - p_{jk})}{n_j}},$$

which may be estimated by

$$\widehat{\sigma}_{jk} = \sqrt{\frac{\widehat{p}_{jk}(1 - \widehat{p}_{jk})}{n_j}}. \tag{5.7}$$

The estimated standard errors must be carefully interpreted, because they are based on the assumption of independence.

5.1.3 Dependent Migrations

The case of dependent rating migrations raises new problems. In this context, $\widetilde{c}_{jk}$ is distributed as sum of n_j correlated Bernoulli variables, see (5.3), indicating for each credit with initial rating j a migration to k by 1. If these Bernoulli variables are pairwise correlated with correlation ρ_{jk}, then the variance σ_{jk}^2 of the unbiased estimator $\widehat{p}_{jk}$ for p_{jk} is (Huschens and Locarek-Junge, 2002, p. 108)

$$\sigma_{jk}^2 = \frac{p_{jk}(1 - p_{jk})}{n_j} + \frac{n_j - 1}{n_j} \rho_{jk} p_{jk}(1 - p_{jk}).$$

The limit

$$\lim_{n_j \to \infty} \sigma_{jk}^2 = \rho_{jk} p_{jk} (1 - p_{jk})$$

shows that the sequence $\widehat{p}_{jk}$ does not obey a law of large numbers for $\rho_{jk} > 0$. Generally, the failing of convergence in quadratic mean does not imply the failing of convergence in probability. But in this case all moments of higher order exist since the random variable $\widehat{p}_{jk}$ is bounded and so the convergence in probability implies the convergence in quadratic mean. For $\rho_{jk} = 0$ the law of large numbers holds. Negative correlations can only be obtained for finite n_j. The lower boundary for the correlation is given by $\rho_{jk} \geq -\frac{1}{n_j - 1}$, which converges to zero when the number of credits n_j grows to infinity.

The law of large numbers fails also if the correlations are different with either a common positive lower bound, or non-vanishing positive average correlation or constant correlation blocks with positive correlations in each block (Finger, 1998, p. 5). This failing of the law of large numbers may not surprise a time series statistician, who is familiar with mixing conditions to ensure mean ergodicity of stochastic processes (Davidson, 1994, chapter 14). In statistical words, in the case of non-zero correlation the relative frequency is not a consistent estimator of the Bernoulli parameter.

The parameters ρ_{jk} may be modeled in consistent way in the framework of a *threshold normal model* with a single parameter ρ (Basel Committee on Banking Supervision, 2006; Gupton et al., 1997; Kim, 1999; McNeil, Frey and Embrechts, 2005). This model specifies a special dependence structure based on a standard multinormal distribution for a vector $(R_1, \ldots, R_n)$ with equicorrelation matrix (Mardia, Kent and Bibby, 1979, p. 461), where R_i $(i = 1, \ldots, n)$ is the standardized asset return of obligor i and n is the number of obligors in the portfolio. The parameter $\rho \geq 0$ may be interpreted as a mean asset return correlation. In this model each pair of variables $(X, Y) = (R_i, R_{i'})$ with $i, i' = 1, \ldots, n$ and $i \neq i'$ is bivariate normally distributed with density function

$$\varphi(x, y; \rho) = \frac{1}{2\pi\sqrt{1 - \rho^2}} \exp\left\{ -\frac{x^2 - 2\rho xy + y^2}{2(1 - \rho^2)} \right\}.$$

The probability $P[(X, Y) \in (a, b)^2]$ is given by

$$\beta(a, b; \rho) = \int_a^b \int_a^b \varphi(x, y; \rho)\, dx\, dy. \tag{5.8}$$

Thresholds for initial rating j are derived from $p_{j1}, \ldots, p_{j,d-1}$ by

$$z_{j0} \stackrel{\text{def}}{=} -\infty,\ z_{j1} \stackrel{\text{def}}{=} \Phi^{-1}(p_{j1}),\ z_{j2} \stackrel{\text{def}}{=} \Phi^{-1}(p_{j1} + p_{j2}), \ldots, z_{jd} \stackrel{\text{def}}{=} +\infty,$$

where Φ is the distribution function of the standard normal distribution and Φ^{-1} it's inverse. Each credit with initial rating j is characterized by a normally distributed variable Z which determines the migration events from j to k by

$$p_{jk} = \mathrm{P}\{Z \in (z_{j,k-1}, z_{jk})\} = \Phi(z_{jk}) - \Phi(z_{j,k-1}).$$

The *simultaneous* transition probabilities of two credits i and i' from category j to k are given by

$$p_{jj:kk} = \mathrm{P}(\widetilde{e}_{i2} = \widetilde{e}_{i'2} = k | \widetilde{e}_{i1} = \widetilde{e}_{i'1} = j) = \beta(z_{j,k-1}, z_{jk}; \rho),$$

i.e., the probability of simultaneous default is

$$p_{jj:dd} = \beta(z_{j,d-1}, z_{jd}; \rho).$$

For a detailed example see Saunders (1999, pp. 122-125). In the special case of independence $p_{jj:kk} = p_{jk}^2$ holds true. Defining a migration from j to k as success correlated Bernoulli variables are obtained with common success parameter p_{jk}, with probability $p_{jj:kk}$ of a simultaneous success, and with the *migration correlation*

$$\rho_{jk} = \frac{p_{jj:kk} - p_{jk}^2}{p_{jk}(1 - p_{jk})}.$$

Note that $\rho_{jk} = 0$ if $\rho = 0$.

Given $\rho \geq 0$ the migration correlation $\rho_{jk} \geq 0$ can be estimated by the restricted Maximum-Likelihood estimator

$$\widehat{\rho}_{jk} = \max\left\{0; \frac{\beta(\widehat{z}_{j,k-1}, \widehat{z}_{jk}; \rho) - \widehat{p}_{jk}^2}{\widehat{p}_{jk}(1 - \widehat{p}_{jk})}\right\} \tag{5.9}$$

with

$$\widehat{z}_{jk} = \Phi^{-1}\left(\sum_{i=1}^{k} \widehat{p}_{ji}\right). \tag{5.10}$$

The estimate

$$\widehat{\sigma}_{jk} = \sqrt{\frac{\widehat{p}_{jk}(1 - \widehat{p}_{jk})}{n_j} + \frac{n_j - 1}{n_j}\widehat{\rho}_{jk}\widehat{p}_{jk}(1 - \widehat{p}_{jk})} \tag{5.11}$$

of the standard deviation

$$\sigma_{jk} = \sqrt{\frac{p_{jk}(1 - p_{jk})}{n_j} + \frac{n_j - 1}{n_j}\rho_{jk}p_{jk}(1 - p_{jk})}$$

is used. The estimator in (5.11) generalizes (5.7), which results in the special case $\rho = 0$.

5.1.4 Computational Aspects

The procedure `RatMigCounts` can be used to compute migration counts from migration events. For the application, the migration events (5.1) have to be stored within an $n \times 2$ data matrix containing n migration events. The result of `RatMigCounts` is the $(d-1) \times d$ matrix of migration counts with elements given in (5.2). An example is given by `RatMigExample1.gau`.

The procedure `RatMigRate` computes migration rates and related estimated standard errors for m periods from an $(d-1) \times d \times m$ array of m-period migration counts and a given non-negative correlation parameter. The calculation uses stochastic integration in order to determine the probability β from (5.8). The accuracy of the applied Monte Carlo procedure is controlled by the input parameter `s`. For $s > 0$ the number m of Monte Carlo replications is at least $(2s)^{-2}$. This guarantees that the user-specified value `s` is an upper bound for the standard deviation of the Monte Carlo estimator for β. Note that with increasing accuracy (i. e. decreasing `s`) the computational effort increases proportionally to m. The output contains matrices `nstart` and `nend` which components are given by (5.4) and (5.5). The matrices `etp`, `emc`, and `esd` contain the $\widehat{p}_{jk}$, $\widehat{\rho}_{jk}$, and $\widehat{\sigma}_{jk}$ from (5.6), (5.9), and (5.11) for $j = 1, \ldots, d-1$ and $k = 1, \ldots, d$. The estimates $\widehat{\rho}_{jk}$ are given only for $\widehat{p}_{jk} > 0$. The matrix `etv` contains the $\widehat{z}_{jk}$ from (5.10) for $j, k = 1, \ldots, d-1$. Note that $z_{j0} = -\infty$ and $z_{jd} = +\infty$. An example is given by `RatMigExample2.gau`.

5.2 Analyzing the Time-Stability of Transition Probabilities

5.2.1 Aggregation over Periods

It is assumed that migration data are given for m periods. This data consist in m matrices of migration counts $\mathbf{C}(t)$ for $t = 1, \ldots, m$ each of type $(d-1) \times d$. The generic element $c_{jk}(t)$ of the matrix $\mathbf{C}(t)$ is the number of migrations from j to k in period t. These matrices may be computed from m data sets of migration events.

An obvious question in this context is whether the transition probabilities can be assumed to be time-invariant or not. A first approach to analyze the time-stability of transition probabilities is to compare the estimated transition probabilities per period for m periods with estimates from pooled data.

The *aggregated migration counts* from m periods are

$$c_{jk}^+ \stackrel{\text{def}}{=} \sum_{t=1}^{m} c_{jk}(t) \tag{5.12}$$

which are combined in the matrix

$$\mathbf{C}^+ \stackrel{\text{def}}{=} \sum_{t=1}^{m} \mathbf{C}(t)$$

of type $(d-1) \times d$. The migration rates computed per period

$$\widehat{p}_{jk}(t) \stackrel{\text{def}}{=} \frac{c_{jk}(t)}{n_j(t)}, \quad t = 1, \ldots, m \tag{5.13}$$

with

$$n_j(t) \stackrel{\text{def}}{=} \sum_{k=1}^{d} c_{jk}(t)$$

have to be compared with the migration rates from the pooled data. Based on the aggregated migration counts the estimated transition probabilities

$$\widehat{p}_{jk}^+ \stackrel{\text{def}}{=} \frac{c_{jk}^+}{n_j{}^+} \tag{5.14}$$

with

$$n_j^+ \stackrel{\text{def}}{=} \sum_{k=1}^{d} c_{jk}^+ = \sum_{t=1}^{m} n_j(t), \quad j = 1, \ldots, d-1$$

can be computed.

5.2.2 Testing the Time-Stability of Transition Probabilities

Under the assumption of *independence* for the migration events the vector of migration counts $(c_{j1}(t), \ldots c_{jd}(t))$ starting from j is in each period t a realization from a multinomial distributed random vector

$$(\widetilde{c}_{j1}(t), \ldots, \widetilde{c}_{jd}(t)) \sim \text{Mult}\{n_j(t); p_{j1}(t), \ldots, p_{jd}(t)\},$$

where $p_{jk}(t)$ denotes the transition probability from j to k in period t. For fixed $j \in \{1, \ldots, d-1\}$ the hypothesis of homogeneity

$$H_0 : p_{j1}(1) = \ldots = p_{j1}(m), p_{j2}(1) = \ldots = p_{j2}(m), \ldots, p_{jd}(1) = \ldots = p_{jd}(m)$$

may be tested (cp. Anderson and Goodman (1957)) with the statistic

$$X_j^2 = \sum_{k=1}^{d} \sum_{t=1}^{m} \frac{\left\{\widetilde{c}_{jk}(t) - n_j(t)\widehat{p}_{jk}^+\right\}^2}{n_j(t)\widehat{p}_{jk}^+}. \tag{5.15}$$

This statistic is asymptotically χ^2-distributed with $(d-1)(m-1)$ degrees of freedom under H_0. H_0 is rejected with approximative level α if the statistic computed from the data is greater than the $(1-\alpha)$-quantile of the χ^2-distribution with $(d-1)(m-1)$ degrees of freedom.

The combined hypothesis of homogeneity

$$H_0 : p_{jk}(t) = p_{jk}(m), \quad t = 1, \ldots, m-1, \quad j = 1, \ldots, d-1, \quad k = 1, \ldots, d$$

means that the matrix of transition probabilities is constant over time. Therefore, the combined null hypothesis may equivalently be formulated as

$$H_0 : \mathbf{P}(1) = \mathbf{P}(2) = \ldots = \mathbf{P}(m),$$

where $\mathbf{P}(t)$ denotes the transition matrix at t with generic element $p_{jk}(t)$. This hypothesis may be tested using the statistic

$$X^2 = \sum_{j=1}^{d-1} X_j^2, \tag{5.16}$$

which is under H_0 asymptotically χ^2-distributed with $(d-1)^2(m-1)$ degrees of freedom. The combined null hypothesis is rejected with approximative level α if the computed statistic is greater than the $(1-\alpha)$-quantile of the χ^2-distribution with $(d-1)^2(m-1)$ degrees of freedom (Bishop, Fienberg and Holland, 1975, p. 265).

This approach creates two problems. Firstly, the two tests are based on the assumption of independence. Secondly, the test statistics are only asymptotically χ^2-distributed. This means that sufficiently large sample sizes are required. A rule of thumb given in the literature is $n_j(t)\widehat{p}_{jk}^+ \geq 5$ for all j and k which is hardly fulfilled in the context of credit migrations.

The two χ^2-statistics in (5.15) and (5.16) are of the Pearson type. Two other frequently used and asymptotically equivalent statistics are the corresponding χ^2-statistics of the Neyman type

$$Y_j^2 = \sum_{k=1}^{d} \sum_{t=1}^{m} \frac{\left\{\widetilde{c}_{jk}(t) - n_j(t)\widehat{p}_{jk}^+\right\}^2}{\widetilde{c}_{jk}(t)}, \quad Y^2 = \sum_{j=1}^{d-1} Y_j^2$$

and the χ^2-statistics

$$G_j^2 = 2 \sum_{k=1}^{d} \sum_{t=1}^{m} \widetilde{c}_{jk}(t) \log \left[\frac{\widetilde{c}_{jk}(t)}{n_j(t)\widehat{p}_{jk}^+} \right], \quad G^2 = \sum_{j=1}^{d-1} G_j^2,$$

which results from Wilks log-likelihood ratio.

Considering the strong assumptions on which these test procedures are based on, a simpler approach, complementing the point estimates $\widehat{p}_{jk}(t)$ by estimated standard errors

$$\widehat{\sigma}_{jk}(t) = \sqrt{\frac{\widehat{p}_{jk}(t)\{1 - \widehat{p}_{jk}(t)\}}{n_j(t)}}$$

for each period $t \in \{1, \ldots, m\}$, may be preferable. For correlated migrations the estimated standard deviation is computed analogously to (5.11).

5.2.3 Example

The following example is based on transition matrices given by Nickell et al. (2000, pp. 208, 213). The data set covers long-term bonds rated by Moody's in the period 1970–1997. Instead of the original matrices of type 8×9 condensed matrices of type 3×4 are used by combining the original data in the $d = 4$ basic rating categories A, B, C, and D, where D stands for the category of defaulted credits.

The aggregated data for the full period from 1970 to 1997 are

$$\mathbf{C} = \begin{bmatrix} 21726 & 790 & 0 & 0 \\ 639 & 21484 & 139 & 421 \\ 0 & 44 & 307 & 82 \end{bmatrix}, \quad \widehat{\mathbf{P}} = \begin{bmatrix} 0.965 & 0.035 & 0 & 0 \\ 0.028 & 0.947 & 0.006 & 0.019 \\ 0 & 0.102 & 0.709 & 0.189 \end{bmatrix},$$

where $\mathbf{C}$ is the matrix of migration counts and $\widehat{\mathbf{P}}$ is the corresponding matrix of estimated transition probabilities. These matrices may be compared with corresponding matrices for three alternative states of the business cycles:

$$\mathbf{C}(1) = \begin{bmatrix} 7434 & 277 & 0 & 0 \\ 273 & 7306 & 62 & 187 \\ 0 & 15 & 94 & 33 \end{bmatrix}, \quad \widehat{\mathbf{P}}(1) = \begin{bmatrix} 0.964 & 0.036 & 0 & 0 \\ 0.035 & 0.933 & 0.008 & 0.024 \\ 0 & 0.106 & 0.662 & 0.232 \end{bmatrix},$$

for the *trough* of the business cycle,

$$\mathbf{C}(2) = \begin{bmatrix} 7125 & 305 & 0 & 0 \\ 177 & 6626 & 35 & 147 \\ 0 & 15 & 92 & 24 \end{bmatrix}, \quad \widehat{\mathbf{P}}(2) = \begin{bmatrix} 0.959 & 0.041 & 0 & 0 \\ 0.025 & 0.949 & 0.005 & 0.021 \\ 0 & 0.115 & 0.702 & 0.183 \end{bmatrix},$$

for the *normal* phase of the business cycle, and

$$
\mathbf{C}(3) = \begin{bmatrix} 7167 & 208 & 0 & 0 \\ 189 & 7552 & 42 & 87 \\ 0 & 14 & 121 & 25 \end{bmatrix}, \quad \widehat{\mathbf{P}}(3) = \begin{bmatrix} 0.972 & 0.028 & 0 & 0 \\ 0.024 & 0.960 & 0.005 & 0.011 \\ 0 & 0.088 & 0.756 & 0.156 \end{bmatrix},
$$

for the *peak* of the business cycle. The three states have been identified by Nickell et al. (2000, Sec. 2.4) as depending on whether the real GDP growth in the country was in the upper, middle or lower third of the growth rates recorded in the sample period.

In the following, these matrices are used for illustrative purposes *as if* data from $m = 3$ periods are given. In order to illustrate the testing procedures presented in Section 5.2.2 the hypothesis is tested that the data from the three periods came from the same theoretical transition probabilities. Clearly, from the construction of the three periods it is to be expected, that the test rejects the null hypothesis. The three χ^2-statistics with $6 = (4-1)(3-1)$ degrees of freedom for testing the equality of the rows of the transition matrices have p-values 0.005, < 0.0001, and 0.697. Thus, the null hypothesis must be clearly rejected for the first two rows at any usual level of confidence while the test for the last row suffers from the limited sample size. Nevertheless, the χ^2-statistic for the simultaneous test of the equality of the transition matrices has $18 = (4-1)^2(3-1)$ degrees of freedom and a p-value < 0.0001. Consequently, the null hypothesis must be rejected at any usual level of confidence.

5.2.4 Computational Aspects

The program `RatMigExample3.gau` computes aggregated migration counts, estimated transition probabilities and χ^2-statistics from a $(d-1) \times d \times m$ array of counts for m periods, compare Section 5.1.4. The output contains matrices `cagg`, `etpagg` and `etp` with components given by (5.12), (5.14) and (5.13). The elements of `esdagg` and `esd` result by replacing $\widehat{p}_{jk}$ in (5.11) by $\widehat{p}_{jk}^{+}$ or $\widehat{p}_{jk}(t)$, respectively. The matrix `chi` contains in the first row the statistics from (5.15) for $j = 1, \ldots, d-1$ and (5.16). The second and third row gives the corresponding degrees of freedom and p-values.

5.3 Multi-Period Transitions

In the multi-period case, transitions in credit ratings are also characterized by rating transition matrices. The m-period transition matrix is labeled $\mathbf{P}^{(m)}$. Its generic element $p_{jk}^{(m)}$ gives the rating transition probability from

rating j to k over the $m \geq 1$ periods. For the sake of simplicity the one-period transition matrix $\mathbf{P}^{(1)}$ is shortly denoted by $\mathbf{P}$ in the following. This transition matrix is considered to be of type $d \times d$. The last row contains $(0, 0, \ldots, 0, 1)$ expressing the absorbing default state. Multi-period transition matrices can be constructed from one-period transition matrices under the assumption of the Markov property.

5.3.1 Homogeneous Markov Chain

Let $\{X(t)\}_{t \geq 0}$ be a discrete-time stochastic process with countable state space. It is called a *first-order Markov chain* if

$$\mathrm{P}\left[(X(t+1) = x(t+1)|X(t) = x(t), \ldots, X(0) = x(0)\right]$$
$$= \mathrm{P}\left[X(t+1) = x(t+1)|X(t) = x(t)\right] \quad (5.17)$$

whenever both sides are well-defined. Further, the process is called a *homogeneous first-order Markov chain* if the right-hand side of (5.17) is independent of t (Brémaud, 1999).

Transferred to rating transitions, homogeneity and the Markov property imply time-invariant one-period transition matrices $\mathbf{P}$. Then the one-period $d \times d$ transition matrix $\mathbf{P}$ contains the non-negative rating transition probabilities

$$p_{jk} = \mathrm{P}(X(t+1) = k|X(t) = j).$$

They fulfill the conditions

$$\sum_{k=1}^{d} p_{jk} = 1$$

and

$$(p_{d1}, p_{d2}, \ldots, p_{dd}) = (0, \ldots, 0, 1).$$

The latter reflects the absorbing default state.

The two-period transition matrix is then calculated by ordinary matrix multiplication, $\mathbf{P}^{(2)} = \mathbf{PP}$. Qualitatively, the composition of the portfolio after one period undergoes the same transitions again. Extended for m periods this reads as

$$\mathbf{P}^{(m)} = \mathbf{P}^{(m-1)}\mathbf{P} = \mathbf{P}^m$$

with non-negative elements

$$p_{jk}^{(m)} = \sum_{i=1}^{d} p_{ji}^{(m-1)} p_{ik}.$$

The recursive scheme can also be applied for non-homogeneous transitions, i.e. for one-period transition matrices being not equal, which is the general case.

5.3.2 Bootstrapping Markov Chains

The one-period transition matrix $\mathbf{P}$ is unknown and must be estimated. The estimator $\widehat{\mathbf{P}}$ is associated with estimation errors which consequently influence the estimated multi-period transition matrices. The traditional approach to quantify this influence turns out to be tedious since it is difficult to obtain the distribution of $(\widehat{\mathbf{P}} - \mathbf{P})$, which could characterize the estimation errors. Furthermore, the distribution of $(\widehat{\mathbf{P}}^{(m)} - \mathbf{P}^{(m)})$, with

$$\widehat{\mathbf{P}}^{(m)} \stackrel{\text{def}}{=} \widehat{\mathbf{P}}^m, \tag{5.18}$$

has to be discussed in order to address the sensitivity of the estimated transition matrix in the multi-period case. It might be more promising to apply resampling methods like the bootstrap combined with Monte Carlo sampling. For a representative review of resampling techniques see Efron and Tibshirani (1993) and Shao and Tu (1995), for bootstrapping Markov chains see Athreya and Fuh (1992), Härdle, Horowitz and Kreiss (2003) and Lahiri (2003).

Assuming a *homogeneous first-order Markov chain* $\{X(t)\}_{t \geq 0}$, the rating transitions are generated from the unknown transition matrix $\mathbf{P}$. In the spirit of the *bootstrap method*, the unknown transition matrix $\mathbf{P}$ is substituted by the estimated transition matrix $\widehat{\mathbf{P}}$, containing transition rates. This then allows to draw a bootstrap sample from the multinomial distribution assuming *independent* rating migrations,

$$(\widetilde{c}_{j1}^*, \ldots, \widetilde{c}_{jd}^*) \sim \text{Mult}(n_j; \widehat{p}_{j1}, \ldots, \widehat{p}_{jd}), \tag{5.19}$$

for all initial rating categories $j = 1, \ldots, d-1$. Here, $\widetilde{c}_{jk}^*$ denotes the bootstrap random variable of migration counts from j to k in one period and $\widehat{p}_{jk}$ is an estimate of the one-period transition probability (the observed transition rate) from j to k.

A bootstrap transition matrix $\widehat{\mathbf{P}}^*$ with generic stochastic elements $\widehat{p}_{jk}^*$ is estimated according to

$$\widehat{p}_{jk}^* = \frac{\widetilde{c}_{jk}^*}{n_j}. \tag{5.20}$$

Obviously, defaulted credits can not upgrade. Therefore, the bootstrap is not necessary for obtaining the last row of $\widehat{\mathbf{P}}^*$, which is $(\widehat{p}_{d1}^*, \ldots, \widehat{p}_{dd}^*) =$

$(0, \ldots, 0, 1)$. Then matrix multiplication gives the m-period transition matrix estimated from the bootstrap sample,

$$\widehat{\mathbf{P}}^{*(m)} = \widehat{\mathbf{P}}^{*m},$$

with generic stochastic elements $\widehat{p}_{jk}^{*(m)}$.

Now the distribution of $\widehat{\mathbf{P}}^{*(m)}$ can be accessed by *Monte Carlo sampling*, e. g. B samples are drawn and labeled $\widehat{\mathbf{P}}_b^{*(m)}$ for $b = 1, \ldots, B$. Then the distribution of $\widehat{\mathbf{P}}^{*(m)}$ estimates the distribution of $\widehat{\mathbf{P}}^{(m)}$. This is justified since the consistency of this bootstrap estimator has been proven by Basawa, Green, McCormick and Taylor (1990). In order to characterize the distribution of $\widehat{\mathbf{P}}^{*(m)}$, the standard deviation $\mathrm{Std}\left(\widehat{p}_{jk}^{*(m)}\right)$ which is the bootstrap estimator of $\mathrm{Std}\left(\widehat{p}_{jk}^{(m)}\right)$, is estimated by

$$\widehat{\mathrm{Std}}\left(\widehat{p}_{jk}^{*(m)}\right) = \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left\{ \widehat{p}_{jk,b}^{*(m)} - \widehat{\mathsf{E}}\left(\widehat{p}_{jk}^{*(m)}\right) \right\}^2} \qquad (5.21)$$

with

$$\widehat{\mathsf{E}}\left(\widehat{p}_{jk}^{*(m)}\right) = \frac{1}{B} \sum_{b=1}^{B} \widehat{p}_{jk,b}^{*(m)}$$

for all $j = 1, \ldots, d-1$ and $k = 1, \ldots, d$. Here, $\widehat{p}_{jk,b}^{*(m)}$ is the generic element of the b-th m-period bootstrap sample $\widehat{\mathbf{P}}_b^{*(m)}$. So (5.21) estimates the unknown standard deviation of the m-period transition rate $\widehat{p}_{jk}^{(m)}$ using B Monte Carlo samples. Please note that $\widehat{p}_{jk}^{(m)}$ is here the generic element of the random matrix $\widehat{\mathbf{P}}^{(m)}$ in (5.18).

5.3.3 Rating Transitions of German Bank Borrowers

In the following the bootstrapping is illustrated in an example. As an estimate of the one-period transition matrix $\mathbf{P}$ the 7×7 rating transition matrix of small and medium-sized *German bank borrowers* from Machauer and Weber (1998, p. 1375) is used, which is shown in Table 5.1. The data cover the period from January 1992 to December 1996.

With this data the m-period transition probabilities are estimated by $\widehat{p}_{jk}^{(m)}$ and the bootstrap estimators of their standard deviations are calculated.

				To k				
From j	1	2	3	4	5	6	Default	n_j
1	0.51	0.40	0.09	0.00	0.00	0.00	0.00	35
2	0.08	0.62	0.19	0.08	0.02	0.01	0.00	103
3	0.00	0.08	0.69	0.17	0.06	0.00	0.00	226
4	0.01	0.01	0.10	0.64	0.21	0.03	0.00	222
5	0.00	0.01	0.02	0.19	0.66	0.12	0.00	137
6	0.00	0.00	0.00	0.02	0.16	0.70	0.12	58
Default	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0

Table 5.1. German rating transition matrix ($d = 7$) and the number of migrations starting from rating $j = 1, \ldots, d$

		Estimates				
From j	$\widehat{p}_{jd}^{(1)}$	$\widehat{Std}\left(\widehat{p}_{jd}^{*(1)}\right)$	$\widehat{p}_{jd}^{(5)}$	$\widehat{Std}\left(\widehat{p}_{jd}^{*(5)}\right)$	$\widehat{p}_{jd}^{(10)}$	$\widehat{Std}\left(\widehat{p}_{jd}^{*(10)}\right)$
1	0.00	0.000	0.004	0.003	0.037	0.015
2	0.00	0.000	0.011	0.007	0.057	0.022
3	0.00	0.000	0.012	0.005	0.070	0.025
4	0.00	0.000	0.038	0.015	0.122	0.041
5	0.00	0.000	0.079	0.031	0.181	0.061
6	0.12	0.042	0.354	0.106	0.465	0.123

Table 5.2. Estimates of the m-period default probabilities and bootstrap estimates of their standard deviations for $m \in \{1, 5, 10\}$ periods

These calculations are done for $m \in \{1, 5, 10\}$ periods and $B = 1000$ Monte Carlo steps. A part of the results is summarized in Table 5.2, where only default probabilities are considered. Note that the probabilities in Table 5.1 are rounded and the following computations are based on integer migration counts.

5.3.4 Portfolio Migration

Based on the techniques presented in the last sections the problem of portfolio migration can be considered, i. e. the distribution of $n(t)$ credits over the d rating categories and its evolution over periods $t \in \{1, \ldots m\}$ can be assessed. Here, a *time-invariant* transition matrix $\mathbf{P}$ is assumed. The randomly changing number of credits in category j at time t is labeled by $\widetilde{n}_j(t)$

and allows to define non-negative *portfolio weights*

$$\widetilde{w}_j(t) \stackrel{\text{def}}{=} \frac{\widetilde{n}_j(t)}{n(t)}, \quad j = 1, \dots, d,$$

which are also random variables. They can be related to migration counts $\widetilde{c}_{jk}(t)$ of period t by

$$\widetilde{w}_k(t+1) = \frac{1}{n(t)} \sum_{j=1}^{d} \widetilde{c}_{jk}(t) \tag{5.22}$$

counting all migrations going from any category to the rating category k. Given $\widetilde{n}_j(t) = n_j(t)$ at t, the migration counts $\widetilde{c}_{jk}(t)$ are binomially distributed

$$\widetilde{c}_{jk}(t)|\widetilde{n}_j(t) = n_j(t) \sim \mathrm{B}\left(n(t)\, w_j(t), p_{jk}\right). \tag{5.23}$$

The non-negative weights are aggregated in a row vector

$$\widetilde{w}(t) = (\widetilde{w}_1(t), \dots, \widetilde{w}_d(t))$$

and sum up to one

$$\sum_{j=1}^{d} \widetilde{w}_j(t) = 1.$$

In the case of *independent* rating migrations, the expected portfolio weights at $t+1$ given the weights at t result from (5.22) and (5.23) as

$$\mathsf{E}[\widetilde{w}(t+1)|\widetilde{w}(t) = w(t)] = w(t)\mathbf{P}$$

and the conditional covariance matrix $V[\widetilde{w}(t+1)|\widetilde{w}(t) = w(t)]$ has elements

$$v_{kl} \stackrel{\text{def}}{=} \begin{cases} \frac{1}{n(t)} \sum_{j=1}^{d} w_j(t) p_{jk}(1 - p_{jk}) & k = l \\ & \text{for} \\ -\frac{1}{n(t)} \sum_{j=1}^{d} w_j(t) p_{jk} p_{jl} & k \neq l. \end{cases} \tag{5.24}$$

For m periods the multi-period transition matrix $\mathbf{P}^{(m)} = \mathbf{P}^m$ has to be used, see Section 5.3.1. Hence, (5.22) and (5.23) are modified to

$$\widetilde{w}_k(t+m) = \frac{1}{n(t)} \sum_{j=1}^{d} \widetilde{c}_{jk}^{(m)}(t)$$

and

$$\widetilde{c}_{jk}^{(m)}(t)|\widetilde{n}_j(t) = n_j(t) \sim \mathrm{B}\left(n(t)\, w_j(t), p_{jk}^{(m)}\right).$$

Here, $\widetilde{c}_{jk}^{(m)}(t)$ denotes the randomly changing number of credits migrating from j to k over m periods starting in t. The conditional mean of the portfolio weights is now given by

$$\mathsf{E}[\widetilde{w}(t+m)|\widetilde{w}(t) = w(t)] = w(t)\mathbf{P}^{(m)}$$

and the elements of the conditional covariance matrix $V[\widetilde{w}(t+m)|\widetilde{w}(t) = w(t)]$ result by replacing p_{jk} and p_{jl} in (5.24) by $p_{jk}^{(m)}$ and $p_{jl}^{(m)}$.

5.3.5 Computational Aspects

Assuming time-invariant transition probabilities, in `RatMigExample4.gau` from a given one-period transition matrix the t-period transition matrices are computed. For a given `m`, all $t = 1, 2, \ldots, $ `m` multi-period transition matrices are computed from the one-period $d \times d$ matrix `p`. Therefore, the output `q` is a $d \times d \times $ `m` array. As an example, the one-year transition matrix given in Nickell et al. (2000, p. 208), which uses Moody's unsecured bond ratings between 31/12/1970 and 31/12/1997, is condensed for simplicity to 4×4 with only 4 basic rating categories, see the example in Section 5.2.3. Again, the last category stands for defaulted credits.

The bootstrapping of Markov chains as described in Section 5.3.2 is performed by the procedure `RatMigRateM`. The bootstrap sample is generated from a $(d-1) \times d$ matrix of migration counts, which is input of the procedure. The result consists of the matrices in the array `btm` which is calculated according (5.19) and (5.20), of the matrix `etm` given in (5.18) and of the matrix `stm` with components given in (5.21). `RatMigExample5.gau` contains an example.

Bibliography

Altman, E. I. and Kao, D. L. (1992). The Implications of Coporate Bond Ratings Drift, *Financial Analysts Journal*, **48**(3): 64–75.

Anderson, T. W. and Goodman, L. A. (1957). Statistical Inference about Markov Chains, *The Annals of Mathematical Statistics*, **28**(1): 89-110.

Athreya, K. B. and Fuh, C. D. (1992). Bootstrapping Markov chains, *in* R. LePage and L. Billard (eds), *Exploring the Limits of Bootstrap*, Wiley, New York, pp. 49–64.

Bangia, A., Diebold, F. X., Kronimus, A., Schagen, C. and Schuermann, T. (2002). Ratings migration and the business cycle, with application to credit portfolio stress testing, *Journal of Banking & Finance* **26**(2-3): 445-474.

Basawa, I. V., Green, T. A., McCormick, W. P. and Taylor, R. L. (1990). Asymptotic bootstrap validity for finite Markov chains, *Communications in Statistics A* **19**: 1493–1510.

Basel Committee on Banking Supervision (2006). International Convergence of Capital Measurement and Capital Standards - A Revised Framework, Comprehensive Version.

Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge.

Brémaud, P. (1999). *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer-Verlag, New York.

Crouhy, M., Galai, D. and Mark, R. (2001). Prototype risk rating system, *Journal of Banking & Finance* **25**: 47–95.

Davidson, J. (1994). *Stochastic Limit Theory*, Oxford University Press, Oxford.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Finger, C. C. (1998). Extended "constant correlations" in CreditManager 2.0, *CreditMetrics Monitor* pp. 5–8. 3rd Quarter.

Gupton, G. M., Finger, C. C. and Bhatia, M. (1997). CreditMetrics - Technical Document, J.P. Morgan.

Härdle, W., Horowitz, J. and Kreiss, J. P. (2003). Bootstrap Methods for Time Series, *International Statistical Review*, **71**(2): 435–459.

Höse, S., Huschens, S. and Wania, R. (2002). Rating Migrations, *in* W. Härdle, T. Kleinow, G. Stahl (eds.), *Applied Quantitative Finance - Theory and Computational Tools*, Springer-Verlag, Berlin, pp. 87–110.

Huschens, S. and Locarek-Junge, H. (2002). Konzeptionelle und statistische Grundlagen der portfolioorientierten Kreditrisikomessung, *in* A. Oehler (ed.), *Kreditrisikomanagement - Kernbereiche, Aufsicht und Entwicklungstendenzen*, 2nd edn., Schäffer-Poeschel Verlag, Stuttgart, pp. 89–114.

Jarrow, R. A., Lando, D. and Turnbull, S. M. (1997). A Markov model for the term structure of credit risk spreads, *The Review of Financial Studies* **10**(2): 481–523.

Kim, J. (1999). Conditioning the transition matrix, *Risk: Credit Risk Special Report*, October: 37–40.

Krüger, U., Stötzel, M. and Trück, S. (2005). Time series properties of a rating system based on financial ratios, *Deutsche Bundesbank, Discussion Paper, Series 2*, No 14/2005.

Lahiri, S. N. (2003). *Resampling Methods for Dependent Data*, Springer-Verlag, New York.

Lancaster, T. (1990). *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Lando, D. and Skødeberg, T. M. (2002). Analyzing rating transitions and rating drift with continuous observations, *Journal of Banking & Finance* **26**(2-3): 423-444.

Machauer, A. and Weber, M. (1998). Bank behavior based on internal credit ratings of borrowers, *Journal of Banking & Finance* **22**: 1355–1383.

Mardia, K. V., Kent, J. T. and Bibby, J. M. (1979). *Multivariate Analysis*, Academic Press, London.

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management – Concepts, Techniques, Tools*, Princeton University Press, Princeton.

Nickell, P., Perraudin, W. and Varotto, S. (2000). Stability of rating transitions, *Journal of Banking & Finance* **24**: 203–227.

Saunders, A. (1999). *Credit Risk Measurement: New Approaches to Value at Risk and Other Paradigms*, Wiley, New York.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*, Springer-Verlag, New York.

6 Cross- and Autocorrelation in Multi-Period Credit Portfolio Models

Christoph K.J. Wagner

6.1 Introduction

For the risk assessment of credit portfolios single-period credit portfolio models are by now widely accepted and used in the practical analysis of loan respectively bonds books in the context of capital modeling. But already Finger (2000) pointed to the role of inter-period correlation in structural models and Thompson, McLeod, Teklos and Gupta (2005) strongly advocated that it is 'time for multi-period capital models'. With the emergence of structured credit products like CDOs the default-times/Gaussian-copula framework became standard for valuing and quoting liquid tranches at different maturities, Bluhm, Overbeck and Wagner (2002). Although it is known that the standard Gaussian-copula-default-times approach has questionable term structure properties the approach is quite often also used for the risk assessment by simply switching from a risk-neutral to a historical or subjective default measure.

From a pricing perspective Andersen (2006) investigates term structure effects and inter-temporal dependencies in credit portfolio loss models as these characteristics become increasingly important for new structures like forward-start CDOs. But the risk assessment is also affected by inter-temporal dependencies. For the risk analysis at different time horizons the standard framework is not really compatible with a single-period correlation structure; Morokoff (2003) highlighted the necessity for multi-period models in that case. Long-only investors in the bespoke tranche market with a risk-return and hold-to-maturity objective have built in the past CDO books with various vintage and maturity years, based only on a limited universe of underlying credits with significant overlap between the pools. A proper assessment of such a portfolio requires a consistent multi-period portfolio framework with reasonable inter-temporal dependence. Similarly, an investor with a large

loan or bond book, enhanced with non-linear credit products, needs a reliable multi-period model with sensible inter-temporal properties as both bond or structured investments display different term structure characteristics.

In the following, we investigate several multi-period models, a CreditMetrics-type approach, i.e. a Markov chain Monte Carlo model with dependency introduced via a Gaussian copula, the well-known model for correlated default times, a continuous threshold model driven by time-changed correlated Wiener processes by Overbeck and Schmidt (2005), and a discrete barrier model (Finger (2000), Hull and White (2001)), also based on a driving Brownian motion. All models meet by construction the marginal default probability term structures. We then investigate the effect of a finer time discretization on the cumulative loss distribution at a given time horizon. The time-changed threshold model is invariant under this operation, whereas the credit migration approach converges to the limit of vanishing cross correlation, i.e. the correlation is 'discretized away'. Thompson et al. (2005) analyse the same problem for the discrete barrier model and observe decreasing loss volatility and tail risk. They conjecture that it converges to the limit of the 'true' portfolio loss distribution. We have similar findings but draw a different conclusion as we attribute the decreasing loss volatility to inherent features of the discrete model. Obviously, it is not congruent with a continuous-time default barrier model like the time-changed threshold model. Hence, the assumption that the time-changed model is the continuous limit of the discrete threshold model is wrong.

These findings imply that these types of credit portfolio models not only have to be calibrated to marginals, but also to a correlation structure for a given time horizon and time discretization in order to yield consistent valuation and risk assessment. We therefore turn to the problem of how to adjust the correlation structure, at least in the credit migration framework, while shortening the time steps such that the cumulative loss distribution is commensurate to a one-period setting at a given time horizon. This approach assumes that we are given a certain correlation structure for a fixed period, e.g. yearly correlations through time series estimation. We show that it suffices to compare the joint default probabilities and adapt the correlation parameter accordingly to obtain commensurate cumulative loss distribution at a given horizon.

Finally, we take a look at the autocorrelation of the different models and briefly highlight the different inter-temporal loss dependency of the models, as this plays an important part in risk-assessing books of CDOs.

6.2 The Models

6.2.1 A Markov-Chain Credit Migration Model

The first model we investigate is essentially a CreditMetrics-type approach in a multi-period setting. A Markov state (rating) $Y \in \{1, \ldots, K\}$ is assigned to each single credit risky entity i, the absorbing state K is the default state. A default probability term-structure $F_i(t)$ exists for each initial credit state i together with a sequence of migration matrices M_{t_k} that is adapted to meet the term-structure. The migration matrix M_{t_n} defines a natural discretization of Y_t, but we can subdivide or refine the discretization arbitrarily through the introduction of a matrix square root $M_t^{1/2} = M_{t/2}$ or a generator matrix Q, $M_t = \exp(tQ)$, see Bluhm et al. (2002) for more details. The discrete Markov process Y_t with time-homogeneous migration matrix does not necessarily meet a given PD-term structure i.e., $\left(M_0^k\right)_{iK} \neq F_i(t_k), k = 1, 2, 3, \ldots,$ (with K as default state). This can easily be rectified by adapting the transition matrices recursively, i.e. the default column of the first matrix is set to the term structure and the remaining entries are renormalized.

With some linear algebra the next matrix can be adjusted accordingly, and so on. These transition matrices are chained together and create a discrete credit migration process for each credit entity, $Y_{t_k}^i$, on a time grid $0 = t_0 < t_1 < t_2 < t_3 < \ldots < t_n$. As migration matrix a Rating Agency's one-year transition matrix is typically used. In the multi-firm context we add a dependency structure between different credit entities, i.e. credit migrations are coupled through a Gaussian copula function with correlation matrix Σ in each step. There is no explicit interdependence between the steps apart from the autocorrelation generated by the migrations. From each migration matrix we can now calculate migration thresholds that separate the transition buckets. For some period t_j the thresholds c_{kl,t_j} are obtained from

$$
\begin{aligned}
c_{kl,t_j} &= \Phi^{-1}\left(\sum_{n=l}^{K} M_{kn,t_j}\right), \quad \text{for } k, l = 1, \ldots, K, \text{ with } \sum_{n=l}^{K} M_{kn,t_j} \neq 0, 1 \\
c_{kl,t_j} &= -\infty, \qquad \text{for} \quad k, l = 1, \ldots, K, \quad \text{with} \quad \sum_{n=l}^{K} M_{kn,t_j} = 0 \\
c_{kl,t_j} &= +\infty, \qquad \text{for} \quad k, l = 1, \ldots, K, \quad \text{with} \quad \sum_{n=l}^{K} M_{kn,t_j} = 1.
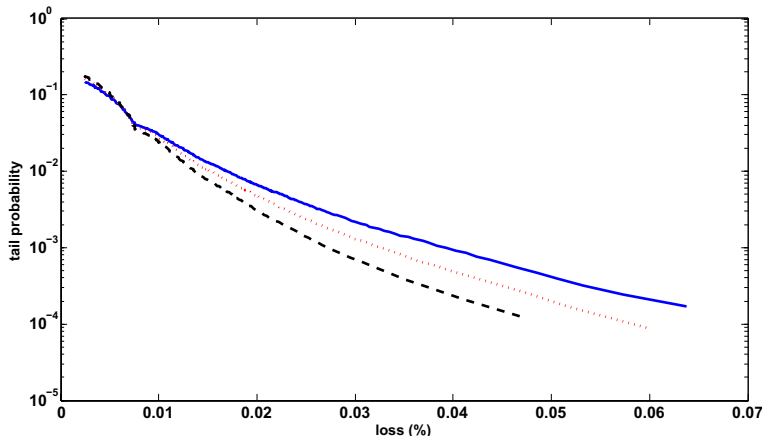\end{aligned}
$$

$$(6.1)$$

Figure 6.1. Refining Time-Discretization, migration model: annual(blue solid), semi-annual(red dotted), and quarterly(black dashed) discretization.

For each period $]t_{j-1}, t_j]$ correlated normal random variables are sampled, $(r_{i,t_j})_{i=1,\ldots,n} \sim \Phi(0, \Sigma)$, and credit i migrates from the initial state l to the final state k if

$$c_{lk-1,t_j} \le r_{i,t_j} < c_{lk,t_j}.$$

As a remark, this type of correlated credit migration model is also the basis of the credit component in Moody's SIV Capital Model (Tabe and Rosa (2004)) and of Moody's KMV CDO Analyzer (Morokoff (2003)). MKMV's CDO Analyzer applies the migration technique to the MKMV Distance-to-Default-Indicator which is far more fine grained than usual rating classes.

But note one problem: The correlation structure of the model is not invariant under the refinement of the time discretization. Denote L the portfoli loss, then Figure 6.1 shows the tail probability $\mathrm{P}(L > x)$ for a sample portfolio with non-vanishing correlation at the one year horizon under annual, semi-annual, and quarterly discretization. For this, we have simply calculated appropriate square-roots of the migration matrices. The fatness in the tail of the loss distribution is significantly reduced for smaller migration intervals. As soon as we introduce correlation to the rating transitions a link between global correlation and discretization is generated. By this we mean that choosing the same local correlation parameter ρ for each time step, the joint arrival probability in the states m, n of two entities at time t, given they start at time 0 in states k, l

$$\mathrm{P}(Y_t^i = m, Y_t^j = n | Y_0^i = k, Y_0^j = l),$$

is a function of how fine we discretize the process, while keeping the local correlation constant. Smaller step-sizes de-correlate the processes Y_t^i and Y_t^j. This can easily be seen by the fact that for smaller step sizes the migration probabilities to the default state get smaller, but since the Gaussian copula has no tail dependence the correlation converges asymptotically to zero as we move the step size to zero. Obviously, this is an unpleasant feature when it comes to practical applications of the model, as e.g. the pricing or risk assessment of correlation sensitive product like a CDO depend then on the time discretization of the implementation.

In order to reconstitute the original correlation over a fixed time interval while halving the time step, one way is to adapt, i.e. increase, the local cross correlation. Suppose

$$P(Y_1^i = K, Y_1^j = K | Y_0^i = k, Y_0^j = l)$$

is the joint default probability for one large step. Cutting the discretization in halves, the joint default probability is now

$$\begin{aligned}
P(Y_1^i = K, Y_1^j = K | Y_0^i = k, Y_0^j = l) = \\
= \sum_{p,q} P(Y_1^i = K, Y_1^j = K | Y_{1/2}^i = p, Y_{1/2}^j = q) \times \\
P(Y_{1/2}^i = p, Y_{1/2}^j = q | Y_0^i = k, Y_0^j = l).
\end{aligned} \tag{6.2}$$

Instead of trying to adjust the correlation for all pairs i, j we confine ourselves to a homogeneous state in the sense of a large pool approximation (Kalkbrener, Lotter and Overbeck (2004)). We obtain one adjustment factor and apply it to all names in the portfolio. For further discretization we simply nest the approach. Figure 6.2 shows the effect of the adjustment for an example. We use an inhomogeneous portfolio of 100 positions with exposures distributed uniformly in $[500, 1500]$, 1-year default probabilities in $[10bp, 100bp]$, and correlation between $[10\%, 30\%]$. As can be seen from the graph both loss distribution are now commensurable. From a risk perspective this degree of similarity seems sufficient, particularely if risk measures like expected shortfall are used. Further improvement can be achieved by computing adjustment factors for each rating state and for each matrix in the sequence of transition matrices (if they are different).

In case of the migration model we have so far chosen independent cross coupling mechanisms at each time step, so autocorrelation is solely introduced through the dispersion of the transition matrices. Defining a copula that couples transitions not only at one step but also between different steps
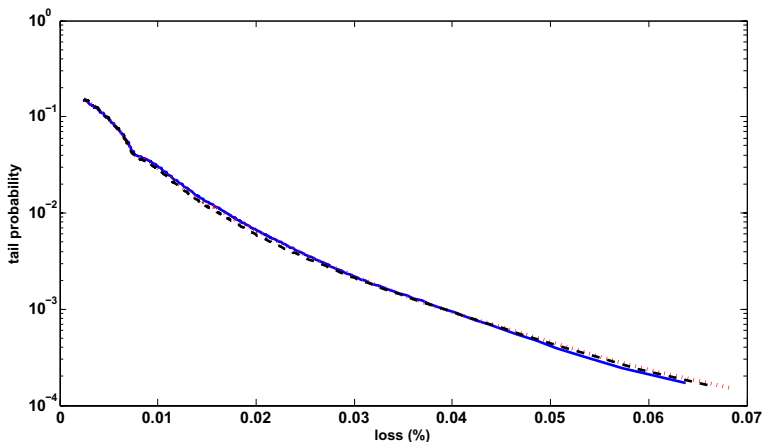
Figure 6.2. Refining Time-Discretization, migration model
with adapted correlation: annual(blue solid), semi-annual(red
dotted), and quarterly(black dashed) discretization.

would introduce explicit autocorrelation to the migration model (see Andersen (2006)), but we run into a heavy calibration problem, since (i) the calibration to the marginal default term structure becomes more involved, and (ii) the adjustment of the correlation structure while refining the discretization is much more difficult. If the cross dependency is formulated via a factor model we can also induce autocorrelation between the time steps by introducing an auto-regressive process for these factors, i.e. the factors Z_n that couple the transitions at each time step n are linked through $Z_{n+1} = \alpha Z_n + \sqrt{1 - \alpha^2} \xi_n$, $Z_1, \xi_n \sim \Phi(0,1)$, independent, and α is some coupling factor.

6.2.2 The Correlated-Default-Time Model

Another wide spread approach for a credit portfolio model is to generate *correlated default times* for the credit securities. This is done in analogy to a one-year-horizon asset value model by taking the credit curves of the securities as cumulative distributions of random default times and coupling these random variables by some copula function, usually the Normal copula, thus generating a multivariate dependency structure for the single default times. It is not by chance that this approach already has been used for the valuation of default baskets as the method focuses only on defaults and not on rating migrations.

From a simulation point of view, the default times approach involves much less random draws than a multi-step approach as we directly model the default times as continuous random variable. Time-consuming calculations in the default times approach could be expected in the part of the algorithm inverting the credit curves $F_i(t)$ in order to calculate default times according to the formula $\tau_i = F_i^{-1}\{\Phi(r_i)\}$. Fortunately, for practical applications the exact time when a default occurs is not relevant. Instead, the only relevant information is if an instrument defaults between two consecutive payment dates. Therefore, the copula function approach for default times can be easily discretized by calculating thresholds at each payment date $t_1 < t_2 < t_3 < \ldots < t_n$ according to

$$c_{i,t_k} \;=\; \Phi^{-1}\{F_i(t_k)\} \; ,$$

where F_i denotes the credit curve for some credit i. Clearly, one has

$$c_{i,t_1} \;<\; c_{i,t_2} \;<\; \ldots \;<\; c_{i,t_n} \;.$$

Setting $c_{i,t_0} = -\infty$, asset i defaults in period $]t_{k-1}, t_k]$ if and only if

$$c_{i,t_{k-1}} \;<\; r_i \;\leq c_{i,t_k} \; ,$$

where $(r_1, ..., r_m) \sim \Phi(0, \Sigma)$ denotes the random vector of standardized asset value log-returns with asset correlation matrix Σ. In a one-factor model setting r_i is typically represented as

$$r_i = \sqrt{\varrho}Y + \sqrt{1-\varrho}Z_i,$$

where $Y, Z_i \sim \Phi(0,1)$ are the systematic and specific risk components of the log-return r_i. Obviously, the discrete implementation of correlated default times is invariant to the refinement of the time discretization by construction.

Note further that the correlated-default-times approach with Gaussian-copula is in fact a static model. For this, we write the conditional joint default probability at different time horizons in a one-factor setting as

$$
\begin{aligned}
\mathrm{P}[\tau_1 < s, \tau_2 < t | Y = y] = {}& \\
= {}& \mathrm{P}[r_1 < \Phi^{-1}\{F_1(s)\}, r_2 < \Phi^{-1}\{F_2(t)\} | Y = y] \\
= {}& \mathrm{P}[Z_1 < \frac{\Phi^{-1}\{F_1(s)\} - \sqrt{\varrho}y}{\sqrt{1-\varrho}}, Z_2 < \frac{\Phi^{-1}\{F_2(t)\} - \sqrt{\varrho}y}{\sqrt{1-\varrho}}] \\
= {}& \Phi\left[\frac{\Phi^{-1}\{F_1(s)\} - \sqrt{\varrho}y}{\sqrt{1-\varrho}}\right] \Phi\left[\frac{\Phi^{-1}\{F_2(t)\} - \sqrt{\varrho}y}{\sqrt{1-\varrho}}\right].
\end{aligned}
\tag{6.3}
$$

The sample of the common factor Y is static for all time horizons, there is no dynamics through time.

6.2.3 A Discrete Barrier Model

Finger (2000) and Hull and White (2001) proposed a discrete multi-period barrier model on a time grid $t_0 < t_1 < \ldots < t_n$ based on correlated Brownian processes B_t^i where the default thresholds $c_i(t_k)$ are decreasing functions of time calibrated to satisfy the marginal term structure $F_i(t_k)$. Credit entity i defaults in period k if, for the first time, $B_{t_k}^i < c_i(t_k)$, i.e.

$$\tau_i = \min\left\{t_k \geq 0 : B_{t_k}^i < c_i(t_k), k = 0, \ldots, n\right\}.$$

The default barriers $c_i(t_k)$ are to be calibrated to match $F_i(t_k)$ such that

$$F_i(t_k) = \mathrm{P}(\tau_i < t_k).$$

Denote $\delta_k = t_k - t_{k-1}$, then from

$$\mathrm{P}\left\{B_{t_1}^i < c_i(t_1)\right\} = F_i(t_1)$$

follows that

$$c_i(t_1) = \sqrt{\delta_1}\Phi^{-1}\left\{F_i(t_1)\right\}.$$

The successive thresholds $c_i(t_k)$ are then found by solving

$$
\begin{aligned}
F_i(t_k) - F_i(t_{k-1}) &= \\
&= \mathrm{P}\left\{B_{t_1}^i > c_i(t_1) \cap \cdots \cap B_{t_{k-1}}^i > c_i(t_{k-1}) \cap B_{t_k}^i < c_i(t_k)\right\} \\
&= \int_{c_i(t_{k-1})}^{\infty} f_i(t_{k-1}, u)\Phi\left[\frac{c_i(t_k) - u}{\sqrt{\delta_k}}\right] du,
\end{aligned}
$$

where $f_i(t_k, x)$ is the density of $B_{t_k}^i$ given $B_{t_j}^i > c_i(t_j)$ for all $j < k$:

$$
\begin{aligned}
f_i(t_1, x) &= \frac{1}{\sqrt{2\pi\delta_1}} \exp\left(-\frac{x^2}{2\delta_1}\right) \\
f_i(t_k, x) &= \int_{c_i(t_{k-1})}^{\infty} f_i(t_{k-1}, u)\frac{1}{\sqrt{2\pi\delta_k}} \exp\left\{-\frac{(x-u)^2}{2\delta_k}\right\} du.
\end{aligned}
$$

Hence, the calibration of the default thresholds is an iterative process and requires the numerical evaluation of integrals with increasing dimension, which renders the model computationally very heavy. Another shortcoming of the model is that it is not invariant under the refinement of the time discretization, Thompson et al. (2005). Figure 6.3 shows the tail probability $\mathrm{P}(L > x)$ of a portfolio loss with different discretization (annual, semi-annual, quarterly) of the model. Obviously, the volatility and tail fatness of the loss distribution decreases with increasing refinement, and it is not clear where the limiting distribution is.
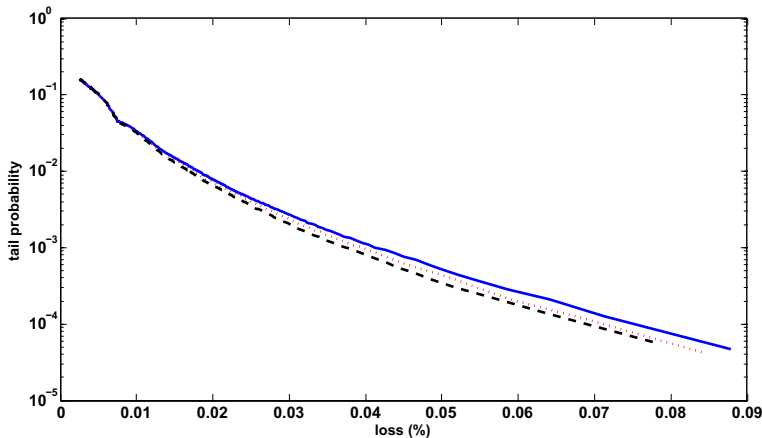
Figure 6.3.   Refining Discretization, Hull-White Model: annual(blue solid), semi-annual(red dotted), and quarterly(black dashed) discretization.

6.2.4 The Time-Changed Barrier Model

The above mentioned discrete barrier model is drawn from a continuous version, i.e. correlated Brownian processes B_t^i with time-dependent barriers $c_i(t)$. The default time of credit i is then the first hitting time of the barrier $c_i(t)$ by the driving process B_t^i:

$$\tau_i = \inf \left\{ t \geq 0 : B_t^i < c_i(t) \right\}.$$

If $c_i(t)$ is absolutely continuous, we can write

$$c_i(t) = c_i(0) + \int_0^t \mu_s^i ds,$$

and the default time τ_i is the first hitting time of the constant barrier $c_i(0)$ by a Wiener process with drift.

$$
\begin{aligned}
Y_t^i &= B_t^i - \int_0^t \mu_s^i ds \\
\tau_i &= \inf \left\{ t \geq 0 : Y_t^i < c_i(0) \right\}.
\end{aligned}
\tag{6.4}
$$

The problem is now to calibrate the model to the prescribed default term structure, $\mathrm{P}[\tau_i < t] = F_i(t)$. To this end, Overbeck and Schmidt (2005) put forward a barrier model based on Brownian processes B_t^i with suitably transformed time scales, (T_t^i), strictly increasing, $T_0^i = 0$. The first passage

time to default τ_i of credit entity i is define through the process

$$Y_t^i = B_{T_t^i}^i$$

and

$$\tau_i = \inf\left\{s \geq 0 : Y_s^i < c_i\right\},$$

with a time independent barrier c_i. From the strong Markov property or the reflection principle of the Brownian motion follows that the first passage time of an untransformed Brownian motion with respect to a constant barrier c

$$\widetilde{\tau} = \inf\left\{t \geq 0 : B_t < c\right\}$$

is distributed as

$$\mathrm{P}(\widetilde{\tau} < t) = \mathrm{P}\left(\min_{0 < s < t} B_s < c\right) = 2\Phi\left(\frac{c}{\sqrt{t}}\right). \tag{6.5}$$

As T_t^i is strictly increasing we find that

$$\begin{aligned} \mathrm{P}(\tau_i < t) &= \mathrm{P}\left(\min_{0 < s < t} B_{T_t^i}^i < c_i\right) = \mathrm{P}\left(\min_{0 < s < T_t^i} B_t^i < c_i\right) \\ &= 2\Phi\left(\frac{c_i}{\sqrt{T_t^i}}\right) \end{aligned} \tag{6.6}$$

Hence, given a default term structure $F_i(t)$ the model is calibrated to the marginals via the time transformation

$$T_t^i = \left[\frac{c_i}{\Phi^{-1}\left\{F_i(t)/2\right\}}\right]^2. \tag{6.7}$$

Since $F(t)$ is strictly increasing this also follows for T_t. The constant default barrier c_i is then obtained by fixing a time t_0 with $T_{t_0}^i = t_0$ which implies

$$c_i = \Phi^{-1}\left\{F_i(t_0)/2\right\}\sqrt{t_0}. \tag{6.8}$$

An obvious, but not necessarily the only sensible choice is to take t_0 as the final maturity. Dependency between credits is introduce here through the (local) instantaneous correlation matrix Σ of the Brownian processes B_t^i. The joint default probabilities $\mathrm{P}[\tau_i < t, \tau_j < t]$ can be written in analytical, but rather technical form, which allows the calibration of the model to prescribed joint default probabilities.

The discretization of the time-changed model for practical applications is straight forward and simply obtained by discretizing the SDE of the correlated Brownian motion while taking into account that the different dimensions evolve at different time scales. Figure 6.4 shows the behavior of the time-changed model under a refinement of the time discretization. Obviously, the model is within sampling errors invariant under this operation.
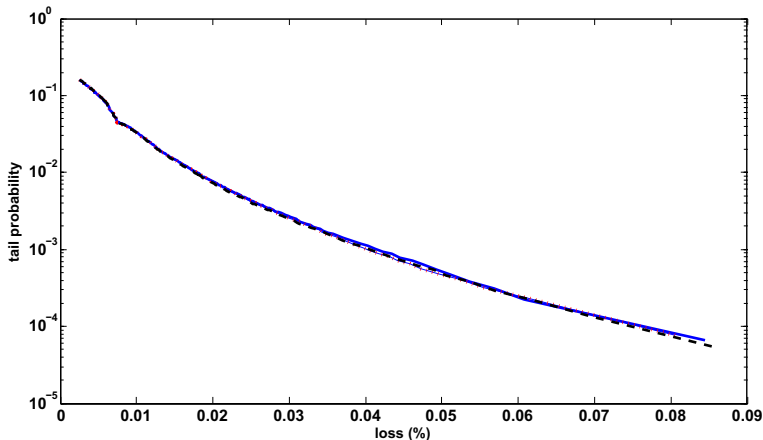
Figure 6.4.   Refining Discretization, time-change Model: annual(blue solid), semi-annual(red dotted), and quarterly(black dashed) discretization.

6.3 Inter-Temporal Dependency and Autocorrelation

Finally, let us take a look at the inter-temporal dependency of the various models. All models are set up to meet by construction the default-term structure, hence produce the same first order loss moments through time, and they are calibrated to the same loss volatility at maturity (4 years). Figures 6.5-6.7 now serve to demonstrate the different inter-temporal characteristics of the four models. The graphs show the different joint loss distributions at the 2- and 4-year horizon, depicted as heat map. The upper triangle is empty as $L(4 \text{ years}) \geq L(2 \text{ years})$. Clearly, the migration model has the least autocorrelation as joint losses accumulate at the edges of the triangle. In contrast, the correlated-default-times model shows the highest inter-temporal dependency between losses, as joint losses accumulate in the middle of the triangle. This comes not as a surprise and reflects the fact that the model is essentially a static one where static factors drive the dependency through the whole time. Due to the driving Brownian motion it is also obvious that the two barrier models show similar inter-temporal dependency that lie somewhere between the first two extreme cases.

The control of inter-temporal dependency is not so much a problem if we only model a single plain vanilla CDO, but as soon as we have a structure with significant default-timing feature or if we want to assess the risk of a

portfolio of non-linear credit products the inter-temporal dependency plays indeed an important role. For risk assessment the dependency through time should also be estimated from credit data, but these estimates seem not to support the high degree of inter-temporal dependency as generated by the barrier models.
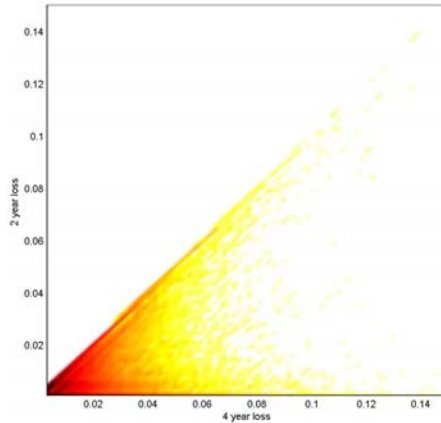

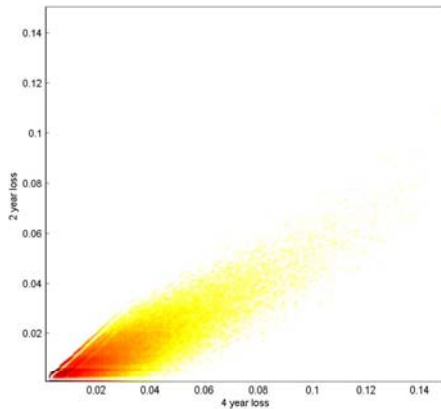Figure 6.5. Joint Loss Distribution (2-4 years), Credit migration model

Figure 6.6. Joint Loss Distribution (2-4 years), Correlated-default-times model
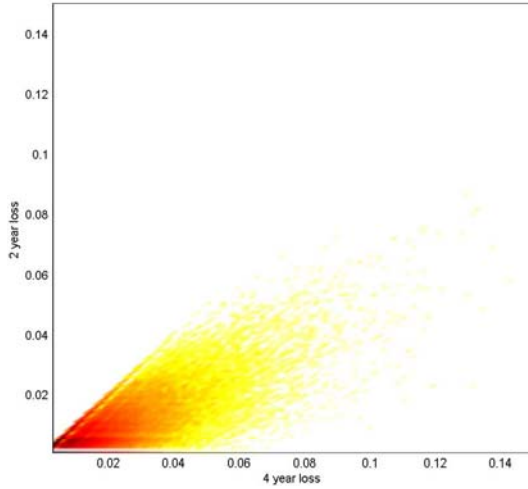
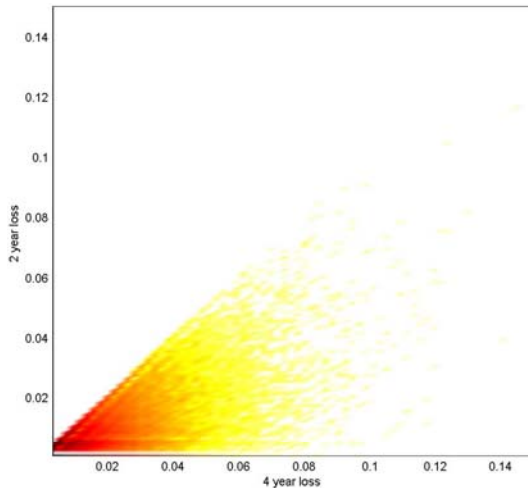Figure 6.7. Joint Loss Distribution (2-4 years), Time-change barrier model

Figure 6.8. Joint Loss Distribution (2-4 years), Discrete barrier model

6.4 Conclusion

For an assessment of a portfolio of structured credit products a multi-period model with known cross- and autocorrelation is necessary. We investigate implementations of four different multi-period credit portfolio model and show

that not for all of the models the correlation structure is invariant under the operation of a refined time discretization. Hence one should not blindly use these tpye of models at different periods and discretizations. In case of the discrete barrier model the continuous limit is unclear, but it is definitely not congruent to the time-changed barrier model. In case of a Markov Chain migration framework we argue that the cumulative loss distribution converges in the limit to a loss distribution with zero correlation as the time discretization is refined towards zero.

We then show how to correct the correlation structure while refining the discretization to obtain a congruent loss distribution at a given horizon. Finally, we analyse the inter temporal dependency of the different models and find that the correlated-default times model has the highest degree of inter-temporal dependency, the migration model relatively little and that the models driven by a Brownian motion are in between these two cases. We therefore conclude that before applying a multi-period model for risk assessment to a structured credit book the properties of the model in terms of inter-temporal and cross correlations should be fully understood, as different models have obviously different properties and will lead to differing results.

Bibliography

Andersen, L. (2006). Portfolio Losses in Factor Models: Term Structures and Inter-temporal Loss Dependence, *working paper*, http://www.defaultrisk.com, 2006.

Bluhm, C., Overbeck, L. and Wagner, C. (2002). *Introduction to Credit Risk Modeling*, Chapman & Hall/CRC.

Bluhm, C. and Overbeck, L. (2006). *Structured Credit Portfolio Analysis, Baskets and CDOs*, Chapman and Hall.

Finger, C. (2000). A comparison of stochastic default rate models, *RiskMetrics Journal*, **1**: 49-73.

Hull, J. and White, A. (2001). Valuing credit default swaps II: Modeling default correlations, *Journal of Derivatives* **8**(3): 12-21.

Kalkbrener, M., Lotter, H. and Overbeck, L. (2004). Sensible and efficient capital allocation for credit portfolios, *RISK* **17**(1): 19-24.

Tabe, R. and Rosa, D. (2004). *Moody's Capital Model*, Moody's.

Thompson, K., McLeod, A., Teklos, P. and Gupta, S. (2005). Time for multi-period capital models, *RISK* 74-78.

Morokoff, W. (2003). Simulation methods for risk analysis of collateralized debt obligations, *Moody's KMV New Product Research Publication*.

Overbeck, L. and Schmidt, W. (2005). Modeling default dependence with threshold models, *Journal of Derivatives* **12**(4): 10-19.

7 Risk Measurement with Spectral Capital Allocation

Ludger Overbeck and Maria Sokolova

Spectral risk measures provide the framework to formulate the risk aversion of a firm specifically for each quantile of the loss distribution of a portfolio. More precisely the risk aversion is codified in a weight function, weighting each quantile. Since the basic coherent building blocks of spectral risk measures are expected shortfall measures, the most intuitive approach comes from combinations of those. For investment decisions the marginal risk or the capital allocation is the sensible approach. Since spectral risk measures are coherent there exists also a sensible capital allocation based on the notion of derivatives or more in the light of the coherency approach as an expectation under a generalized maximal scenario.

7.1 Introduction

Portfolio modeling has two main objectives: the quantification of portfolio risk, which is usually expressed as the economic capital of the portfolio, and its allocation to subportfolios and individual transactions. The standard approach in credit portfolio modeling is to define the economic capital in terms of a quantile of the portfolio loss distribution

$$q_\alpha(L) \;=\; F_L^{-1}(\alpha).$$

The capital charge of an individual transaction is traditionally based on a covariance technique and called volatility contribution. We refer to Bluhm et al. (2002) and Crouhy et al. (2000) for a survey on credit portfolio modeling and capital allocation.

Since the work by Artzner et al (1997) coherent risk measures are discussed intensively in finance and risk management. More recent is the question of a more coherent capital allocation. Especially the use of expected shortfall allocation as an allocation rule is recommend in Overbeck (2000),Denault

(2001), Bluhm et al. (2002), Kurth and Tasche (2003) and Kalkbrener et al. (2004).

Expected shortfall measures

$$\mathsf{E}\,S_\alpha(L) \;=\; \frac{1}{1-\alpha} \int_\alpha^1 q_u(L)du$$

are the building blocks of more general coherent risk measures, the spectral risk measure ρ. These are convex mixtures of expected shortfall measures. They can be represented by their spectral measure μ through

$$\rho = \rho_\mu = \int_0^t \mathsf{E}\,S_\alpha(1-\alpha)\mu(da) \tag{7.1}$$

or as a weighted sum of quantiles with $w(\alpha) = \mu([0, \alpha])$,

$$\rho = \rho_\mu = \rho_w = \int_0^1 q_\alpha(\cdot)w(\alpha)d\alpha. \tag{7.2}$$

In this paper we apply the allocation rules associated with a spectral risk measure to a credit portfolio and point out, which consequences to risk management the choice of the weight function w, the spectral measure μ or the measure

$$\widetilde{\mu} \stackrel{\text{def}}{=} (1-\alpha)\mu(d\alpha),$$

which we call mixing measure and thought to be the most easily one to calibrate and implement. The theoretical basis of the approach can be found in the basic papers Kalkbrener (2002), Kalkbrener et al (2004) and the explicit application to spectral capital allocation is provided by Overbeck (2005). We will first present the theoretical foundation of the proposed risk and allocation measures and then discuss general impact of the choice of the weight or mixing function and finally exhibits the differences on a concrete credit portfolio example.

7.2 Review of Coherent Risk Measures and Allocation

7.2.1 Coherent Risk Measures

It is well-known that the following four conditions define a coherent risk measure, Artzner et al (1997, 1999), Delbaen (2000).

Formally, a risk measure is nothing else as a positive real valued function r defined on the set of random variable (potential losses) V. The number $r(X)$ denotes the risk in portfolio X. r is called coherent if it obeys the following 4 rules.

⊡ Subadditivity (Diversification)

$$r(X + Y) < r(X) + r(Y)$$

⊡ Positive homogenous (Scaling)

$$r(aX) = ar(X), a > 0$$

⊡ Monotone

$$r(X) < r(Y) \text{ if } X < Y \text{ (almost surely)}$$

⊡ Translation property

$$r(X + a) = r(X) - a$$

Convex analysis gives already that a sub-additive positive homogenous function r can be point wise written as the maximal value of all linear functions which are below r (Delbaen (2000), Kalkbrener (2002), Kalkbrener et al (2004)). For risk measures this means that the first two axioms above lead to the following representation

$$r(X) \;=\; \max\{l(X) \mid l < r, \; l \text{ linear function }\} \tag{7.3}$$

The risk measure evaluate at a loss variable X takes the same value as the largest value of all linear function which lies below r on V evaluated on X.

Conceptually, this is similar to the gradient of the function r evaluated at the point X or as the best linear approximation of r which coincides with r at the point X. We will later see that this intuition gives rise to a sensible capital allocation.

A typical linear function for random variable is the expectation operator. Hence the basic result by Artzner et al (1997), Delbaen (2000)

$$r(X) \;=\; \sup\{E_Q[X] \mid Q \in \mathcal{Q}\} \tag{7.4}$$

$\mathcal{Q}, = \mathcal{Q}_r$, a suitable set of probability measures of absolutely continuous probability measures $Q << P$ with density dQ/dP, is similar to the representation (7.3).

The set $\mathcal{Q}$ is called the generalized scenarios associated with r. If the supremum is actually taken at some probability measure, this probability measure or its density with respect to P is called the generalized scenario associated with r. These approach also fits into the intuitive feature of risk measurement, namely scenario or stress analysis. For the interpretation in terms of scenarios the formulation with probability measure is more natural, but for the axiomatic approach to capital allocation the representation (7.3) is very useful.

The currently most prominent example of a coherent risk measure is Expected Shortfall (sometimes called Conditional VaR /tail conditional expectation). It is denoted by $\mathsf{E}\,S_\alpha$ and measures the average loss above the α-quantile of the loss distribution. The associated generalized scenarios can be explained as follows:

To each loss variable Y define the scenario as the "historical" calibrated objective scenario constraint on the condition that the loss variable exceeded its quantile. The expected shortfall coincides with the largest mean loss in these scenarios. Intuitively,

$$\mathsf{E}\{L|L > q_\alpha(L)\} = \max\{E\{L|Y > q_\alpha(Y)\}| \text{ all } Y \in L_\infty\}$$

Even if generalized scenarios are defined as a supremum, in the case of Expected Shortfall we can identify the density of the maximal "scenario". For this we need the formally correct definition of Expected Shortfall at level α. The problem with the intuitive definition above is the possible positive mass at the quantile itself. The exact definition of the Expected Shortfall at level α is therefore (Acerbi and Tasche (2002), Kalkbrener et al (2004):

DEFINITION 7.1

$$\mathsf{E}\,S_\alpha(L) \overset{\text{def}}{=} (1-\alpha)^{-1}\big(\mathsf{E}[L\,\mathbf{1}\{L > \mathsf{q}_\alpha(L)\}] + \mathsf{q}_\alpha(L) \cdot [\mathsf{P}\{L \le \mathsf{q}_\alpha(L)\} - \alpha]\big).$$

Here we take the quantile defined by

$$q_u(L) = \inf\{x|P(L \le x) \ge u\}$$

the smallest u-quantile

Since $\mathsf{E}\,S_\alpha(L) = \mathsf{E}\{Lg_\alpha(L)\}$ with the function

$$g_\alpha(Y) \overset{\text{def}}{=} (1-\alpha)^{-1}[\mathbf{1}\{Y > \mathsf{q}_\alpha(Y)\} + \beta_Y\mathbf{1}\{Y = \mathsf{q}_\alpha(Y)\}], \qquad (7.5)$$

where β_Y is a real number and

$$\beta_Y \overset{\text{def}}{=} \frac{\mathsf{P}\{Y \le \mathsf{q}_\alpha(Y)\} - \alpha}{\mathsf{P}\{Y = \mathsf{q}_\alpha(Y)\}} \ \text{ if } \ \mathsf{P}\{Y = \mathsf{q}_\alpha(Y)\} > 0.$$

the density of the associated maximal scenario turns out to be the function g_α. Note that $\mathsf{E}\, S_\alpha(Y) = \mathsf{E}\{Y \cdot g(Y)\}$ and $\mathsf{E}\, S_\alpha(X) \geq \mathsf{E}\{X \cdot g(Y)\}$ for every $X, Y \in V$.

7.2.2 Spectral Risk Measures

For the interpretation of this density function (7.5) in terms of risk aversion as outlined in Acerbi (2002), let us reformulate the expected shortfall as an integral over the quantile function, the inverse of the distribution of L. It is well-known that

$$\mathsf{E}\, S_\alpha(L) = (1 - \alpha)^{-1} \int_\alpha^1 q_u(L) du.$$

The implicit risk aversion with expected shortfall is, that all quantiles below α or all losses below the α quantile have no weights, i.e. there is no risk aversion and all losses above the α-quantile have the same risk aversion. Therefore the risk aversion weight function associated with $\mathsf{E}\, S_\alpha$ turns out to be

$$w_{\mathsf{E}\, S_\alpha}(u) \;=\; (1 - \alpha)^{-1}\mathbf{1}(u > \alpha). \tag{7.6}$$

From a risk management point of view there might be many other weights given to some confidence levels u. If the weight function is increasing, which is reasonable since higher losses should have larger risk aversion weight, then we arrive at spectral risk measures.

DEFINITION 7.2 *Let w be an increasing function from $[0, 1]$ such that $\int_0^1 w(u) du = 1$, then the map r_w defined by*

$$r_w(L) = \int_0^1 w(u) q_u(L) du$$

is called a spectral risk measure with weight function w.

The name spectral risk measure comes from the representation

$$r_w(X) = \int_0^1 \mathsf{E}\, S_\alpha(1 - \alpha)\mu_u(da) \tag{7.7}$$

$$\text{with the spectral measure } \mu([0, b]) = w(b). \tag{7.8}$$

This representation is very useful when we want to find the scenario function representing a spectral risk measure r_w.

PROPOSITION 7.1 *The density of the scenario associated with the risk measure equals*

$$L_w \overset{\text{def}}{=} g_w(L) \overset{\text{def}}{=} \int_0^1 g_\alpha(L)(1-\alpha)\mu(d\alpha). \tag{7.9}$$

Here $g_\alpha(L)$ is defined in formula (7.5). In particular

$$r_w(L) = \mathsf{E}(LL_w) \tag{7.10}$$

Proof: We have

$$
\begin{aligned}
r_w(L) &= \int_0^1 \mathsf{E}\, S_\alpha(L)(1-\alpha)\mu(d\alpha) \\
&= \int_0^1 \mathsf{E}(LL_\alpha)(1-\alpha)\mu(d\alpha) \\
&= \int_0^1 \max[\mathsf{E}\{Lg_\alpha(Y)\}|Y \in L_\infty](1-\alpha)\mu(d\alpha) \\
&\geq \max[\int_0^1 \mathsf{E}\{L \int_0^1 g_\alpha(Y)(1-\alpha)\mu(d\alpha)\}|Y \in L_\infty] \\
&= \max[\mathsf{E}\{Lg_w(Y)\}|\forall Y \in L_\infty] \\
&\geq \mathsf{E}\{Lg_w(L)\}
\end{aligned}
$$

Hence

$$r_w(L) = \max[\mathsf{E}\{Lg_w(Y)\}|\forall Y \in L_\infty] = \mathsf{E}\{Lg_w(L)\}$$

$$\diamond.$$

7.2.3 Coherent Allocation Measures

Starting with the representation (7.3) one can now find for each Y a linear function $h_Y = h_Y^r$ which satisfies

$$r(Y) = h_Y(Y) \text{ and } h_Y(X) \leq r(X), \forall X. \tag{7.11}$$

A "diversifying" capital allocation associated with r is given by

$$\Lambda_r(X,Y) = h_Y(X). \tag{7.12}$$

The function Λ_r is then *linear* in the first variable and *diversifying* in the sense that the capital allocated to a portfolio X is always bounded by the capital of X viewed as its own subportfolio

$$\Lambda(X,Y) \leq \Lambda(X,X). \tag{7.13}$$

$\Lambda(X, X)$ can be called the standalone capital or risk measure of X. In general we have the following two theorems: A linear and diversifying capital allocation Λ, which is continuous, i.e. $\lim_{\epsilon \to 0} \Lambda(X, Y + \epsilon X) = \Lambda(X, Y) \; \forall X$, at a portfolio Y, is uniquely determined by its associated risk measure, i.e. the diagonal values of Λ. More specifically, given the portfolio Y then the capital allocated to a subportfolio X of Y is the derivative of the associated risk measure ρ at Y in the direction of X.

PROPOSITION 7.2 *Let Λ be a linear, diversifying capital allocation. If Λ is continuous at $Y \in V$ then for all $X \in V$*

$$\Lambda(X, Y) = \lim_{\epsilon \to 0} \frac{r(Y + \epsilon X) - \rho(Y)}{\epsilon}.$$

The following theorem states the equivalence between positively homogeneous, sub-additive risk measures and linear, diversifying capital allocations.

PROPOSITION 7.3 (a) *If there exists a linear, diversifying capital allocation Λ with associated risk measure r, i.e. $r(X) = \Lambda(X, X)$, then r is positively homogeneous and sub-additive.*
(b) *If r is positively homogeneous and sub-additive then Λ_r as defined in (7.12) is a linear, diversifying capital allocation with associated risk measure r.*

7.2.4 Spectral Allocation Measures

Since in the case of spectral risk measures r_w the maximal linear functional in (7.11) can be identified as an integration with respect to the probability measure with density (7.9) from Theorem 1, we obtain $h_Y(X) = \mathsf{E}\{X g_w(Y)\}$ and therefore the following capital allocation

$$\Lambda_w(X, Y) \;=\; \mathsf{E}\{X g_w(Y)\} = \int_0^1 \mathsf{E}\, SC_\alpha(X, Y)(1 - \alpha)\mu(d\alpha) \quad (7.14)$$

$$= \int_0^1 \mathsf{E}\, SC_\alpha(X, Y)\widetilde{\mu}(d\alpha) \qquad (7.15)$$

$$\text{where} \qquad \mathsf{E}\, SC_\alpha(X, Y) = \mathsf{E}\{X g_\alpha(Y)\} \qquad (7.16)$$

is the Expected Shortfall Contribution and $\widetilde{\mu}$ is defined in (7.17). Intuitively, the capital allocated to transaction or subportfolio X in a portfolio Y equals its expectation under the generalized maximal scenario associated with w.

7.3 Weight Function and Mixing Measure

One might try to base the calibration or determination of the spectral risk measure based on the spectral measure μ or the weight function w. Since the weight function w is nothing else as the distribution function of μ, there is also a 1-1 correspondence to the more intuitive mixing measure

$$\widetilde{\mu}(d\alpha) \;=\; (1-\alpha)\mu(d\alpha). \tag{7.17}$$

If we define more generally for an arbitrary measure $\widetilde{\mu}$ the functional

$$\widetilde{\rho} \;=\; \int_0^1 \mathsf{E}\, S_\alpha \widetilde{\mu}(da) \tag{7.18}$$

then $\widetilde{\rho}$ is coherent iff $\widetilde{\mu}$ is a probability measure. Since

$$
\begin{aligned}
1 \;&=\; \widetilde{\mu}([0,1]) = \int_0^1 (1-u)\mu(du) \\
&= \int_0^1 \int_0^1 \mathbf{1}[u,1](v)\,dv\,\mu(du) = \int_0^1 \int_0^1 \mathbf{1}[0,v](u)\mu(du)\,dv \\
&= \int_0^1 w(v)\,dv.
\end{aligned}
$$

If we have now a probability measure $\widetilde{\mu}$ on $[0,1]$ the representing μ and w in (7.1,7.2) can be obtained by

$$\frac{d\mu}{d\widetilde{\mu}} \;=\; \frac{1}{1-\alpha} \tag{7.19}$$

$$w(b) \;=\; \mu([0,b]) = \int_0^b \frac{1}{1-\alpha}\widetilde{\mu}(d\alpha). \tag{7.20}$$

7.4 Risk Aversion

If we assume a discrete measure

$$\widetilde{\mu} \;=\; \sum_{i=1}^n p_i \delta_{\alpha_i} \tag{7.21}$$

then the risk aversion function w is an increasing step function with step size of $p_i/(1 - \alpha_i)$ at the points α_i

$$w(b) = \sum_{\alpha_i \leq b} \frac{p_i}{1 - \alpha_i}. \tag{7.22}$$

This has to be kept in mind. If we assume equal weights for the two expected shortfall at 99% and 90% then the increase in risk aversion at the first quantile 90% is $0.5/0.1 = 5$ and $0.5/0.01 = 50$. The risk aversion against losses above the 99% is therefore 11 times higher than against those between the 90% and 99% quantile. It is therefore sensible to assume quite small weights on $\mathsf{E}\, S_\alpha$ with large αs.

7.5 Implementation

There are several ways to implement a spectral contribution in a portfolio model. According to Acerbi(2002) a Monte-Carlo-based implementation of the spectral risk measure would work as follows:

Let L^n be the n-th realization of the portfolio loss. If we have generated N loss distribution scenario, let us denote by $n : N$ index of the n-th largest loss which itself is then denote by $L^{n:N}$, i.e. the indices $1 : N, 2 : N, .., N : N \in \mathsf{N}$ are defined by the property that

$$L^{1:N} < L^{2:N} < ... < L^{N:N}$$

The approximative spectral risk measure is then defined by

$$\sum_{n=1}^{N} L^{n:N} w(n/N) / \sum_{k=1}^{N} w(k/N)$$

Therefore a natural way to approximate the spectral contribution of another random variable L_i, which specifically might be a transaction in the portfolio represented by L or a subportfolio of L, is

$$\sum_{n=1}^{N} L_i^{n:N} \frac{w(n/N)}{\sum_{k=1}^{N} w(k/N)}, \tag{7.23}$$

where $L_i^{n,N}$ denotes the loss in transaction i in the scenario $n : N$, i.e. in the scenario where the portfolio loss was the n-th largest. It is then expected

that

$$\mathsf{E}(L_i L_w) = \lim_{N \to \infty} \sum_{n=1}^{N} L_i^{n:N} \frac{w(n/N)}{\sum_{k=1}^{N} w(k/N)}.$$

As in most applications we assume that

$$L = \sum_i L_i$$

with the transaction loss variable L_i and in the example later we will actually calculate within a multi-factor Merton-type credit portfolio model.

7.5.1 Mixing Representation

Let us review the standard implementation of the expected shortfall contribution. In the setting of the previous setting we can see that for $w(u) = \frac{1}{1-\alpha} \mathbf{1}[\alpha, 1](u)$ the weights for all scenarios with $\frac{n}{N} < \alpha$ is 0 and for all others it is

$$\frac{\frac{1}{1-\alpha}}{\sum_{k=\{(\alpha)N\}}^{N} \frac{1}{1-\alpha}}$$

$$\cong \quad \frac{1}{(1-\alpha)N}$$

(Here $[\cdot]$ denote the Gauss brackets.) Therefore the expected shortfall contribution equals

$$\frac{1}{\{(1-\alpha)N\}} \sum_{n=(\alpha N)}^{N} L_i^{n:N} \qquad (7.24)$$

or more intuitively the average of the counterparty i losses in all scenarios where the portfolio losses was higher or equal than the $[\alpha N]$ largest portfolio loss.

Due to the fact that we have chosen a finite convex combination of Expected Shortfall , i.e. the mixing measure

$$\widetilde{\mu}(du) \quad = \quad \sum_{k=1}^{K} p_i \delta_{\alpha_i}$$

and formulae (7.24) and (7.18) we will take for a transaction L_i the approximation

$$\mathrm{SCA}(L_i, L)_{\mathrm{vecp, vec\alpha}, N} \quad = \quad \sum_{k=1}^{K} p_i \left[\frac{1}{\{(1-\alpha_i)N\}} \sum_{n=[\alpha_i N]}^{N} L_i^{n:N} \right] \qquad (7.25)$$

as the Spectral Capital Allocation with discrete mmixing measure μ represented by the vectors $\text{vec} p = (p_1, .., p_K), \text{vec} \alpha = (\alpha_1, .., \alpha_K)$ for a Monte-Carlo-Sample of length N.

7.5.2 Density Representation

Another possibility is to rely on the approximation of the Expected Shortfall Contribution as in Kalkbrener et al (2004) and to integrate over the spectral measure μ:

$$\mathsf{E}(L_i L_w) = \lim_{N \to \infty} \int_0^1 \left\{ \sum_{n=1}^N L_i^{n:N} \frac{w_\alpha(i/N)}{\sum_{k=1}^N w_\alpha(k/N)} (1-\alpha) \right\} \mu(da) \quad (7.26)$$

If L has a continuous distribution than we have that

$$\mathsf{E}(L_i L_w) = \mathsf{E}\{L_i \int_0^1 L_\alpha \mu(d\alpha)\}$$

$$= \int_0^1 \mathsf{E}[L_i \mathbf{1}\{L > q_\alpha(L)\}](1-\alpha)^{-1} \mu(d\alpha)$$

$$= \lim_{N \to \infty} N^{-1} \sum_{n=1}^N L_i^n \int_0^1 \mathbf{1}\{L^n > q_\alpha(L)\}(1-\alpha)^{-1} \mu(d\alpha) \quad (7.27)$$

If L has not a continuous distribution we have to use the density function (7.9) and might approximate the spectral contribution by

$$\mathsf{E}(L_i L_w) \sim N^{-1} \sum_{n=1}^N L_i^n g_w(L^n). \quad (7.28)$$

The actual calculation of the density g_w in (7.28) might be quite involved. On the other hand the integration with respect to μ in (7.26) and (7.27) is also not easy. If w is a step function as in the example 1 above, then μ is a sum of weighted Dirac-measure and the implementation of spectral risk measure as in (7.23) is straightforward.

7.6 Credit Portfolio Model

In the examples below we apply the presented concepts to a standard default only type model with a normal copula based on an industry and region factor

model, with 27 factors mainly based on MSCI equity indices. We assume fixed recovery and exposure-at-default. For a specification of such a model, we could refer to Bluhm et al. (2002) or other text books on credit risk modeling.

7.7 Examples

7.7.1 Weighting Scheme

Lets take 5 quantile 50%, 90%, 95%, 99%, 99.9% and the 99.98% quantile. We like now to find weighting scheme for Expected Shortfall, which still gives a nice risk aversion function. Or inversely we start with a sensible risk aversion as in (7.29) and then solve for the suitable convex combination of expected shortfall measures.

As a first step in the application of spectral risk measures one might think to give to different loss probability levels different weight. This is a straight-forward extension of expected shortfall. One might view Expected Shortfall at the 99%-level view as a risk aversion which ignores losses below the 99%-quantile and all losses above the 99%-quantile have the same influence. From an investors point of view this means that only senior debts are cushioned by risk capital. One might on the other hand also be aware of losses which occur more frequently, but of course with a lower aversion than those appearing rarely.

As a concrete example one might set that losses up to the 50% confidence level should have zero weights, losses between 50% and 99% should have a weight w_0 and losses above the 99%-quantile should have a weight of $k_1 w_0$ and above the 99.9% quantile it should have a weight of $k_2 w_0$. The first tranch from 50% to 99% correspond to an investor in junior debt, and the tranch from 99% to 99.9% to a senior investor and above the 99.9% a super senior investor or the regulators are concerned. This gives a step function for w:

$$w(u) \;=\; w_0 \mathbf{1}(0.99 > u > 0.5) + k_1 w_0 \mathbf{1}(0.999 > u > 0.99)$$
$$+ k_2 w_0 \mathbf{1}(1 > u > 0.999) \tag{7.29}$$

The parameter w_0 should be chosen such that the integral over w is still 1.

7.7.2 Concrete Example

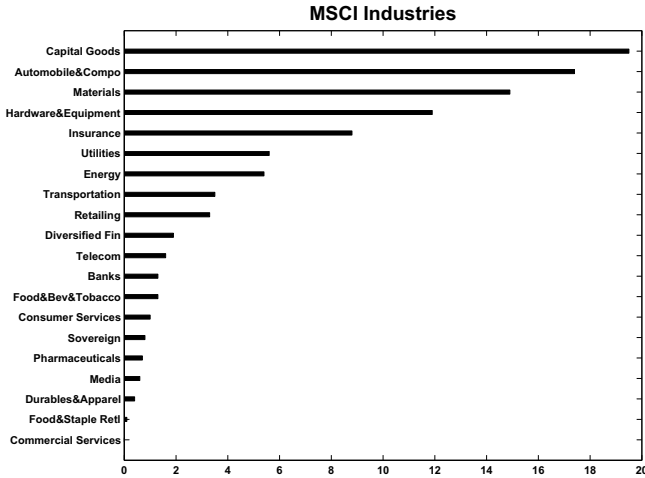The portfolio consists of 279 assets with total notional EUR 13.7bn and the following industry and regions breakdown:

MSCI Industries

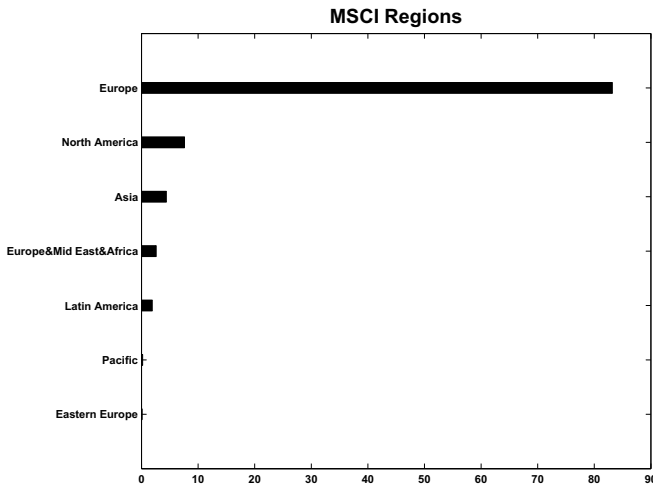Figure 7.1. MSCI industry breakdown ⊙ XFGIndustryBreakdown

MSCI Regions

Figure 7.2. MSCI region breakdown ⊙ XFGRegionsBreakdown

The portfolio correlation structure is obtained from the R^2 and the correlation structure of the industry and regional factors. The R^2 is the R^2 of the

one-dimensional regression of the asset returns with respect to its composite factor, modeled as the sum of industry and country factor. The underlying factor model is based on 24 MSCI Industries and 7 MSCI Regions. The weighted average R^2 is 0.5327.
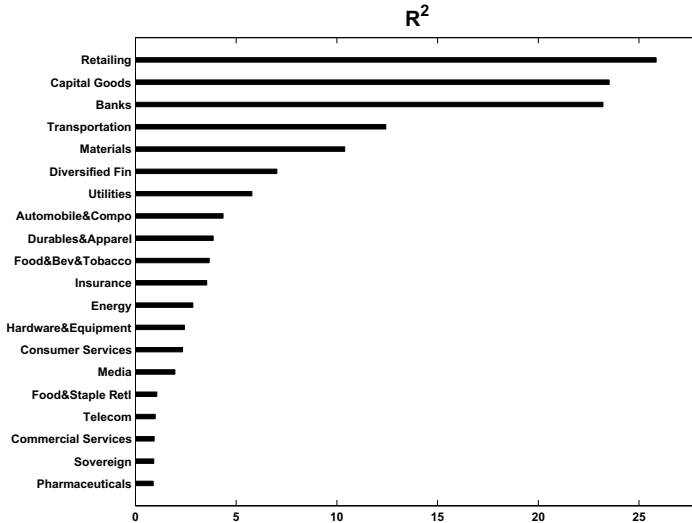

Figure 7.3. R^2 values of different MSCI industries.
⊡ XFGRsquared

The risk contributions are calculated at quantiles $50\%, 90\%, 95\%, 99\%, 99.9\%$ and 99.98%.

Figure 7.4 shows the total Expected Shortfall Contributions allocated to the industries normalized with respect to automobile industry risk contributions and ordered by $\mathrm{ESC}_{99\%}$.

In order to capture all risks of the portfolio a risk measure, which combines few quantile levels, is needed. As one can see, Hardware and Materials have mainly tail exposure (largest consumption of ESC at the 99.98%-quantile), where Transportation, Diversified Finance and Sovereign have the second to fourth largest consumption of ESC at the 50%-quantile, i.e. are considerable more exposed to events happening roughly every second year as Hardware and Materials.

The spectral risk measure as a convex combination of Expected Shortfall risk measures at the following quantiles $50\%, 90\%, 95\%, 99\%, 99.9\%$ and 99.98% can capture both effects, at the tail and at the median of the loss distribution.

Four spectral risk measures are calculated. The first three are calibrated in
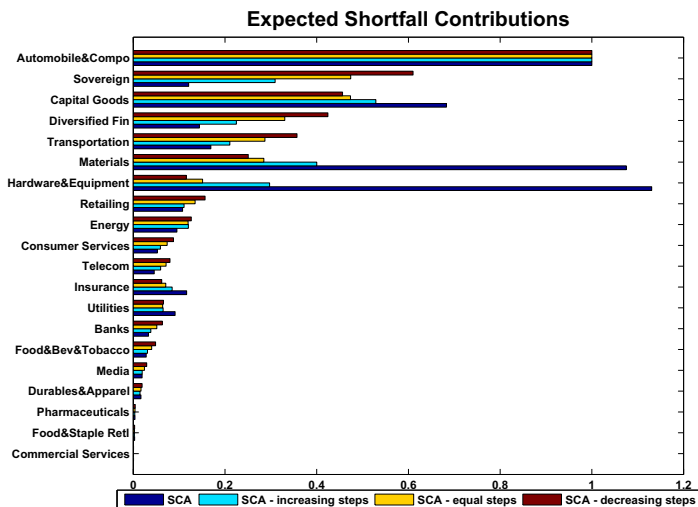
Expected Shortfall Contributions

Figure 7.4. Expected shortfall contributions for different industries at different quantiles. **Q** XFGESC

terms of increase of the risk aversion function at each considered quantile as in Figure 7.5. The least conservative one is "SCA - decreasing steps" in which the risk aversion increases at each quantile by half the size it has increased at the quantile before. "SCA -equal steps" increases in risk aversion by the same amount at each quantile, "SCA -increasing steps" increases in risk aversion at each quantile by doubling the increase at each quantile. The last most conservative one is SCA - 0.1/0.1/0.1/0.15/0.15/0.4, in which the weights of $\widetilde{\mu}$ are directly set to 0.1 at the 50%, 90%, 95%- quantiles, 0.15 at the 99% and 99.9%- quantiles and 0.4 at the 99.98%-quantile as in Figure 7.6. The last one has a very steep increase in the risk aversion at the extreme quantiles.

As a comparison to the expected shortfall, the chart below shows the Spectral risk allocation allocated to industries ordered by SCA - equal steps and normalized with respect to automobile industry SCA as in Figure 7.7.

All tables so far were based on the risk allocated to the industries. Much of the displayed effects are just driven by exposure, i.e. "Automotive" is by far the largest exposure in that portfolio and all sensible risk measure should mirror this concentration. Interestingly enough the most tail emphasizing measures are the exceptions. There the largest contributors Hardware and Materials have actually less than 10% of the entire exposure.

Usually one uses as well percentage figures and risk return figures for portfolio management. On the chart "RC/TRC" the percentage of total risk (TRC)
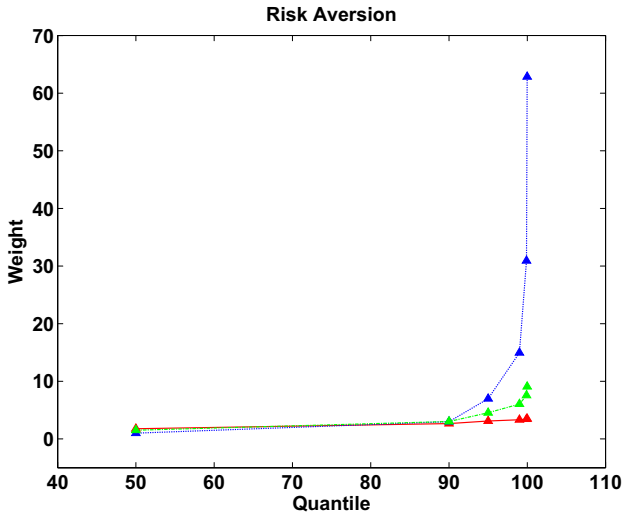
Figure 7.5.   Risk aversion calculated with respect to different methods.   The dotted blue, dashed-dotted and solid lines represent "SCA - decreasing steps", "SCA - equal steps" and "SCA - incresing steps" correspondingly.
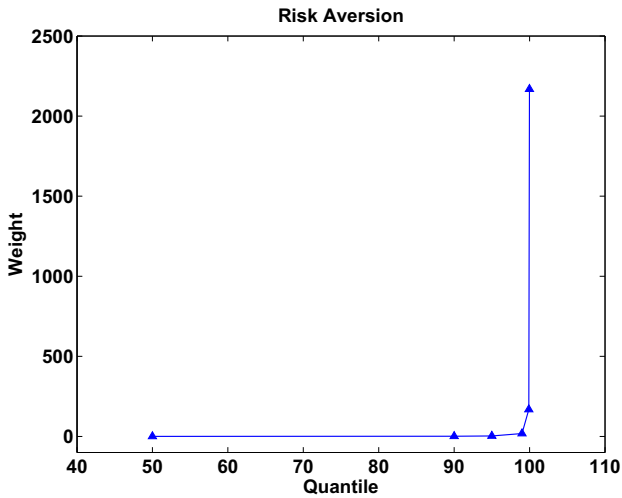◨ XFGriskaversion

Figure 7.6.   Risk aversion when the weights are directly set to 0.1 at the 50%, 90%, 95%- quantiles, 0.15 at the 99% and 99.9%- quantiles and 0.4 at the 99.98%-quantile.
◨ XFGriskaversion2

Figure 7.7. Different risk contributions with respect to different SCA methods. XFGSCA

allocated to the specific industries is displayed in Figure 7.8.

Figure 7.8. Total risk contributions with respect to different SCA methods. XFGRCTRC

For the risk management the next table showing allocated risk capital per exposure is very useful. It compares the riskiness of the industry normalized

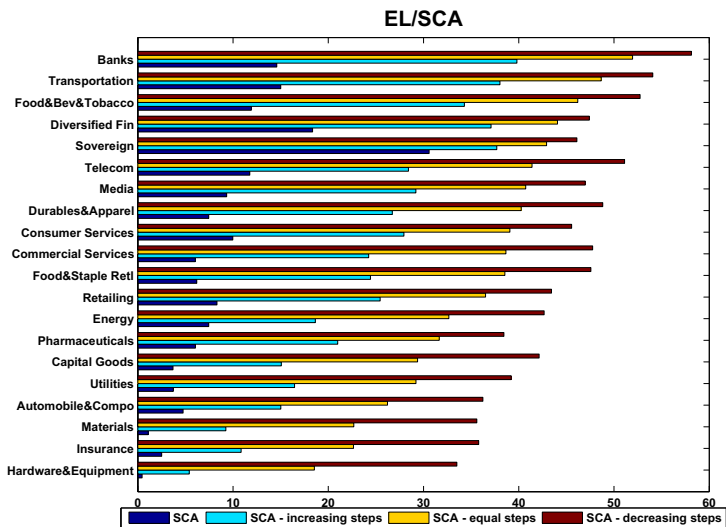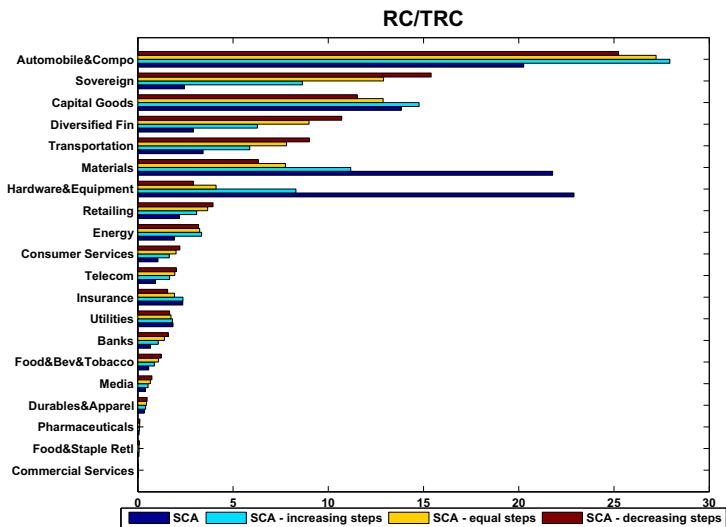by their exposure. Intuitively it means that if you increase the exposure in "transportation" by a small amount like 100.000 Euro than the additionally capital measured by SCA-increasing steps will increase by 2.5%, i.e. by 2.5000 Euro. In that sense it gives the marginal capital rate in each industry class. Here the sovereign class is the most risky one. In that portfolio the sovereign exposure was a single transaction with a low rated country and it is therefore no surprise that "sovereign" performance worst in all risk measures.
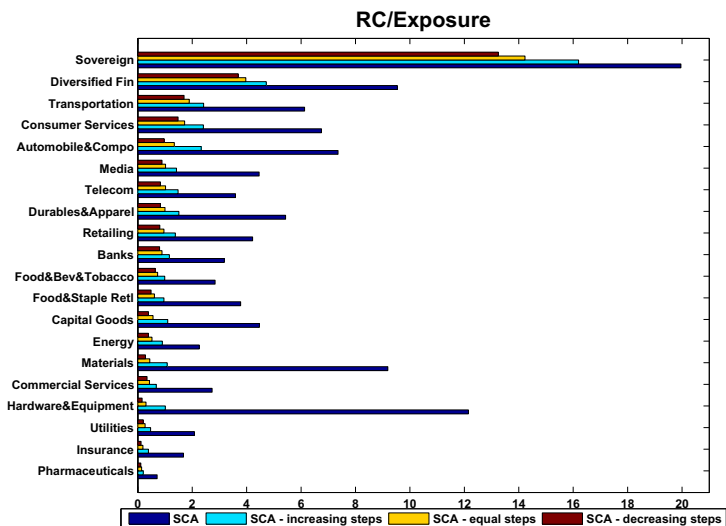

Figure 7.9. Allocated risk capital per exposure with respect to different SCA methods. ⊙XFGRCExposure

With that information one should now be in the position to judge about the possible choice of the most sensible spectral risk measure among the four presented. The measure denoted by SCA based on the weights 0.1,0.1,0.1, 0.15,0.15, 0.4, overemphasis tail risk and ignores volatility risk like the 50%-quantile. From the other three spectral risk measures, also the risk aversion function of the one with increasing steps, does emphasis too much the higher quantiles. SCA decreasing steps seems to punished counterparties with a low rating very much, it seems to a large extend expected loss driven, which can be also seen in the following table on the RAROC-type Figures 7.10. On that table "decreasing steps" does not show much dispersion. One could in summary therefore recommend SCA-equal steps.

For information purpose we have also displayed the Expected Loss/Risk Ratio for the Expected Shortfall Contribution in Figure 7.11. Here the dispersion for the ESC at the 50% quantile is even lower as for the SCA-decreasing steps.
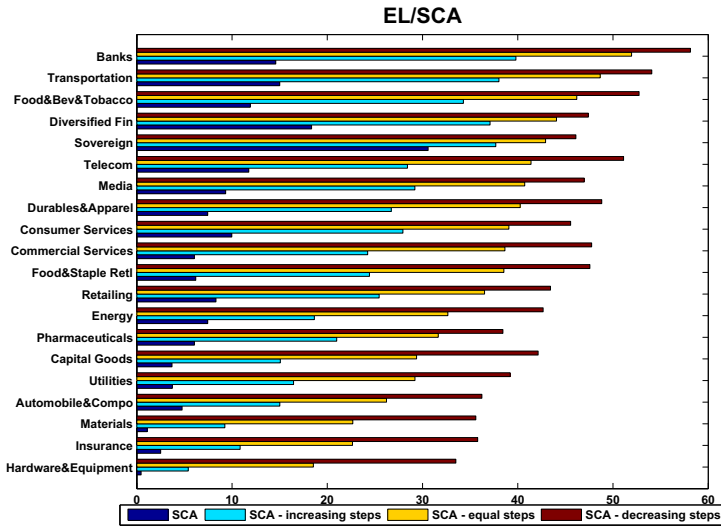
EL/SCA

Figure 7.10. EL/SCA with respect to different SCA methods.
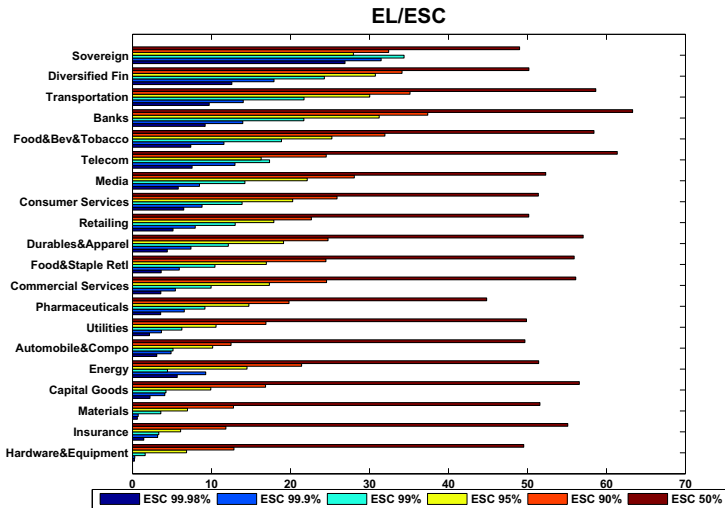⌀ XFGELSCA

EL/ESC

Figure 7.11. Expected Loss/Risk Ratio for the Expected Shortfall Contribution at different quantiles. ⌀ XFGELESC

7.8 Summary

In order to combine different loss levels in one risk measure spectral risk measures provide a sensible tool. Weighting of the quantiles is usually be done by the risk aversion function. Starting from an implementation point of view it looks more convenient to write a spectral risk measure as a convex combination of expected shortfall measures. However one has to be careful in the effects on the risk aversion function. All this holds true and become even more important if capital allocation is considered, which finally serves as a decision tool to differentiate sub-portfolios with respect to their riskiness. We analyze an example portfolio with respect to the risk impact of the industries invested in. Our main focus are the different specification of the spectral risk measure and we argue in favour for the spectral risk measure based on a risk aversion which has the same magnitude of increase at each considered quantile, namely the 50%, 90%, 95%, 99%, 99.9%, and 99.98% quantile. This risk measure exhibits a proper balance between tail risk and more volatile risk.

Bibliography

C. Acerbi (2002). Spectral measures of risk: A coherent representation of subjective risk aversion, *Journal of Banking and Finance* **26**, 1505-1518.

C. Acerbi and D. Tasche (2002). On the coherence of expected shortfall, *Journal of Banking and Finance*, **26**: 1487-1503.

P. Artzner, F. Delbaen, J.-M. Eber and D. Heath (1997). Thinking coherently, *RISK*, **Nov**: 68-71.

P. Artzner, F. Delbaen, J.-M. Eber and D. Heath (1999). Coherent measures of risk, *Mathematical Finance*, **9**: 203-228.

C. Bluhm, L. Overbeck and C. Wagner (2002). An Introduction to Credit Risk Modeling, CRC Press/Chapman & Hall.

M. Crouhy, D. Galai and R. Mark (1999). A comparative analysis of current credit risk models, *Journal of Banking and Finance*, **24**.

F. Delbaen, (2000). Coherent risk measures, *Lecture notes, Scuola Normale Superiore di Pisa*.

M. Denault, (2001). Coherent allocation of risk capital, *Journal of Risk* **4**, 1.

M. Kalkbrener, (2002). An axiomatic approach to capital allocation, *Technical Document, Deutsche Bank AG, Frankfurt*.

M. Kalkbrener, H. Lotter and L. Overbeck (2004). Sensible and efficient capital allocation, *Risk*, **Jan**.

A. Kurth and D. Tasche (2003). Contributions to credit risk, *Risk*, **March**: 84-88.

L. Overbeck (2000). Allocation of economic capital in loan portfolios. Proceedings "Measuring Risk in Complex Stochastic Systems", Berlin 1999. Editors: Stahl/Härdle, *Lecture Notes in Statistics*, Springer.

L. Overbeck and G. Stahl. (2003). Stochastic essentials for the risk management of credit portfolios, *Kredit und Kapital*, **1**.

L. Overbeck (2004). Spectral Capital Allocation, RISK Books, London.

R.T. Rockafellar and S. Uryasev. (2000). Optimization of conditional value-at-risk, *Journal of Risk*, **2**: 21-41.

8 Valuation and VaR Computation for CDOs Using Stein's Method

Nicole El Karoui, Ying Jiao, David Kurtz

8.1 Introduction

8.1.1 A Primer on CDO

Collateralized debt obligations (CDOs) are an innovation in the structured finance market that allow investors to invest in a diversified portfolio of assets at different risk attachment points to the portfolio. The basic concept behind a CDO is the redistribution of risk: some securities backed by a pool of assets in a CDO will be higher rated than the average rating of the portfolio and some will be lower rated.

Generally, CDOs take two forms, cash flow or synthetic. For a cash flow vehicle, investor capital is used directly to purchase the portfolio collateral and the cash generated by the portfolio is used to pay the investors in the CDO. Synthetic CDOs are usually transactions that involve an exchange of cash flow through a credit default swap or a total rate of return swap. The CDO basically sells credit protection on a reference portfolio and receives all cash generated on the portfolio. In these types of transaction, the full capital structure is exchanged and there is no correlation risk for the CDO issuer.

In this study, we are primarily interested in valuing (synthetic) single tranche CDO. It is very important to note that these products are exposed to correlation risk. In practice the CDO issuer sells protection on a portion of the capital structure on a reference portfolio of names. In exchange, he receives a running spread, usually paid quarterly, which value depends on the risk of the individual issuers in the reference portfolio and on a correlation hypothesis between those names. For liquid reference portfolios (indices) like Trac-X and iBoxx there exists now a liquid market for these single tranche CDOs and as a consequence for the correlation.

We now describe mathematically the payoff of a single tranche CDO on a reference portfolio of size n and maturity T. Let τ_i denote the default time of the ith name in the underlying portfolio and N_i be its notional value. The total notional is $N = \sum_{i=1}^{n} N_i$. We use ω_i to represent the weight of the ith name in the portfolio i.e. $\omega_i = N_i/N$. Let R_i be the recovery rate of name i. The cumulative loss process is given by $L_t = \sum_{i=1}^{n} N_i(1 - R_i)\mathbf{1}_{\{\tau_i \leq t\}}$ and the percentage loss process is

$$ l_t = L_t/N = \sum_{i=1}^{n} \omega_i(1 - R_i)\mathbf{1}_{\{\tau_i \leq t\}}. $$

Usually the capital structure is decomposed in the following way: let us write the interval $(0, 1]$ as the unions of the non-overlapping interval $(\alpha_{j-1}, \alpha_j]$ where $0 = \alpha_0 < \alpha_1 < \cdots < \alpha_k = 1$. The points α_{j-1} and α_j are called, respectively, the attachment and detachment points of the jth tranche. At time t, the loss of the jth tranche is given as a call spread i.e. $l_t^{(j)} = (l_t - \alpha_{j-1})^+ - (l_t - \alpha_j)^+$.

The cash flows of a single tranche CDO are as follows: The protection seller, on one hand, receives at times $\{t_1, \cdots, t_M = T\}$ the coupon $\kappa_j c_{t_u}^j$ ($u = 1, \cdots, M$) where κ_j is called the spread of the tranche and $c_{t_u}^j = 1 - l_{t_u}^{(j)}/(\alpha_j - \alpha_{j-1})$ is the outstanding notional of the tranche at time t_u. The protection buyer, on the other hand, receives at each default time t that occurs before the maturity the amount $\Delta l_t^{(j)} l_t^{(j)} - l_{t-}^{(j)}$.

From the point of view of pricing, for the jth tranche of the CDO, our objective is to find the value of the spread κ_j. From now on, we shall consider a continuously compounded CDO of maturity T. The value of the default leg and the premium leg are given respectively by the following formulas:

$$ \text{Default Leg} = -(\alpha_j - \alpha_{j-1}) \int_0^T B(0, t) q(\alpha_{j-1}, \alpha_j, \mathrm{d}t), $$

$$ \text{Premium Leg} = \kappa_j \times (\alpha_j - \alpha_{j-1}) \int_0^T B(0, t) q(\alpha_{j-1}, \alpha_j, t) \mathrm{d}t $$

where $B(0, t)$ is the value at time 0 of a zero coupon maturing at time t assuming deterministic interest rates and $q(\alpha_{j-1}, \alpha_j, t) := \mathsf{E}(c_t^j)$ is the tranche survival probability at time t computed under a given risk-neutral probability. Thanks to the integration by part formula, the fair spread κ_j is then computed

as

$$\kappa_j = \frac{1 - B(0,T)q(\alpha_{j-1}, \alpha_j, T) + \int_0^T q(\alpha_{j-1}, \alpha_j, t)B(0, \mathrm{d}t)}{\int_0^T B(0,t)q(\alpha_{j-1}, \alpha_j, t)\mathrm{d}t}. \tag{8.1}$$

To obtain the value of the preceding integrals, the key term to compute is the functions q, which can be expressed as a linear combination of call prices of the form

$$C(t,k) = \mathsf{E}\{(l_t - k)_+\}. \tag{8.2}$$

8.1.2 Factor Models

The main element in computing CDO value is the distribution of the percentage loss l. As mentioned earlier, this distribution depends in a critical manner on the spread (or market implied default probabilities) of the individual names and their correlation as quoted for instance in the liquid tranche market. As a consequence, we need a way to model the correlation between default times of individual names. In practice and in order to obtain tractable results, the market adopts a simplified approach - the factor models.

The main characteristic of the factor models, e.g. see Andersen and Sidenius and Basu (2003), is the conditional independence between the default times $\tau_1, \cdots, \tau_n$. In this framework, the market is supposed to contain some latent factors which impact all concerning firms at the same time. Conditionally on these factors, denoted by U (and we may assume U is uniformly distributed on $(0, 1)$ without loss of generality), the default events $E_i = \{\tau_i \leq t\}$ are supposed to be independent. To define the correlation structure using the factor framework, it is sufficient to define the conditional default probabilities. In a nutshell, this tantamounts to choose a function F such that $0 \leq F \leq 1$ and

$$\int_0^1 F(p, u)\mathrm{d}u = p, \quad 0 \leq p \leq 1.$$

If $p_i = \mathrm{P}(E_i)$, the function $F(p_i, u)$ is to be interpreted as $\mathrm{P}(E_i|U = u)$.

The standard Gaussian copula case with correlation ρ corresponds to the function F defined by

$$F(p, u) = \Phi\left\{\frac{\Phi^{-1}(p) - \sqrt{\rho}\Phi^{-1}(u)}{\sqrt{1 - \rho}}\right\}$$

where $\Phi(x)$ is the distribution function of the standard normal distribution $N(0,1)$. Other copula functions, which corresponds to different types of correlation structure, can be used in a similar way. The main drawback of the Gaussian correlation approach is the fact that one cannot find a unique model parameter ρ able to price all the observed market tranches on a given basket. This phenomenon is referred to as *correlation skew* by the market practitioners. One way to take into account this phenomenon is to consider that the correlation ρ is itself dependent on the factor value. See Burtschell, Gregory and Laurent (2007) for a discussion of this topic.

In the factor framework, the conditional cumulative loss l can be written as a sum of independent random variables given U. It is then possible to calculate (8.2) by analytical or numerical methods:

⊡ Firstly, calculate the conditional call value using exact or approximated numerical algorithms,

⊡ secondly, integrate the result over the factor U.

In the sequel, we will explore new methodologies to compute approximations of the conditional call value in an accurate and very quick manner.

8.1.3 Numerical Algorithms

The challenge for the practitioners is to compute quickly prices for their (usually large) books of CDOs in a robust way.

Several methods are proposed to speed up the numerical calculations, such as the recursive method: Hull and White (2004), Brasch (2004), saddle-point method: Martin, Thompson and Browne (2001), Antonov, Mechkov and Misirpashaev (2005) and the Gaussian approximation method: Vasicek (1991). In this paper, we propose a new numerical method which is based on the Stein's method and the zero-bias transformation.

Stein's method is an efficient tool to estimate the approximation errors in the limit theorem problems. We shall combine the Stein's method and the zero bias transformation to propose first-order approximation formulas in both Gauss and Poisson cases. The error estimations of the corrected approximations are obtained. We shall compare our method with other methods numerically. Thanks to the simple closed-form formulas, we reduce largely the computational burden for standard single tranche deals.

In financial problems, the binomial-normal approximation has been studied in different contexts. In particular, Vasicek (1991) has introduced the normal

approximation to a homogeneous portfolio of loans. In general, this approximation is of order $\mathcal{o}(1/\sqrt{n})$. The Poisson approximation, less discussed in the financial context, is known to be robust for small probabilities in the approximation of binomial laws. (One usually asserts that the normal approximation remains robust when $np \geq 10$. If np is small, the binomial law approaches a Poisson law.) In our case, the size of the portfolio is fixed for a standard synthetic CDO tranche and $n \approx 125$. In addition, the default probabilities are usually small. Hence we may encounter both cases and it is mandatory to study the convergence speed since n is finite.

The rest of this study is organized as follows: We present in Section 8.2 the theoretical results; Section 8.3 is devoted to numerical tests; finally Section 8.4 explores real life applications, namely, efficient pricing of single tranche CDO and application of this new methodology to VaR computation.

8.2 First Order Gauss–Poisson Approximations

8.2.1 Stein's Method - the Normal Case

Stein's method is an efficient tool to study the approximation problems. In his pioneer paper, Stein (1972) first proposed this method to study the normal approximation in the central limit theorem. The method has been extended to the Poisson case later by Chen (1975).

Generally speaking, the zero bias transformation is characterized by some functional relationship implied by the reference distributions, normal or Poisson, such that the "distance" between one distribution and the reference distribution can be measured by the "distance" between the distribution and its zero biased distribution.

In the framework of Stein's method, the zero bias transformation in the normal case is introduced by Goldstein and Reinert (1997), which provides practical and concise notation for the estimations. In the normal case, the zero biasing is motivated by the following observation of Stein: a random variable Z has the centered normal distribution $N(0, \sigma^2)$ if and only if $\mathsf{E}\{Zf(Z)\} = \sigma^2 \mathsf{E}\{f'(Z)\}$ for all regular enough functions f. In a more general context, Goldstein and Reinert propose to associate with any random variable X of mean zero and variance $\sigma^2 > 0$ its zero bias transformation random variable X^* if the following relationship (8.3) holds for any function f of C^1-type,

$$\mathsf{E}\{Xf(X)\} = \sigma^2 \mathsf{E}\{f'(X^*)\}. \tag{8.3}$$

The distribution of X^* is unique with density function given by $p_{X^*}(x) = \sigma^{-2}\mathsf{E}(X\mathbf{1}_{\{X>x\}})$.

The centered normal distribution is invariant by the zero bias transformation. In fact, X^* and X have the same distribution if and only if X is a centered Gaussian variable.

We are interested in the error of the normal approximation $\mathsf{E}\{h(X)\} - \mathsf{E}\{h(Z)\}$ where h is some given function and Z is a centered normal r.v. with the same variance σ^2 of X. By *Stein's equation*:

$$xf(x) - \sigma^2 f'(x) = h(x) - \Phi_\sigma(h) \tag{8.4}$$

where $\Phi_\sigma(h) = \mathsf{E}\{h(Z)\}$. We have that

$$\begin{aligned}
\mathsf{E}\{h(X)\} - \Phi_\sigma(h) &= \mathsf{E}\{Xf_h(X) - \sigma^2 f'_h(X)\} = \sigma^2\mathsf{E}\{f'_h(X^*) - f'_h(X)\} \\
&\leq \sigma^2\|f''_h\|_{\sup}\mathsf{E}(|X^* - X|)
\end{aligned} \tag{8.5}$$

where f_h is the solution of (8.4). Here the property of the function f_h and the difference between X and X^* are important for the estimations.

The Stein's equation can be solved explicitly. If $h(t)\exp(-\frac{t^2}{2\sigma^2})$ is integrable on $\mathbb{R}$, then one solution of (8.4) is given by

$$f_h(x) = \frac{1}{\sigma^2\phi_\sigma(x)}\int_x^\infty \{h(t) - \Phi_\sigma(h)\}\phi_\sigma(t)dt \tag{8.6}$$

where $\phi_\sigma(x)$ is the density function of $N(0,\sigma^2)$. The function f_h is one order more differentiable than h. Stein has established that $\|f''_h\|_{\sup} \leq 2\|h'\|_{\sup}/\sigma^2$ if h is absolutely continuous.

For the term $X - X^*$, the estimations are easy when X and X^* are independent by using a symmetrical term $X^s = X - \widetilde{X}$ where $\widetilde{X}$ is an independent duplicate of X:

$$\mathsf{E}(|X^* - X|) = \frac{1}{4\sigma^2}\mathsf{E}\left(|X^s|^3\right), \quad \mathsf{E}(|X^* - X|^k) = \frac{1}{2(k+1)\sigma^2}\mathsf{E}\left(|X^s|^{k+2}\right). \tag{8.7}$$

When it concerns dependent random variables, a typical example is the sum of independent random variables. We present here a construction of zero biased variable introduced in Goldstein and Reinert (1997) using a random index to well choose the weight of each summand variable.

Proposition 8.1 *Let X_i $(i = 1, \ldots, n)$ be independent zero-mean r.v. of finite variance $\sigma_i^2 > 0$ and X_i^* having the X_i-zero normal biased distribution.*

We assume that $(\bar{X}, \bar{X}^*) = (X_1, \ldots, X_n, X_1^*, \ldots, X_n^*)$ are independent r.v. Let $W = X_1 + \cdots + X_n$ and denote its variance by σ_W^2. Let $W^{(i)} = W - X_i$ and I be an random index which is independent of $(\bar{X}, \bar{X}^*)$ such that $\mathrm{P}(I = i) = \sigma_i^2/\sigma_W^2$. Then $W^* = W^{(I)} + X_I^*$ has the W-zero biased distribution.

Although W and W^* are dependent, the above construction based on a random index choice enables us to obtain the estimation of $W - W^*$, which is of the same order of $X - X^*$ in the independent case:

$$\mathsf{E}\left(|W^* - W|^k\right) = \frac{1}{2(k+1)\sigma_W^2} \sum_{i=1}^{n} \mathsf{E}\left(|X_i^s|^{k+2}\right), \quad k \geq 1. \qquad (8.8)$$

8.2.2 First-Order Gaussian Approximation

In the classical binomial-normal approximation, as discussed in Vasicek (1991), the expectation of functions of conditional losses can be calculated using a Gaussian expectation. More precisely, the expectation $\mathsf{E}\{h(W)\}$ where W is the sum of conditional independent individual loss variables can be approximated by $\Phi_{\sigma_W}(h)$ where

$$\Phi_{\sigma_W}(h) = \frac{1}{\sqrt{2\pi}\sigma_W} \int_{-\infty}^{\infty} h(u) \exp\left(-\frac{u^2}{2\sigma_W^2}\right) \mathrm{d}u$$

and σ_W is the standard deviation of W. The error of this zero-order approximation is of order $\mathcal{O}(1/\sqrt{n})$ by the well-known Berry-Esseen inequality using the Wasserstein distance, e.g. Petrov (1975), Chen and Shao (2005), except in the symmetric case.

We shall improve the approximation quality by finding a correction term such that the corrected error is of order $\mathcal{O}(1/n)$ even in the asymmetric case. Some regularity condition is required on the considered function. Notably, the call function, not possessing second order derivative, is difficult to analyze. In the following theorem, we give the explicit form of the corrector term alongside the order of the approximation.

PROPOSITION 8.1 *Let $X_1, \ldots, X_n$ be independent random variables of mean zero such that $\mathsf{E}(X_i^4)$ $(i = 1, \ldots, n)$ exists. Let $W = X_1 + \cdots + X_n$ and $\sigma_W^2 = \mathsf{Var}(W)$. For any function h such that h'' is bounded, the normal approximation $\Phi_{\sigma_W}(h)$ of $\mathsf{E}\{h(W)\}$ has the corrector:*

$$C_h = \frac{\mu_{(3)}}{2\sigma_W^4}\Phi_{\sigma_W}\left\{\left(\frac{x^2}{3\sigma_W^2} - 1\right)xh(x)\right\} \qquad (8.9)$$

where $\mu_{(3)} = \sum_{i=1}^{n} \mathsf{E}(X_i^3)$. *The corrected approximation error is bounded by*

$$\left| \mathsf{E}\{h(W)\} - \Phi_{\sigma_W}(h) - C_h \right|$$

$$\leq \|f_h^{(3)}\|_{\sup} \left\{ \frac{1}{12} \sum_{i=1}^{n} \mathsf{E}\left(|X_i^s|^4\right) + \frac{1}{4\sigma_W^2} \left| \sum_{i=1}^{n} \mathsf{E}(X_i^3) \right| \sum_{i=1}^{n} \mathsf{E}\left(|X_i^s|^3\right) \right.$$

$$\left. + \frac{1}{\sigma_W} \sqrt{\sum_{i=1}^{n} \sigma_i^6} \right\}.$$

Proof:

By taking first order Taylor expansion, we obtain

$$\mathsf{E}\{h(W)\} - \Phi_{\sigma_W}(h) = \sigma_W^2 \, \mathsf{E}\{f_h'(W^*) - f_h'(W)\}$$

$$= \sigma_W^2 \, \mathsf{E}\{f_h''(W)(W^* - W)\} + \sigma_W^2 \, \mathsf{E}\left[f_h^{(3)}\{\xi W + (1-\xi)W^*\} \xi (W^* - W)^2 \right] \tag{8.10}$$

where ξ is a random variable on $[0,1]$ independent of all X_i and X_i^*. First, we notice that the remaining term is bounded by

$$\mathsf{E}\left[|f_h^{(3)}\{\xi W + (1-\xi)W^*\}| \xi (W^* - W)^2 \right] \leq \frac{\|f_h^{(3)}\|_{\sup}}{2} \mathsf{E}\{(W^* - W)^2\}.$$

Then we have

$$\sigma_W^2 \left| \mathsf{E}\left[f_h^{(3)}\{\xi W + (1-\xi)W^*\} \xi (W^* - W)^2 \right] \right| \leq \frac{\|f_h^{(3)}\|_{\sup}}{12} \sum_{i=1}^{n} \mathsf{E}\left(|X_i^s|^4\right). \tag{8.11}$$

Secondly, we consider the first term in the right-hand side of (8.10). Since X_I^* is independent of W, we have

$$\mathsf{E}\{f_h''(W)(W^* - W)\} = \mathsf{E}\{f_h''(W)(X_I^* - X_I)\}$$
$$= \mathsf{E}(X_I^*) \, \mathsf{E}\{f_h''(W)\} - \mathsf{E}\{f_h''(W)X_I\}. \tag{8.12}$$

For the second term $\mathsf{E}\{f_h''(W)X_I\}$ of (8.12), since

$$\mathsf{E}\{f_h''(W)X_I\} = \mathsf{E}\left\{ f_h''(W) \, \mathsf{E}(X_I | \bar{X}, \bar{X}^*) \right\},$$

we have using the conditional expectation that

$$\left| \mathsf{E}\{f_h''(W)X_I\} \right| \leq \frac{1}{\sigma_W^2} \sqrt{\mathsf{Var}\{f_h''(W)\}} \sqrt{\sum_{i=1}^{n} \sigma_i^6}. \tag{8.13}$$

Notice that

$$\mathsf{Var}\{f_h''(W)\} = \mathsf{Var}\{f_h''(W) - f_h''(0)\} \le \mathsf{E}[\{f_h''(W) - f_h''(0)\}^2] \le \|f_h^{(3)}\|_{\mathrm{sup}}^2 \sigma_W^2.$$

Therefore

$$\left| \mathsf{E}\{f_h''(W)X_I\} \right| \le \frac{\|f_h^{(3)}\|_{\mathrm{sup}}}{\sigma_W} \sqrt{\sum_{i=1}^{n} \sigma_i^6}.$$

For the first term $\mathsf{E}(X_I^*)\,\mathsf{E}\{f_h''(W)\}$ of (8.12), we write it as the sum of two parts

$$\mathsf{E}(X_I^*)\,\mathsf{E}\{f_h''(W)\} = \mathsf{E}(X_I^*)\Phi_{\sigma_W}(f_h'') + \mathsf{E}(X_I^*)\,\mathsf{E}\{f_h''(W) - \Phi_{\sigma_W}(f_h'')\}.$$

The first part is the candidate for the corrector. We apply the zero order estimation to the second part and get

$$\left| \mathsf{E}(X_I^*)\Big[\mathsf{E}\{f_h''(W)\} - \Phi_{\sigma_W}(f_h'')\Big] \right| \le \frac{\|f_h^{(3)}\|_{\mathrm{sup}}}{4\sigma_W^4}\Big|\sum_{i=1}^{n}\mathsf{E}(X_i^3)\Big|\sum_{i=1}^{n}\mathsf{E}\left(|X_i^s|^3\right). \tag{8.14}$$

Then, it suffices to write

$$\begin{aligned}
&\mathsf{E}\{h(W)\} - \Phi_{\sigma_W}(h) \\
&= \sigma_W^2\Big[\mathsf{E}(X_I^*)\Phi_{\sigma_W}(f_h'') + \mathsf{E}(X_I^*)\big[\mathsf{E}\{f_h''(W)\} - \Phi_{\sigma_W}(f_h'')\big] - \mathsf{E}\{f_h''(W)X_I\}\Big] \\
&\quad + \sigma_W^2\,\mathsf{E}\left[f_h^{(3)}\{\xi W + (1-\xi)W^*\}\xi(W^* - W)^2\right].
\end{aligned} \tag{8.15}$$

Combining (8.11), (8.13) and (8.14), we let $C_h = \sigma_W^2\,\mathsf{E}(X_I^*)\Phi_{\sigma_W}(f_h'')$ and we deduce the error bound. Finally, we use the invariant property of the normal distribution under zero bias transformation and the Stein's equation to obtain (8.9).   □

The corrector is written as the product of two terms: the first one depends on the moments of X_i up to the third order and the second one is a normal expectation of some polynomial function multiplying h. Both terms are simple to calculate, even in the inhomogeneous case.

To adapt to the definition of the zero biasing random variable, and also to obtain a simple representation of the corrector, the variables X_i's are set to be of expectation zero in Theorem 8.1. This condition requires a normalization step when applying the theorem to the conditional loss. A useful example concerns the centered Bernoulli random variables which take two real values and are of expectation zero.

Note that the moments of X_i play an important role here. In the symmetric case, we have $\mu_{(3)} = 0$ and as a consequence $C_h = 0$ for any function h. Therefore, C_h can be viewed as an asymmetric corrector in the sense that, after correction, the approximation realizes the same error order as in the symmetric case.

To precise the order of the corrector, let us consider the normalization of an homogeneous case where X_i's are i.i.d. random variables whose moments may depend on n. Notice that

$$\Phi_{\sigma_W}\left\{\left(\frac{x^2}{3\sigma_W^2} - 1\right)xh(x)\right\} = \sigma_W\Phi_1\left\{\left(\frac{x^2}{3} - 1\right)xh(\sigma_W x)\right\}.$$

To ensure that the above expectation term is of constant order, we often suppose that the variance of W is finite and does not depend on n. In this case, we have $\mu_{(3)} \sim o(1/\sqrt{n})$ and the corrector C_h is also of order $o(1/\sqrt{n})$. Consider now the percentage default indicator variable $\mathbf{1}_{\{\tau_i \leq t\}}/n$, whose conditional variance given the common factor equals to $p(1-p)/n^2$ where p is the conditional default probability of ith credit, identical for all in the homogeneous case. Hence, we shall fix p to be zero order and let $X_i = (\mathbf{1}_{\{\tau_i \leq t\}} - p)/\sqrt{n}$. Then σ_W is of constant order as stated above. Finally, for the percentage conditional loss, the corrector is of order $o(1/n)$ because of the remaining coefficient $1/\sqrt{n}$.

The X_i's are not required to have the same distribution: we can handle easily different recovery rates (as long as they are independent r.v.) by computing the moments of the product variables $(1 - R_i)\mathbf{1}_{\{\tau_i \leq t\}}$. The corrector depends only on the moments of R_i up to the third order. Note however that the dispersion of the recovery rates, also of the nominal values can have an impact on the order of the corrector.

We now concentrate on the call function $h(x) = (x - k)_+$. The Gauss approximation corrector is given in this case by

$$C_h = \frac{\mu_{(3)}}{6\sigma_W^2}k\phi_{\sigma_W}(k) \tag{8.16}$$

where $\phi_\sigma(x)$ is the density function of the distribution $N(0, \sigma^2)$. When the strike $k = 0$, the corrector $C_h = 0$. On the other hand, the function $k\exp\left(-\frac{k^2}{2\sigma_W^2}\right)$ reaches its maximum and minimum values when $k = \sigma_W$ and $k = -\sigma_W$, and then tends to zero quickly.

The numerical computation of this corrector is extremely simple since there is no need to take expectation. Observe however that the call function is a Lipschitz function with $h'(x) = \mathbf{1}_{\{x > k\}}$ and h'' exists only in the distribution

sense. Therefore, we can not apply directly Theorem 8.1 and the error estimation deserves a more subtle analysis. The main tool we used to establish the error estimation for the call function is a concentration inequality in Chen and Shao (2001). For detailed proof, interested reader may refer to El Karoui and Jiao (2007).

We shall point out that the regularity of the function h is essential in the above result. For more regular functions, we can establish correction terms of corresponding order. However, for the call function, the second order correction can not bring further improvement to the approximation results in general.

8.2.3 Stein's Method - the Poisson Case

The Poisson case is parallel to the Gaussian one. Recall that Chen (1975) has observed that a non-negative integer-valued random variable Λ of expectation λ follows the Poisson distribution if and only if $\mathsf{E}\{\Lambda g(\Lambda)\} = \lambda \mathsf{E}\{g(\Lambda + 1)\}$ for any bounded function g. Similar as in the normal case, let us consider a random variable Y taking non-negative integer values and $\mathsf{E}(Y) = \lambda < \infty$. A r.v. Y^* is said to have the *Y-Poisson zero biased distribution* if for any function g such that $\mathsf{E}\{Yg(Y)\}$ exists, we have

$$\mathsf{E}\{Yg(Y)\} = \lambda \mathsf{E}\{g(Y^* + 1)\}. \tag{8.17}$$

Stein's Poisson equation is also introduced in Chen (1975):

$$yg(y) - \lambda g(y + 1) = h(y) - \mathcal{P}_\lambda(h) \tag{8.18}$$

where $\mathcal{P}_\lambda(h) = \mathsf{E}\{h(\Lambda)\}$ with $\Lambda \sim P(\lambda)$. Hence, for any non-negative integer-valued r.v. V with expectation λ_V, we obtain the error of the Poisson approximation

$$\mathsf{E}\{h(V)\} - \mathcal{P}_\lambda(h) = \mathsf{E}\left\{Vg_h(V) - \lambda_V g_h(V+1)\right\} = \lambda_V \mathsf{E}\left\{g_h(V^*+1) - g_h(V+1)\right\} \tag{8.19}$$

where g_h is the solution of (8.18) and is given by

$$g_h(k) = \frac{(k-1)!}{\lambda^k} \sum_{i=k}^{\infty} \frac{\lambda^i}{i!} \{h(i) - \mathcal{P}_\lambda(h)\}. \tag{8.20}$$

It is unique except at $k = 0$. However, the value $g(0)$ does not enter into our calculations afterwards.

We consider now the sum of independent random variables. Let $Y_i\,(i = 1, \cdots, n)$ be independent non-negative integer-valued r.v. with positive expectations λ_i and let Y_i^* have the Y_i-Poisson zero biased distribution. Assume that Y_i and Y_i^* are mutually independent. Denote by $V = Y_1 + \cdots + Y_n$ and $\lambda_V = \mathsf{E}(V)$. Let I be a random index independent of $(\bar{Y}, \bar{Y}^*)$ satisfying $P(I = i) = \lambda_i/\lambda_V$. Then $V^{(I)} + Y_I^*$ has the V-Poisson zero biased distribution where $V^{(i)} = V - Y_i$.

For any integer $l \geq 1$, assume that Y and Y_i have to up $(l+1)$-order moments. Then

$$\mathsf{E}(|Y^* - Y|^l) = \frac{1}{\lambda}\,\mathsf{E}\left(Y|Y^s - 1|^l\right), \ \ \mathsf{E}(|V^* - V|^l) = \frac{1}{\lambda_V}\sum_{i=1}^{n}\mathsf{E}\left(Y_i|Y_i^s - 1|^l\right).$$

Finally, recall that Chen has established $\|\Delta g_h\|_{\sup} \leq 6\|h\|_{\sup}\min\left(\lambda^{-\frac{1}{2}}, 1\right)$ with which we obtain the following zero order estimation

$$|\,\mathsf{E}\{h(V)\} - \mathcal{P}_{\lambda_V}(h)| \leq 6\|h\|_{\sup}\min\left(\frac{1}{\sqrt{\lambda_V}}, 1\right)\sum_{i=1}^{n}\mathsf{E}\left(Y_i|Y_i^s - 1|\right). \quad (8.21)$$

There also exist other estimations of error bound (see e.g. Barbour and Eagleson (1983)). However we here are more interested in the order than the constant of the error.

8.2.4 First-Order Poisson Approximation

We now present the first-order Poisson approximation following the same idea as in the normal case. Firstly, recall the zero-order approximation formula. If V is a random variable taking non-negative integers with expectation λ_V, then we may approximate $\mathsf{E}\{h(V)\}$ by a Poisson function

$$\mathcal{P}_{\lambda_V}(h) = \sum_{m=0}^{n}\frac{\lambda_V^m}{m!}e^{-\lambda_V}h(m).$$

The Poisson approximation is efficient under some conditions, for example, when $V \sim B(n, p)$ and $np < 10$. We shall improve the Poisson approximation by presenting a corrector term as above. We remark that due to the property that a Poisson distributed random variable takes non-negative integer values, the variables Y_i's in Theorem 8.2 are discrete integer random variables.

PROPOSITION 8.2 *Let $Y_1, \ldots, Y_n$ be independent random variables taking non-negative integer values such that $\mathsf{E}(Y_i^3)\ (i = 1, \ldots, n)$ exist. Let*

$V = Y_1 + \cdots + Y_n$ with expectation $\lambda_V = \mathsf{E}(V)$ and variance $\sigma_V^2 = \mathsf{Var}(V)$. Then, for any bounded function h defined on $\mathbb{N}_+$, the Poisson approximation $\mathcal{P}_{\lambda_V}(h)$ of $\mathsf{E}\{h(V)\}$ has the corrector:

$$C_h^{\mathcal{P}} = \frac{\sigma_V^2 - \lambda_V}{2} \mathcal{P}_{\lambda_V}(\Delta^2 h) \tag{8.22}$$

where $\mathcal{P}_\lambda(h) = \mathsf{E}\{h(\Lambda)\}$ with $\Lambda \sim \mathcal{P}(\lambda)$ and $\Delta h(x) = h(x+1) - h(x)$. The corrected approximation error is bounded by

$$\left| \mathsf{E}\{h(V)\} - \mathcal{P}_{\lambda_V}(h) - \lambda_V \mathcal{P}_{\lambda_V}\{\Delta g_h(x+1)\} \mathsf{E}(Y_I^* - Y_I) \right|$$

$$\leq 2\|\Delta g_h\|_{\sup} \sum_{i=1}^n \lambda_i \mathsf{E}\left\{ |Y_i^* - Y_i| \left(|Y_i^* - Y_i| - 1\right) \right\}$$

$$+ 6\|\Delta g_h\|_{\sup} \left\{ \sum_{i=1}^n \mathsf{E}(Y_i | Y_i^s - 1|) \right\}^2$$

$$+ \mathsf{Var}\{\Delta g_h(V+1)\}^{\frac{1}{2}} \left\{ \sum_{i=1}^n \lambda_i^2 \mathsf{Var}(Y_i^* - Y_i) \right\}^{\frac{1}{2}}.$$

Proof:
Let us first recall the discrete Taylor formula. For any integers x and any positive integer $k \geq 1$,

$$g(x+k) = g(x) + k\Delta g(x) + \sum_{j=0}^{k-1} (k - 1 - j)\Delta^2 g(x+j).$$

Similar as in the Gaussian case, we apply the above formula to right-hand side of $\mathsf{E}\{h(V)\} - \mathcal{P}_{\lambda_V}(h) = \lambda_V \mathsf{E}\{g_h(V^* + 1) - g_h(V + 1)\}$ and we shall make decompositions. Since $V^* - V$ is not necessarily positive, we take expansion around $V^{(i)}$ for the following three terms respectively and obtain

$$\mathsf{E}\left\{ g_h(V^* + 1) - g_h(V + 1) - \Delta g_h(V + 1)(V^* - V) \right\} = \sum_{i=1}^n \frac{\lambda_i}{\lambda_V} \cdot$$

$$\left[\mathsf{E}\left\{ g_h(V^{(i)} + 1) + Y_i^* \Delta g_h(V^{(i)} + 1) + \sum_{j=0}^{Y_i^* - 1} (Y_i^* - 1 - j)\Delta^2 g_h(V^{(i)} + 1 + j) \right\} \right.$$

$$- \mathsf{E}\left\{ g_h(V^{(i)} + 1) + Y_i \Delta g_h(V^{(i)} + 1) + \sum_{j=0}^{Y_i - 1} (Y_i - 1 - j)\Delta^2 g_h(V^{(i)} + 1 + j) \right\}$$

$$\left. - \mathsf{E}\left\{ \Delta g_h(V^{(i)} + 1)(Y_i^* - Y_i) + \sum_{j=0}^{Y_i - 1} (Y_i^* - Y_i)\Delta^2 g_h(V^{(i)} + 1 + j) \right\} \right]$$

which implies that the remaining term is bounded by

$$\left| \mathsf{E}\left\{ g_h(V^* + 1) - g_h(V + 1) - \Delta g_h(V + 1)(V^* - V) \right\} \right|$$
$$\leq \|\Delta^2 g_h\|_{\sup} \sum_{i=1}^{n} \frac{\lambda_i}{\lambda_V} \left[\mathsf{E}\left\{ \binom{Y_i^*}{2} + \binom{Y_i}{2} \right\} + \mathsf{E}\left\{ |Y_i(Y_i^* - Y_i)| \right\} \right].$$

We then make decomposition

$$\mathsf{E}\left\{ \Delta g_h(V + 1)(V^* - V) \right\}$$
$$= \mathcal{P}_{\lambda_V}\{\Delta g_h(x + 1)\} \mathsf{E}(Y_I^* - Y_I) + \mathrm{Cov}\left\{ Y_I^* - Y_I, \Delta g_h(V + 1) \right\} \quad (8.23)$$
$$+ \left[\mathsf{E}\{\Delta g_h(V + 1)\} - \mathcal{P}_{\lambda_V}\{\Delta g_h(x + 1)\} \right] \mathsf{E}(Y_I^* - Y_I).$$

Similar as in the Gaussian case, the first term of (8.23) is the candidate of the corrector. For the second term, we use again the technique of conditional expectation and obtain

$$\mathrm{Cov}\left\{ \Delta g_h(V + 1), Y_I^* - Y_I \right\} \leq \frac{1}{\lambda_V} \mathsf{Var}\{\Delta g_h(V + 1)\}^{\frac{1}{2}} \left\{ \sum_{i=1}^{n} \lambda_i^2 \mathsf{Var}(Y_i^* - Y_i) \right\}^{\frac{1}{2}}.$$

For the last term of (8.23), we have by the zero order estimation

$$\left[\mathsf{E}\{\Delta g_h(V + 1)\} - \mathcal{P}_{\lambda_V}\{\Delta g_h(x + 1)\} \right] \mathsf{E}(Y_I^* - Y_I) \leq \frac{6\|\Delta g_h\|_{\sup}}{\lambda_V} \times$$
$$\left\{ \sum_{i=1}^{n} \mathsf{E}(Y_i|Y_i^s - 1|) \right\}^2.$$

It remains to observe that $\mathcal{P}_{\lambda_V}\{\Delta g_h(x+1)\} = \frac{1}{2}\mathcal{P}_{\lambda_V}(\Delta^2 h)$ and let the corrector to be

$$C_h^{\mathcal{P}} = \frac{\lambda_V}{2} \mathcal{P}_{\lambda_V}(\Delta^2 h) \mathsf{E}(Y_I^* - Y_I).$$

Combining all these terms, we obtain

$$\left| \mathsf{E}\{h(V)\} - \mathcal{P}_{\lambda_V}(h) - C_h^{\mathcal{P}} \right|$$
$$\leq \|\Delta^2 g_h\|_{\sup} \sum_{i=1}^{n} \lambda_i \mathsf{E}\left\{ |Y_i^* - Y_i| \left(|Y_i^* - Y_i| - 1 \right) \right\}$$
$$+ \mathsf{Var}\{\Delta g_h(V + 1)\}^{\frac{1}{2}} \left\{ \sum_{i=1}^{n} \lambda_i^2 \mathsf{Var}(Y_i^* - Y_i) \right\}^{\frac{1}{2}} \quad (8.24)$$
$$+ 6\|\Delta g_h\|_{\sup} \left\{ \sum_{i=1}^{n} \mathsf{E}(Y_i|Y_i^s - 1|) \right\}^2.$$

□

The Poisson corrector $C_h^{\mathcal{P}}$ is of similar form with the Gaussian one and contains two terms as well: one term depends on the moments of Y_i and the other is a Poisson expectation.

Since Y_i's are $\mathbb{N}_+$-valued random variables, they can represent directly the default indicators $\mathbf{1}_{\{\tau_i \le t\}}$. This fact limits however the recovery rate to be identical or proportional for all credits. We now consider the order of the corrector. Suppose that λ_V does not depend on n to ensure that $\mathcal{P}_{\lambda_V}(\Delta^2 h)$ is of constant order. Then in the homogeneous case, the conditional default probability $p \sim o(1/n)$. For the percentage conditional losses, as in the Gaussian case, the corrector is of order $o(1/n)$ with the coefficient $1/n$.

Since $\Delta^2 h(x) = \mathbf{1}_{\{x=k-1\}}$ for the call function, its Poisson approximation corrector is given by

$$C_h^{\mathcal{P}} = \frac{\sigma_V^2 - \lambda_V}{2(\lfloor k \rfloor - 1)!} e^{-\lambda_V} \lambda_V^{\lfloor k \rfloor - 1} \tag{8.25}$$

where $\lfloor k \rfloor$ is the integer part of k. The corrector vanishes when the expectation and the variance of the sum variable V are equal. The difficulty here is that the call function is not bounded. However, we can prove that Theorem 8.2 holds for any function of polynomial increasing speed El Karoui and Jiao (2007).

8.3 Numerical Tests

Before exploring real life applications, we would like in this section to perform some basic testing of the preceding formulae. In the sequel, we consider the call value $\mathsf{E}\{(l - k)_+\}$ where $l = n^{-1} \sum_{i=1}^{n}(1 - R_i)\xi_i$ and the ξ_i's are independent Bernoulli random variables with success probability equal to p_i.

8.3.1 Validity Domain of the Approximations

We begin by testing the accuracy of the corrected Gauss and Poisson approximations for different values of $np = \sum_{i=1}^{n} p_i$ in the case $R_i = 0$, $n = 100$ and for different values of k such that $0 \le k \le 1$. The benchmark value is obtained through the recursive methodology well known by the practitioners which computes the loss distribution by reducing the portfolio size by one name at each recursive step.
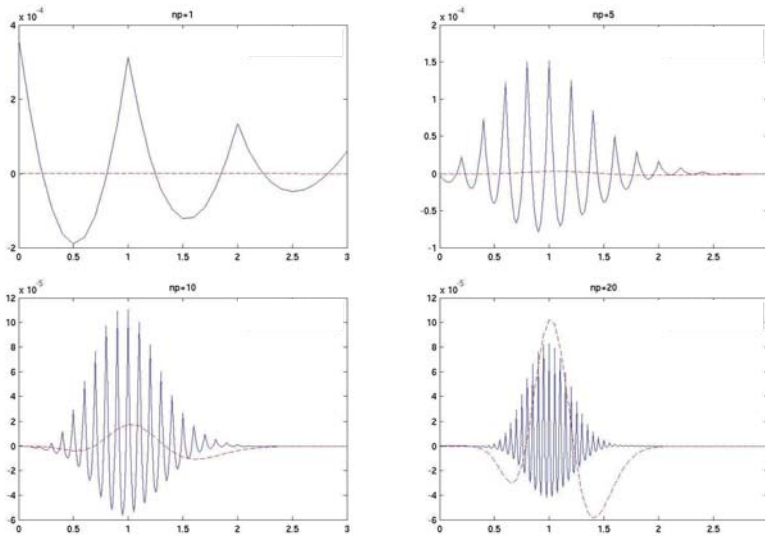
Figure 8.1. Gauss and Poisson approximation errors for various values of np as a function of the strike over the expected loss, with line curve for Gaussian errors and dotted curve for Poisson errors. XFGgperror

In Figure 8.1 are plotted the differences between the corrected Gauss approximation and the benchmark (Error Gauss) and the corrected Poisson approximation and the benchmark (Error Poisson) for different values of np as a function of the call strike over the expected loss. Note that when the tranche strike equals the expected loss, the normalized strike value in the Gaussian case equals zero due to the centered random variables, which means that the correction vanishes. We observe in Figure 8.1 that the Gaussian error is maximal around this point.

We observe on these graphs that the Poisson approximation outperforms the Gaussian one for approximately $np < 15$. On the contrary, for large values of np, the Gaussian approximation is the best one. Because of the correction, the threshold between the Gauss-Poisson approximation is higher than the classical one $np \approx 10$. In addition, the threshold may be chosen rather flexibly around 15. Combining the two approximations, the minimal error of the two approximations is relatively larger in the overlapping area when np is around 15. However, we obtain satisfactory results even in this case. In all the graphs presented, the error of the mixed approximation is inferior than

1 bp.

Our tests are made with inhomogeneous p_i's obtained as

$$p_i = p \exp(\sigma W_i - 0.5\sigma^2)$$

(log-normal random variable with expectation p and volatility σ) where W_i is a family of independent standard normal random variables and values of σ ranging from 0% to 100%. Qualitatively, the results were not affected by the heterogeneity of the p_i's.

Observe that there is oscillation in the Gaussian approximation error, while the Poisson error is relatively smooth. This phenomenon is related to the discretization impact of discrete laws.

As far as a unitary computation is concerned (one call price), the Gaussian and Poisson approximation perform much better than the recursive methodology: we estimate that these methodologies are 200 times faster. To be fair with the recursive methodology one has to recall that by using it we obtain not only a given call price but the whole loss distribution which correspond to about 100 call prices. In that case, our approximations still outperform the recursive methodology by a factor 2.

8.3.2 Stochastic Recovery Rate - Gaussian Case

We then consider the case of stochastic recovery rate and check the validity of the Gauss approximation in this case. Following the standard in the industry (Moody's assumption), we will model the R_i's as independent beta random variables with expectation 50% and standard deviation 26%.

An application of Theorem 8.1 is used so that the first order corrector term takes into account the first three moments of the random variables R_i. To describe the obtained result let us first introduce some notations. Let $\mu_{R_i}, \sigma^2_{R_i}$ and $\gamma^3_{R_i}$ be the first three centered moments of the random variable R_i, namely

$$\mu_{R_i} = \mathsf{E}(R_i), \quad \sigma^2_{R_i} = \mathsf{E}\{(R_i - \mu_{R_i})^2\}, \quad \gamma^3_{R_i} = \mathsf{E}\{(R_i - \mu_{R_i})^3\}.$$

We also define $X_i = n^{-1}(1 - R_i)\xi_i - \mu_i$ where $\mu_i = n^{-1}(1 - \mu_{R_i})p_i$ and $p_i = \mathsf{E}(\xi_i)$. Let W be $\sum_{i=1}^{n} X_i$. We have

$$\sigma^2_W = \mathsf{Var}(W) = \sum_{i=1}^{n} \sigma^2_{X_i} \quad \text{where} \quad \sigma^2_{X_i} = \frac{p_i}{n^2}\left\{\sigma^2_{R_i} + (1 - p_i)(1 - \mu_{R_i})^2\right\}.$$
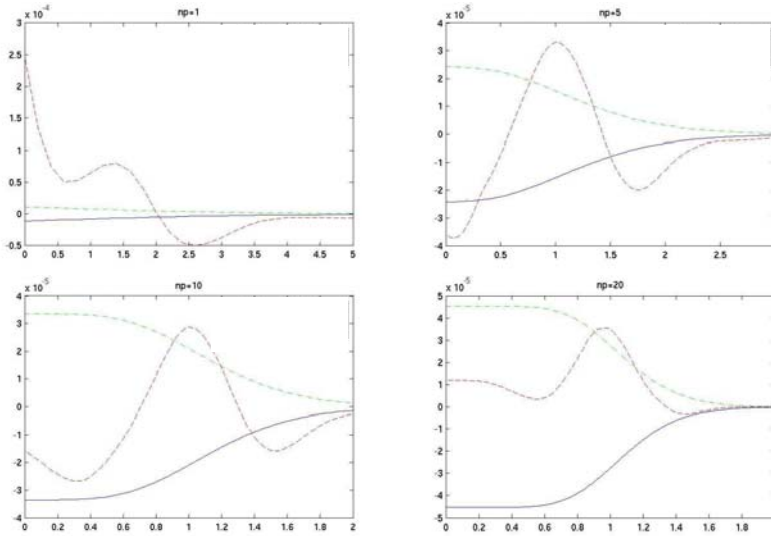
Figure 8.2. Gaussian approximation errors in the stochastic recovery case for various values of np as a function of the strike over the expected loss, compared with upper and lower 95% confidence interval bounds of Monte Carlo 1,000,000 simulations. **Q** `XFGstoerror`

Finally, if $\widetilde{k} = k - \sum_{i=1}^{n} \mu_i$, we have the following approximation

$$\mathsf{E}\{(l-k)_+\} \approx \Phi_{\sigma_W}(\cdot - k)_+ + \frac{1}{6}\frac{1}{\sigma_W^2} \sum_{i=1}^{n} \mathsf{E}\big(X_i^3\big)\widetilde{k}\phi_{\sigma_W}(\widetilde{k})$$

where

$$\mathsf{E}(X_i^3) = \frac{p_i}{n^3}\Big\{(1-\mu_{R_i})^3\big(1-p_i\big)(1-2p_i) + 3(1-p_i)(1-\mu_{R_i})\sigma_{R_i}^2 - \gamma_{R_i}^3\Big\}.$$

The benchmark is obtained using standard Monte Carlo integration with 1,000,000 simulations. We display, in Figure 8.2, the difference between the approximated call price and the benchmark as a function of the strike over the expected loss. We also consider the lower and upper 95% confidence interval for the Monte Carlo results. As in the standard case, one observes that the greater the value of np the better the approximation. Furthermore, the stochastic recovery brings a smoothing effect since the conditional loss no longer follows a binomial law.

The Poisson approximation, due to constraint of integer valued random variables, can not treat directly the stochastic recovery rates. We can however take the mean value of R_i's as the uniform recovery rate especially for low value of np without improving the results except for very low strike (equal to a few bp).

8.3.3 Sensitivity Analysis

We are finally interested in calculating the sensitivity with respect to p_j. As for the Greek of the classical option theory, direct approximations using the finite difference method implies large errors. We hence propose the following procedure.

Let $l_t^j = \frac{1-R}{n}\mathbf{1}_{\{\tau_j \leq t\}}$. Then for all $j = 1, \cdots, n$,

$$(l_t - k)_+ = \mathbf{1}_{\{\tau_j \leq t\}}\Big(\sum_{i \neq j} l_t^i + \frac{1-R}{n} - k\Big)_+ + \mathbf{1}_{\{\tau_j > t\}}\Big(\sum_{i \neq j} l_t^i - k\Big)_+.$$

As a consequence, we may write

$$\mathsf{E}\{(l_t - k)_+|U\} = F(p_j, U)\,\mathsf{E}\Big\{\Big(\sum_{i \neq j} l_t^i + \frac{1-R}{n} - k\Big)_+\Big|U\Big\}$$

$$+ \{1 - F(p_j, U)\}\,\mathsf{E}\Big\{\Big(\sum_{i \neq j} l_t^i - k\Big)_+\Big|U\Big\}.$$

Since the only term which depends on p_j is the function $F(p_j, U)$, we obtain that $\partial_{p_j} C(t, k)$ can be calculated as

$$\int_0^1 \mathrm{d}u\,\partial_1 F(p_j, u)\,\mathsf{E}\Big[\Big\{\sum_{i \neq j} l_t^i + \omega_j(1 - R_j) - k\Big\}_+ - \Big(\sum_{i \neq j} l_t^i - k\Big)_+\Big|U = u\Big] \quad (8.26)$$

where we compute the call spread using the mixed approximation for the partial total loss.

We test this approach in the case where $R_i = 0$ on a portfolio of 100 names such that one fifth of the names has a default probability of 25 bp, 50 bp, 75 bp, 100 bp and 200 bp respectively for an average default probability of 90 bp. We compute call prices derivatives with respect to each individual name probability according to the formula (8.26) and we benchmark this result by the sensitivities given by the recursive methodology.

In Figure 8.3, we plot these derivatives for a strike value of 3% computed using the recursive and the approximated methodology. Our finding is that
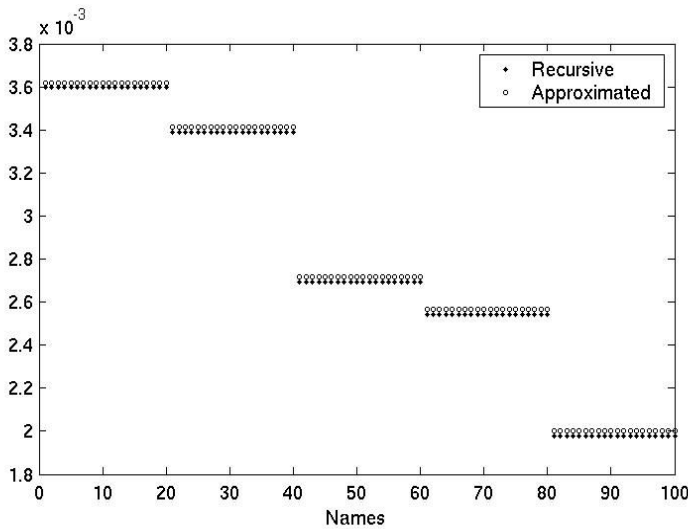
Figure 8.3. Sensitivity with respect to individual default probability by the approximated and the recursive methodology, for 5 types of 100 total names. **Q XFGsensibility**

in all tested cases (strike ranging from 3% to 20%), the relative errors on these derivatives are less than 1% except for strike higher than 15%, in which case the relative error is around 2%. Note however that in this case, the absolute error is less than 0.1 bp for derivatives whose values are ranging from 2 bp to 20 bp. We may remark that the approximated methodology always overvalues the derivatives value. However in the case of a true mezzanine tranche this effect will be offset. We consider these results as very satisfying.

8.4 Real Life Applications

After recalling the main mathematical results, we use them on two real life applications: valuation of single tranche CDOs and computing VaR figures in a timely manner.

8.4.1 Gaussian Approximation

Let μ_i and σ_i be respectively the expectation and standard deviation of the random variable $\chi_i = n^{-1}(1-R)\mathbf{1}_{\{\tau_i \leq t\}}$. Let $X_i = \chi_i - \mu_i$ and $W = \sum_{i=1}^{n} X_i$,

so that the expectation and standard deviation of the random variable W are 0 and $\sigma_W = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$ respectively. Let also p_i be the default probability of issuer i. We want to calculate

$$C(t, k) = \mathsf{E}\{(l_t - k)_+\} = \mathsf{E}\{(W - \widetilde{k})_+\}$$

where $\widetilde{k} = k - \sum_{i=1}^{n} \mu_i$.

Assuming that the random variables X_i's are mutually independent, the result of Theorem 8.1 may be stated in the following way

$$C(t, k) \approx \int_{-\infty}^{+\infty} \mathrm{d}x \, \phi_{\sigma_W}(x)(x - \widetilde{k})_+ + \frac{1}{6}\frac{1}{\sigma_W^2} \sum_{i=1}^{n} \mathsf{E}(X_i^3) \widetilde{k} \phi_{\sigma_W}(\widetilde{k}) \qquad (8.27)$$

where $\mathsf{E}(X_i^3) = \frac{(1-R)^3}{n^3} p_i(1 - p_i)(1 - 2p_i)$. The first term on the right-hand side of (8.27) is the Gaussian approximation that can be computed in closed form thanks to Bachelier formula whereas the second term is a correction term that accounts for the non-normality of the loss distribution.

In the sequel, we will compute the value of the call option on a loss distribution by making use of the approximation (8.27). In the conditionally independent case, one can indeed write

$$\mathsf{E}(l_t - k)_+ = \int \mathrm{P}_U(\mathrm{d}u) \, \mathsf{E}\{(l_t - k)_+|U = u\}$$

where U is the latent variable describing the general state of the economy. As the default time are conditionally independent upon the variable U, the integrand may be computed in closed form using (8.27).

We note finally that in the real life test, we model U in a non-parametric manner such that the base correlation skew of the market can be reproduced.

8.4.2 Poisson Approximation

Recall that $\mathcal{P}_\lambda$ is the Poisson measure of intensity λ. Let $\lambda_i = p_i$ and $\lambda_V = \sum_{i=1}^{n} \lambda_i$ where now $V = \sum_{i=1}^{n} Y_i$ with $Y_i = \mathbf{1}_{\{\tau_i \leq t\}}$. We want to calculate

$$C(t, k) = \mathsf{E}\{(l_t - k)_+\} = \mathsf{E}\{(n^{-1}(1 - R)V - k)_+\}.$$

Recall that the operator Δ is such that $(\Delta f)(x) = f(x + 1) - f(x)$. We also let the function h be defined by $h(x) = \{n^{-1}(1 - R)x - k\}_+$.

Assuming that the random variables Y_i's are mutually independent, we may write according to the results of theorem 8.2 that

$$C(t,k) \approx \mathcal{P}_{\lambda_V}(h) - \frac{1}{2}\left(\sum_{i=1}^{n} \lambda_i^2\right)\mathcal{P}_{\lambda_V}(\Delta^2 h) \tag{8.28}$$

where

$$\mathcal{P}_{\lambda_V}(\Delta^2 h) = n^{-1}(1-R)\mathrm{e}^{-\lambda_V}\frac{\lambda_V^{\lfloor m \rfloor - 1}}{(\lfloor m \rfloor - 1)!}$$

where $m = nk/(1-R)$. The formula (8.28) may be used to compute the unconditional call price in the same way as in the preceding subsection.

8.4.3 CDO Valuation

In this subsection, we finally use both Gaussian and Poisson first order approximations to compute homogeneous single tranche CDO value and break even as described in formula (8.1). As this formula involves conditioning on the latent variable U, we are either in the validity domain of the Poisson approximation or in the validity domain of the Gaussian approximation. Taking into account the empirical facts underlined in Section 8.3, we choose to apply the Gaussian approximation for the call value as soon as $\sum_i F(p_i, u) > 15$ and the Poisson approximation otherwise. All the subsequent results are benchmarked using the recursive methodology.

Our results for the quoted tranches are gathered in the following table. Level represents the premium leg for the spread of 1 bp and break even is the spread of CDO in (8.1).

In the following table are gathered the errors on the break even expressed in bp. We should note that in all cases, the error is less than 1.15 bp which is below the market uncertainty that prevails on the bespoke CDO business. We observe furthermore that the error is maximal for the tranche 3%-6% which correspond to our empirical finding (see Figure 8.1) that the approximation error is maximal around the expected loss of the portfolio (equal here to 4.3%).

Trying to understand better these results, we display now in the following two tables the same results but for equity tranches.

Attach	Detach	Output	REC	Approx.
0%	3%	Default Leg	2.1744%	2.1752%
		Level	323.2118%	323.2634%
		Break Even	22.4251%	22.4295%
3%	6%	Default Leg	0.6069%	0.6084%
		Level	443.7654%	443.7495%
		Break Even	4.5586%	4.5702%
6%	9%	Default Leg	0.1405%	0.1404%
		Level	459.3171%	459.3270%
		Break Even	1.0197%	1.0189%
9%	12%	Default Leg	0.0659%	0.0660%
		Level	462.1545%	462.1613%
		Break Even	0.4754%	0.4758%
12%	15%	Default Leg	0.0405%	0.0403%
		Level	463.3631%	463.3706%
		Break Even	0.2910%	0.2902%
15%	22%	Default Leg	0.0503%	0.0504%
		Level	464.1557%	464.1606%
		Break Even	0.1549%	0.1552%
0%	100%	Default Leg	3.1388%	3.1410%
		Level	456.3206%	456.3293%
		Break Even	1.1464%	1.1472%

Table 8.1. Break even values for the quoted tranches, by recursive method and our approximation method respectively

	Error
0-3	0.44
3-6	1.15
6-9	- 0.08
9-12	0.04
12-15	- 0.08
15-22	0.02
0-100	0.08

Table 8.2. Break even errors for the quoted tranches compared to the recursive method, expressed in bp

Attach	Detach	Output	REC	Mixte
0%	3%	DL	2.1744%	2.1752%
		Level	323.2118%	323.2634%
		BE	22.4251%	22.4295%
0%	6%	DL	2.7813%	2.7836%
		Level	383.4886%	383.5114%
		BE	12.0878%	12.0969%
0%	9%	DL	2.9218%	2.9240%
		Level	408.7648%	408.7853%
		BE	7.9422%	7.9476%
0%	12%	DL	2.9877%	2.9900%
		Level	422.1122%	422.1302%
		BE	5.8984%	5.9025%
0%	15%	DL	3.0282%	3.0303%
		Level	430.3624%	430.3788%
		BE	4.6909%	4.6940%
0%	22%	DL	3.0785%	3.0807%
		Level	441.1148%	441.1280%
		BE	3.1723%	3.1744%
0%	100%	DL	3.1388%	3.1410%
		Level	456.3206%	456.3293%
		BE	1.1464%	1.1472%

Table 8.3. Break even values for the equity tranches, by recursive method and our approximation method respectively

	Error
0-3	0.44
0-6	0.92
0-9	0.55
0-12	0.41
0-15	0.31
0-22	0.21
0-100	0.08

Table 8.4. Break even errors for the equity tranches compared to the recursive method, expressed in bp

8.4.4 Robustness of VaR Computation

In this section, we consider the VaR computation for a given CDOs book and show that the use of the Gaussian first order approximation as in subsection

8.4.1 can speed up substantially credit derivatives VaR computation without loss of numerical accuracy. We restrict our attention on the Gaussian approximation as we want to be able to consider non-homogeneous reference portfolio. We study the approximation effect on VaR computation by using a stylized portfolio which strikes and maturities are distributed such that the resulting book is reasonably liquid and diversified.

Our finding is that we may safely use this approximation without a significant loss of accuracy for our stylized portfolio and this could lead, according to our estimation, to a reduction of 90% of VaR computation time as compared with the recursive methodology. The production of the VaR in due time for financial institution will then still be possible even if its business on single tranche increases steadily.

To test the robustness of the proposed approximation in VaR computation, we decide to study the accuracy (as compared by a full recursive valuation) of differences of the form

$$\Delta^{\omega}(T, K) = \mathrm{E}^{\omega}\{(l_T - K)_+\} - \mathrm{E}^{\omega_0}\{(l_T - K)_+\}$$

for various (spreads and correlation) VaR scenarios ω randomly generated. Here ω_0 denotes the initial scenario.

Generating VaR Scenarios

We aim here at generating by a Monte Carlo procedure a family of scenarios for spreads and the base correlation that we will assume constant in this set of tests.

We choose the following dynamic for the daily variation of the spreads of the common reference portfolio

$$\frac{\Delta s_i}{s_i} = 50\%\big(\sqrt{30\%}\varepsilon + \sqrt{70\%}\varepsilon_i\big)\sqrt{\Delta t}$$

where $\varepsilon, \varepsilon_1, \ldots, \epsilon_M$ are independent standard Gaussian random variables and $\Delta t 1/252$. In other words, we assume a joint log-normal dynamic with volatility 50% and correlation 30%.

We then assume that the shocks on the base correlation are normally distributed with initial value 30% and annual volatility of 15%.

In the sequel and for our testing, we will use a sample of 1000 such scenarios of spreads and correlation daily moves.

Stylized Portfolio Description

We start from the stylized distribution of a single tranche CDO portfolio. The resulting position is chosen so that is is reasonably liquid and diversified in term of maturity, strike and credit risk.

Each strike (expressed in expected loss unit) and maturity will be assigned a positive and a negative weight according to the corresponding notional in position. Hence, we come up with two positive normalized (=unity total mass) measures μ_+ and μ_- that reflects the book repartition in terms of strike (expressed in expected loss) and maturity. We also let $\mu = \mu_+ - \mu_-$ and $\widetilde{\mu} = |\mu|/2 = (\mu_+ + \mu_-)/2$.

We give below an example to explain more precisely. Let us consider, for instance, a protection buyer CDO position with maturity T, with expected loss $\mathsf{E}(l_T)$, with notional N and strikes A and B expressed in percentage. We also define $a(T) = A/\mathsf{E}(l_T)$ and $b(T) = B/\mathsf{E}(l_T)$. Using the following approximate formulas for the payout of the default and premium legs

$$\text{Default Leg} = N \times \big[\{l_T - a(T)\,\mathsf{E}(l_T)\}_+ - \{l_T - b(T)\,\mathsf{E}(l_T)\}_+\big],$$

$$\text{Premium Leg} = N \times \text{Spread} \times \frac{T}{2} \times$$
$$\big[(B - A) - \{l_{T/2} - a(T/2)\,\mathsf{E}(l_{T/2})\}_+ + \{l_{T/2} - b(T/2)\,\mathsf{E}(l_{T/2})\}_+\big],$$

we observe that this deal will contribute for a positive amount of N on the point $\{a(T), T\}$, a negative amount of $-N$ on the point $\{b(T), T\}$, a positive amount of $N \times \text{Spread} \times T/2$ on the point $\{a(T/2), T/2\}$ and a negative amount of $-N \times \text{Spread} \times T/2$ on the point $\{b(T/2), T/2\}$.

Error Computation

Let $\Delta_{GA}^\omega(T, K)$ and $\Delta_{REC}^\omega(T, K)$ be the value of the difference

$$\mathsf{E}^\omega\{(l_T - K)_+\} - \mathsf{E}^{\omega_0}\{(l_T - K)_+\}$$

as given respectively by the Gaussian approximation and a full recursive valuation.

We are interested in different types of errors that will allow us to assess the robustness of the proposed approximation for VaR computation purposes. The algebraic average error (see Figure 8.4) arising from the use of the approximation on the book level and expressed in spread term may be defined as

$$\text{Algebraic Average Error}(\omega) \int \frac{\mu(\mathrm{d}k, \mathrm{d}T)}{T} \big\{\Delta_{GA}^\omega(T, kE_T) - \Delta_{REC}^\omega(T, kE_T)\big\}$$
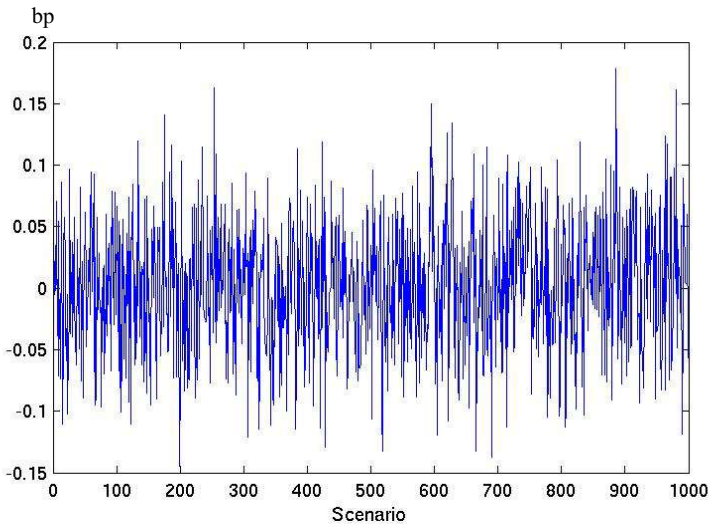
Figure 8.4. Algebraic Average Error of VaR per Scenario,
expressed in bp. ◘ XFGalgerror

where $E_T = \mathsf{E}(l_T)$. The maximum algebraic average error on the book in
spread term is defined as

$$\text{Max Algebraic Error} = \max_{\omega} |\text{Algebraic Average Error}(\omega)| .$$

Note that this way of computing the error allows the offset of individual er-
rors due to the book structure. It is reasonable to take these effects into
account when one tries to degrade numerical computation for VaR compu-
tation purposes. However, we will also compute the more stringent absolute
average error (see Figure 8.5) on the book in spread term which is defined as

$$\text{Absolute Average Error}(\omega) \int \frac{\widetilde{\mu}(\mathrm{d}k, \mathrm{d}T)}{T} |\Delta_{GA}^{\omega}(T, kE_T) - \Delta_{REC}^{\omega}(T, kE_T)|.$$

The maximum absolute average error on the book in spread term is then
defined as

$$\text{Max Absolute Error} = \max_{\omega} \text{Absolute Average Error}(\omega).$$
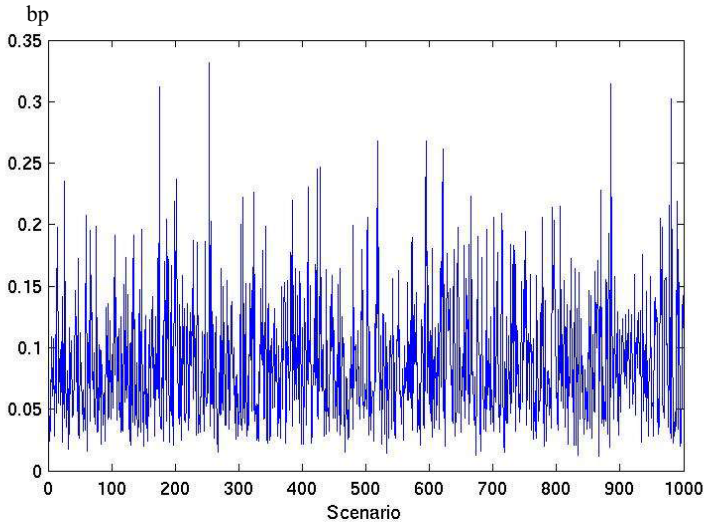
Our main results are

Figure 8.5. Absolute Average Error of VaR per Scenario, expressed in bp. **XFGabsolute**

Max Algebraic Error = 0.1785 bp,

Max Absolute Error = 0.3318 bp.

As expected the maximum algebraic error is half the maximum absolute error as we allow the offsetting of the error due to the book structure.

These results are quite satisfying and justify the use of this approach for VaR computations in an industrial setting.

Bibliography

Antonov, A., Mechkov, S. and Misirpashaev, T. (2005). Analytical techniques for synthetic CDOs and credit default risk measures. *Working Paper*, NumeriX.

Andersen, L., Sidenius, J. and Basu, S. (2003). All hedges in one basket, *Risk*, 16:67–72.

Barbour A. D. and Eagleson, G. K. (1983). Poisson approximation for some statistics based on exchangeable trials, *Advances in Applied Probability*, 15(3):585–600.

Brasch, H.-J. (2004). A note on efficient pricing and risk calculation of credit basket products, *Working paper*, TD Securities.

Burtschell, X., Gregory, J. and Laurent J.-P. (2007). Beyond the Gaussian copula: stochastic and local correlation, *Journal of Credit Risk*, 3(1),31-62.

Chen, L.H.Y. (1975). Poisson approximation for dependent trials, *Annals of Probability*, 3:534–545.

Chen, L. H. Y. and Shao, Q.-M. (2001). A non-uniform Berry-Esseen bound via Stein's method, *Probability Theory and Related Fields*, 120:236–254.

Chen, L. H. Y. and Shao, Q.-M. (2005). Stein's method for normal approximation. *An Introduction to Stein's Method*, volume 4,1-59, *Lecture Notes Series, IMS, National University of Singapore*, Singapore University Press and World Scientific Publishing.

Diener, F. and Diener, M. (2004). Asymptotics of price oscillations of a european Call option in a tree model, *Mathematical Finance*, 14(2):271–293.

El Karoui, N. and Jiao, Y. (2007). Stein's method and zero bias transformation: application to large loss approximation and CDO pricing, *Working paper*, Ecole Polytechnique.

Goldstein, L. and Reinert, G. (1997). Stein's method and the zero bias transformation with application to simple random sampling, *Annals of Applied Probability*, 7:935–952.

Hull, J. and White, A. (2004). Valuation of a CDO and an nth-to-default CDS without Monte Carlo simulation, *Working paper*, University of Toronto.

Martin, R., Thompson, K.and Browne, C. (2001). Taking to the saddle, *Risk*, June, 14(6).

Petrov, V. V. (1975). *Sums of Independent Random Variables*, Springer-Verlag.

Stein, C. (1972). A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *Proc, Sixty Berkeley Symp. Math. Statist. Probab.*, 583-602. Univ. California Press, Berkeley.

Vasicek, O. (1991). Limiting loan loss probability distribution, *Working paper*, Moody's KMV, 1991.

Part III

Implied Volatility

9 Least Squares Kernel Smoothing of the Implied Volatility Smile

Matthias R. Fengler and Qihua Wang

9.1 Introduction

Functional flexibility is the cornerstone for model building and model selection in quantitative finance, for it is often difficult, if not impossible, to justify a specific parametric form of an economic relationship on theoretical grounds. Furthermore, in a dynamic context, the economic structure may be subject to changes and fluctuations. Hence, estimation techniques that do not impose a priori restrictions on the estimate, such as non- and semiparametric methods, are increasingly popular.

In finance, a common challenge is to the implied volatility smile function. Based on the assumption of a geometric Brownian motion governing the stock price dynamics, an unknown volatility parameter is implied from observed option prices using the Black and Scholes (1973) formula. By theory the resulting function should be constant in strike prices and dates of maturity. Yet, as a matter of fact, one typically observes a curved and 'smiley' functional pattern across different strikes for a fixed maturity which is called the implied volatility smile.

Although a large number of alternative pricing models were proposed in the literature, it remains difficult to fully explain the shape of the smile with option standard option pricing models, see Bergomi (2004) for an assessment. Therefore, it is common practice to fit a parametric function directly to the observed implied volatility smile, which is used by market makers at plain vanilla desks to manage their positions. As a pathway for more flexibility one combines both the Black and Scholes (BS) model and nonparametric smoothing of the smile. For instance, Ait-Sahalia and Lo (1998) suggest to replace the constant volatility parameter by a semiparametric function. Alternatively, Fengler, Härdle and Mammen (2007a) propose to model implied volatility as a latent factor process fitted by semiparametric methods.

These approaches share in common that they first derive implied volatilities by equating the BS formula with observed market prices to solve for the diffusion coefficient and that the actual fitting algorithm is applied in a second step. As an alternative one can directly base the estimate on a least squares kernel estimator that takes observed option prices, e.g. call prices $\widetilde{C}_t$, as input parameters. This estimator was proposed by Gouriéroux, Monfort and Tenreiro (1994) and Gouriéroux, Monfort and Tenreiro (1995) to predict a stochastic volatility process in a latent factor model; we shall employ it here for smoothing the smile. It is based on the representation

$$
\begin{aligned}
\widehat{\sigma}(\kappa_t, \tau) \;=\; & \arg\min_{\sigma} \sum_{i=1}^{n} \left\{ \widetilde{C}_{t_i} - C^{BS}(\cdot, \sigma) \right\}^2 w(\kappa_{t_i}) \\
& \times \; K_{(1)}\left(\frac{\kappa_t - \kappa_{t_i}}{h_{1,n}} \right) K_{(2)}\left(\frac{\tau - \tau_i}{h_{2,n}} \right) \;,
\end{aligned}
$$

where $C^{BS}(\cdot, \sigma)$ denotes the BS price for calls, $\kappa_t \stackrel{\text{def}}{=} K/S_t$ is moneyness with K denoting the strike and S_t the current asset price, and $\tau \stackrel{\text{def}}{=} T - t$ time to maturity, where T is the expiry date. $K_{(1)}(\cdot)$ and $K_{(2)}(\cdot)$ are kernel functions and $w(\cdot)$ denotes a uniformly continuous and bounded weight function, which allows for different weights of observed option prices. It can be used to give less weight to in-the-money options which a may contain a liquidity premium.

An advantage of this estimator of the smile is that it allows for constructing pointwise confidence intervals which explicitly take the nonlinear transformation from option prices to implied volatility into account. These confidence intervals can be used to derive bid-ask-spreads in a statistically based fashion, or to support trading decisions of statistical arbitrage models.

9.2 Least Squares Kernel Smoothing of the Smile

European style calls are contingent claims on an asset S_t (for simplicity, paying no dividends), which yield at a given expiry day T a pay-off $\max(S_T - K, 0)$. The strike price is denoted by K. Assuming that the asset price process S_t follows a geometric Brownian motion with a constant diffusion coefficient

σ, the BS formula is given by

$$
\begin{aligned}
C^{BS}(S_t, t, K, T, r, \sigma) &= S_t \Phi(d_1) - e^{-r\tau} K \Phi(d_2), & (9.1) \\
d_1 &\overset{\text{def}}{=} \frac{\log(S_t/K) + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}, \\
d_2 &\overset{\text{def}}{=} d_1 - \sigma\sqrt{\tau},
\end{aligned}
$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of a standard normal random variable, r the risk-free interest rate, $\tau \overset{\text{def}}{=} T - t$ time to maturity. Since the actual volatility σ of the underlying price process cannot be observed directly, one studies the volatility that is implied in option prices observed on markets $\widetilde{C}_t$. Implied volatility $\widehat{\sigma}$ is defined as:

$$
\widehat{\sigma}: \qquad C_t^{BS}(S_t, t, K, T, r, \widehat{\sigma}) - \widetilde{C}_t = 0, \qquad (9.2)
$$

By monotonicity in volatility, the solution $\widehat{\sigma} > 0$ is unique. The purpose is to estimate the function $\widehat{\sigma}: (K, T) \rightarrow \widehat{\sigma}(K, T)$. More convenient is the representation $\widehat{\sigma}: (\kappa, \tau) \rightarrow \widehat{\sigma}(\kappa, \tau)$ expressed in relative terms by moneyness $\kappa_t \overset{\text{def}}{=} K/S_t$ and time to maturity τ.

By homogeneity in K and S, the BS formula can be rewritten in terms of moneyness:

$$
C^{BS}(S_t, t, K, T, r, \sigma) = S_t\, c^{BS}(\kappa_t, \tau, r, \sigma) \qquad (9.3)
$$

where $c^{BS}(\kappa_t, \tau, r, \sigma) = \Phi(d_1) - e^{-r\tau}\kappa_t\Phi(d_2)$ and $d_1 = \frac{-\log\kappa_t + (r + \frac{1}{2}\sigma^2)\tau}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$ (as before). As observed by Gouriéroux et al. (1995), this allows for weaker assumptions on the estimator. The least squares kernel estimator is then defined by:

$$
\begin{aligned}
\widehat{\sigma}(\kappa_t, \tau) &= \arg\min_{\sigma} \sum_{i=1}^{n} \left\{ \widetilde{c}_{t_i} - c^{BS}(\kappa_{t_i}, \tau_i, r, \sigma) \right\}^2 w(\kappa_{t_i}) \\
&\times\; K_{(1)}\left(\frac{\kappa_t - \kappa_{t_i}}{h_{1,n}}\right) K_{(2)}\left(\frac{\tau - \tau_i}{h_{2,n}}\right), & (9.4)
\end{aligned}
$$

where $\widetilde{c}_t \overset{\text{def}}{=} \widetilde{C}_t/S_t$. The kernel functions are denoted by $K_{(1)}(\cdot)$ and $K_{(2)}(\cdot)$, $w(\cdot)$ is a weight function, and $i = 1, \ldots, n$ a numbering of the option data.

In (9.4), we minimize the pricing error between observed option prices and the BS formula where the volatility is replaced by an unknown nonparametric function. We point out that we do not interpret the pricing error as a mispricing which could be exploited by an arbitrage strategy, but rather follow the notion of pricing errors developed in Renault (1997). We perceive the

error term as an option pricing error due to a neglected heterogeneity factor. Hence, the econometric specification is with respect to another martingale measure equivalent to the actual pricing measure. This notion allows us to stay within Harrison and Kreps (1979) framework.

The least squares kernel estimator is based on the following assumptions:

(A1) $\mathsf{E}_t \kappa_t^4 < \infty$, where E_t denotes the expectation operator with respect to time-t information;

(A2) $w(\cdot)$ is a uniformly continuous and bounded weight function;

(A3) $K_{(1)}(\cdot)$ and $K_{(2)}(\cdot)$ are bounded probability density kernel functions with bounded support;

(A4) interest rate r is constant.

Assumption (A1) is a weak assumption. It is justified, since by the institutional arrangements at futures exchanges, options at new strikes are always launched in the neighborhood of S_t. To understand this assumption note that S_t is measureable with respect to the information at time t and that by simple no-arbitrage considerations, we have $0 \leq \mathsf{E}_t \widetilde{C}_t \leq S_t$. (A2) is usually satisfied by weight functions. In Section 9.3 we will discuss possible choices for $w(\cdot)$. Condition (A3) is met by a lot of kernels used in nonparametric regression, such as the quartic or Epanechnikov kernel functions. Assumption (A4) is needed to derive the BS formula. It can be justified by the empirical observation that asset price variability largely outweighs changes in the interest rate, Bakshi, Cao and Chen (1997). Nevertheless, the impact from changing interest rates can be substantial for options with a very long time to maturity.

Given assumptions (A1) to (A4) one obtains consistency:

PROPOSITION 9.1 *Let $\sigma(\kappa_t, \tau)$ be the solution of* $\mathsf{E}_t[\{\widetilde{c}_{t_1} - c^{BS}(\kappa_t, \tau, r, \sigma)\} w(\kappa_t)] = 0$. *If conditions (A1), (A2), (A3) and (A4) are satisfied, we have:*

$$\widehat{\sigma}(\kappa_t, \tau) \xrightarrow{p} \sigma(\kappa_t, \tau)$$

as $nh_{1,n}h_{2,n} \to \infty$.

Furthermore, we introduce the notations:

$$A_i(\kappa_t, \tau, r, \sigma) \stackrel{\text{def}}{=} \widetilde{c}_{t_i} - c^{BS}(\kappa_{t_i}, \tau_i, r, \sigma) \,,$$

$$B(\kappa_t, \tau, r, \sigma) \stackrel{\text{def}}{=} \frac{\partial c^{BS}(\kappa_t, \tau, r, \sigma)}{\partial \sigma} = S_t^{-1} \frac{\partial C^{BS}(\cdot)}{\partial \sigma} = \sqrt{\tau} \phi(d_1) \,, \quad (9.5)$$

$$D(\kappa_t, \tau, r, \sigma) \stackrel{\text{def}}{=} \frac{\partial^2 c^{BS}(\kappa_t, \tau, r, \sigma)}{\partial \sigma^2} = S_t^{-1} \frac{\partial^2 C^{BS}(\cdot)}{\partial \sigma^2}$$

$$= \sqrt{\tau} \phi(d_1) d_1 d_2 \sigma^{-1} \,, \quad (9.6)$$

where $\phi(u) = \frac{1}{\sqrt{2\pi}} e^{-u^2/2}$. In financial language, B, the sensitivity of the option price with respect to volatility, is called 'vega'. Its second derivative D is termed 'volga'. Then one can establish

PROPOSITION 9.2 *Under conditions (A1), (A2), (A3), and (A4), if* $\mathsf{E}_t\{B^2(\kappa_t, \tau, r, \sigma)w(\kappa_t)\} \neq \mathsf{E}_t\{A(\kappa_t, \tau, r, \sigma)D(\kappa_t, \tau, r, \sigma)w(\kappa_t)\}$, *we have*

$$\sqrt{nh_{1n}h_{2n}}\{\widehat{\sigma}(\kappa_t, \tau) - \sigma(\kappa_t, \tau)\} \stackrel{\mathcal{L}}{\longrightarrow} N(0, \gamma^{-2}\nu^2),$$

where

$$\gamma^2 \stackrel{\text{def}}{=} \left[\mathsf{E}_t\{-B^2(\kappa_t, \tau, r, \sigma)w(\kappa_t) + A(\kappa_t, \tau, r, \sigma)D(\kappa_t, \tau, r, \sigma)w(\kappa_t)\} \right]^2 f_t(\kappa_t, \tau)$$

$$\nu^2 \stackrel{\text{def}}{=} \mathsf{E}_t\{A^2(\kappa_t, \tau, r, \sigma)B^2(\kappa_t, \tau, r, \sigma)w^2(\kappa_t)\} \int K_{(1)}^2(u)K_{(2)}^2(v)\,dudv \,,$$

and $f_t(\kappa_t, \tau)$ *is the joint (time-t conditional) probability density function of* κ_t *and* τ *respectively.*

The proof can be found in Gouriéroux et al. (1994) and in the appendix. The asymptotic distribution depends on first and second order derivatives, and the weight function. Nevertheless an approximation is simple, since first and second order derivatives have the analytical expressions given in (9.5) and (9.6).

9.3 Application

9.3.1 Weighting Functions, Kernels, and Minimization Scheme

In order to obtain a good forecast of asset price variability, the early literature on implied volatility discusses different weighting schemes of implied volatility

the observations intensively. In principle, this results in scalar estimates of the form

$$\widehat{\sigma}^* = \arg \min_{\sigma} \sum_{i=1}^{n} w_i \{\widetilde{C}_i - C^{BS}(\sigma)\}^2 / \sum_{i=1}^{n} w_i , \qquad (9.7)$$

where w_i is a weight. Schmalensee and Trippi (1978) and Whaley (1982) argue in favor of unweighted averages, i.e. $w_i = 1$. Beckers (1981) suggests the vega as weights, $w_i = \partial C_i / \partial \sigma$, while Latané and Rendelman (1976) propose squared vega as weights. Chiras and Manaster (1978) employ the elasticity with respect to volatility, i.e. $w_i = \frac{\partial C_i}{\partial \sigma} \frac{\sigma}{C_i}$.

The vega is a Gaussian shaped function in the underlying centered approximately at-the-money (ATM), cf. Equation (9.5). Elasticity is a decreasing function in the underlying for calls. Thus, obvious concern of these weighting procedures is to give low weight to in-the-money (ITM) options, and highest weight to ATM or out-of-the-money (OTM) options. Due to the lower trading volume of the first, ITM options are suspected to trade at a liquidity premium which may ensue biased estimates of volatility. In contrast to this argument, one could also give little weight to ATM options and much bigger to ITM and OTM options, e.g. by using the inverse of the squared vega $w_i = (\partial C_i / \partial \sigma)^{-2}$. The rationale of this choice is that inserting the inverse of the squared vega into the call price smoother (9.4) implies smoothing in the implied volatility domain at first order.

As kernel functions we use quartic kernels,

$$K(u) = \frac{15}{16}(1 - u^2)^2 \mathbf{1}(|u| \leq 1) ,$$

where $\mathbf{1}(\mathcal{A})$ denotes the indicator function of the event $\mathcal{A}$. In practice, the choice of the kernel function has little impact on the estimates Härdle (1990). Given global convexity (see appendix), we use the Golden section search implemented in the statistical software package XploRe. The estimation tolerance is fixed at 10^{-5}.

9.3.2 Data Description and Empirical Demonstration

The data used contain the January and February 2001 tick statistics of the DAX futures contracts and DAX index options and is provided by the futures exchange EUREX, Frankfurt/Main. Both futures and option data are contract based data, i.e. each single contract is registered together with its price, contract size, and the time of settlement up to a second. Interest rate data in daily frequency, i.e. 1, 3, 6, 12 months EURIBOR rates, are gathered from

Observation date	Time to expiry (days)	min	max	mean	standard deviation	total number of observations	calls
Jan. 02, 2001	17	0.1711	0.3796	0.2450	0.0190	1219	561
Feb. 02, 2001	14	0.1199	0.4615	0.1730	0.0211	1560	813

Table 9.1. Implied volatility data obtained by inverting the BS formula separately for each observation in the sense of two-step estimators.

Thomson Financial Datastream, and linearly interpolated to approximate the riskless interest rate for the option specific time to maturity.

For our application, we use data from January 02 and February 02, 2001. For data preparation we apply a scheme described in more detail in Hafner and Wallmeier (2001). In a first step, we recover the DAX index values. To this end, we group to each option price observation the futures price F_t of the nearest available futures contract, which was traded within a one minute interval around the observed option price. The futures price observation is taken from the most heavily traded futures contract on the particular day, which is the March 2001 contract. The no-arbitrage price of the underlying index in a frictionless market without dividends is given by $S_t = F_t e^{-r_{T_F,t}(T_F-t)}$, where S_t and F_t denote the index and the futures price respectively, T_F the futures contract's maturity date, and $r_{T,t}$ the interest rate with maturity $T - t$. In the case of a capital weighted performance index as is the DAX index, Deutsche Börse (2006), dividends less corporate tax are reinvested into the index. Thus, the dividend yield can be assumed to be zero.

In Table 12.1, we give an overview of the data set. We present summary statistics in the form of the implied volatility data rather than in form of the option price data. The corresponding option data (each option price divided by the discounted future) can be seen in the top panel of Figure 9.1. Since the data are transaction data containing potential misprints, we apply a filter in deleting all observations whose implied volatility is bigger than 0.7 and less than 0.1. The settlement price of the March 2001 futures contract was 6340 EUR at a volume of 30 088 contracts on Jan. 02, 2001, and 6669.5 EUR and 34 244 contracts on Feb. 02, 2001.

The plots are displayed in Figures 9.1 and 9.2. The top panel shows the observed option prices given on the moneyness scale, while the lower panel demonstrates the estimate (with equal weighting) together with a pointwise confidence interval. For the construction we used a kernel estimate of the density based on Silverman's rule of thumb Härdle, Müller, Sperlich and Werwatz (2004). Naturally, the interval broadens in the wings of the smile, when data become scarce. In a trading context, these confidence intervals allow for sta-

tistically well defined bid-ask spread. They could be used to support trading decisions, for instance, in statistical arbitrage models. Extensions of these ideas for multi-asset equity options and correlation products can be found in Fengler and Schwendner (2004) and Fengler, Pilz and Schwendner (2007b).
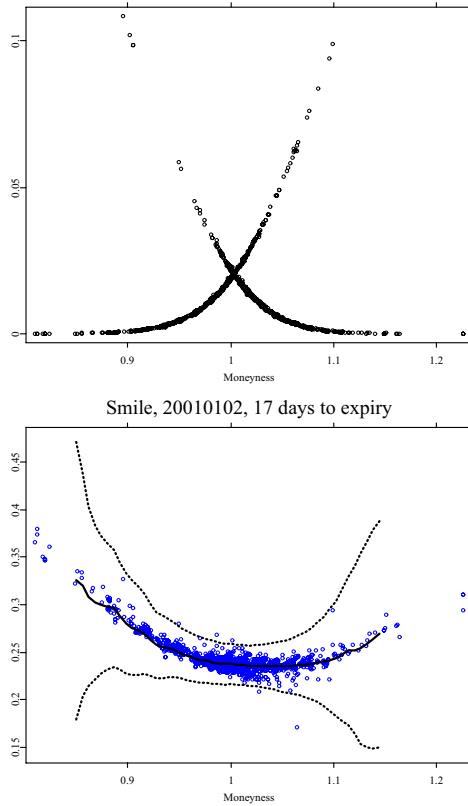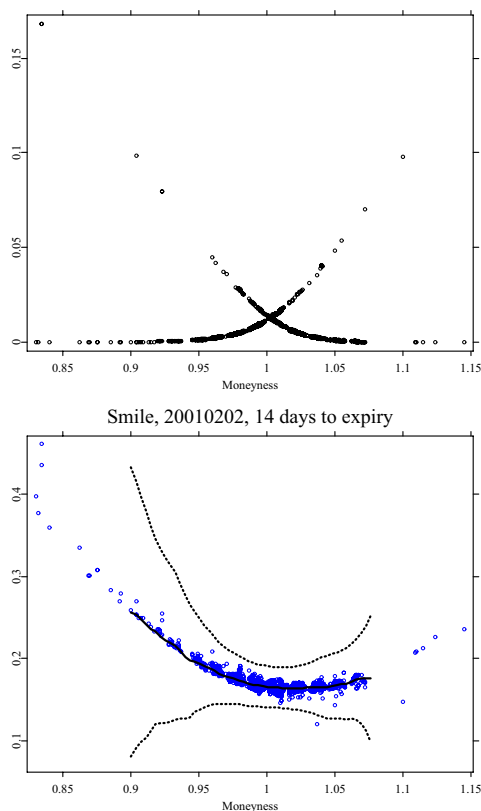

Figure 9.1. Upper panel: observed option price data on Jan 02, 2001. From lower left to upper right put prices, from upper left to lower right (normalized) call prices. Lower panel: least squares kernel smoothed implied volatility smile for 17 days to expiry on Jan 02, 2001. Bandwidth $h_1 = 0.025$. Dotted lines are the 95% confidence intervals for $\widehat{\sigma}$. Single dots are implied volatility data obtained by inverting the BS formula.
Q XFGLSK

Figure 9.2. Upper panel: observed option price data on Feb. 02, 2001. From lower left to upper right put prices, from upper left to lower right (normalized) call prices. Lower panel: least squares kernel smoothed implied volatility smile for 14 days to expiry on Jan 02, 2001. Bandwidth $h_1 = 0.015$. Dotted lines are the 95% confidence intervals for $\widehat{\sigma}$. Single dots are implied volatility data obtained by inverting the BS formula.
 XFGLSK

Bibliography

Ait-Sahalia, Y. and Lo, A. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices, *Journal of Finance* **53**: 499–548.

Bakshi, G., Cao, C. and Chen, Z. (1997). Empirical performance of alternative option pricing models, *Journal of Finance* **52**(5): 2003–2049.

Beckers, S. (1981). Standard deviations implied in option prices as predictors of future stock price variability, *Journal of Banking and Finance* **5**: 363–382.

Bergomi, L. (2004). Smile dynamics, *RISK* **17**(9): 117–123.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy* **81**: 637–654.

Chiras, D. P. and Manaster, S. (1978). The information content of option prices and a test for market efficiency, *Journal of Financial Economics* **6**: 213–234.

Deutsche Börse (2006). *Guide to the Equity Indices of Deutsche Börse*, 5.12 edn, Deutsche Börse AG, 60485 Frankfurt am Main, Germany.

Fengler, M. R. and Schwendner, P. (2004). Quoting multiasset equity options in the presence of errors from estimating correlations, *Journal of Derivatives* **11**(4): 43–54.

Fengler, M. R., Härdle, W. and Mammen, E. (2007a). A semiparametric factor model for implied volatility surface dynamics, *Journal of Financial Econometrics* **5**(2): 189–218.

Fengler, M. R., Pilz, K. and Schwendner, P. (2007b). Basket volatility and correlation, *in* I. Nelken (ed.), *Volatility as an asset class*, Risk Books, Haymarket, London, pp. 95–131.

Gouriéroux, C., Monfort, A. and Tenreiro, C. (1994). Nonparametric diagnostics for structural models, *Document de travail 9405*, CREST, Paris.

Gouriéroux, C., Monfort, A. and Tenreiro, C. (1995). Kernel M-estimators and functional residual plots, *Document de travail 9546*, CREST, Paris.

Hafner, R. and Wallmeier, M. (2001). The dynamics of DAX implied volatilities, *International Quarterly Journal of Finance* **1**(1): 1–27.

Härdle, W. (1990). *Applied Nonparametric Regression*, Cambridge University Press, Cambridge, UK.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer-Verlag, Berlin, Heidelberg.

Harrison, J. and Kreps, D. (1979). Martingales and arbitrage in multiperiod securities markets, *Journal of Economic Theory* **20**: 381–408.

Latané, H. A. and Rendelman, J. (1976). Standard deviations of stock price ratios implied in option prices, *Journal of Finance* **31**: 369–381.

Renault, E. (1997). Econometric models of option pricing errors, *in* D. M. Kreps and K. F. Wallis (eds), *Advances in Economics and Econometrics, Seventh World Congress*, Econometric Society Monographs, Cambridge University Press, pp. 223–278.

Schmalensee, R. and Trippi, R. R. (1978). Common stock volatility expectations implied by option premia, *Journal of Finance* **33**: 129–147.

Whaley, R. (1982). Valuation of American call options on dividend-paying stocks: Empirical tests, *Journal of Financial Economics* **10**: 29–58.

9.4 Proofs

PROOF OF THEOREM (9.1):

For notational simplicity, we introduce:

$$Z(x, y) \stackrel{\text{def}}{=} w(x)\, K_{(1)}\left(\frac{\kappa_t - x}{h_{1,n}}\right) K_{(2)}\left(\frac{\tau - y}{h_{2,n}}\right) , \tag{9.8}$$

and

$$\widehat{L}_n(\sigma) \stackrel{\text{def}}{=} \frac{1}{nh_{1,n}h_{2,n}} \sum_{i=1}^{n} \{\widetilde{c}_{t_i} - c^{BS}(\kappa_{t_i}, \tau_i, r, \sigma)\}^2\, Z(\kappa_{t_i}, \tau_i) . \tag{9.9}$$

We also drop in the following the dependence of the option prices on r.

As a first step, let us prove

$$\widehat{L}_n(\sigma) \xrightarrow{p} L(\sigma) \stackrel{\text{def}}{=} \mathsf{E}_t\left[\{\widetilde{c}_t - c^{BS}(\kappa_t, \tau, \sigma)\}^2 w(\kappa_t)\right] . \tag{9.10}$$

It is observed that

$$
\begin{aligned}
\widehat{L}_n(\sigma) &= \frac{1}{nh_{1,n}h_{2,n}} \sum_{i=1}^{n} \Big\{ \{\widetilde{c}_{t_i} - c^{BS}(\kappa_{t_i}, \tau_i, \sigma)\}^2\, Z(\kappa_{t_i}, \tau_i) \\
&\quad - \mathsf{E}_t\left[\{\widetilde{c}_{t_i} - c^{BS}(\kappa_{t_i}, \tau_i, \sigma)\}^2\, Z(\kappa_{t_i}, \tau_i)\right] \Big\} \\
&\quad + \frac{1}{h_{1,n}h_{2,n}} \mathsf{E}_t\left[\{\widetilde{c}_{t_1} - c^{BS}(\kappa_{t_1}, \tau_1, \sigma)\}^2\, Z(\kappa_{t_1}, \tau_1)\right] \\
&\stackrel{\text{def}}{=} \alpha_n + \beta_n .
\end{aligned}
\tag{9.11}
$$

Standard arguments can be used to prove

$$\mathsf{E}_t\alpha_n^2 = \mathcal{O}\left((nh_{1,n}h_{2,n})^{-1}\right) \tag{9.12}$$

by conditions (A1) and (A2).

By Taylor's expansion, we have

$$
\begin{aligned}
\beta_n &= \frac{1}{h_{1,n}h_{2,n}} \mathsf{E}_t \int \{\widetilde{c}_{t_1} - c^{BS}(x, y, \sigma)\}^2\, Z(x, y)\, dx\, dy \\
&= \mathsf{E}_t \int \{\widetilde{c}_t - c^{BS}(\kappa_t - h_{1,n}u, \tau - h_{2,n}v, \sigma)\}^2 \\
&\quad \times\, w(\kappa_t - h_{1,n}u) K_{(1)}(u) K_{(2)}(v)\, du\, dv \xrightarrow{p} L(\sigma) . \tag{9.13}
\end{aligned}
$$

Equations (9.12) and (9.13) together prove (9.10).

In a second step, we have, recalling the definition of $\sigma(\kappa_t, \tau)$:

$$
\begin{aligned}
\left.\frac{\partial L(\sigma)}{\partial \sigma}\right|_{\sigma=\sigma(\kappa_t,\tau)} &= -2\,\mathsf{E}_t \widetilde{c}_t w(\kappa_t)\left.\frac{\partial}{\partial \sigma}c^{BS}(\kappa_t,\tau,\sigma)\right|_{\sigma=\sigma(\kappa_t,\tau)} \\
&\quad + 2\,\mathsf{E}_t c^{BS}(\kappa_t,\tau,\sigma(\kappa_t,\tau))w(\kappa_t)\left.\frac{\partial}{\partial \sigma}c^{BS}(\kappa_t,\tau,\sigma)\right|_{\sigma=\sigma(\kappa_t,\tau)} \\
&= 0\,, \tag{9.14}
\end{aligned}
$$

and

$$
\begin{aligned}
\left.\frac{\partial^2 L(\sigma)}{\partial \sigma^2}\right|_{\sigma=\sigma(\kappa_t,\tau)} &= -2\,\mathsf{E}_t \widetilde{c}_t w(\kappa_t)\left.\frac{\partial^2}{\partial \sigma^2}c^{BS}(\kappa_t,\tau,\sigma)\right|_{\sigma=\sigma(\kappa_t,\tau)} \\
&\quad + 2\,\mathsf{E}_t w(\kappa_t)\left(\left.\frac{\partial}{\partial \sigma}c^{BS}(\kappa_t,\tau,\sigma)\right|_{\sigma=\sigma(\kappa_t,\tau)}\right)^2 \\
&\quad + 2\,\mathsf{E}_t w(\kappa_t)c^{BS}(\kappa_t,\tau,\sigma(\kappa,\tau))\left.\frac{\partial^2}{\partial \sigma^2}c^{BS}(\kappa_t,\tau,\sigma)\right|_{\sigma=\sigma(\kappa,\tau)} \\
&= 2\,\mathsf{E}_t w(\kappa_t)\left(\left.\frac{\partial}{\partial \sigma}c^{BS}(\kappa_t,\tau,\sigma)\right|_{\sigma=\sigma(\kappa_t,\tau)}\right)^2\,. \tag{9.15}
\end{aligned}
$$

This together with (9.10) proves that $\widehat{L}_n(\sigma)$ converges in probability to a convex function with a unique minimum at $\sigma = \sigma(\kappa_t, \tau)$. $\widehat{\sigma}_n(\kappa_t, \tau) \xrightarrow{p} \sigma(\kappa_t, \tau)$ is proved.

PROOF OF THEOREM (9.2):

Recalling the definition of $\widehat{\sigma}(\kappa_t, \tau)$, it follows that $\widehat{\sigma}(\kappa_t, \tau)$ is the solution of the following equation:

$$
\begin{aligned}
U_n(\sigma) &\overset{\text{def}}{=} \frac{1}{nh_{1,n}h_{2,n}}\sum_{i=1}^{n}A_i(\kappa_{t_i},\tau_i,\sigma)B_i(\kappa_{t_i},\tau_i,\sigma)\,Z(\kappa_{t_i},\tau_i) \\
&= 0\,. \tag{9.16}
\end{aligned}
$$

By Taylor's expansion, we get

$$
0 = U_n(\widehat{\sigma}(\kappa_t,\tau)) = U_n(\sigma(\kappa_t,\tau)) + U_n'(\sigma^*)\Big(\widehat{\sigma}_t(\kappa_t,\tau) - \sigma(\kappa_t,\tau)\Big)\,, \tag{9.17}
$$

where σ^* lies between σ and $\widehat{\sigma}$ and $U_n'(\sigma^*) \overset{\text{def}}{=} \frac{\partial}{\partial \sigma}U_n(\sigma)|_{\sigma=\sigma^*}$.

From (9.17), we have

$$\widehat{\sigma}(\kappa_t, \tau) - \sigma(\kappa_t, \tau) = -\{U_n'(\sigma^*)\}^{-1} U_n(\sigma) . \tag{9.18}$$

By some algebra, we obtain

$$
\begin{aligned}
U_n'(\sigma) \;=\; & \frac{1}{nh_{1,n}h_{2,n}} \sum_{i=1}^{n} \Bigg(\Big\{ \big(\frac{\partial}{\partial\sigma} A_i(\kappa_{t_i}, \tau_i, \sigma)\big) B_i(\kappa_{t_i}, \tau_i, \sigma) \\
& + \; A_i(\kappa_{t_i}, \tau_i, \sigma)\big(\frac{\partial}{\partial\sigma} B_i(\kappa_{t_i}, \tau_i, \sigma)\big) \Big\} Z(\kappa_{t_i}, \tau_i) \\
& - \; \mathsf{E}_t \Big[\Big\{ \big(\frac{\partial}{\partial\sigma} A_i(\kappa_{t_i}, \tau_i, \sigma)\big) B_i(\kappa_{t_i}, \tau_i, \sigma) \\
& + \; A_i(\kappa_{t_i}, \tau_i, \sigma)\frac{\partial}{\partial\sigma} B_i(\kappa_{t_i}, \tau_i, \sigma) \Big\} Z(\kappa_{t_i}, \tau_i) \Big] \Bigg) \\
& + \; \frac{1}{nh_{1,n}h_{2,n}} \sum_{i=1}^{n} \mathsf{E}_t \Big[\Big\{ \big(\frac{\partial}{\partial\sigma} A_i(\kappa_{t_i}, \tau_i, \sigma)\big) B_i(\kappa_{t_i}, \tau_i, \sigma) \\
& + \; A_i(\kappa_{t_i}, \tau_i, \sigma)\frac{\partial}{\partial\sigma} B_i(\kappa_{t_i}, \tau_i, \sigma) \Big\} Z(\kappa_{t_i}, \tau_i) \Big] \\
\overset{\text{def}}{=}\; & \triangle_{1,n} + \triangle_{2,n} .
\end{aligned}
\tag{9.19}
$$

Inspect first $\triangle_{1,n}$ in Equation (9.19): by some algebra, we get

$$
\begin{aligned}
\mathsf{E}_t \triangle_{1,n}^2 \;\le\; & \frac{1}{n^2 h_{1,n}^2 h_{2,n}^2} \sum_{i=1}^{n} \mathsf{E}_t \Bigg[\Big\{ \big(\frac{\partial}{\partial\sigma} A_i(\kappa_{t_i}, \tau_i, \sigma)\big) B_i(\kappa_{t_i}, \tau_i, \sigma) \\
& + \; A_i(\kappa_{t_i}, \tau_i, \sigma)\frac{\partial}{\partial\sigma} B_i(\kappa_{t_i}, \tau_i, \sigma) \Big\} Z(\kappa_{t_i}, \tau_i) \Bigg]^2 \\
=\; & \frac{f_t^2(\kappa_t, \tau) \int K_{(1)}^2(u)\, du \int K_{(2)}^2(v)\, dv}{nh_{1,n}h_{2,n}} \mathsf{E}_t \Bigg[\Big\{ \big(\frac{\partial}{\partial\sigma} A_1(\kappa_t, \tau, \sigma) B_1(\kappa_t, \tau, \sigma) \\
& + \; A_1(\kappa_t, \tau, \sigma)\frac{\partial}{\partial\sigma} B_1(\kappa_t, \tau, \sigma)\big)^2 \Big\} w(\kappa_t) \Bigg] \\
& + \; o\Big(\frac{1}{nh_{1,n}h_{2,n}}\Big) \;\longrightarrow\; 0 ,
\end{aligned}
\tag{9.20}
$$

as $nh_{1,n}h_{2,n} \to \infty$. The joint (time-$t$ conditional) probability density function of κ_t and τ is denoted by $f_t(\kappa_t, \tau)$.

To consider $\triangle_{2,n}$ in Equation (9.19), denote $D(\kappa_t, \tau, \sigma) \overset{\text{def}}{=} \frac{\partial}{\partial\sigma} B(\kappa_t, \tau, \sigma)$, for

simplicity. Note that $\frac{\partial}{\partial \sigma} A(\kappa_t, \tau, \sigma) = -B(\kappa_t, \tau, \sigma)$. Thus, we have:

$$
\begin{aligned}
\triangle_{2,n} &= \frac{1}{h_{1,n} h_{2,n}} \, \mathsf{E}_t \Big\{ \int \Big(-B^2(x, y, \sigma) + A(x, y, \sigma) D(x, y, \sigma) \Big) \\
&\qquad \times \; Z(x, y) f_t(x, y) \, dx \, dy \Big\} \\
&= \mathsf{E}_t \int \Big\{ -B^2(\kappa_t - h_{1,n} u, \tau - h_{2,n} v, \sigma) \\
&\qquad + \; A(\kappa_t - h_{1,n} u, \tau - h_{2,n} v, \sigma) D(\kappa_t - h_{1,n} u, \tau - h_{2,n} v, \sigma) \Big\} \\
&\qquad \times \; w(\kappa_t) \, f_t(\kappa_t - h_{1,n} u, \tau - h_{2,n} v) K_{(1)}(u) K_{(2)}(v) \, du \, dv \\
&\longrightarrow \; \Big[-\mathsf{E}_t \Big\{ B^2(\kappa_t, \tau, \sigma) w(\kappa_t) \Big\} \\
&\qquad + \; \mathsf{E}_t \Big\{ A(\kappa_t, \tau, \sigma) D(\kappa_t, \tau, \sigma) w(\kappa_t) \Big\} \Big] f_t(\kappa_t, \tau) \, . \qquad (9.21)
\end{aligned}
$$

Equations (9.19), (9.20), (9.21) and the fact $U_n'(\sigma^*) - U_n'(\sigma) \to 0$ together prove:

$$
\begin{aligned}
U_n'(\sigma^*) \; \xrightarrow{\; p \;} \; \Big[\mathsf{E}_t \Big\{ -B^2(\kappa_t, \tau, \sigma) w(\kappa_t) \Big\} \\
+ \; \mathsf{E}_t \Big\{ A(\kappa_t, \tau, \sigma) D(\kappa_t, \tau, \sigma) w(\kappa_t) \Big\} \Big] f_t(\kappa_t, \tau) \, . \\
(9.22)
\end{aligned}
$$

Now, let

$$
u_{ni} \overset{\text{def}}{=} \frac{1}{h_{1,n} h_{2,n}} A(\kappa_{t_i}, \tau_i, \sigma) B(\kappa_{t_i}, \tau_i, \sigma) \, Z(\kappa_{t_i}, \tau_i) \, . \qquad (9.23)
$$

For some $\delta > 0$, we have:

$$
\begin{aligned}
\mathsf{E}_t |u_{ni}|^{2+\delta} &= \frac{1}{h_{1,n}^{2+\delta} h_{2,n}^{2+\delta}} \mathsf{E}_t A^{2+\delta}(\kappa_{t_i}, \tau_i, \sigma) B^{2+\delta}(\kappa_{t_i}, \tau_i, \sigma)^{2+\delta} Z^{2+\delta}(\kappa_{t_i}, \tau_i) \\
&= \frac{1}{h_{1,n}^{1+\delta} h_{2,n}^{1+\delta}} \mathsf{E}_t \left[\int A^{2+\delta}(\kappa_t - h_n u, \tau - h_n v, \sigma) \right. \\
&\quad \times \left. B^{2+\delta}(\kappa_t - h_n u, \tau - h_n u, \sigma) \right. \\
&\quad \times \left. Z^{2+\delta}(\kappa_t - h_{1,n} u, \tau - h_{2,n} v) \, du \, dv \right] \\
&= \frac{f_t(\kappa_t, \tau) \int K_{(1)}^{2+\delta}(u) \, du \int K_{(2)}^{2+\delta}(v) \, dv}{h_{1,n}^{1+\delta} h_{2,n}^{1+\delta}} \\
&\quad \times \mathsf{E}_t \left[A^{2+\delta}(\kappa_t, \tau, \sigma) B^{2+\delta}(\kappa_t, \tau, \sigma) w^{2+\delta}(\kappa_t) \right] \\
&\quad + \mathcal{O}\left(\frac{1}{h_{1,n}^{1+\delta} h_{2,n}^{1+\delta}} \right).
\end{aligned}
\tag{9.24}
$$

Similarly, we get:

$$
\begin{aligned}
\mathsf{E}_t u_{ni}^2 &= \frac{f_t(\kappa_t, \tau) \int K_{(1)}^2(u) \, du \int K_{(2)}^2(v) \, dv}{h_{1,n} h_{2,n}} \\
&\quad \times \mathsf{E}_t \{ A^2(\kappa_t, \tau, \sigma) B^2(\kappa, \tau, \sigma) w^2(\kappa_t) \} \\
&\quad + \mathcal{O}\left(\frac{1}{h_{1,n} h_{2,n}} \right).
\end{aligned}
\tag{9.25}
$$

(9.24) and (9.25) together prove

$$
\frac{\sum_{i=1}^n \mathsf{E}_t |u_{ni}|^{2+\delta}}{\left(\sum_{i=1}^n \mathsf{E}_t |u_{ni}|^2 \right)^{\frac{2+\delta}{2}}} = \mathcal{O}((nh_{1,n}h_{2,n})^{-\frac{\delta}{2}}) = o(1)
\tag{9.26}
$$

as $nh_{1,n}h_{2,n} \to 0$.

Applying the Liapounov central limit theorem, we get

$$
\sqrt{nh_{1,n}h_{2,n}} \, U_n(\sigma) \xrightarrow{\mathcal{L}} \mathrm{N}\left(0, f_t(\kappa_t, \tau) \, \nu^2 \right),
\tag{9.27}
$$

where

$$
\nu^2 \overset{\text{def}}{=} \mathsf{E}_t \{ A^2(\kappa_t, \tau, \sigma) B^2(\kappa_t, \tau, \sigma) w^2(\kappa_t) \} \int K_{(1)}^2(u) K_{(2)}^2(v) \, du \, dv.
\tag{9.28}
$$

By (9.22) and (9.27), Theorem (9.2) is proved.

10 Numerics of Implied Binomial Trees

Wolfgang Härdle and Alena Myšičková

For about 20 years now, discrepancies between market option prices and Black and Scholes (BS) prices have widened. The observed market option price showed that the BS implied volatility, computed from the market option price by inverting the BS formula varies with strike price and time to expiration. These variations are known as "the volatility smile (skew)" and volatility term structure, respectively.

In order to capture the dependence on strike and time to maturity, various smile-consistent models (based on an arbitrage-free approach), have been proposed in the literature. One approach is to model the volatility as a stochastic process, see Hull and White (1987) or Derman and Kani (1998); another works with discontinuous jumps in the stock price, see Merton (1976). However, these extensions cause several practical difficulties such as the violation of the risk-neutrality or no-arbitrage. In contrast, more recent publications proposed by Rubinstein (1994), Derman and Kani (1994), Dupire (1994), and Barle and Cakici (1998) have introduced a locally deterministic volatility function that varies with market price and time. These models independently construct a discrete approximation to the continuous risk neutral process for the underlying assets in the form of binomial or trinomial trees. These deterministic volatility models have both practical and theoretical advantages: they are easily realisable and preserve the no-arbitrage idea inherent in the BS model.

The implied binomial tree (IBT) method constructs a numerical procedure which is consistent with the smile effect and the term structure of the implied volatility. The IBT algorithm is a data adaptive modification of the Cox, Ross and Rubinstein (1979)(CRR) method where the stock evolves along a risk neutral binomial tree with constant volatility.

The following three requirements should be minimally satisfied by an IBT:

- ⊡ correct reproduction of the volatility smile
- ⊡ node transition probabilities lying in $[0, 1]$-interval only
- ⊡ risk neutral branching process (forward price of the underlying asset equals the conditional expected value of itself) at each step.

The last two conditions also guarantee no-arbitrage; should the stock price fall below or above its corresponding forward price, the transition probability would exceed the $[0, 1]$-interval.

The basic aim of the IBT is the estimation of implied probability distributions, or state price densities (SPD), and local volatility surfaces. Furthermore, the IBT may evaluate the future stock price distributions according to the BS implied volatility surfaces which are calculated from observed daily market European option prices.

In this chapter, we describe the numerical construction of the IBT and compare the predicted implied price distributions. In Section 10.1, a detailed construction of the IBT algorithm for European options is presented. First, we introduce the Derman and Kani (1994) (DK) algorithm and show its possible drawbacks. Afterwards, we follow an alternative IBT algorithm by Barle and Cakici (1998) (BC), which modifies the DK method by a normalisation of the central nodes according to the forward price in order to increase its stability in the presence of high interest rates. In Section 10.2 we compare the SPD estimations with simulated conditional density from a diffusion process with a non-constant volatility. In the last section, we apply the IBT to a real data set containing underlying asset price, strike price, time to maturity, interest rate, and call/put option price from EUREX (Deutsche Börse Database). We compare the SPD estimated by real market data with those predicted by the IBT.

10.1 Construction of the IBT

In the early 1970s, Black and Scholes presented the Geometric Brownian Motion (GBM) model, where the stock price S_t is a solution of the stochastic differential equation (SDE):

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t \ , \tag{10.1}$$

with a standard Wiener process W_t and the constant instantaneous drift μ. The constant instantaneous volatility function σ measures the return variability around its expectation μ. Using a risk neutral measure Q, see Fengler (2005), the BS pricing formulae for european call and put options are:

$$
\begin{aligned}
C_t &= e^{-r\tau}\mathsf{E}_\mathsf{Q}\{\max(S_T - K, 0)\} \tag{10.2}\\
P_t &= e^{-r\tau}\mathsf{E}_\mathsf{Q}\{\max(K - S_T, 0)\} \ . \tag{10.3}
\end{aligned}
$$

Under these relations the underlying at the expiration date follows a conditional lognormal distribution with density:

$$q(S_T|S_t, r, \tau, \sigma) = \frac{1}{S_T\sqrt{2\pi\sigma^2\tau}} \exp\left[-\frac{\left\{\log\left(\frac{S_T}{S_t}\right) - (r - \frac{\sigma^2}{2})\tau\right\}^2}{2\sigma^2\tau}\right] . \quad (10.4)$$

In the upper equations T is the expiration date, S_t is the stock price at time t, $\tau = T - t$ is time to maturity, K is the strike price and r is the riskless interest rate.

Looking at a general SDE for an underlying asset price process:

$$\frac{dS_t}{S_t} = \mu(S_t, t)dt + \sigma(S_t, t, \cdot)dW_t , \quad (10.5)$$

we can differentiate the following three concepts of volatility, see Fengler (2005):

Instantaneous volatility $\sigma(S_t, t, \cdot)$

⊡ measures the instantaneous standard deviation of $\log S_t$

⊡ depends on the current level of the asset price S_t, time t and possibly on other state variables denoted with '·'.

Implied volatility $\hat{\sigma}_t(K, T)$

⊡ the BS option price implied measure of volatility, the instantaneous standard deviation of $\log S_t$

⊡ the volatility parameter corresponds to the BS price and a particular observed market option price

⊡ depends on the strike K, the expiration date T and time t.

Local volatility $\sigma_{K,T}(S_t, t)$

⊡ expected instantaneous volatility conditional on a particular level of the asset price $S_T = K$ at $t = T$

⊡ In a deterministic model we can write $\sigma_{K,T}(S_t, t) = \sigma(K, T)$.

The CRR binomial tree is constructed as a discrete approximation of a GBM process with a constant instantaneous volatility $\sigma_t(S_t, t) = \sigma$. Analogously, the IBT can be viewed as a discretization of an instantaneous volatility model:

$$\frac{dS_t}{S_t} = \mu_t dt + \sigma(S_t, t)dW_t, \quad (10.6)$$

where $\sigma(S_t, t)$ depends on both the underlying price and time. The purpose of the IBT is to construct a discrete implementation of the extended BS model based on the observed option prices yielding the variable volatility $\sigma(S_t, t)$. In addition, the IBT may reflect a non-constant drift μ_t. After the construction of the IBT, we are able to estimate a local volatility from underlying stock prices and transition probabilities.

In the IBT construction, only observable data (market option prices, underlying prices, interest rate) are used, it is therefore nonparametric in nature. Several alternative studies based on the kernel method, Aït-Sahalia and Lo (1998), or nonparametric constrained least squares, Yatchew and Härdle (2006), and curve-fitting methods, Jackwerth and Rubinstein (1996) have been published in recent years.

10.1.1 The Derman and Kani Algorithm

In the DK IBT approach, stock prices, transition probabilities and Arrow-Debreu prices (discounted risk neutral probabilities) are calculated iteratively level by level, starting in the level zero.
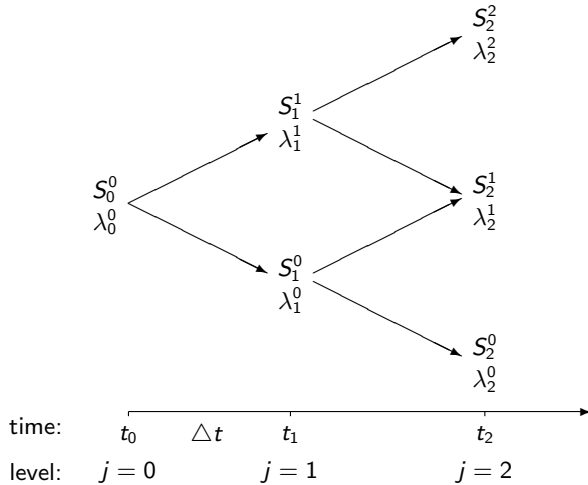
Figure 10.1. Construction of an implied binomial tree.

Figure 10.1 illustrates the construction of the first two nodes of an IBT. We build the IBT on the time interval $[0, T]$ with $j = 0, 1, 2, \ldots, n$ equally spaced

levels, $\triangle t$ apart. We start at zero level with $t = 0$, here the stock price equals the current price of the underlying: $S_0^0 = S$. There are $n+1$ nodes at the nth level of the tree, we indicate the stock price of the ith node at the nth level by S_n^i, and the forward price at level $n+1$ of S_n^i at level n by $F_n^i = e^{r \triangle t} S_n^i$. The conditional probability $p_{i+1}^n = \mathrm{P}(S_{n+1} = S_{n+1}^{i+1} | S_n = S_n^i)$ is the transition probability of making a transition from node (n, i) to node $(n+1, i+1)$.

The forward price $F_{n,i}$ is required to satisfy the risk neutral condition:

$$F_n^i = p_{i+1}^n S_{n+1}^{i+1} + (1 - p_{i+1}^n) S_{n+1}^i . \qquad (10.7)$$

Thus we obtain the transition probability from the following equation:

$$p_{i+1}^n = \frac{F_n^i - S_{n+1}^i}{S_{n+1}^{i+1} - S_{n+1}^i} . \qquad (10.8)$$

The Arrow-Debreu price is the price of an option which pays 1 unit payoff if the stock price S_t at time t attains the value S_n^i, and 0 otherwise. The Arrow-Debreu price in the state i at level n can be computed as the expected discounted value of its payoff: $\lambda_n^i = \mathsf{E}[e^{-rt} \mathbf{1}(S_t = S_n^i) | S_0 = S_0^0]$. In general, Arrow-Debreu prices can be obtained by the iterative formula, where $\lambda_0^0 = 1$ as a definition.

$$
\begin{aligned}
\lambda_{n+1}^0 &= e^{-r \triangle t} \left\{ \lambda_n^0 (1 - p_1^n) \right\} \\
\lambda_{n+1}^{i+1} &= e^{-r \triangle t} \left\{ \lambda_n^i p_{i+1}^n + \lambda_n^{i+1} (1 - p_{i+2}^n) \right\}, \quad 0 \le i \le n-1 \qquad (10.9) \\
\lambda_{n+1}^{n+1} &= e^{-r \triangle t} \left\{ \lambda_n^n p_{n+1}^n \right\}
\end{aligned}
$$

To illustrate the calculation of the Arrow-Debreu prices, we provide an example with a construction of a CRR binomial tree. Let us assume that the current value of the underlying $S = 100$, time to maturity $\tau = T = 2$ years, $\triangle t = 1$ year, constant volatility $\sigma = 10\%$, and riskless interest rate $r = 0.03$. The Arrow-Debreu price tree shown in the Figure 10.3 can be calculated from the stock price tree in the Figure 10.2.

Using the CRR method, the stock price at the lower node at the first level equals $S_1^0 = S_0^0 \cdot e^{-\sigma \triangle t} = 100 \cdot e^{-0.1} = 90.52$, and at the upper node $S_1^1 = S_0^0 \cdot e^{\sigma \triangle t} = 110.47$. The transition probability $p_1^0 = 0.61$ is obtained by the formula (10.8) with $F_0^0 = S_0^0 e^{0.03} = 103.05$. Now, we calculate λ_1^i for $i = 0, 1$, according to the formula (10.9): $\lambda_1^0 = e^{-r \triangle t} \cdot \lambda_0^0 \cdot (1 - p_1^0) = 0.36$ and $\lambda_1^1 = e^{-r \triangle t} \cdot \lambda_0^0 \cdot p_1^0 = 0.61$. At the second level, we calculate the stock prices according to the corresponding nodes at the first level, for example: $S_2^0 = S_1^0 \cdot e^{-\sigma \triangle t} = 81.55$, $S_2^1 = S_0^0 = 100$ and $S_2^2 = S_1^1 \cdot e^{\sigma \triangle t} = 122.04$.
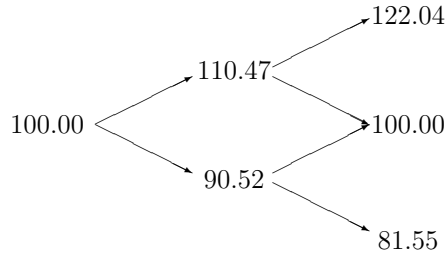
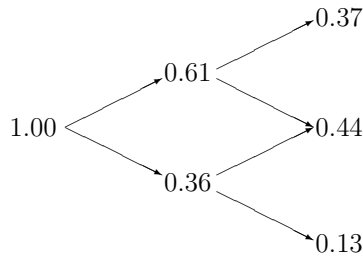Figure 10.2. CRR binomial tree for stock prices with $T = 2$ years, $\triangle t = 1$, $\sigma = 0.1$ and $r = 0.03$. $\blacksquare$ XFGIBT01

Figure 10.3. CRR binomial tree for Arrow-Debreu prices with $T = 2$ years, $\triangle t = 1$, $\sigma = 0.1$ and $r = 0.03$. $\blacksquare$ XFGIBT01

The corresponding Arrow-Debreu prices λ_2^i for $i = 0, 1, 2$ are obtained by the substitution in the formula 10.9:

$$\lambda_2^0 = e^{-r \triangle t} \cdot \lambda_1^0 \cdot (1 - p_1^1) = 0.13$$
$$\lambda_2^1 = e^{-r \triangle t} \cdot \{\lambda_1^0 \cdot p_1^1 + \lambda_1^1 \cdot (1 - p_2^1) = 0.44\}$$
$$\lambda_2^2 = e^{-r \triangle t} \cdot \lambda_1^1 \cdot p_2^1 = 0.37 \ .$$

In the BS model with the state price density (SPD) from 10.4, the option

prices are given by:

$$C(K, \tau) = e^{-r\tau} \int_0^{+\infty} \max(S_T - K, 0)\, q(S_T | S_t, r, \tau) dS_T, \qquad (10.10)$$

$$P(K, \tau) = e^{-r\tau} \int_0^{+\infty} \max(K - S_T, 0)\, q(S_T | S_t, r, \tau) dS_T, \qquad (10.11)$$

where $C(K, \tau)$ and $P(K, \tau)$ denote call option price and put option price respectively, and K is the strike price. In the IBT, option prices are calculated in discrete time intervals $\tau = n\triangle t$ using the Arrow-Debreu prices,

$$C(K, n\triangle t) = \sum_{i=0}^{n} \lambda_{n+1}^{i+1} \max(S_{n+1}^{i+1} - K, 0), \qquad (10.12)$$

$$P(K, n\triangle t) = \sum_{i=0}^{n} \lambda_{n+1}^{i+1} \max(K - S_{n+1}^{i+1}, 0). \qquad (10.13)$$

Using the risk neutral condition (10.7) and the discrete option price calculation from (10.12) or (10.13), one obtains the iteration formulae to construct the IBT.

Let us assume the strike price is equal to the known stock price: $K = S_n^i = S$. Then the contribution from the transition to the first in-the-money upper node can be separated from the other contributions. Using the iterative formulae for the Arrow-Debreu prices (10.9) in the equation (10.12):

$$e^{r\triangle t}C(S, n\triangle t) = \lambda_n^0(1 - p_1^n)\max(S_{n+1}^0 - S, 0) + \lambda_n^n p_{n+1}^n \max(S_{n+1}^{n+1} - S, 0)$$

$$+ \sum_{j=0}^{n-1}\left\{\lambda_n^j p_{j+1}^n + \lambda_n^{j+1}(1 - p_{j+2}^n)\right\}\max(S_{n+1}^{j+1} - S, 0)$$

$$= \left\{\lambda_n^i p_{i+1}^n + \lambda_n^{i+1}(1 - p_{i+2}^n)\right\}(S_{n+1}^{i+1} - S) + \lambda_n^n p_{n+1}^n(S_{n+1}^{n+1} - S)$$

$$+ \sum_{j=i+1}^{n-1}\left\{\lambda_n^j p_{j+1}^n + \lambda_n^{j+1}\left(1 - p_{j+2}^n\right)\right\}(S_{n+1}^{j+1} - S)$$

$$= \lambda_n^i p_{i+1}^n(S_{n+1}^{i+1} - S)$$

$$+ \sum_{j=i+1}^{n-1}\lambda_n^j p_{j+1}^n(S_{n+1}^{j+1} - S) + \lambda_n^n p_{n+1}^n(S_{n+1}^{n+1} - S)$$

$$+ \lambda_n^{i+1}(1 - p_{i+2}^n)(S_{n+1}^{i+1} - S) + \sum_{j=i+2}^{n}\lambda_n^j(1 - p_{j+1}^n)(S_{n+1}^{j} - S)$$

$$= \lambda_n^i p_{i+1}^n(S_{n+1}^{i+1} - S)$$

$$+ \sum_{j=i+1}^{n}\lambda_n^j\left\{\left(1 - p_{j+1}^n\right)(S_{n+1}^{j} - S) + p_{j+1}^n(S_{n+1}^{j+1} - S)\right\}\ .$$

Entering the risk neutral condition (10.7) in the last term, one obtains:

$$e^{r\triangle t}C(S, n\triangle t) = \lambda_n^i p_{i+1}^n\left(S_{n+1}^{i+1} - S\right) + \sum_{j=i+1}^{n}\lambda_n^j\left(F_n^j - S\right)\ . \tag{10.14}$$

Now, the stock price for the upper node can be rewritten in terms of the known Arrow-Debreu prices λ_n^i, the known stock prices S_n^i and the known forwards F_n^i:

$$S_{n+1}^{i+1} = \frac{S_{n+1}^i\left\{C\left(S_n^i, n\triangle t\right)e^{r\triangle t} - \rho_u\right\} - \lambda_n^i S_n^i\left(F_n^i - S_{n+1}^i\right)}{C\left(S_n^i, n\triangle t\right)e^{r\triangle t} - \rho_u - \lambda_n^i\left(F_n^i - S_{n+1}^i\right)}\ , \tag{10.15}$$

where ρ_u denotes the following summation term:

$$\rho_u = \sum_{j=i+1}^{n}\lambda_n^j(F_n^j - S_n^i)\ . \tag{10.16}$$

The transition from the nth to the $(n + 1)$th level of the tree is defined by $(2n + 3)$ parameters, i.e. $(n + 2)$ stock prices of the nodes at the $(n + 1)$th

level, and $(n + 1)$ transition probabilities (when the IBT starts at the zero-level). Suppose $(2n+1)$ parameters corresponding to the nth level are known, the stock prices S_{n+1}^i and transition probabilities p_{i+1}^n at all nodes above the centre of the tree corresponding to the $(n+1)$th level can be found iteratively using the equations (10.15) and (10.8) as follows:

We always start from the central nodes, if n is odd, define $S_{n+1}^i = S_0^0 = S$, for $i = (n + 1)/2$. If n is even, we start from the two central nodes just below and above the centre of the level, S_{n+1}^i and S_{n+1}^{i+1} for $i = n/2$, and set $S_{n+1}^i = (S_n^i)^2 / S_{n+1}^{i+1} = S^2 / S_{n+1}^{i+1}$, which adjusts the logarithmic CRR centring spacing between S_n^i and S_{n+1}^{i+1} to be the same as that between S_n^i and S_{n+1}^i. Substituting this relation into (10.15) one gets the formula for the upper of the two central nodes for the odd levels:

$$S_{n+1}^{i+1} = \frac{S \left\{ C\left(S, n\triangle t\right) e^{r\triangle t} + \lambda_n^i S - \rho_u \right\}}{\lambda_n^i F_n^i - e^{r\triangle t} C\left(S, n\triangle t\right) + \rho_u} \qquad \text{for } i = \frac{n}{2}\,. \qquad (10.17)$$

Once we have the initial nodes' stock prices, according to the relationships among the different parameters, we can repeat the process to calculate those at higher nodes $(n + 1, j)$, $j = i + 2, \ldots n + 1$ one by one.

Similarly, we can calculate the parameters at lower nodes $(n + 1, j)$, $j = i - 1, \ldots, 1$ at the $(n + 1)$th level by using the known put prices $P(K, n\triangle t)$ for $K = S_n^i$.

$$S_{n+1}^i = \frac{S_{n+1}^{i+1} \left\{ e^{r\triangle t} P(S_n^i, n\triangle t) - \rho_l \right\} - \lambda_n^i S_n^i (F_n^i - S_{n+1}^{i+1})}{e^{r\triangle t} P \left\{ S_n^i, (n + 1)\triangle t \right\} - \rho_l + \lambda_n^i (F_n^i - S_{n+1}^{i+1})}\,, \qquad (10.18)$$

where ρ_l denotes the sum over all nodes below the one with price S_n^i:

$$\rho_l = \sum_{j=0}^{i-1} \lambda_n^j (S_n^i - F_n^j)\,. \qquad (10.19)$$

Transition probabilities and Arrow-Debreu prices are obtained by (10.8) and (10.9), respectively.

$C(K, \tau)$ and $P(K, \tau)$ in (10.15) and (10.18) are the interpolated values for a call or put struck today at strike price K and time to maturity τ. In the DK construction, they are obtained by the CRR binomial tree with constant parameters $\sigma = \sigma_{imp}(K, \tau)$, calculated from the known market option prices. In practice, calculating interpolated option prices by the CRR method is computationally intensive.

10.1.2 Compensation

The transition probability p_i^n at any node should lie between 0 and 1, this condition avoids the riskless arbitrage: if $p_{i+1}^n > 1$, the stock price S_{n+1}^{i+1} would fall below the forward price F_n^i, similarly, if $p_{i+1}^n < 0$, the strike price S_{n+1}^i would fall above the forward price F_n^i. Therefore it is useful to limit the estimated stock prices by the neighbouring forwards from the previous level:

$$F_n^i < S_{n+1}^{i+1} < F_n^{i+1} . \qquad (10.20)$$

If the stock price does not fulfil the above inequality condition, we redefine it by assuming that the logaritmic difference between the stock prices at this node and its adjacent is equal to the logaritmic difference between the corresponding stock prices at the two nodes at the previous level, i.e., $\log(S_{n+1}^{i+1}/S_{n+1}^i) = \log(S_n^i/S_n^{i-1})$. Sometimes, the obtained price still does not satisfy inequality (10.20), then we substitute the stock price S_{n+1}^{i+1} by the average of F_n^i and F_n^{i+1}.

As used in the construction of the IBT in (10.12) or (10.13), the implied conditional distribution, the SPD $q(S_T|S_t, r, \tau)$, could be estimated at discrete time $\tau = n\triangle t$ by the product of the Arrow-Debreu prices λ_{n+1}^i at the $(n+1)$th level with the influence of the interest rate $e^{rn\triangle t}$. To fulfill the risk-neutrality condition (10.7), the conditional expected value of the underlying log stock price in the following $(n + 1)$th level, given the stock price at the nth level is defined as:

$$M = \mathsf{E}_{\mathsf{Q}}\{\log(S_{n+1})|S_n = S_n^i\} = p_{i+1}^n \log(S_{n+1}^{i+1}) + (1 - p_{i+1}^n) \log(S_{n+1}^i) . \quad (10.21)$$

We can specify such a condition also for the conditional second moments of $\log(S_{n+1})$ at $S_n = S_n^i$, which is the implied local volatility $\sigma^2(S_n^i, n\triangle t)$ during the time period $\triangle t$:

$$
\begin{aligned}
\sigma^2(S_n^i, \triangle t) &= \mathsf{Var}_{\mathsf{Q}}\{\log(S_{n+1})|S_n = S_n^i\} \\
&= p_{i+1}^n \{\log(S_{n+1}^{i+1}) - M\}^2 + (1 - p_{i+1}^n)\{\log(S_{n+1}^i) - M\}^2 \\
&= 2\log\left(\frac{S_{n+1}^{i+1}}{S_{n+1}^i}\right)\{p_{i+1}^n(1 - p_{i+1}^n)\} .
\end{aligned} \qquad (10.22)
$$

After the construction of an IBT, all stock prices, transition probabilities, and Arrow-Debreu prices at any node in the tree are known. We are thus able to calculate the local volatility $\sigma(S_n^i, m\triangle t)$ at any level m.

In general, the instantaneous volatility function used in the diffusion model (10.6) is different from the local volatility function derived in (10.22), only in the BS model are they identical. Additional, the BS implied volatility

$\widehat{\sigma}(K, \tau)$, which assumes the Black-Scholes model at least locally, differs from the local volatility $\sigma(s, \tau)$, they describe different characteristics of the second moment using different parameters.

If we choose $\triangle t$ small enough, we obtain the estimated SPD at fixed time to maturity, and the distribution of local volatility $\sigma(S, \tau)$.

10.1.3 Barle and Cakici Algorithm

Barle and Cakici (1998) (BC) suggest an improvement of the DK construction. The first major modification is the choice of the strike price in which the option should be evaluated (as in 10.14). In the BC algorithm, the strike price K is chosen to be equal to the forward price F_n^i, and similarly to the DK construction, using the discrete approximation (10.12) we get:

$$
\begin{aligned}
e^{r\triangle t}C(F_n^i, n\triangle t) &= \sum_{j=0}^{n} \lambda_{n+1}^{j+1} \max(S_{n+1}^{j+1} - F_n^i, 0) \\
&= \{\lambda_n^i p_{i+1}^n + \lambda_n^{i+1}(1 - p_{i+2}^n)\}(S_{n+1}^{i+1} - F_n^i) + \lambda_n^n p_{n+1}^n(S_{n+1}^{n+1} - F_n^i) \\
&\quad + \sum_{j=i+1}^{n-1} \{\lambda_n^j p_{j+1}^n + \lambda_n^{j+1}\left(1 - p_{j+2}^n\right)\}(S_{n+1}^{j+1} - F_n^i) \\
&= \lambda_n^i p_{i+1}^n(S_{n+1}^{i+1} - F_n^i) \\
&\quad + \sum_{j=i+1}^{n} \lambda_n^j \left\{\left(1 - p_{j+1}^n\right)(S_{n+1}^j - F_n^i) + p_{j+1}^n(S_{n+1}^{j+1} - F_n^i)\right\} \ .
\end{aligned}
$$

Entering the risk neutral condition again (10.7) one obtains:

$$
e^{r\triangle t}C(F_n^i, n\triangle t) = \lambda_n^i p_{i+1}^n \left(S_{n+1}^{i+1} - F_n^i\right) + \sum_{j=i+1}^{n} \lambda_n^j \left(F_n^j - F_n^i\right) \ . \tag{10.23}
$$

Identify the upper sum as:

$$
\varrho_u = \sum_{j=i+1}^{n} \lambda_n^j \left(F_n^j - F_n^i\right) \ , \tag{10.24}
$$

and using the equation for the transition probability (10.8) we can write the recursion relation for the stock price in the upper node as follows:

$$
S_{n+1}^{i+1} = \frac{S_{n+1}^i \left\{C\left(F_n^i, n\triangle t\right)e^{r\triangle t} - \varrho_u\right\} - \lambda_n^i F_n^i \left(F_n^i - S_{n+1}^i\right)}{C\left(F_n^i, n\triangle t\right)e^{r\triangle t} - \varrho_u - \lambda_n^i \left(F_n^i - S_{n+1}^i\right)} \ . \tag{10.25}
$$

Analogous to the DK construction, we start from the central nodes of the binomial tree, but in contrast with the DK construction the BC construction takes the riskless interest rate into account. If $(n + 1)$ is even, the price of the central node $S_{n+1}^i = S_0^0 e^{r \triangle t}$ for $i = (n + 1)/2$. If $(n + 1)$ is odd, the two central nodes must satisfy $S_{n+1}^i \cdot S_{n+1}^{i+1} = (F_n^i)^2$. Adding this condition to the equation (10.25) the lower central node can be calculated as:

$$S_{n+1}^i = F_n^i \frac{\lambda_n^i F_n^i - \{e^{r \triangle t} C(F_n^i, n \triangle t) - \varrho_u\}}{\lambda_n^i F_n^i + \{e^{r \triangle t} C(F_n^i, n \triangle t) - \varrho_u\}} \quad \text{for} \quad i = 1 + n/2, \qquad (10.26)$$

the upper one is then: $S_{n+1}^{i+1} = (F_n^i)^2 / S_{n+1}^i$.

After stock prices of the central nodes are obtained, we repeat the recursion equation (10.25) to calculate the stock prices at higher nodes $(n + 1, j), j = i + 2, \ldots, n + 1$. The transition probabilities and Arrow-Debreu prices are calculated through (10.8) and (10.9), respectively.

Similarly, an analogous recursion relation for the stock prices at lower nodes can be found by using put option prices at strike F_n^i:

$$S_{n+1}^i = \frac{S_{n+1}^{i+1} \{P(F_n^i, n \triangle t) e^{r \triangle t} - \varrho_l\} \lambda_n^i F_n^i (S_{n+1}^{i+1} - F_n^i)}{P(F_n^i, n \triangle t) e^{r \triangle t} - \varrho_l - \lambda_n^i (S_{n+1}^{i+1} - F_n^i)}, \qquad (10.27)$$

where where ϱ_l denotes the lower sum:

$$\varrho_l = \sum_{j=0}^{i-1} \lambda_n^j (F_n^i - F_n^j) \, .$$

Notice that BC use the Black-Scholes call and put option prices $C(K, \tau)$ and $P(K, \tau)$, which makes the calculation faster than the interpolation technique based on the CRR method.

The balancing inequality (10.20), to avoid negative transition probabilities, and therewith the arbitrage is still used in the BC algorithm: they re-estimate S_{n+1}^{i+1} by the average of F_n^i and F_n^{i+1}, though the choice of any point between these forward prices is sufficient.

10.2 A Simulation and a Comparison of the SPDs

The following detailed example illustrates the construction of the tree from the smile, using the DK algorithm first, and the BC algorithm afterwards.

Let us assume that the current value of the underlying stock $S = 100$, with no dividend and the annually compounded riskless interest rate $r = 3\%$ per year for all time expirations. For the implied volatility function, we use a convex function:

$$\widehat{\sigma} = \frac{-0.2}{\{\log(K/S_t)\}^2 + 1} + 0.3 \,, \tag{10.28}$$

taken from Fengler (2005). For simplicity, we do not model a term structure of the implied volatility. The BS option prices needed for growing the tree are calculated from this implied volatility function. We construct the IBTs with time to maturity $T = 1$ year discretized in five time steps.

10.2.1 Simulation Using the DK Algorithm

Using the assumption on the BS implied volatility surface described above, we obtain the one year stock price implied binomial tree (Figure 10.4), the upward transition probability tree (Figure 10.5), and the Arrow-Debreu price tree (Figure 10.6).
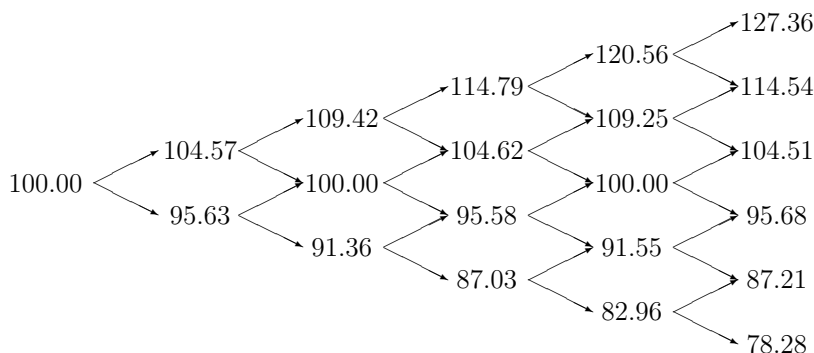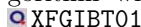
Figure 10.4. Stock price tree calculated with the DK algorithm with $S_0^0 = 100$, $r = 0.03$ and $T = 1$ year.
XFGIBT01

All the IBTs correspond to time to maturity $\tau = 1$ year, and $\triangle t = 1/5$ year. Figure 10.4 shows the estimated stock prices starting at the zero level with $S_0^0 = S = 100$. The elements in the j-th column correspond to the $(j-1)$th level of the stock price tree. Figure 10.5 shows the transition probabilities, its element (n, j) represents the transition probability from the node $(n-1, j-1)$ to the node (n, j). The third tree displayed in Figure 10.6 contains the Arrow-Debreu prices. Its elements in the j-th column match the Arrow-Debreu
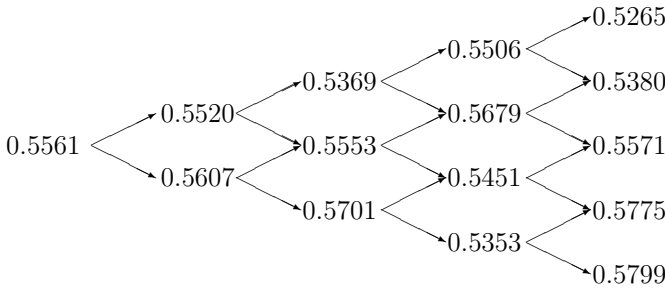
Figure 10.5. Transition probability tree calculated with the DK algorithm with $S_0^0 = 100$, $r = 0.03$ and $T = 1$ year. ▣ XFGIBT01
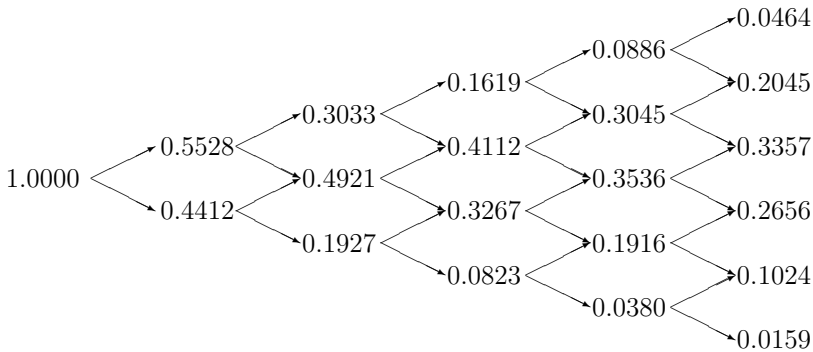

Figure 10.6. Arrow-Debreu price tree calculated with the DK algorithm with $S_0^0 = 100$, $r = 0.03$ and $T = 1$ year. ▣ XFGIBT01

prices in the $(j-1)$ th level. Using the stock prices together with Arrow-Debreu prices of the nodes at the final level, a discrete approximation of the implied price distribution can be obtained. Notice that by the definition of the Arrow-Debreu price, the risk neutral probability corresponding to each node should be calculated as the product of the Arrow-Debreu price and the factor $e^{rj\triangle t}$ in the level j.

Choosing the time steps small enough, we obtain more accurate estimation of the implied price distribution and the local volatility surface $\sigma(S, \tau)$. We

still use the same implied volatility function from (10.28), and assume $S_0^0 = 100$, $r = 0.03$, $T = 5$ years.

SPD estimation arising from fitting the implied five-year tree with 40 levels is shown in Figure 10.7. Local volatility surface computed from the implied tree at different times to maturity and stock price levels is shown in Figure 10.8. Obviously, the local volatility captures the volatility smile, which decreases with the strike price and increases with the time to maturity.
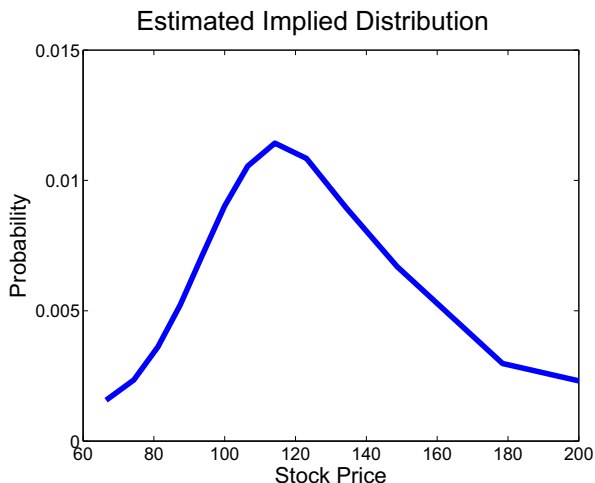

Figure 10.7. SPD estimation by the DK IBT computed with $S_0^0 = 100$, $r = 0.03$ and $T = 5$ years. Q XFGIBT02

10.2.2 Simulation Using the BC Algorithm

The BC algorithm can be applied in analogy to the DK technique. The computing part is replaced by the BC algorithm, we are using the implied volatility function from (10.28) as in the DK algorithm. Figures 10.9 - 10.11 show the one-year stock price tree with five steps, transition probability tree, and Arrow-Debreu tree. Figure 10.12 presents the plot of the estimated SPD by fitting a five year implied binomial tree with 40 levels using BC algorithm. Figure 10.13 shows the characteristics of the local volatility surface of the generated IBT, the local volatility follows the "volatility smile", which decreases with the stock price and increases with time.
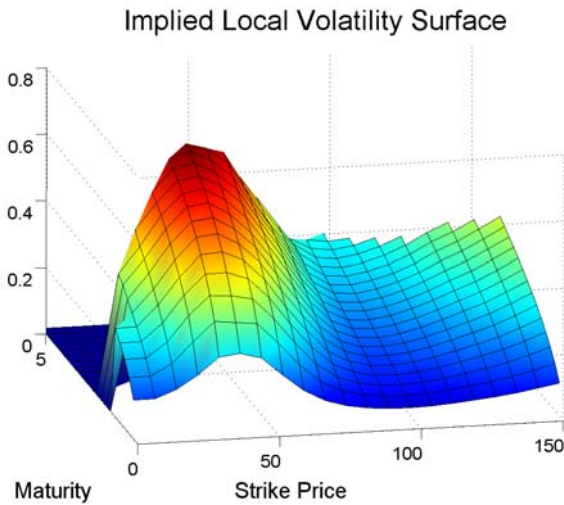
Implied Local Volatility Surface

Figure 10.8.   Implied local volatility surface estimated by the DK IBT with $S_0^0 = 100$, $r = 0.03$ and $T = 5$ years. **Q** XFGIBT02 .
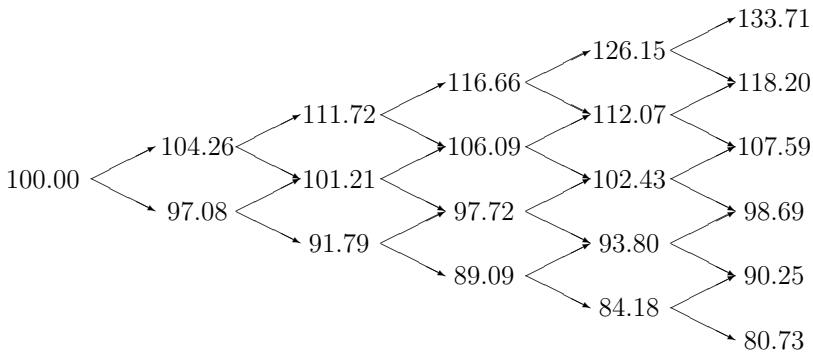

Figure 10.9. Stock price tree calculated with the BC algorithm with $S_0^0 = 100$, $r = 0.03$ and $T = 1$ year. **Q** XFGIBT01

10.2.3 Comparison with the Monte-Carlo Simulation

We now compare the SPD estimation obtained by the two IBT methods with the estimated density function of a simulated process S_t generated from
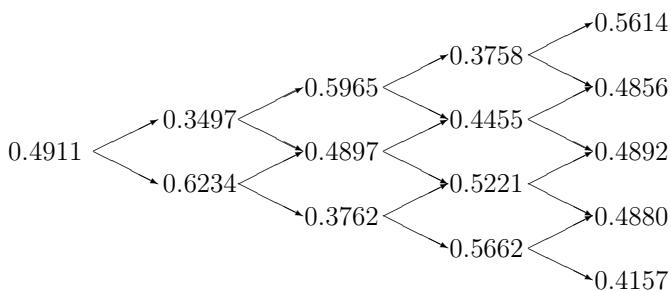
Figure 10.10. Transition probability tree calculated with the BC algorithm with $S_0^0 = 100$, $r = 0.03$ and $T = 1$ year.
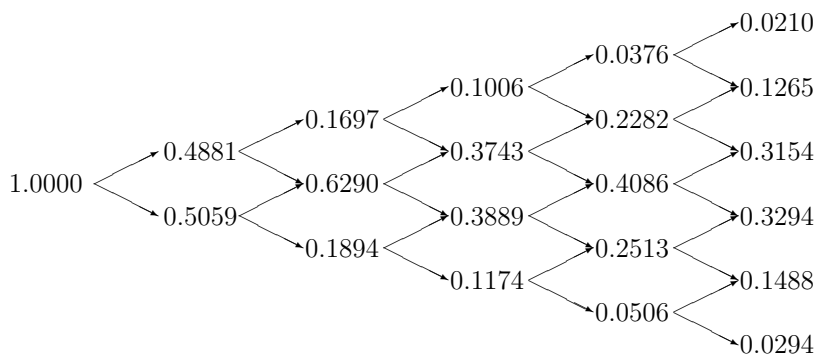Ⓠ XFGIBT01

Figure 10.11. Arrow-Debreu price tree calculated with the BC algorithm with $S_0^0 = 100$, $r = 0.03$ and $T = 1$ year.
Ⓠ XFGIBT01

the diffusion process (10.6). To perform a discrete approximation of this diffusion process, we use the Euler scheme with time step $\delta = 1/1000$, the constant drift $\mu_t = r = 0.03$ and the volatility function $\sigma(S_t, t) = \left[\frac{-0.2}{\{\log(K/S_t)\}^2 + 1} + 0.3 \right]$.

Compared to Sections 10.2.2 and 10.2.2 where we started from the BS implied volatility surface, here we construct the IBTs direct from the simulated option price function. In the construction of the IBTs, we calculate the option prices
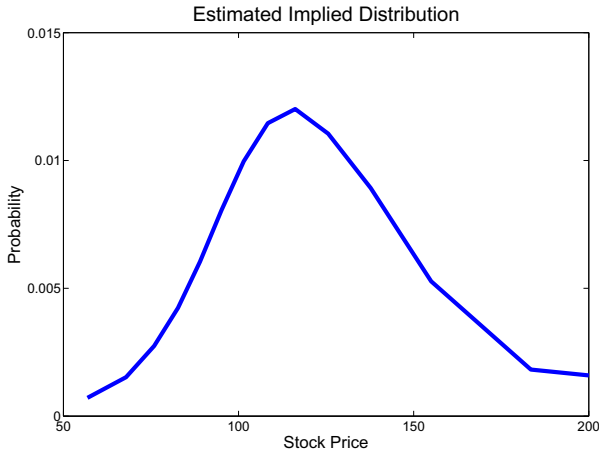
Figure 10.12. SPD estimation by the BC IBT computed with
$S_0^0 = 100$, $r = 0.03$ and $T = 5$ years. **Q** XFGIBT02
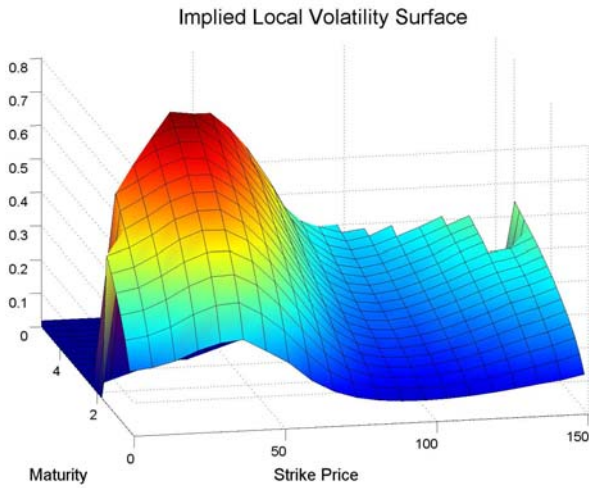

Figure 10.13. Implied local volatility surface estimated by the
BC IBT with $S_0^0 = 100$, $r = 0.03$ and $T = 5$ years. **Q** XFGIBT02

corresponding to each node at the implied tree according to their theoretical
definitions (10.3) and (10.3) from the simulated asset prices S_t. We simulate
S_t for $t = i/4$ year, $i = 1, \ldots, 50$ in the diffusion model (10.6) with the
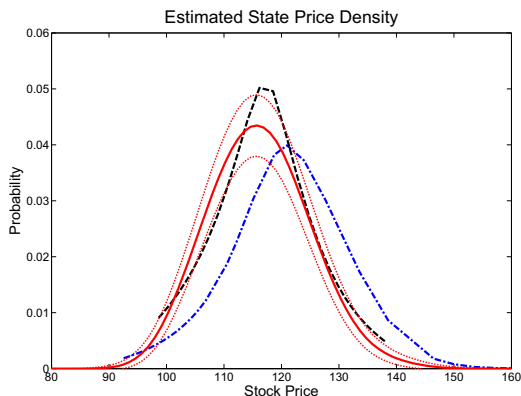
Monte-Carlo simulation method.

Figure 10.14. SPD estimation from the DK IBT (blue dashed line) and from the BC IBT (black dashed line) compared to the estimation by Monte-Carlo simulation with its 95% confidence band (red lines). Level = 50, $T = 5$ years, $\triangle t = 0.1$ year. **Q** XFGIBT03

From the estimated distribution shown in Figure 10.14, we observe small deviations of the SPDs obtained from the two IBT methods from the estimation obtained by the Monte-Carlo simulation. The SPD estimation by the BC algorithm coincides substantially better with the estimation from the simulated process than the estimation by the DK algorithm, which shows a shifted mean of its SPD.

As above, we can also estimate the local volatility surface from the both implied binomial trees. Compare Figure 10.15 with Figure 10.16 and notice that some edge values cannot be obtained directly from the five-year IBT. However, both local volatility surface plots actually coincide with the volatility smile characteristic, the implied local volatility of the out-the-money options decreases with the increasing stock price, and increases with time.

10.3 Example – Analysis of EUREX Data

In the following example we use the IBTs to estimate the price distribution of the real stock market data. We use underlying asset prices, strike prices, time to maturity, interest rates, and call/put option prices from EUREX at 19
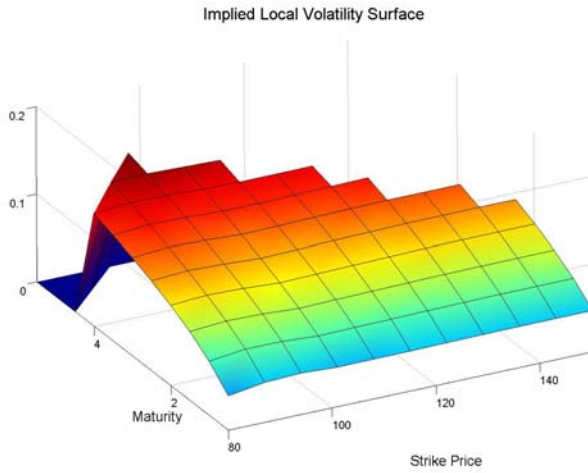
Figure 10.15. Implied local volatility surface of the simulated
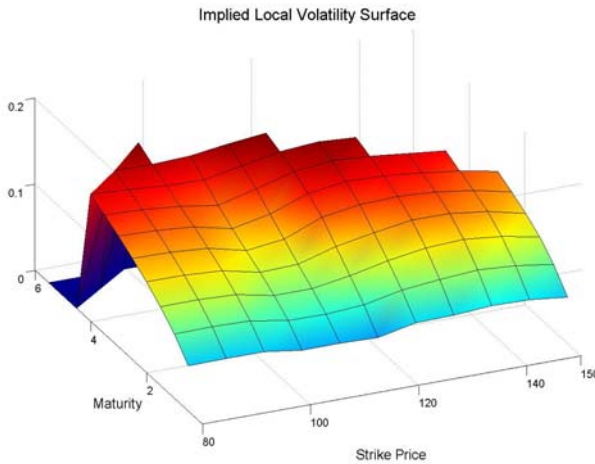model, calculated from DK IBT. XFGIBTcdk

Figure 10.16. Implied local volatility surface of the simulated
model, calculated from BC IBT. XFGIBTcbc

March, 2007, taken from the database of German stock exchange. First, we
estimate the BS implied volatility surface from the data set with the technique
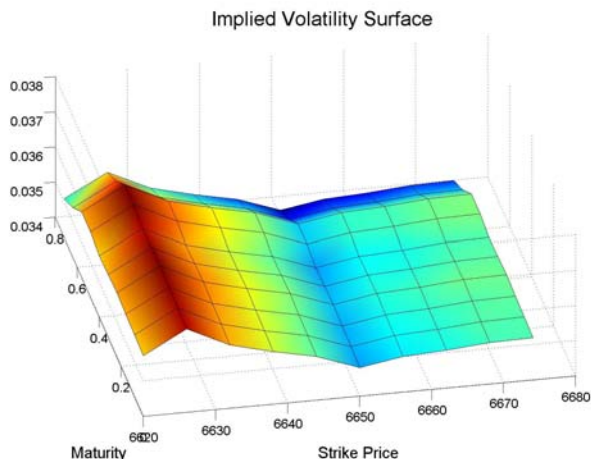
Figure 10.17. BS implied volatility surface estimated from real stock and option prices. **Q** XFGIBT05

of Fengler, Härdle and Villa (2003). Figure 10.17 shows the estimated implied volatility surface, which reflects the characteristics that the implied volatility decreases with the strike price and increases with time to maturity.

Now we construct the IBTs, where we calculate the interpolated option prices with the CRR binomial tree method using the estimated implied volatility. Fitting the function of option prices directly from the market option prices causes difficulties since the function approaches a value of zero for very high strike prices which would violate no-arbitrage conditions.

The estimated stock price distribution, obtained by the BC and the DK IBT with 40 levels, for $\tau = 0.5$ year, is shown in Figure 10.18. Obviously, the both estimated SPDs are nearly identical. The SPDs do not show any deviations from the log-normal characteristics according to their skewness and kurtosis.

From the simulations and real data example, we conclude that the implied binomial tree is a simple smile-consistent method to assess the future stock prices. Still, some limitations of the algorithms remain. With an increasing interest rate or with a small time step, negative transition probabilities occur more often. When the interest rate is high, the BC algorithm is a better choice. The DK algorithm cannot handle with higher interest rates such as $r = 0.2$, in this case the BC algorithm still can be used. In addition, the negative probabilities appear more rarely in the BC algorithm than in
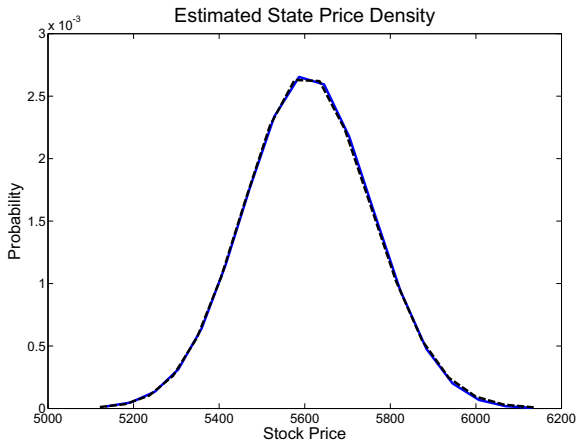
Figure 10.18. SPD estimation by the BC IBT (black dashed line) and by the DK IBT (blue solid line) from the EUREX data, $\tau = 0.5$ year, level $= 25$. ◙ XFGIBT05

the DK construction, even though most of them appear at the edge of the trees. But, by modifying these values we are effectively losing the information about the volatility behavior at the corresponding nodes. This deficiency is a consequence of our condition that continuous diffusion process is modeled as a discrete binomial process. Improving of this requirement leads to a transition to multinomial or varinomial trees which have a drawback of more complicated models with difficult realization.

Besides its basic function to price derivatives in consistency with market prices, IBTs are also useful for hedging, calculating local volatility surfaces or estimation of the future price distribution according to the historical data. In the practical application, the reliability of the approach depends critically on the quality of the dynamics estimation of the underlying process, such as of the BS implied volatility surface obtained from the market option prices.

Bibliography

Ait-Sahalia, Y. and Lo, A. (1998). Nonparametric Estimation of State-Price Densities Implicit in Financial Asset Prices, *Journal of Finance*, **53**: 499–547.

Ait-Sahalia, Y. , Wang, Y. and Yared, F.(2001). Do Option Markets Correctly Price the Probabilities of Movement of the Underlying Asset? *Journal of Econometrics*, **102**: 67–110.

Barle, S. and Cakici, N. (1998). How to Grow a Smiling Tree *The Journal of Financial Engineering*, **7**: 127–146.

Bingham, N.H. and Kiesel, R. (1998). *Risk-neutral Valuation: Pricing and Hedging of Financial Derivatives*, Springer Verlag, London.

Cox, J., Ross, S. and Rubinstein, M. (1979). Option Pricing: A simplified Approach, *Jouranl of Financial Economics* **7**: 229–263.

Derman, E. and Kani, I. (1994). The Volatility Smile and Its Implied Tree *http://www.gs.com/qs/*

Derman, E. and Kani, I. (1998). Stochastic Implied Trees: Arbitrage Pricing with Stochastic Term and Strike Structure of Volatility, *International Journal of Theroetical and Applied Finance* **1**: 61–110.

Dupire, B. (1994). Pricing with a Smile, *Risk* **7**: 18–20.

Fengler, M. R.(2005). *Semiparametric Modeling of Implied Volatility*, Springer Verlag, Heidelberg.

Fengler, M. R., Härdle, W. and Villa, Chr. (2003). The Dynamics of Implied Volatilities: A Common Principal Components Approach, *Review of Derivative Research* **6**: 179–202.

Härdle, W., Hlávka, Z. and Klinke, S. (2000). *XploRe Application Guide*, Springer Verlag, Heidelberg.

Hui, E.C. (2006). An enhanced implied tree model for option pricing: A study on Hong Kong property stock options, *International Review of Economics and Finance* **15**: 324–345.

Hull, J. and White, A. (1987). The Pricing of Options on Assets with Stochastic Volatility, *Journal of Finance* **42**: 281–300.

Jackwerth, J. (1999). Optional-Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review, *Journal of Finance* **51**: 1611–1631.

Jackwerth, J. and Rubinstein, M. (1996). Recovering Probability Distributions from Option Prices, *Journal of Finance* **51**: 1611–1631.

Kim, I.J. and Park, G.Y. (2006). An empirical comparison of implied tree models for KOSPI 200 index options, *International Review of Economics and Finance* **15**: 52–71.

Kloeden, P., Platen, E. and Schurz, H. (1994). *Numerical Solution of SDE Through Computer Experiments*, Springer Verlag, Heidelberg.

Merton, R. (1976). Option Pricing When Underlying Stock Returns are Discontinuous, *Journal of Financial Economics* **January-March**: 125–144.

Moriggia, V., Muzzioli,S. and Torricelli, C. (2007). On the no-arbitrage condition in option implied trees, *European Journal of Operational Research* forthcoming.

Muzzioli,S. and Torricelli, C. (2005). The pricing of options on an interval binomial tree. An application to the DAX-index option market, *European Journal of Operational Research* **163**: 192–200.

Rubinstein, M. (1994). Implied Binomial Trees. *Journal of Finance* **49**: 771–818.

Yatchew, A. and Härdle,W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap, *Journal of Econometrics* **133**: 579–599.

11 Application of Extended Kalman Filter to SPD Estimation

Zdeněk Hlávka and Marek Svojik

The state price density (SPD) carries important information on the behavior and expectations of the market and it often serves as a base for option pricing and hedging. Many commonly used SPD estimation technique are based on the observation (Breeden and Litzenberger, 1978) that the SPD $f(.)$ may be expressed as

$$f(K) = \exp\{r(T-t)\}\frac{\partial^2 C_t(K,T)}{\partial K^2}, \tag{11.1}$$

where $C_t(K,T)$ is a price of European call option with strike price K at time t expiring at time T and r denotes the risk free interest rate. An overview of estimation techniques is given in Jackwerth (1999). Kernel smoothers were in this framework applied by Aït-Sahalia and Lo (1998), Aït-Sahalia and Lo (2000), or Huynh, Kervella, and Zheng (2002). Some modifications of the nonparametric smoother allowing to apply no-arbitrage constraints were proposed, e.g., by Aït-Sahalia and Duarte (2003), Bondarenko (2003), or Yatchew and Härdle (2006). Apart of the choice of a suitable estimation method, Härdle and Hlávka (2005) show that the covariance structure of the observed option prices carries additional important information that should to be considered in the estimation procedure. Härdle and Hlávka (2005) suggest a simple and easily applicable approximation of the covariance. A more detailed discussion of option price errors may be found in Renault (1997).

In this chapter, we will estimate the SPD from observed call option prices using the well-known Kalman filter, invented already in the early sixties and marked by Harvey (1989). Kalman filter may be shortly described as a statistical method used for estimation of the non-observable component of a state-space model and it already became an important econometric tool for financial and economic estimation problems in continuous time finance. More precisely, the Kalman filter is a recursive procedure for computing the optimal estimator of the state vector $\boldsymbol{\xi}_i$ at time i, based on information available at time i. For derivation of the Kalman filter, we focus on the general system

characterized by a *measurement equation*

$$\mathbf{y}_i = \mathbf{B}_i(\boldsymbol{\psi})\boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i(\boldsymbol{\psi}), \qquad i = 1, \ldots, n, \tag{11.2}$$

and a *transition equation*

$$\boldsymbol{\xi}_i = \boldsymbol{\Phi}_i(\boldsymbol{\psi})\boldsymbol{\xi}_{i-1} + \boldsymbol{\eta}_i(\boldsymbol{\psi}), \qquad i = 1, \ldots, n, \tag{11.3}$$

where $\mathbf{y}_i$ is the g-dimensional vector of the observable variables and $\boldsymbol{\xi}_i$ denotes the unobservable k-dimensional *state vector*, with unknown parameters $\boldsymbol{\psi}$, a known matrix $\mathbf{B}_i(\boldsymbol{\psi})$, and a noise term $\boldsymbol{\varepsilon}_i(\boldsymbol{\psi})$ of serially uncorrelated disturbances with zero mean and variance matrix $\mathbf{H}_i(\boldsymbol{\psi})$. The symbols used in the transition equation (11.3) are the *transition matrix* $\boldsymbol{\Phi}_i(\boldsymbol{\psi})$ and a zero mean Gaussian noise term $\boldsymbol{\eta}_i(\boldsymbol{\psi})$ with a known variance matrix $\mathbf{Q}_i(\boldsymbol{\psi})$. The specification of the state space model is completed by assuming independence between the error terms $\boldsymbol{\varepsilon}_t(\boldsymbol{\psi})$ and $\boldsymbol{\eta}_t(\boldsymbol{\psi})$. Additionally, we assume that these error terms are uncorrelated with the normally distributed *initial state vector* $\boldsymbol{\xi}_0$ having expected value $\boldsymbol{\xi}_{0|0}$ and variance matrix $\boldsymbol{\Sigma}_{0|0}$.

The state-space model (11.2)–(11.3) is suitable for situations in which, instead of being able to observe the state vector $\boldsymbol{\xi}_i$ directly, we can only observe some noisy function $\mathbf{y}_i$ of $\boldsymbol{\xi}_i$. The problem of determining the state of the system from noisy measurements $\mathbf{y}_i$ is called *estimation*. *Filtering* is a special case of estimation with the objective of obtaining an estimate of $\boldsymbol{\xi}_i$ given observations up to time i. It can be shown that the optimal estimator of $\boldsymbol{\xi}_i$, i.e., minimizing the mean squared error (MSE), is the mean of the conditional distribution of the state vector $\boldsymbol{\xi}_i$. When estimating $\boldsymbol{\xi}_i$ using information up to time s, we denote the conditional expectation of $\boldsymbol{\xi}_i$ given $\mathcal{F}_s$ for convenience by $\boldsymbol{\xi}_{i|s} = \mathsf{E}[\boldsymbol{\xi}_i|\mathcal{F}_s]$. The conditional variance matrix of $\boldsymbol{\xi}_i$ given $\mathcal{F}_s$ is denoted as $\boldsymbol{\Sigma}_{i|s} = \mathsf{Var}[\boldsymbol{\xi}_i|\mathcal{F}_s]$.

In our case, we will see that the relationship between the state and observed variables is nonlinear and the problem has to be linearized by Taylor expansion.

11.1 Linear Model

Let us remind that the payoff for a call option is given by

$$(S_T - K)_+ = \max(S_T - K, 0).$$

Let $\mathbf{C}_t(K, T)$ be the call pricing function of a European call option with strike price K observed at time t and expiring at time T. We consider a call option

with this payoff. Let S_T denotes the price of the underlying asset at T, and r the risk free interest rate. Then, the fair price $\mathbf{C}_t(K,T)$ of a European call option at the current time t may be expressed as the discounted expected value of the payoff $(S_T - K)_+$ with respect to the SPD $f(.)$, i.e.,

$$\mathbf{C}_t(K,T) = e^{-r(T-t)} \int_0^{+\infty} (S_T - K)_+ f(S_T) dS_T. \qquad (11.4)$$

Clearly, the call pricing function $\mathbf{C}_t(K,T)$ is monotone decreasing and convex in K.

In the rest of this chapter, we will assume that the discount factor $e^{-r(T-t)}$ in (11.4) is equal to 1. In practice, this may be easily achieved by dividing the observed option prices by this known discount factor.

In (11.1), we have already seen that the SPD may be expressed as the discounted second derivative of the call pricing function $\mathbf{C}_t(K,T)$ with respect to the strike price K. We will use this relationship to construct an SPD estimator based on the observed call option prices.

11.1.1 Linear Model for Call Option Prices

On a fixed day t, the i-th observed option price corresponding to the time of expiry T will be denoted as $\mathbf{C}_i = \mathbf{C}_{t,i}(\mathbf{K}_i, T)$, where K_i denotes the corresponding strike price. The vector of all observed option prices will be denoted as $\boldsymbol{\mathcal{C}} = (\mathbf{C}_1, \ldots, \mathbf{C}_n)^\top$. Without loss of generality, we assume that the corresponding vector of the strike prices $\mathcal{K} = (K_1, \ldots, K_n)^\top$ has the following structure

$$\mathcal{K} = \begin{pmatrix} \mathbf{K}_1 \\ \mathbf{K}_2 \\ \vdots \\ \mathbf{K}_n \end{pmatrix} = \begin{pmatrix} k_1 1_{n_1} \\ k_2 1_{n_2} \\ \vdots \\ k_n 1_{n_p} \end{pmatrix},$$

where $k_1 < k_2 < \cdots < k_p$ are the p distinct values of the strike prices, 1_{n_j} denotes a vector of ones of length n_j, and $n_j = \sum_{i=1}^{n} \mathbf{1}(\mathbf{K}_i = k_j)$.

The further assumptions and constraints that have to be satisfied by the developing the linear model are largely taken from Härdle and Hlávka (2005). We impose only constraints that guarantee that the estimated function is probability density, i.e., it is positive and it integrates to one. The SPD is parameterized by assuming that for a fixed day t and time to maturity $\tau = T - t$, the i-th observed option price $\mathbf{C}_i$ corresponding to strike price $\mathbf{K}_i$, the option prices $\mathbf{C}_i = \mathbf{C}_{t,i}(\mathbf{K}_i, T)$ follows the linear model

$$\mathbf{C}_{t,i}(\mathbf{K}_i, T) = \mu(\mathbf{K}_i) + \epsilon_i, \qquad (11.5)$$

where $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)^\top \sim N(\mathbf{0}, \Sigma)$ is random vector of correlated normally distributed random errors.

In the next section, we will parameterize the vector of the mean option prices $\mu(.)$ in terms of the state price density. This parameterization will allow us to derive SPD estimators directly from the linear model (11.5).

11.1.2 Estimation of State Price Density

In Härdle and Hlávka (2005), it was suggested to rewrite the vector of the conditional means $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_p)^\top$ in terms of the parameters $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{p-1})^\top$ as

$$\boldsymbol{\mu} = \boldsymbol{\Delta\beta}, \tag{11.6}$$

where

$$\boldsymbol{\Delta} = \begin{pmatrix} 1 & \Delta_p^1 & \Delta_{p-1}^1 & \Delta_{p-2}^1 & \cdots & \Delta_3^1 & \Delta_2^1 \\ 1 & \Delta_p^2 & \Delta_{p-1}^2 & \Delta_{p-2}^2 & \cdots & \Delta_3^2 & 0 \\ \vdots & & & & & & \vdots \\ 1 & \Delta_p^{p-2} & \Delta_{p-1}^{p-2} & 0 & \cdots & 0 & 0 \\ 1 & \Delta_p^{p-1} & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \tag{11.7}$$

and $\Delta_j^i = \max(k_j - k_i, 0)$ denotes the positive part of the distance between k_i and k_j, i.e. the i-th and the j-th ($1 \leq i \leq j \leq p$) sorted distinct observed values of the strike price.

The vector of parameters β in (11.6) may be interpreted as an estimate of the second derivatives of the call pricing function and consequently, according to (11.1), also as the estimator of the state price density.

The constraints on the conditional means μ_j such as positivity, monotonicity and convexity can be reexpressed in terms of β_j—it suffices to request that $\beta_j > 0$ for $j = 0, \ldots, p-1$ and that $\sum_{j=2}^{p-1} \beta_j \leq 1$.

Using this notation, the linear model for the observed option prices $\mathbf{C}$ is obtained by

$$\mathbf{C}(\mathcal{K}) = \boldsymbol{\mathcal{X}_\Delta \beta} + \boldsymbol{\epsilon}, \tag{11.8}$$

where $\boldsymbol{\mathcal{X}_\Delta}$ is the design matrix obtained by repeating each row of the matrix $\boldsymbol{\Delta}$ n_i-times for $i = 1, \ldots, p$.

11.1.3 State-Space Model for Call Option Prices

In order to apply Kalman filter without any constraints on the resulting SPD estimates $(\beta_i, \; i = 0, \ldots, p-1)$, we rewrite the linear model (11.8) in a state-space form for the i-th observation on a fixed day t:

$$\mathbf{C}_i(\mathcal{K}) \;=\; \boldsymbol{\mathcal{X}_\Delta}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \tag{11.9}$$
$$\boldsymbol{\beta}_i \;=\; \boldsymbol{\beta}_{i-1} + \boldsymbol{\eta}_i, \tag{11.10}$$

where $\boldsymbol{\mathcal{X}_\Delta}$ is the design matrix from (11.8) and $\boldsymbol{\varepsilon}_i \sim \mathrm{N}\left(\mathbf{0}, \sigma^2\mathbf{I}\right)$ and $\boldsymbol{\eta}_i \sim \mathrm{N}\left(\mathbf{0}, \nu^2\delta_i\mathbf{I}\right)$ are uncorrelated random vectors. We assume that the variance of η_i depends linearly on the time δ_i between the i-th and the $(i-1)$-st trade.

In the following, we determine the Kalman filter in a standard way. The standard approach has to be only slightly modified as in every step i we observe only option price $C_i(K_i)$ corresponding to only one strike price K_i.

Prediction step   In the prediction step, we forecast the state vector by calculating the conditional moments of the state variables given the information up to time $t-1$ to obtain the *prediction equations*

$$\boldsymbol{\beta}_{i|i-1} \;=\; \mathsf{E}(\boldsymbol{\beta}_i|\mathcal{F}_{i-1}) = \boldsymbol{\beta}_{i-1|i-1}, \tag{11.11}$$
$$\boldsymbol{\Sigma}_{i|i-1} \;=\; \boldsymbol{\Sigma}_{i-1|i-1} + \nu^2\delta_i\mathbf{I}. \tag{11.12}$$

Updating step   Denoting by $\boldsymbol{\Delta}_i$ the i-th row of the design matrix $\boldsymbol{\mathcal{X}_\Delta}$, i.e., the row corresponding to the i-th observed strike price K_i, we arrive to the *updating equations*

$$\boldsymbol{\beta}_{i|i} \;=\; \boldsymbol{\beta}_{i|i-1} + \mathbf{K}_i I_i, \tag{11.13}$$
$$\boldsymbol{\Sigma}_{i|i} \;=\; \left(\mathbf{I} - \mathbf{K}_i\boldsymbol{\Delta}_i\right)\boldsymbol{\Sigma}_{i|i-1}, \tag{11.14}$$

where

$$I_i = C_i(K_i) - C_{i|i-1}(K_i) = C_i(K_i) - \boldsymbol{\Delta}_i\boldsymbol{\beta}_{i|i-1}$$

is the *prediction error* with variance $F_{i|i-1} = \mathsf{Var}(I_i|\mathcal{F}_{i-1}) = \sigma^2 + \boldsymbol{\Delta}_i\boldsymbol{\Sigma}_{i|i-1}\boldsymbol{\Delta}_i^\top$ and $\mathbf{K}_i = \boldsymbol{\Sigma}_{i|i-1}\boldsymbol{\Delta}_i^\top F_{i|i-1}^{-1}$ is the *Kalman gain*.

The prediction and updating equations (11.11)–(11.14) jointly constitute the linear Kalman filter for a European call option. Unfortunately, in this case the practical usefulness of the linear Kalman filter is limited as the resulting SPD estimator does not have to be probability density. A more realistic nonlinear model is presented in the following Section 11.2.

11.2 Extended Kalman Filter and Call Options

In the following, we constrain the vector of parameters $\boldsymbol{\beta}_i = (\beta_0, \ldots, \beta_{p-1})^\top$ so that it may always be reasonably interpreted as a probability density. We propose a reparameterization of the model in terms of parameters $\boldsymbol{\xi}_i = (\xi_0, \ldots, \xi_{p-1})^\top$ via a smooth function $\mathbf{g}_i(\cdot) = \big(g_0(\cdot), \ldots, g_{p-1}(\cdot)\big)^\top$ by setting

$$\beta_0 = g_0(\boldsymbol{\xi}_i) = \exp(\xi_0), \tag{11.15}$$

$$\beta_k = g_k(\boldsymbol{\xi}_i) = S^{-1}\exp(\xi_k), \quad \text{for } k = 1, \ldots, p-1, \tag{11.16}$$

where $S = \sum_{j=1}^{p-1} \exp(\xi_j)$ simplifies the notation.

Obviously, $\sum_{j=1}^{p-1} \beta_j = 1$ and $\beta_j > 0$, $j = 0, \ldots, p-1$. This means that the parameters β_j, $j = 1, \ldots, p-1$ are positive and integrate to one and may be interpreted as a reasonable estimates of the values of the SPD.

The linear model for the option prices (11.8) rewritten in terms of $\boldsymbol{\xi}_i$ leads a nonlinear state space model given by the measurement equation

$$\mathbf{C}_i(\boldsymbol{\mathcal{K}}) = \boldsymbol{\mathcal{X}}_{\boldsymbol{\Delta}} \mathbf{g}_i(\boldsymbol{\xi}_i) + \boldsymbol{\varepsilon}_i, \tag{11.17}$$

and the transition equation

$$\boldsymbol{\xi}_i = \boldsymbol{\xi}_{i-1} + \boldsymbol{\eta}_i, \tag{11.18}$$

where $\boldsymbol{\varepsilon}_i \sim \mathrm{N}\big(\mathbf{0}, \sigma^2\mathbf{I}\big)$ and $\boldsymbol{\eta}_i \sim \mathrm{N}\big(\mathbf{0}, \nu^2\delta_i\mathbf{I}\big)$ satisfy the same assumptions as in Section 11.1.3.

The extended Kalman filter for the above problem may be linearized by Taylor expansion using the Jacobian matrix $\mathbf{B}_{i|i-1}$ computed in $\boldsymbol{\xi}_i = \boldsymbol{\xi}_{i|i-1}$:

$$\mathbf{B}_{i|i-1} = \left.\frac{\partial \mathbf{g}_i(\boldsymbol{\xi}_i)}{\partial \boldsymbol{\xi}_i^\top}\right|_{\boldsymbol{\xi}_i = \boldsymbol{\xi}_{i|i-1}} \tag{11.19}$$

$$= \frac{1}{S^2}\begin{pmatrix} S^2 e^{\xi_0} & 0 & 0 & \cdots & 0 \\ 0 & e^{\xi_1}\big(S - e^{\xi_1}\big) & -e^{\xi_1+\xi_2} & \cdots & -e^{\xi_1+\xi_{p-1}} \\ 0 & -e^{\xi_2+\xi_1} & e^{\xi_2}\big(S - e^{\xi_2}\big) & \cdots & -e^{\xi_2+\xi_{p-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & -e^{\xi_{p-1}+\xi_1} & -e^{\xi_{p-1}+\xi_2} & \cdots & e^{\xi_{p-1}}\big(S - e^{\xi_{p-1}}\big) \end{pmatrix}.$$

Now, the linearized version of the Kalman filter algorithm for model (11.17)–(11.18) is straightforward. Similarly as in Section 11.1.3, we obtain *extended prediction equations*

$$\boldsymbol{\xi}_{i|i-1} = \boldsymbol{\xi}_{i-1|i-1}, \tag{11.20}$$

$$\boldsymbol{\Sigma}_{i|i-1} = \boldsymbol{\Sigma}_{i-1|i-1} + \nu^2 \delta_i \mathbf{I}, \tag{11.21}$$

and *extended updating equations*

$$\boldsymbol{\xi}_{i|i} = \boldsymbol{\xi}_{i|i-1} + \mathbf{K}_i I_i, \tag{11.22}$$

$$\boldsymbol{\Sigma}_{i|i} = \big(\mathbf{I} - \mathbf{K}_i \boldsymbol{\Delta}_i \mathbf{B}_{i|i-1}\big) \boldsymbol{\Sigma}_{i|i-1}, \tag{11.23}$$

where $I_i = C_i(K_i) - \boldsymbol{\Delta}_i \mathbf{g}_i(\boldsymbol{\xi}_{i|i-1})$ is the prediction error, $F_{i|i-1} = \mathsf{Var}(I_i | \mathcal{F}_{i-1}) = \sigma^2 + \boldsymbol{\Delta}_i \mathbf{B}_{i|i-1} \boldsymbol{\Sigma}_{i|i-1} \mathbf{B}_{i|i-1}^\top \boldsymbol{\Delta}_i^\top$ its variance, and $\mathbf{K}_i = \boldsymbol{\Sigma}_{i|i-1} \mathbf{B}_{i|i-1}^\top \boldsymbol{\Delta}_i^\top F_{i|i-1}^{-1}$ the Kalman gain.

The recursive equations (11.20)–(11.23) form the extended Kalman filter recursions and lead the vector $\mathbf{g}_i(\boldsymbol{\xi}_i) = \boldsymbol{\beta}_i$ representing estimates of the SPD.

11.3 Empirical Results

In this section, the extended Kalman filter is used to estimate SPD from DAX call option prices. In other words, our objective is to estimate the call function $\mathbf{C}_t(K, T)$ subject to monotonicity and convexity constraints, i.e., the constraint that the implied SPD is non-negative and it integrates to one.

We choose data over a sufficiently brief time span so that the time to maturity τ, the interest rate r, and both the current time t and the time of expiry T may be considered as constant. The full data set contains observed call and put option prices for various strike prices and maturities τ. From now on, for each trading day, we consider only a subset containing the call options $\mathbf{C}_{t,i}(K_i, T)$, $i = 1, \ldots, n$ with the shortest time to expiry $\tau = T - t$. In 1995, we have few hundreds such observations each day. In 2003, the number of daily observations increases to thousands.

Apart of the strike prices K_i and option prices $\mathbf{C}_{t,i}(K_i, T)$, the data set contains also information on the risk-free interest rate r, the time of trade (given in seconds after midnight), the current value of the underlying asset (DAX), time to expiry, and type of the option (Call/Put).

As the risk-free interest rate r and the time to expiry $T - t$ are known and given in our data set, we may work with option prices corrected by the

known discount factor $e^{-r(T-t)}$. This modification guarantees that the second derivative of the discounted call pricing function is equal to the state price density.

11.3.1 Extended Kalman Filtering in Practice

In order to implement the Kalman filter in practice, we need to:

1. set the *initial values* of unknown parameters,
2. *estimate the unknown parameters* from data.

Initialization In order to use the extended version Kalman filter, we have to choose initial values $\mathbf{\Sigma}_{0|0}$ and $\boldsymbol{\beta}_{0|0}$ and variances of both error terms $\boldsymbol{\varepsilon}_i$ and $\boldsymbol{\eta}_i$. We choose initial $\mathbf{\Sigma}_{0|0} = \mathbf{I}$ and

$$
\boldsymbol{\beta}_{0|0} = \left(\underbrace{\widehat{\mathsf{E}}\{C(k_p)\}}_{\beta_0}, \underbrace{\frac{1}{p-1}, \ldots, \frac{1}{p-1}}_{p-1} \right),
$$

i.e., β_0 is set as the sample mean of option prices corresponding to the largest strike price k_p. The remaining values, defining the initial distribution of the SPD, are set uniformly.

The parameter σ^2 may be interpreted as the standard error of the option price in Euros. The interpretation of the parameter ν^2 is more difficult and it depends on the time intervals between consecutive trades and on the range of the observed strike prices. For the first run of the algorithm, we set the variance matrices as $\mathsf{Var}[\boldsymbol{\varepsilon}_i] = \sigma^2 \mathbf{I}$ and $\mathsf{Var}[\boldsymbol{\eta}_i] = \nu^2 \delta_i \mathbf{I}$, with $\sigma^2 = 1$ and

$$
\nu^2 = 1/[\{(\max_{i=1,\ldots,n} K_i - \min_{i=1,\ldots,n} K_i)/2\}^2 \min_{i=1,\ldots,n} \delta_i].
$$

This choice is quite arbitrary but it reflects that the parameter σ^2 should be small (in Euros) and that the parameter ν^2 is related to the time and to the range of the observed strike prices. Note that these are only initial values and more realistic estimates are obtained in the next iterations of the extended Kalman filter.

Extended Kalman filter Given the starting values $\boldsymbol{\beta}_{0|0}$, $\mathbf{\Sigma}_{0|0}$, σ^2, and ν^2, the extended Kalman filter is given by equations (11.20)–(11.23). The non-linear projections $\mathbf{g}_i(.)$ guarantee that the state vector $\boldsymbol{\beta}_i = \mathbf{g}_t(\boldsymbol{\xi}_i)$ satisfies the required constraints.

Parameter estimation   The unknown parameters σ^2 and ν^2 are estimated by Maximum Likelihood (ML) method. More precisely, we use the *prediction error decomposition of the likelihood function* described in Kellerhals (2001, Chapter 5), the resulting log-likelihood is then maximized numerically. Note that another approach to parameter estimation based on the Kalman smoother and EM-algorithm is described in Harvey (1989, Section 4.2.4).

The behavior of the extended Kalman filter depends also on the choice of the starting value $\boldsymbol{\beta}_{0|0}$. Assuming that the shape of the SPD doesn't change too much during the day, we may improve on the initial "uniform" SPD by taking $\boldsymbol{\beta}_{n|n}$, shift the corresponding SPD by the difference of the value of the underlying asset, and by using the resulting set of parameters as the starting value $\boldsymbol{\beta}_{0|0}$. In practice, one might use the final estimator $\beta_{n|n}$ from day t as the initial estimator on the next day $t+1$

Kalman filter iterations   Combining the initial parameter choice, the Kalman filter, and the parameter estimation, we obtain the following iterative algorithm:

1. Choose the initial values.
2. Run the extended Kalman filter (11.20)–(11.23) with current values of the parameters.
3. Use the Kalman filter predictions to estimate the parameters σ^2 and ν^2 by numerical maximization of the log-likelihood and update the initial values $\boldsymbol{\beta}_{0|0}$ and $\boldsymbol{\Sigma}_{0|0}$ using $\boldsymbol{\beta}_{n|n}$ and $\boldsymbol{\Sigma}_{n|n}$.
4. Either stop the algorithm or return to step 2 depending on the chosen stopping rule.

In practice, the stopping rule for the above iterative algorithm may be based on the values of the log-likelihood obtained in step 3 of the iterative algorithm. In the following real life examples, we will run fixed number of iterations as an illustration.

11.3.2 SPD Estimation in 1995

The first example is using data from two trading days in 1995; these are the two data sets as in Härdle and Hlávka (2005). The call option prices observed on 11th (January 14th) and 12th trading day (January 15th) in 1995 are plotted on the left-hand side graphics in Figures 11.1 and 11.2. The main difference between these two trading days is that the strike prices traded on 15th January cover larger range of strike prices. This means that
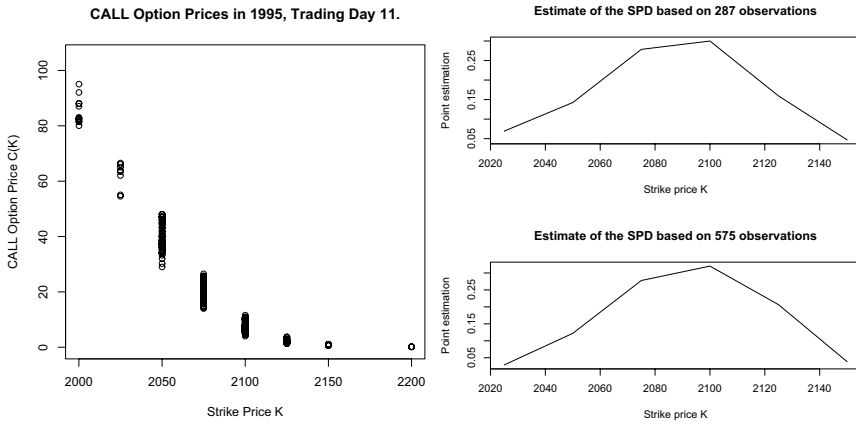
Figure 11.1. European call option prices with shortest time to expiry plotted against strike price K (left) and two of the filtered SPD estimates (right) on JAN-14-1995. **Q** XFGKF1995a

also the support of the estimated SPDs will be larger on January 15th than on January 14th.

JAN-14-1995 Using the data from January 14th, 1995, we ran 10 iterations of the algorithm described in Section 11.3.1. The resulting parameter estimates, $\widehat{\sigma}^2 = 0.0111$ and $\widehat{\nu}^2 = 2.6399$, seem to be stable. In the last four iterations, estimates of σ^2 vary between 0.0081 and 0.0115 and estimates of ν^2 are varying between 2.4849 and 2.6399.

The Kalman filter provides SPD estimate in each time $i = 1, \ldots, n$ and we thus obtain altogether $n = 575$ estimates of $\boldsymbol{\beta}_i$ during this one day. Two of these filtered SPD estimates on JAN-14-1995 are displayed on the on the right-hand side of Figure 11.1; the upper plot shows the estimator at time $i_1 = 287 \doteq n/2$ (12:45:44.46) and the lower plot the estimator at time $i_2 = n = 575$ (15:59:59.78), i.e., The lower plot contains the estimator of the SPD at the end of this trading day.

Both estimates look very similar but the latter one is shifted a bit towards higher values. This shift is clearly due to a change in the value of the underlying asset (DAX) from 2087.691 to 2090.17 during the same time period.

JAN-15-1995 Next, the same technique is applied to data observed on January 15th, 1995, see Figure 11.2. Two of the resulting filtered SPD estimates
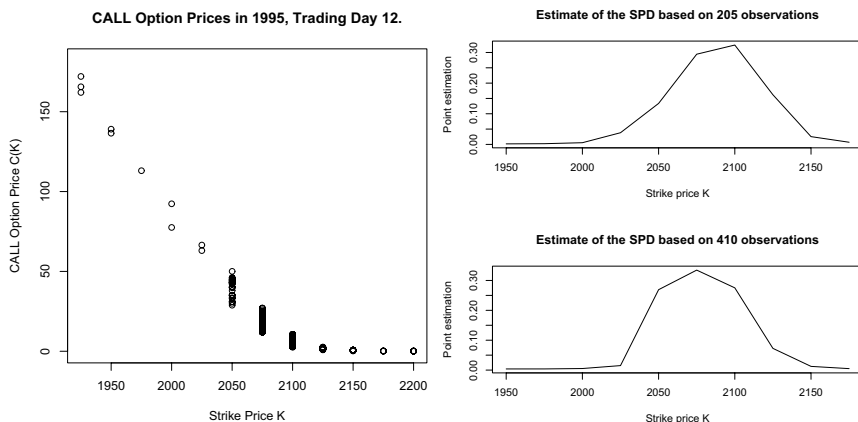
Figure 11.2. European call option prices with the shortest time to expiry plotted against strike price K (left) and two of the filtered SPD estimates (right) on JAN-15-1995.
 XFGKF1995b

are plotted in the graphics on the right-hand side of Figure 11.2. The SPD estimator calculated at the time $i_1 = n/2 = 205$ (12:09:14.65) is almost identical to the final estimate from January 14th; the most visible difference is the larger support for the estimated SPD on JAN-15-1995. At the end of this trading day, for $i_2 = n = 410$ (15:59:52.14), the estimate is shifted a bit to the left and more concentrated. The shift to the left corresponds again to a decrease in the value of the DAX from 2089.377 to 2075.989.

The parameter estimates obtained after 10 iterations of the algorithm described in Section 11.3.1 are $\widehat{\sigma}^2 = 0.0496$ and $\widehat{\nu}^2 = 0.621$. Smaller value of $\widehat{\nu}^2$ seemingly suggests that the SPD was changing more slowly on JAN-15-1995 but this parameter must be interpreted with a caution as its scale depends also on the size of the time interval between consecutive trades and on the range of the observed strike prices.

11.3.3 SPD Estimation in 2003

The next example is using the most recent data set in our database. On February 25th, 2003, we observe altogether 1464 call option prices with the shortest time to expiry. Compared to the situation in 1995, the option markets in 2003 are more liquid and the number of distinct strike prices included in the data set is larger than in 1995. Our data set contains 30 distinct strike

prices on FEB-25-2003 compared to 8 on JAN-14-1995 and 12 on JAN-15-1995.

The call option prices observed on FEB-25-2003 are plotted as a function of their strike price on the left-hand side plot in Figure 11.3.
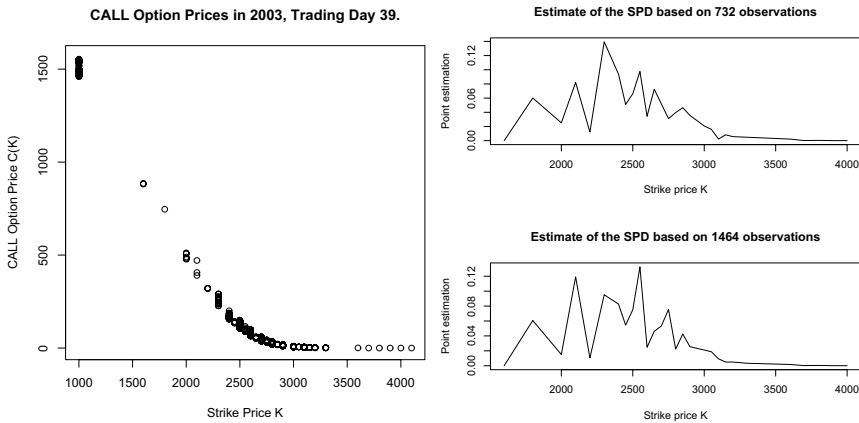

Figure 11.3. European call option prices with shortest time to expiry plotted against strike price K on FEB-25-2003, $n = 1464$ observed prices (left) and the resulting SPD estimates after 10 iterations (right). **Q** XFGKF2003

After ten iterations of the iterative extended Kalman filtering algorithm described in Section 11.3.1, we obtain parameter estimates $\widehat{\sigma}^2 = 0.0324$ and $\widehat{\nu}^2 = 3.1953$. The corresponding SPD estimates for times $i_1 = n/2 = 732$ and $i_2 = n = 1464$ are plotted on the right-hand side of Figure 11.3.

On FEB-25-2003, the resulting estimates do not look very much like a typical (smooth and unimodal) probability densities. Instead, we observe a lot of spikes and valleys. This is due to the fact that the algorithm does not penalize non-smoothness and the reparameterization (11.15)–(11.16) guarantees only that the resulting SPD estimates are positive and integrate to one.

In order to obtain more easily interpretable results, the resulting estimates may be smoothed using, e.g., the Nadaraya-Watson kernel regression estimator (Nadaraya, 1964; Watson, 1964). As the smoothing of the vector $\boldsymbol{\beta}_{n|n}$ corresponds to a multiplication with a (smoothing) matrix, say $\boldsymbol{S}$, the smoothing step may be implemented after the Kalman filtering, see Härdle (1991) or Simonoff (1996) for more details on kernel regression.

Using the variance matrix $\boldsymbol{\Sigma}_{n|n}$ from the filtering step of the extended Kalman

filtering algorithm, we calculate the variance matrix of $\boldsymbol{\xi}^{smooth}_{n|n} = \boldsymbol{S}\boldsymbol{\xi}_{n|n}$ as $\mathsf{Var}\boldsymbol{\xi}^{smooth}_{n|n} = \boldsymbol{S}\boldsymbol{\Sigma}_{n|n}\boldsymbol{S}^{\top}$. This leads an approximation of the asymptotic variance of the smooth SPD estimate $\boldsymbol{\beta}^{smooth}_{n|n} = \boldsymbol{g_n}(\boldsymbol{\xi}^{smooth}_{n|n})$ as $\mathsf{Var}\boldsymbol{\beta}^{smooth}_{n|n} = \boldsymbol{B_n}\boldsymbol{S}\boldsymbol{\Sigma}_{n|n}\boldsymbol{S}^{\top}\boldsymbol{B}_n^{\top}$, where $\boldsymbol{B_n}$ now denotes the Jacobian matrix (11.19) calculated in $\boldsymbol{\xi}^{smooth}_{n|n}$.

The resulting smooth SPD estimate at the end of the trading day (time n) $\boldsymbol{\beta}^{smooth}_{n|n}$ with pointwise asymptotic 95% confidence intervals obtained as $\boldsymbol{\beta}^{smooth}_{n|n} \pm 1.96\sqrt{\mathrm{diag}(\mathsf{Var}\boldsymbol{\beta}^{smooth}_{n|n})}$ is plotted in Figure 11.4.
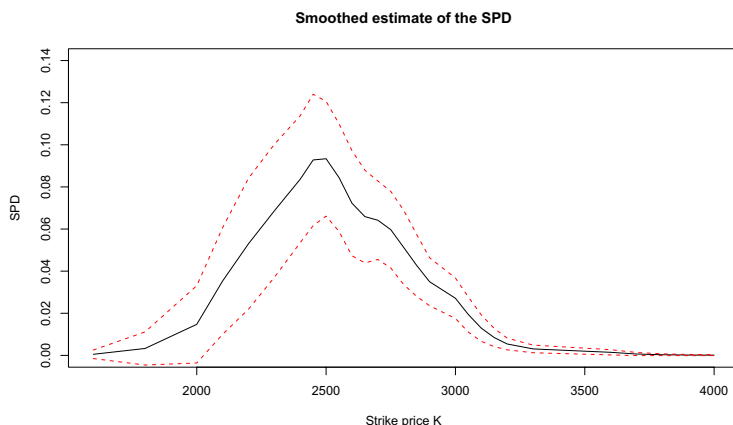
Figure 11.4. Smoothed SPD estimate on FEB-25-2003, $n = 1464$ with pointwise asymptotic confidence intervals.
XFGKF2003

11.4 Conclusions

We presented and illustrated the application of extended Kalman filtering towards arbitrage free SPD estimation.

An application of the extended Kalman filtering methodology on real-world data sets in Section 11.3 shows that this method provides very good results for data sets with small number of distinct strike prices, see Figures 11.1 and 11.2.

When the number of distinct strike price increases, the linear model becomes overparameterized, and the resulting SPD estimators are not smooth

anymore, see Figure 11.3. However, even in this case, the SPD estimator captures quite well the general shape of the SPD and smooth SPD estimator may be obtained by applying, e.g., the Nadaraya-Watson kernel regression estimator allowing also easy calculation of pointwise asymptotic confidence intervals.

Compared to other commonly used estimation techniques, the extended Kalman filtering methodology is able to capture the intra-day development of the SPD and it allows to update the estimates dynamically whenever new information becomes available. The extended Kalman filtering methodology combined with kernel smoothing is fast, easily applicable, and it provides interesting insights.

Bibliography

Aït-Sahalia, Y. and Duarte, J., 2003, Nonparametric option pricing under shape restrictions, *Journal of Econometrics* 116, 9–47.

Aït-Sahalia, Y. and Lo, A.W., 1998, Nonparametric estimation of state-price densities implicit in financial asset prices, *Journal of Finance* 53, 499–547.

Aït-Sahalia, Y. and Lo, A.W., 2000, Nonparametric risk management and implied risk aversion, *Journal of Econometrics* 94, 9–51.

Bondarenko, O., 2003, Estimation of risk-neutral densities using positive convolution approximation, *Journal of Econometrics* 116, 85–112.

Breeden, D. and Litzenberger, R., 1978, Prices of State-Contingent Claims Implicit in Option Prices, *Journal of Business* 51, 621–651.

Franke, J., Härdle, W. and Hafner, Ch., 2008, Statistics of Financial Markets. Springer, Berlin.

Härdle, W., 1991, Applied Nonparametric Regression. Cambridge University Press, Cambridge.

Härdle, W. and Hlávka, Z., 2005, Dynamics of State Price Densities. *Sonderforschungsbereich 649 Discussion Paper 2005-021, Humboldt-Universität zu Berlin.*

Harvey, A. C., 1989, Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Huynh, K., Kervella, P., and Zheng, J., 2002, Estimating state-price densities with nonparametric regression. In Härdle, Kleinow and Stahl, eds., Applied quantitative finance, Springer Verlag, Heidelberg 171–196.

Jackwerth, J.C., 1999, Option-implied risk-neutral distributions and implied binomial trees: a literature review, *Journal of Derivatives* 7, 66–82.

Kellerhals, B. P., 2001, Financial Pricing Models in Continuous Time and Kalman Filtering. Springer, Heidelberg.

Nadaraya, E. A., 1964, On estimating regression, *Theory of Probability and its Applications* 9 (1), 141–142.

Renault, E., 1997, Econometric models of option pricing errors. In Kreps and Wallis, eds., Advances in Economics and Econometrics: Theory and Applications, Seventh World Congress, Volume III, Cambridge University Press, Cambridge, 223–278.

Simonoff, J. S., 1996, Smoothing Methods in Statistics. Springer, New York.

Svoj´ik, M., 2007, Application of Kalman Filtering, Diploma Thesis, Charles University in Prague, Faculty of Mathematics and Physics.

Watson, G. S., 1964, Smooth regression analysis, *Sankhya, Ser. A* 26, 359–372.

Yatchew, A. and Härdle, W., 2006, Nonparametric state price density estimation using constrained least squares and the bootstrap, *Journal of Econometrics* 133(2), 579-599.

12 Stochastic Volatility Estimation Using Markov Chain Simulation

Nikolaus Hautsch and Yangguoyi Ou

Stochastic volatility (SV) models are workhorses for the modelling and prediction of time-varying volatility on financial markets and are essential tools in risk management, asset pricing and asset allocation. In financial mathematics and financial economics, stochastic volatility is typically modeled in a continuous-time setting which is advantageous for derivative pricing and portfolio optimization. Nevertheless, since data is typically only observable at discrete points in time, in empirical applications, discrete-time formulations of SV models are equally important.

SV models can be economically motivated by the mixture-of-distribution hypothesis (MDH) postulated by Clark (1973), whereby asset returns follow a mixture of normal distributions with a mixing process depending on the (unobservable) information arrival process. If the mixing process is positively autocorrelated, the resulting return process reveals volatility clustering which is a well-known and typical feature of financial return series. The MDH gives rise to the idea that asset return volatility follows its own stochastic process which is updated by unobservable innovations. This is in contrast to an autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982), where the conditional variance given the available information set is a function of past observations. Denote h_t as the time-t conditional variance of asset return y_t with conditional mean μ_t and $y_t - \mu_t = h_t^{1/2} z_t$, $z_t \sim IID(0,1)$, and let $\mathcal{F}_t$ denote the time-t information set. Then, ARCH processes imply $\mathsf{Var}[h_t | \mathcal{F}_{t-1}] = 0$, i.e., the variance is conditionally deterministic given the (observable) history of the process. Conversely, SV models can be characterized by the property $\mathsf{Var}[h_t | \mathcal{F}_{t-1}] \neq 0$, i.e., there is an unpredictable component in h_t.

A main difficulty of the SV framework compared to the widely used (Generalized) ARCH model is that the likelihood of SV models is not directly available. This requires the use of simulation techniques, like simulated maximum likelihood, method of simulated moments or Markov chain Monte Carlo (MCMC) techniques. Because of the computational costs, SV models are

still less popular in financial practice. Nevertheless, increasing computer power and the further development of efficient sampling techniques weaken this drawback noticeably. Furthermore, recent literature on the estimation of realized volatility confirms the idea of the MDH that log returns follow a normal - log normal mixture (see, e.g., Andersen, Bollerslev, Diebold and Labys (2003)) and thus strengthens the economic foundation of the SV model. Finally, SV models provide a natural framework to accommodate specific properties of financial return processes such as fat-tailedness, leverage effects and the occurrence of jumps.

The main objective of this chapter is to present the most important specifications of discrete-time SV models, to illustrate the major principles of Markov Chain Monte Carlo (MCMC) based statistical inference, and to show how to implement these techniques to estimate SV models. In this context, we provide a hands-on approach which is easily extended in various directions. Moreover, we will illustrate empirical results based on different SV specifications using returns on stock indices and foreign exchange rates.

In Section 12.1, we will introduce the standard SV model. Section 12.2 presents several extended SV models. MCMC based Bayesian inference is discussed in Section 12.3, whereas empirical illustrations are given in Section 12.4.

12.1 The Standard Stochastic Volatility Model

The standard stochastic volatility model as introduced by Taylor (1982) is given by

$$
\begin{aligned}
y_t &= \exp(h_t/2)u_t, & u_t &\sim \mathrm{N}(0,1), & &\text{(12.1a)} \\
h_t &= \mu + \phi(h_{t-1} - \mu) + \eta_t, & \eta_t &\sim \mathrm{N}(0,\sigma_\eta^2), & &\text{(12.1b)}
\end{aligned}
$$

where y_t denotes the log return at time t, $t = 1,\ldots,T$, and h_t is the log volatility which is assumed to follow a stationary AR(1) process with persistence parameter $|\phi| < 1$. The error terms u_t and η_t are Gaussian white noise sequences. The unconditional distribution of h_t is given by

$$
h_t \sim \mathrm{N}\left(\mu_h, \sigma_h^2\right), \quad \mu_h = \mu, \quad \sigma_h^2 = \frac{\sigma_\eta^2}{1 - \phi^2}, \qquad \text{(12.2)}
$$

where μ_h and σ_h^2 denote the unconditional mean and variance of returns, respectively.

Under the assumption that $\mathsf{E}[y_t^4] < \infty$, the first two even moments of y_t are given by

$$\mathsf{E}[y_t^2] = \mathsf{E}[\exp(h_t)]\,\mathsf{E}[u_t^2] \quad = \exp(\mu_h + \sigma_h^2/2), \tag{12.3}$$
$$\mathsf{E}[y_t^4] = \mathsf{E}[\exp(2h_t)]\,\mathsf{E}[u_t^4] = 3\exp(2\mu_h + 2\sigma_h^2). \tag{12.4}$$

Consequently, the kurtosis is

$$\mathrm{K}(y_t) \overset{\mathrm{def}}{=} \frac{\mathsf{E}[y_t^4]}{\mathsf{E}[y_t^2]^2} = 3\exp(\sigma_h^2) = 3\exp\left(\frac{\sigma_\eta^2}{1 - \phi^2}\right) \tag{12.5}$$

with $\mathrm{K}(y_t) > 3$ as long as $\sigma_\eta^2 > 0$. Hence, the kurtosis generated by SV processes increases with σ_η^2 and $|\phi|$ (given $|\phi| < 1$).

The autocorrelation function (ACF) of y_t^2 is computed as

$$\mathrm{Corr}(y_t^2, y_{t-\tau}^2) = \frac{\exp(\sigma_h^2 \phi^\tau) - 1}{3\exp(\sigma_h^2) - 1}, \quad \tau = 1, 2, \ldots, \tag{12.6}$$

and thus decays exponentially in τ. Consequently, for $\phi \in (0, 1)$, squared returns are positively autocorrelated.

The estimation of SV models is not straightforward since the likelihood cannot be computed in closed form. Let θ denote the collection of all model parameters, e.g., $\theta = (\mu, \phi, \sigma_\eta^2)$ for the standard SV model. Then, the likelihood function is defined by

$$p(y|\theta) \overset{\mathrm{def}}{=} \int_h p(y|h, \theta) p(h|\theta) dh, \tag{12.7}$$

where $y = (y_1, \ldots, y_T)$ and $h = (h_1, \ldots, h_T)$ are the vectors of returns and latent volatility states, respectively. The so-called full-information likelihood, corresponding to the conditional probability density function (p.d.f.), $p(y|h, \theta)$, is specified by (12.1a), whereas the conditional p.d.f. of the volatility states, $p(h|\theta)$, is given by (12.1b). The likelihood function (12.7) is an analytically intractable T-dimensional integral with respect to the unknown latent volatilities. In the econometric literature, several estimation methods have been proposed, including generalized method of moments (Melino and Turnbull, 1990), quasi-maximum likelihood estimation (Harvey, Ruiz, and Shephard, 1994), efficient method of moments (Gallant, Hsie, and Tauchen, 1997), simulated maximum likelihood (Danielsson, 1994) and efficient importance sampling (Liesenfeld and Richard, 2003). Markov Chain Monte Carlo (MCMC) techniques have been introduced by Jacquier, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1998). More details on MCMC-based inference will be given in Section 12.3.

12.2 Extended SV Models

12.2.1 Fat Tails and Jumps

Though the standard SV model is able to capture volatility clustering typically exhibited by financial and economic time series, the model implied kurtosis is often far too small to match the sample kurtosis observed in most financial return series. See, for example, Liesenfeld and Jung (2000) and Chib, Nardari, and Shephard (2002). An obvious reason is that a normal - log normal mixture as implied by the standard SV model is not flexible enough to capture the fat-tailedness commonly observed in financial return distributions. A further reason is that the basic SV model cannot account for potential jumps in the return process.

In this section, we discuss two SV specifications taking into account both pitfalls. The first one is an extension of the standard SV model allowing the error term u_t to be Student-t distributed resulting in the so-called SVt model. In the second approach, a jump component is introduced in the measurement equation in (12.1). This will lead to the so-called SVJ model.

The SVt Model

The SVt model is specified by

$$y_t = \exp(h_t/2)u_t, \qquad\qquad u_t \sim t_v, \qquad\qquad (12.8a)$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \qquad (12.8b)$$

where u_t follows a standardized t-distribution with $v > 2$ degrees of freedom. The model can be alternatively represented by a scale mixture of normal distributions. Let λ_t denote an i.i.d. random variable following an inverse-gamma distribution. Then, the SVt model can be rewritten as

$$y_t = \exp(h_t/2)\sqrt{\lambda_t}u_t, \qquad u_t \sim \mathrm{N}(0,1), \qquad (12.9a)$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \qquad (12.9b)$$
$$\lambda_t \sim \text{Inv-Gamma}(v/2, v/2), \qquad v > 2, \qquad (12.9c)$$

where λ_t itself is a latent variable. The representation of the SVt model in terms of a scale mixture is particularly useful in an MCMC context since it converts a non-log-concave sampling problem into a log-concave one. This allows for sampling algorithms which guarantee convergence in finite time, see ,e.g., Frieze, Kannan and Polson (1994).

Allowing log returns to be Student-t distributed naturally changes the behavior of the stochastic volatility process. In the standard SV model, large values of $|y_t|$ induce large values of h_t. In contrast, with an additional source of flexibility, λ_t, the SVt model can caputure large values of $|y_t|$ without necessarily increasing h_t. A tpyical consequence is that SVt models imply a higher persistence in volatility dynamics than the standard SV model.

Employing simulated maximum likelihood methods Liesenfeld and Jung (2000) provide an estimate $\widehat{\nu} = 6.31$ for the USD/DM foreign exchange (FX) rate from 1980 to 1990, and a value of 6.30 for the USD/JPY FX rate over 5 years from 1981 to 1985. Chib et al. (2002) estimate the SVt model based on MCMC techniques and report an estimate $\widehat{\nu} = 12.53$ for daily S&P 500 returns between July 1962 and August 1997.

The SV Model with Jump Components

The question of to which extent asset return processes are driven by continuous and/or jump components is an ongoing topic in the current literature. Both (G)ARCH and standard SV models rest on the assumption of a continuous price process and thus are not able to accommodate jumps in returns. The latter is particularly important during periods of news arrivals when the market gets under stress and becomes less liquid. However, the SV framework allows for a natural inclusion of a jump component in the return process. This yields the SVJ model given by

$$y_t = k_t q_t + \exp(h_t/2)u_t, \qquad\qquad u_t \sim \mathrm{N}(0,1), \qquad\qquad (12.10)$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad\qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \qquad\qquad (12.11)$$
$$k_t \sim \mathrm{N}(\alpha_k, \beta_k), \qquad\qquad\qquad\qquad\qquad\qquad\qquad (12.12)$$
$$q_t \sim \mathbb{B}(\kappa), \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (12.13)$$

where q_t is a Bernoulli random variable taking on the value one whenever a jump occurs with probability κ, and is zero otherwise. The jump size is represented by the time-varying random variable k_t which is assumed to follow a normal distribution with mean α_k and variance β_k. Both q_t and k_t are latent variables. Then, the model is based on three latent components, h_t, q_t, and k_t.

As in the SVt model, the inclusion of a jump component influences the properties of the stochastic volatility process. Large values of $|y_t|$ are now attributed rather to the the jump component than to the volatility process. As in the SVt model this typically induces a higher persistence in the volatility process.

Eraker, Johannes, and Polson (2003) estimate the number of jumps in returns

to be approximately 1.5 per year for daily S&P 500 returns from 1980 to 1999, and 4.4 per year for NASDAQ 100 index returns from 1985 to 1999. Chib et al. (2002) estimate 0.92 jumps per year for daily S&P 500 returns covering a period from 1962 to 1997.

Similarly, jump components can be also included in the volatility process in order to capture instantaneous movements in volatility. Bates (2000) and Duffie, Pan, and Singleton (2000) provide evidence that both jumps in returns and volatilities are important to appropriately capture the dynamics in financial return processes. For S&P 500 returns from 1980 to 1999, Eraker et al. (2003) estimate 1.4 volatility jumps per year.

12.2.2 The Relationship Between Volatility and Returns

Studying the relation between expected stock returns and expected variance is a fundamental topic in financial economics. Though a positive relationship between expected returns and expected variances is consistent with the notion of rational risk-averse investors requiring higher expected returns as a risk premium during volatile market periods, it is not consistently supported by empirical research. Whereas French, Schwert, and Stambaugh (1987) and Campbell and Hentschel (1992) find positive relationships between expected risk premia and conditional volatility, several other studies find converse dependencies. In fact, there is evidence that unexpected returns and innovations to the volatility process are negatively correlated. This can be explained either by the volatility feedback theory by French et al. (1987), or by the well-known leverage effect discussed by Black (1976).

In this section, we will discuss two types of SV models allowing the return and volatility process to be correlated, namely the SV-in-Mean (SVM) model and the Asymmetric SV (ASV) model. While the SVM model includes the volatility component directly in the mean equation, the ASV model allows for mutual correlations between return and volatility innovations.

The SV-in-Mean Model

The SV-in-Mean (SVM) model is given by

$$y_t = d \cdot h_t + \exp(h_t/2)u_t, \qquad u_t \sim N(0,1), \qquad (12.14a)$$

$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2), \qquad (12.14b)$$

where the parameter d captures the relationship between returns and both expected as well as unexpected volatility components. This can be seen by

rewriting (12.14a) as

$$y_t = d \cdot h_{t|t-1} + d\left(h_t - h_{t|t-1}\right) + \exp(h_t/2)u_t, \qquad (12.15)$$

where $h_{t|t-1}$ denotes the expected volatility defined by the conditional variance at time t given the information available at time $t-1$. Accordingly, the term $(h_t - h_{t|t-1})$ gives the innovation to the volatility process.

French et al. (1987) regress monthly excess returns of U.S. stock portfolios on both expected and unexpected volatility components stemming from ARMA models based on daily data. Excluding the unexpected volatility component results in a weakly positive relationship between excess returns and volatility. In contrast, including both volatility components does not only result in a significantly negative impact of the volatility innovation but also reverses the sign of the ex ante relationship. Hence, the negative relationship between unexpected returns and innovations to the volatility process seems to dominate the weaker, presumably positive, relation between the expected components.

The Asymmetric SV Model

Empirical evidence for 'good' and 'bad' news having different effects on the future volatility is typically referred to as the leverage or asymmetric effect. According to the leverage effect, an unexpected drop in prices ('bad' news) increases the expected volatility more than an unexpected increase ('good' news) of similar magnitude. According to Black (1976) this is due to asymmetric effects of changes of the firm's financial leverage ratio. In SV models, leverage effects are captured by allowing the observation error u_t and the future process error η_{t+1} to be correlated. Then, the ASV model is specified by

$$y_t = \exp(h_t/2)u_t, \qquad (12.16a)$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad (12.16b)$$
$$\begin{pmatrix} u_t \\ \eta_{t+1} \end{pmatrix} \sim \mathrm{N}\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma_\eta \\ \rho\sigma_\eta & \sigma_\eta \end{pmatrix} \right\}, \qquad (12.16c)$$

where ρ denotes the correlation between u_t and η_{t+1}.

The ASV model has been extensively studied in the literature. Harvey and Shephard (1996) estimate the model using quasi-maximum likelihood providing $\widehat{\rho} = -0.66$ for daily U.S. stock returns ranging from 1962 to 1987. Based on the same data, Sandmann and Koopman (1998) and Jacquier, Polson, and Rossi (2004) estimate an ASV specification, where the *contemporaneous* return and volatility are correlated. Using simulated MLE methods and MCMC

based Bayesian inference, the two studies provide estimates of $\widehat{\rho} = -0.38$ and $\widehat{\rho} = -0.48$, respectively.

12.2.3 The Long Memory SV Model

In the previous sections, we have considered a first order autoregressive process for the log volatility h_t. This induces that the autocorrelations of h_t decay geometrically and volatility is said to exhibit short memory. However, empirical autocorrelations for absolute and squared returns typically decay more slowly and thus are not geometrically bounded. This implies so-called long range dependence or long memory effects. See, for example, Bollerslev and Mikkelsen (1996). One possibility to capture such effects is to allow for fractionally integrated processes, which have been developed and extensively studied over the last 25 years, see, e.g., Granger and Joyeux (1980), and Beran (1994), among others. Long memory SV models have been introduced by Breidt, Carto, and de Lima (1998), Harvey (1998), and Arteche (2004). Then, the log volatility process follows an ARFIMA(p, d, q) process given by

$$y_t = \exp(h_t/2)u_t, \qquad\qquad u_t \sim \mathrm{N}(0, 1), \qquad (12.17)$$

$$\phi(L)(1 - L)^d(h_t - \mu) = \theta(L)\eta_t, \qquad \eta_t \sim \mathrm{N}(0, \sigma_\eta^2), \qquad (12.18)$$

where d denotes the fractional differencing parameter and L denotes the lag operator with

$$\phi(L) = 1 - \sum_{i=1}^{p} \phi_i L^i, \quad \theta(L) = 1 + \sum_{i=1}^{q} \theta_i L^i, \qquad (12.19)$$

and the roots of the polynomials $\phi(\cdot)$ and $\theta(\cdot)$ lying strictly outside the unit circle. If $d \in (-0.5, 0.5)$, the volatility process reveals long memory and is weakly stationary. The fractional differencing operator $(1 - L)^d$ can be expressed in terms of the series expansion

$$(1 - L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(d+1)}{\Gamma(k+1)\Gamma(d-k+1)}(-1)^k L^k, \qquad (12.20)$$

with $\Gamma(\cdot)$ denoting the gamma function (see, e.g., Beran (1994)).

The autocorrelation of $\log h_t^2$ is derived, e.g., by Baillie (1996), Breidt et al. (1998), or Harvey (1998). It is asymptotically proportional to π^{2d-1}, as long as $d \in (-0.5, 0.5)$. Similar asymptotic results are applicable to $|y_t|$ and y_t^2.

Breidt et al. (1998) estimate the Fractionally Integrated SV (FISV) model by maximizing the spectral quasi-likelihood and obtain estimates of $d = 0.44$ and

$\phi = 0.93$ for daily returns of a value-weighted market portfolio of U.S. stocks between 1962 and 1989. Gallant et al. (1997) use efficient method of moments techniques to provide estimates of d ranging between 0.48 and 0.55 for a series of daily returns from the S&P composite price index ranging from 1928 to 1987. Brockwell (2005) develops an MCMC sampling algorithm for the estimation of the FISV model and provides $d = 0.42$ for daily ASD-USD FX rates between 1999 and 2004.

12.3 MCMC-Based Bayesian Inference

In this section, we will give a brief review of MCMC-based Bayesian inference and will illustrate its application to estimate the standard SV model. For an introduction to Bayesian econometrics, see, for example, Koop (2006) and Greenberg (2008).

12.3.1 Bayes' Theorem and the MCMC Algorithm

Let θ denote a vector of model parameters including all latent variables, and let y collect the observed data. By considering θ to be a random vector, its inference is based on the posterior distribution, $p(\theta|y)$, which can be represented by Bayes' theorem

$$p(\theta|y) \propto p(y|\theta)p(\theta), \tag{12.21}$$

where $p(y|\theta)$ denotes the likelihood function depending on the model parameters and the data y. Correspondingly, $p(\theta)$ defines the prior distribution reflecting subjective prior beliefs on the distribution of θ. Consequently, the posterior distribution $p(\theta|y)$ can be viewed as a combination of objective and subjective information. If the prior is noninformative, Bayesian inference for the parameter vector θ is equivalent to likelihood-based inference.

The principle of MCMC-based Bayesian inference is to simulate $p(\theta|y)$ based on a Markov chain of random draws stemming from a family of candidate-generating densities from which it is easy to sample. Let $x \in \mathbb{R}^d$ denote a random variable (in the given context it corresponds to θ) following a Markov chain with transition kernel $p(x, y)$ corresponding to the conditional density of y given x. The invariant distribution is given by $\pi^*(y) = \int_{\mathbb{R}^d} p(x, y)\pi^*(x)dx$. An important result in Markov chain theory is that if $p(x, y)$ satisfies the reversibility condition

$$f(x)p(x, y) = f(y)p(y, x), \tag{12.22}$$

- For $g = 1, \ldots, G$:
    1. Generate Y from $q(x^{(j)}, y)$ and U from $\mathbb{U}[0, 1]$.
    2. If $U \leq \alpha(x^{(j)}, Y) = \min\left\{\frac{f(Y)q(Y, x^{(j)})}{f(x^{(j)})q(x^{(j)}, Y)}, 1\right\}$
       Set $x^{(j+1)} = Y$.
    Else
       Set $x^{(j+1)} = x^{(j)}$.
    3. Return $\{x^{(1)}, x^{(2)}, \ldots, x^{(G)}\}$.

Figure 12.1. The Metropolis-Hasings Sampling Algorithm

then, $f(\cdot)$ is the invariant density for the kernel $p(\cdot)$, i.e., $f(\cdot) = \pi^*(\cdot)$.

An important MCMC technique is the Metropolis-Hastings (M-H) algorithm as developed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) and generalized by Hastings (1970). The major idea is to build on (12.22) and finding a reversible kernel whose invariant distribution equals the target distribution $f(\cdot)$. This is performed by starting with an irreversible kernel (proposal density) $q(y, x)$ for which $f(x)q(x, y) > f(y)q(y, x)$, i.e., loosely speaking, the process moves from x to y too often and from y to x too rarely. This can be corrected by introducing a probability $\alpha(x, y) < 1$ that the move is made. I.e., we choose $\alpha(x, y)$ such that

$$f(x)\alpha(x, y)q(x, y) = f(y)\alpha(y, x)q(y, x). \tag{12.23}$$

It is easily shown that this relationship is fulfilled for

$$\alpha(x, y) = \begin{cases} \min\left\{\frac{f(y)q(y, x)}{f(x)q(x, y)}, 1\right\}, & \text{if } f(x)q(x, y) \neq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{12.24}$$

This yields a transition kernel $q_{MH}(x, y)$ satisfying the reversibility condition and is defined by

$$q_{MH}(x, y) \stackrel{\text{def}}{=} q(x, y)\alpha(x, y), \quad x \neq y. \tag{12.25}$$

The resulting M-H sampling algorithm is summarized by Figure 12.1.

A crucial issue is an appropriate choice of the family of candidate-generating densities. Depending on the form and the complexity of the sampling problem, various techniques have been proposed in the literature. The probably most straightforward technique is proposed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) suggesting a random walk chain, where $q(x, y) = q_0(y - x)$, and $q_0(\cdot)$ is a multivariate density. Then, y is drawn from $y = x + z$ with z following q_0. If q_0 is symmetric around zero, we

have $q(x, y) = q(y, x)$ and thus $\alpha(x, y) = f(y)/f/(x)$. A further simple choice of candidate-generating densities is proposed by Hastings (1970) and is given by $q(x, y) = q_0(y)$, i.e., y is sampled independently from x resulting in an independence chain. Then, $\alpha(x, y) = f(y)/f(x) \cdot q(x)/q(y)$. A popular and more efficient method is the acceptance-rejection (A-R) M-H sampling method which is available whenever the target density is bounded by a density from which it is easy to sample. If the target density is fully bounded, the M-H algorithm is straightforwardly combined with an acceptance-rejection step. This principle will be illustrated in more detail in the next section in order to sample the latent volatility states h_t. A more sophisticated M-H A-R algorithm which does not need a blanketing function but only a *pseudo-dominating* density is proposed by Tierney (1994).

If the dimension of x is high, the M-H algorithm is facilitated by applying it to blocks of parameters. For instance, if the target density can be expressed in terms of two blocks of variables, i.e., $f(x_1, x_2)$, the M-H algorithm allows to sample from each block x_i given the other block x_j, $j \neq i$. Then, the probability for moving from x_1 to the candidate value Y_1 given x_2 is

$$\alpha(x_1, Y_1 | x_2) = \frac{f(Y_1, x_2) q_1(Y_1, x_1 | x_2)}{f(x_1, x_2) q_1(x_1, Y_1 | x_2)}. \tag{12.26}$$

If the kernel $q_1(x_1, Y_1 | x_2)$ is the conditional distribution $f(x_1 | x_2)$, then

$$\alpha(x_1, Y_1 | x_2) = \frac{f(Y_1, x_2) f(x_1 | x_2)}{f(x_1, x_2) f(Y_1 | x_2)} = 1 \tag{12.27}$$

since $f(Y_1 | x_2) = f(Y_1, x_2)/f(x_2)$ and $f(x_1 | x_2) = f(x_1, x_2)/f(x_2)$. If $f(x_1 | x_2)$ is available for direct sampling, the resulting algorithm is referred to as the *Gibbs sampler*, see (Geman and Geman, 1984).

Applying the M-H (or Gibbs) algorithm to sub-blocks of the vector x is a common proceeding in Bayesian statistics if the posterior distribution is of high dimension. This is particularly true for SV models where θ also includes the unobservable volatility states. In this context, the posterior distribution $p(\theta|y)$ is broken up into its complete conditional distributions $p(\theta_i | \theta_{-i}, y), i = 1, \ldots, N$, where N is the number of conditional distributions, θ_i denotes the i-th block of parameters and θ_{-i} denotes all elements of θ excluding θ_i. The theoretical justification for this proceeding is given by the theorem by Hammersley and Clifford (71) which is proven by Besag (1974). The intuition behind this theorem is that the knowledge of the complete set

of conditional posterior distributions,

$$p(\theta_1|\theta_2, \theta_3, \ldots, \theta_k, y),$$
$$p(\theta_2|\theta_1, \theta_3, \ldots, \theta_k, y),$$
$$\vdots$$
$$p(\theta_k|\theta_1, \theta_2, \ldots, \theta_{k-1}, y),$$

up to a constant of proportionality, is equivalent to the knowledge of the posterior distribution $p(\theta_1, \ldots, \theta_k|y)$. This allows applying the M-H algorithm to sub-blocks of θ leading to the Gibbs sampler if the individual conditional posterior distributions $p(\theta_i|\theta_{-i}, y)$ are directly available for sampling. In practice, Gibbs and M-H algorithms are often combined resulting in "hybrid" MCMC procedures as also illustrated in the next section.

The implementation of MCMC algorithms involves two steps. In the first step, M-H algorithms generate a sequence of random variables, $\{\theta^{(i)}\}_{i=1}^{G}$, converging to the posterior distribution $p(\theta|y)$. The algorithm is applied until convergence is achieved. In practice, the convergence of the Markov chain can be checked based on trace plots, autocorrelation plots or convergence tests, such as Geweke's Z-score test, Heidelberg-Welch's stationarity test and the half-width test, see, e.g., Cowles and Carlin (1996). In the second step, Monte Carlo methods are employed to compute the posterior mean of the parameters. In particular, given the generated Markov chain, $\{\theta^{(g)}\}_{g=1}^{G}$, the population mean $\mathsf{E}[f(\theta)|y] = \int f(\theta)p(\theta|y)d\theta$ can be consistently estimated by the sample mean

$$\frac{1}{G - g_1} \sum_{g=g_1+1}^{G} f(\theta^{(g)}), \qquad (12.28)$$

where g_1 is the number of burn-in periods which are discarded to reduce the influence of initial values $(\theta^{(0)})$. The length of the burn-in period typically consists of $10\% - 15\%$ of all MCMC iterations.

Consequently, the implementation of MCMC techniques requires both the convergence of the Markov chain and the convergence of the sample average. If the Markov chain is irreducible, aperiodic and positive recurrent, the Markov chain $\{\Theta^{(g)}\}_{g=1}^{G}$ generated from the MCMC algorithm converges to its invariant distribution, i.e.

$$\theta^{(g)} \xrightarrow{\mathcal{L}} \theta \quad \text{for } g \to \infty, \qquad (12.29)$$

where $\theta \sim p(\theta|y)$. For more details, see, e.g.,Tierney (1994) or Greenberg (2008).

The convergence of the sample average of a function $m(\cdot)$ of $\{\Theta^{(g)}\}_{g=1}^{G}$ to its population counterpart,

$$\frac{1}{G}\sum_{g=1}^{G} m(\theta^{(g)}) \overset{\text{a.s.}}{\rightarrow} E[m(\theta)|y] \quad \text{for } G \to \infty \tag{12.30}$$

is ensured by the ergodicity of the Markov chain. As shown by Tierney (1994), the latter property is sufficient to ensure also the convergence of the Markov chain to its invariant distribution.

12.3.2 MCMC-Based Estimation of the Standard SV Model

In this section, we will illustrate the estimation of the standard SV model using the M-H algorithm. For convenience, we restate model (12.1) as given by

$$y_t = \exp(h_t/2)u_t, \qquad\qquad u_t \sim \text{N}(0,1), \tag{12.31a}$$
$$h_t = \mu + \phi(h_{t-1} - \mu) + \eta_t, \qquad \eta_t \sim \text{N}(0, \sigma_\eta^2) \tag{12.31b}$$

with $\theta = (\mu, \phi, \sigma_\eta^2)$ and $h = (h_1, \cdots, h_T)$. Applying Bayes' theorem we have

$$p(\theta, h|y) \propto p(y|\theta, h)p(h|\theta)p(\theta). \tag{12.32}$$

Bayesian inference for the model parameters θ and the volatility states h is based on the posterior distribution $p(\theta, h|y)$ which is proportional to the product of the likelihood function $p(y|\theta, h)$ specified by (12.31a), the conditional distribution of the volatility states $p(h|\theta)$ given by (12.31b), and the prior distribution $p(\theta)$.

The model is completed by specifying the prior distributions for θ. We assume that the model parameters are a priori independently distributed as follows:

$$p(\mu) = \text{N}(\alpha_\mu, \beta_\mu^2), \tag{12.33a}$$
$$p(\phi) = \text{N}(\alpha_\phi, \beta_\phi^2)\mathbf{1}(-1, +1)(\phi), \tag{12.33b}$$
$$p(\sigma_\eta^2) = \mathbb{IG}(\alpha_\sigma, \beta_\sigma), \tag{12.33c}$$

where $\mathbb{IG}(\cdot, \cdot)$ denotes an inverse-gamma distribution and $\text{N}(a, b)\mathbf{1}(-1, +1)(x)$ defines a normal distribution with mean a, variance b, which is truncated between -1 and 1. This rules out near unit-root behavior of ϕ. The parameters $\alpha_{(\cdot)}$ and $\beta_{(\cdot)}$, characterizing the prior distributions, are called hyperparameters, which are specified by the researcher.

- Initialize $h^{(0)}, \mu^{(0)}, \phi^{(0)}$ and $\sigma_\eta^{2(0)}$.
- For $g = 1, \ldots, G$:
    1. For $t = 1, \ldots, T$:
        Sample $h_t^{(g)}$ from $p(h_t | y, h_{<t}^{(g)}, h_{>t}^{(g-1)}, \mu^{(g-1)}, \phi^{(g-1)}, \sigma_\eta^{2(g-1)})$.
    2. Sample $\sigma_\eta^{2(g)}$ from $p(\sigma_\eta^2 | y, h^{(g)}, \mu^{(g-1)}, \phi^{(g-1)})$.
    3. Sample $\phi^{(g)}$ from $p(\phi | y, h^{(g)}, \sigma_\eta^{2(g)}, \mu^{(g-1)})$.
    4. Sample $\mu^{(g)}$ from $p(\mu | y, h^{(g)}, \phi^{(g)}, \sigma_\eta^{2(g)})$.

Figure 12.2. Single-move Gibbs sampler for the standard SV model

Given the prior distributions, the conditional posteriors for the model parameters are derived as

$$p(\mu | y, h, \phi, \sigma_\eta^2) \propto p(y | h, \mu, \phi, \sigma_\eta^2) p(h | \mu, \phi, \sigma_\eta^2) p(\mu), \qquad (12.34a)$$

$$p(\phi | y, h, \sigma_\eta^2, \mu) \propto p(y | h, \mu, \phi, \sigma_\eta^2) p(h | \mu, \phi, \sigma_\eta^2) p(\phi), \qquad (12.34b)$$

$$p(\sigma_\eta^2 | y, h, \mu, \phi) \propto p(y | h, \mu, \phi, \sigma_\eta^2) p(h | \mu, \phi, \sigma_\eta^2) p(\sigma_\eta^2). \qquad (12.34c)$$

Since the volatility states h subsume all information about $(\mu, \phi, \sigma_\eta^2)$, the full information likelihood function $p(y | h, \mu, \phi, \sigma_\eta^2)$ is a constant with respect to the model parameters, and thus can be omitted.

By successively conditioning we get

$$p(h | \mu, \phi, \sigma_\eta^2) = p(h_1 | \mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1} | h_t, \mu, \phi, \sigma_\eta^2), \qquad (12.35)$$

where $p(h_{t+1} | h_t, \mu, \phi, \sigma_\eta^2)$ is specified according to (12.31b). Moreover, inserting $p(\sigma_\eta^2), p(\phi), p(\mu)$, given by (12.33), and $p(h | \mu, \phi, \sigma_\eta^2)$, given by (12.35), into (12.34), the full conditional posteriors can be reformulated, after eliminating constant terms, as (for details, see Appendix 12.5.1)

$$p(\sigma_\eta^2 | y, h, \mu, \phi) \propto \mathbb{IG}(\widehat{\alpha}_\sigma, \widehat{\beta}_\sigma), \qquad (12.36)$$

$$p(\phi | y, h, \sigma_\eta^2, \mu) \propto \mathrm{N}(\widehat{\alpha}_\phi, \widehat{\beta}_\phi^2) \mathbf{1}(-1, +1)(\phi), \qquad (12.37)$$

$$p(\mu | y, h, \phi, \sigma_\eta^2) \propto \mathrm{N}(\widehat{\alpha}_\mu, \widehat{\beta}_\mu^2), \qquad (12.38)$$

where the hyper-parameters are estimated by

$$\widehat{\alpha}_\sigma = \alpha_\sigma + \frac{T}{2}, \tag{12.39}$$

$$\widehat{\beta}_\sigma = \beta_\sigma + \frac{1}{2} \left\{ \sum_{t=1}^{T-1}(h_{t+1} - \mu - \phi(h_t - \mu))^2 + (h_1 - \mu)^2(1 - \phi^2) \right\}, \tag{12.40}$$

$$\widehat{\alpha}_\phi = \widehat{\beta}_\phi^2 \left\{ \frac{\sum_{t=1}^{T-1}(h_{t+1} - \mu)(h_t - \mu)}{\sigma_\eta^2} + \frac{\alpha_\phi}{\beta_\phi^2} \right\}, \tag{12.41}$$

$$\widehat{\beta}_\phi^2 = \left\{ \frac{\sum_{t=1}^{T-1}(h_t - \mu)^2 - (h_1 - \mu)^2}{\sigma_\eta^2} + \frac{1}{\beta_\phi^2} \right\}^{-1}, \tag{12.42}$$

$$\widehat{\alpha}_\mu = \widehat{\beta}_\mu^2 \left\{ \frac{h_1(1 - \phi^2) + (1 - \phi)\sum_{t=1}^{T-1}(h_{t+1} - \phi h_t)}{\sigma_\eta^2} + \frac{\alpha_\mu}{\beta_\mu^2} \right\}, \tag{12.43}$$

$$\widehat{\beta}_\mu^2 = \left\{ \frac{1 - \phi^2 + (T - 1)(1 - \phi)^2}{\sigma_\eta^2} + \frac{1}{\beta_\mu^2} \right\}^{-1}. \tag{12.44}$$

Since it is possible to directly sample from the conditional posteriors, we obtain a straightforward (single-move) Gibbs sampler which breaks the joint posterior $p(\theta, h, y)$ into $T + 3$ univariate conditional posteriors. The resulting Gibbs algorithm is summarized in Figure 12.2, where the subscripts of $h_{<t}^{(\cdot)}$ and $h_{>t}^{(\cdot)}$ denote the periods before and after t respectively.

The most difficult part of the estimation of SV models is to effectively sample the latent states h_t from their full conditional posterior. In this context, an M-H A-R algorithm can be applied. Below we briefly illustrate a sampling procedure which is also used by Kim et al. (1998). In this context, Bayes' theorem implies

$$p(h_t|y, h_{-t}, \theta) \propto p(y_t|h_t, \theta)p(h_t|h_{-t}, \theta), \tag{12.45}$$

$$= \frac{1}{\sqrt{2\pi \exp(h_t)}} \exp\left\{ -\frac{y_t^2}{2\exp(h_t)} \right\} p(h_t|h_{-t}, \theta), \tag{12.46}$$

$$= f^*(y_t, h_t, \theta)p(h_t|h_{-t}, \theta), \tag{12.47}$$

where, h_{-t} denotes all elements of $h = (h_1, \cdots, h_T)$ excluding h_t. Exploiting the Markovian structure of the SV model we can derive

$$p(h_t|h_{-t}, \theta) = p(h_t|h_{t-1}, h_{t+1}, \theta) = p_N(h_t|\alpha_t, \beta^2), \tag{12.48}$$

where, $p_N(x|a, b)$ denotes the normal density function with mean a and vari-

ance b, and

$$\alpha_t = \mu + \frac{\phi\{(h_{t-1} - \mu) + (h_{t+1} - \mu)\}}{(1 + \phi^2)}, \quad \beta^2 = \frac{\sigma_\eta^2}{1 + \phi^2}. \tag{12.49}$$

An acceptance-rejection step is implemented exploiting the fact that $\exp(-h_t)$ is bounded by a linear function in h_t. By applying a Taylor expansion for $\exp(-h_t)$ around α_t we obtain

$$\log f^*(y_t, h_t, \theta) \leq -\frac{1}{2}\log(2\pi) - \frac{1}{2}h_t - \frac{y_t^2}{2}[\exp(-\alpha_t)\{1 + \alpha_t - h_t \exp(-\alpha_t)\}] \tag{12.50}$$

$$\overset{\text{def}}{=} \log g^*(y_t, h_t, \theta). \tag{12.51}$$

Since $p(h_t | h_{-t}, \theta) = p_N(h_t | \alpha_t, \beta^2)$, we have

$$p(h_t | h_{-t}, \theta) f^*(y_t, h_t, \theta) \leq p_N(h_t | \alpha_t, \beta^2) g^*(y_t, h_t, \theta). \tag{12.52}$$

Then, the right-hand side of (12.52), after eliminating constant terms, can be represented by

$$p_N(h_t | \alpha_t, \beta^2) g^*(y_t, h_t, \theta) = k \cdot p_N(h_t | \alpha_t^*, \beta^2), \tag{12.53}$$

where k is a real valued constant, and $p_N(h_t | \alpha_t^*, \beta^2)$ denotes a normal density with mean $\alpha_t^* = \alpha_t + \frac{\beta^2}{2}(y_t^2 \exp\{-\alpha_t\} - 1)$ and variance β^2.

Hence, since the target distribution, $p(h_t | h_{-t}, \theta) f^*(y_t, h_t, \theta)$, is bounded by $p_N(h_t | \alpha_t^*, \beta^2)$ up to a constant k, the acceptance-rejection method can be applied to sample h_t from $p(h_t | y, h_{-t}, \theta)$ with acceptance probability

$$P\left\{U \leq \frac{f^{**}(y_t, h_t, \theta) p(h_t | h_{-t}, \theta)}{k p_N(h_t | \alpha_t^*, \beta^2)}\right\} = \frac{f^{**}(y_t, h_t, \theta)}{g^{**}(y_t, h_t, \theta)}$$

where $U \sim \mathbb{U}[0,1]$. Figure 12.3 summarizes the A-R algorithm to sample the latent volatility states h_t.

12.4 Empirical Illustrations

12.4.1 The Data

Below we will illustrate estimations of the standard SV model, the SVt model and the SVJ model based on time series of the DAX index, the Dow Jones

- For $t = 1, \cdots, T$:
    1. Draw h_t^* from $p_N(h_t | \alpha_t^*, \beta^2)$.
    2. Draw U from $\mathbb{U}[0,1]$.
    3. If $U \leq f^*(y_t, h_t^*, \theta)/g^*(y_t, h_t^*, \theta)$
        set $h_t = h_t^*$.
    Else
        go to step 1.

Figure 12.3. A-R method to sample the volatility states h_t

	Mean	SD	Median	0.1-q	0.9-q	Skewness	Kurtosis
DAX	3.7e-04	0.013	5.0e-4	-0.021	0.006	-0.295	7.455
Dow Jones	3.6e-04	0.009	3.0e-4	-0.009	0.008	-0.230	8.276
GBP/USD	3.6e-06	0.005	<1.0e-9	-0.006	0.009	-0.126	5.559

Table 12.1. Summary statistics for daily returns of the DAX index, the Dow Jones index, and the GBP/USD exchange rate from 01/01/1991 to 21/03/2007. **Q** XFGsummary

index and the GBP/USD FX rate. All time series cover the period from 1 January, 1991 to 21 March, 2007. We use daily continuously compounded returns yielding 4,231 observations. Table 12.1 reports the mean, standard deviation, median, 10%- and 90%-quantiles, and the empirical skewness as well as kurtosis of the three series. All series reveal negative skewness and overkurtosis which is a common finding for financial returns.

12.4.2 Estimation of SV Models

The standard SV model is estimated by running the Gibbs and A-R M-H algorithm based on 25,000 MCMC iterations, where $5,000$ iterations are used as burn-in period. Table 12.2 displays the choice of the prior distributions and the hyper-parameters as well as the resulting prior mean and standard deviation.

Table 12.3 shows the sample mean (MEAN), the sample standard deviation (SD), the time-series standard errors (ts-SE), and the 95%-credibility interval (CI) based on $G = 20,000$ MCMC replications. The time-series standard errors give an estimate of the variation that is expected in computing the

Prior Distribution	Hyper-Parameters		Mean	S.D.
$p(\mu) = \mathrm{N}(\alpha_\mu, \beta_\mu^2)$	$\alpha_\mu = 0$	$\beta_\mu = 100$	0	10
$p(\phi) = \mathrm{N}(\alpha_\phi, \beta_\phi^2)I_{(-1,+1)}(\phi)$	$\alpha_\mu = 0$	$\beta_\mu = 100$	0	1
$p(\sigma_\eta^2) = \mathcal{IG}(\alpha_\sigma, \beta_\sigma)$	$\alpha_\sigma = 2.5$	$\beta_\sigma = 0.025$	0.167	0.024

Table 12.2. Prior distributions, hyper-parameters, and implied prior means as well as standard deviations for the standard SV model. ◘XFGprior

Parameter	Mean	SD	ts-SE	95% CI
DAX				
μ	-8.942	0.192	1.5e-3	(-9.327,-8.565)
ϕ	0.989	0.002	2.0e-4	(0.983, 0.994)
σ_η	0.115	0.009	1.0e-3	(0.096, 0.137)
Dow Jones				
μ	-9.471	0.171	1.3e-3	(-9.810,-9.142)
ϕ	0.990	0.003	2.0e-4	(0.984, 0.995)
σ_η	0.087	0.010	1.1e-3	(0.069, 0.108)
GBP/USD				
μ	-10.238	0.649	4.3e-3	(-10.519,-9.997)
ϕ	0.993	0.002	2.0e-4	(0.988, 0.997)
σ_η	0.041	0.006	8.0e-4	(0.029, 0.054)

Table 12.3.   Estimation results for the standard SV model. ◘XFGparameter

mean of the MC replications and is computed as $SD/\sqrt{n}$. As a rule of thumb, Geweke (1992) suggests to choose G such that the time series standard error is less than approximately 5% of the sample standard deviation.

Since the three time series reveal similar properties, we concentrate on the results for DAX index returns. The volatility process is highly persistent as indicated by an estimate of ϕ of 0.989. This near-to-unit-root behavior is a quite typical finding for financial return series and is consistent with the commonly observed volatility clustering. The estimated (smoothed) volatility

states are computed by

$$\widehat{h}_t = \frac{1}{G - g_1} \sum_{g=g_1+1}^{G} \exp(h_t^{(g)}/2), \tag{12.54}$$

where $h_t^{(g)}$ denotes the realizations of the Markov chain stemming from the M-H A-R algorithm illustrated in the previous section, and g_1 is the burn-in period. The resulting plots of the smoothed volatilities are shown in Figure 12.4. It is nicely illustrated that the estimated latent volatility closely mimics the movements of $|y_t|$ supporting the idea of using absolute or squared returns as (noisy) proxies for h_t.
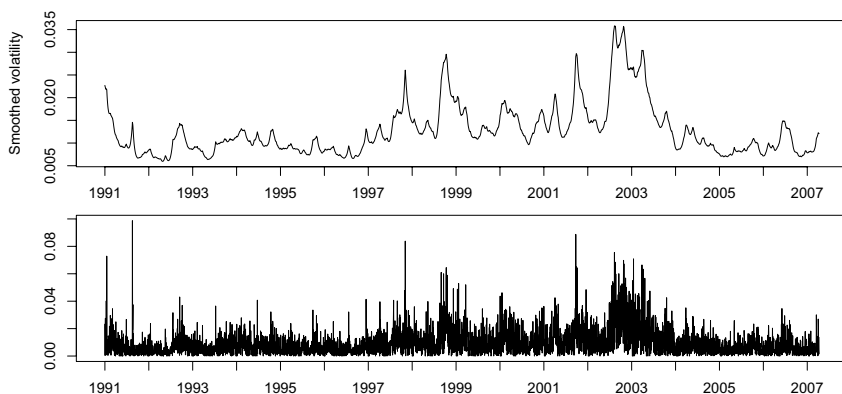
Figure 12.4. Top: Smoothed estimates of h_t. Bottom: Absolute returns, $|y_t|$. ◘ XFGvolabs

Misspecification tests are implemented based on the standardized innovations, $y_t \exp(-\widehat{h}_t/2)$ which should be i.i.d. Applying Ljung-Box tests and ARCH tests (Engle, 1982) shown in Figure 12.5 yield p-values of 0.094 and 0.023, respectively. For the BDS independence test we find a p-value of 0.011. The corresponding plot of the standardized innovations as well as ACF plots of standardized innovations and squared standardized innovations are given by graphs (a), (c) and (d), respectively, in Figure 12.5. The standardized innovations reveal a big outlier on 19/08/1991 where the DAX index dropped from 1653.33 to 1497.93. Such a behavior is not easily captured by a continuous distribution for h_t and requires accounting for jumps. Nevertheless,

though it is evident that the model is obviously not flexible enough to completely explain the volatility dynamics, the diagnostics indicate a quite satisfying dynamic performance. This is particularly true when the parameter parsimony of the model is taken into account.

It is not surprising that the model is unable to capture the distributional properties of the returns. We observe that the standard SV model with a model implied kurtosis of 5.74 is not able to fully explain the over-kurtosis in the data. This is confirmed by the Jarque-Bera normality test and the QQ plot revealing departures from normality mainly stemming from extreme innovations.
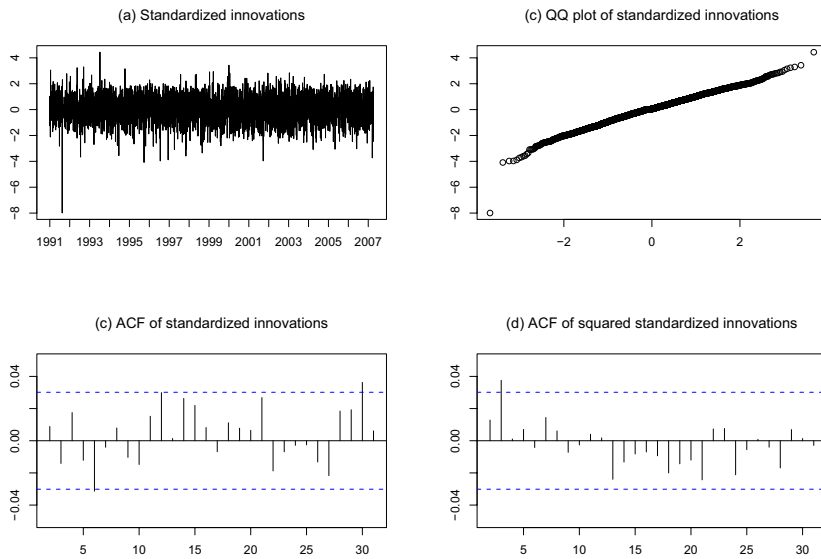
Figure 12.5. Time series plot, QQ plot and autocorrelogram of (squared) standardized innovations. ◐ XFGstdinnov

Finally, the results of convergence diagnostics are reported in Table 12.4. All parameters pass both the Geweke's z-scores test and the Heidelberg-Welch's stationarity and half-width tests indicating a proper convergence of the Markov chain to its invariant distribution.

Table 12.5 shows the estimation results based on the SVt and SVJ model. For the sake of brevity and given that we have qualitatively similar findings

Parameter	Z-score Test		Stationarity and Half-Width Test			
	z-score	p-value	p-value	Mean	Half-width	Ratio
μ^{SV}	0.199	0.843	0.645	-8.895	0.003	-0.001
ϕ^{SV}	0.032	0.972	0.897	0.928	0.001	0.001
σ_u^{SV}	-0.413	0.686	0.979	0.329	0.003	0.009

Table 12.4.   Convergence Diagnostics.   The half-width test is passed if the corresponding ratio is less than 0.01. ⧉ XFGconvergence

Parameter	Mean	SD	ts-SE	95% CI
The SVt model:				
μ	-9.201	0.230	2.3e-3	(-9.663,-8.752)
ϕ	0.991	0.002	1.0e-4	(0.985, 0.995)
σ_η	0.117	0.012	1.1e-3	(0.095, 0.145)
ν	12.443	1.812	2.3e-1	(9.600,16.923)
The SVJ model:				
μ	-9.107	2.3e-01	1.8e-03	(-9.568,-8.663)
ϕ	0.991	2.7e-03	2.1e-04	(0.984, 0.995)
σ_η	0.124	1.3e-02	1.4e-03	(0.101, 0.153)
α_k	-0.005	2.9e-05	1.8e-07	(-0.005,-0.004)
$\sqrt{\beta_k}$	0.029	6.5e-03	8.7e-04	(0.020, 0.045)
κ	0.010	3.9e-03	3.1e-04	(0.003, 0.019)

Table 12.5. Estimation results for the SVt and SVJ model based on DAX index returns. ⧉ XFGsvtjparameter

for the other return series, we focus only on DAX index returns. We obtain an estimate of the degrees of freedom in the SVt model of about $\widehat{\nu} = 12.44$ indicating the presence of fat-tailedness in the data and a clear misspecification of the standard (Gaussian) SV model. The estimates for the SVJ model reveal a daily average jump size of about $\widehat{\alpha}_k = 0.005\%$ with estimated standard deviation $\sqrt{\widehat{\beta}_k} = 0.029$. Estimates of κ reveal an average probability of observing a jump of about 1% on a daily basis. This implies that on average a jump in returns may occur on average every 100 trading days.

Figure 12.6 depicts the QQ plots of the normalized innovations based on the standard SV model (left), the SVt model (middle), and the SVJ model (right).

It is shown that the inclusion of Student-t errors improves the distributional properties of the model only slightly. Actually, we observe that both the basic SV and the SVt model are not able to capture extreme observations in the tails of the distribution. In contrast, the SVJ model turns out to be more appropriate to accommodate outliers. This result indicates the importance of allowing returns to be driven by a jump component.
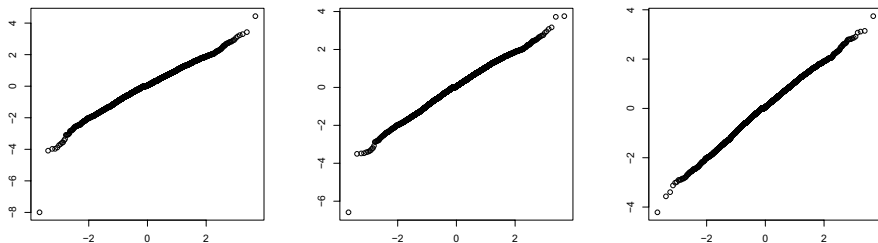

Figure 12.6. QQ plots of normalized innovations based on the standard SV model (left), the SVt model (middle), and the SVJ model (right). ⊞ XFGsvtsvjqq

12.5 Appendix

12.5.1 Derivation of the Conditional Posterior Distributions

Using Bayes' theorem, the conditional posterior distribution of σ_η^2 is given by

$$p(\sigma_\eta^2|y, h, \mu, \phi) \propto p(y|h, \mu, \phi, \sigma_\eta^2)p(h|\mu, \phi, \sigma_\eta^2)p(\sigma_\eta^2).$$

By assuming σ_η^2 to follow an inverse-gamma distribution and successively conditioning on $p(h|\mu, \phi, \sigma_\eta^2)$, we obtain

$$p(\sigma_\eta^2|y, h, \mu, \phi) \propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2)\mathbb{IG}(\sigma_\eta^2|\alpha_\sigma, \beta_\sigma),$$

where the density function $p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2)$ is given by (12.1b).

After eliminating all constant terms with respect to σ_η^2, we obtain

$$p(\sigma_\eta^2|y, h, \mu, \phi)$$

$$\propto \exp\left[-\frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{T-1}\{h_{t+1} - \mu - \phi(h_t - \mu)\}^2}{2\sigma_\eta^2}\right]$$

$$\times \left(\frac{1}{\sigma_\eta^2}\right)^{\frac{T}{2}} \frac{(\beta_\sigma)^{\alpha_\sigma} e^{-\beta_\sigma/\sigma_\eta^2}}{\Gamma(\alpha_\sigma)(\sigma_\eta^2)^{\alpha_\sigma+1}}$$

$$\propto \exp\left[-\frac{\beta_\sigma + \frac{1}{2}(h_0 - \mu)^2(1 - \phi^2) + \frac{1}{2}\sum_{t=1}^{T-1}\{h_{t+1} - \mu - \phi(h_t - \mu)\}^2}{\sigma_\eta^2}\right]$$

$$\times \left(\frac{1}{\sigma_\eta^2}\right)^{(\alpha_\sigma+\frac{T}{2})+1}.$$

It is easy to see that the posterior density $p(\sigma_\eta^2|y, h, \mu, \phi)$ is proportional to an inverse-gamma density. Consequently, we have

$$p(\sigma_\eta^2|y, h, \mu, \phi) \propto \mathbb{IG}(\widehat{\alpha}_\sigma, \widehat{\beta}_\sigma),$$

where,

$$\widehat{\alpha}_\sigma = \alpha_\sigma + \frac{T}{2},$$

$$\widehat{\beta}_\sigma = \beta_\sigma + \frac{1}{2}(h_1 - \mu)^2(1 - \phi^2) + \frac{1}{2}\sum_{t=1}^{T-1}\{h_{t+1} - \mu - \phi(h_t - \mu)\}^2.$$

Mimicking the proceeding for σ_η^2 we can derive the conditional posteriors for μ and ϕ in a similar way. Then, we obtain

$$p(\mu|y, h, \phi, \sigma_\eta^2) \propto p(h|\mu, \phi, \sigma_\eta^2)p(\mu),$$

$$\propto p(h_1|\mu, \phi, \sigma_\eta^2)\prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2)\mathrm{N}(\alpha_\mu, \beta_\mu),$$

$$\propto \exp\left(-\frac{1}{2}\left[\mu^2\underbrace{\left\{\frac{1 - \phi^2 + (T - 1)(1 - \phi)^2}{\sigma_\eta^2} + \frac{1}{\beta_\mu^2}\right\}}_{A}\right.\right.$$

$$\left.\left.- 2\mu\underbrace{\left\{\frac{h_1(1 - \phi^2) + (1 - \phi)\sum_{t=1}^{T-1}(h_{t+1} - \phi h_t)}{\sigma_\eta^2} + \frac{\alpha_\mu}{\beta_\mu^2}\right\}}_{B}\right]\right),$$

$$\propto \mathrm{N}\left(\frac{B}{A}, \frac{1}{A}\right)$$

and

$$p(\phi|y, h, \sigma_\eta^2, \mu) \propto p(h|\mu, \phi, \sigma_\eta^2)p(\phi),$$

$$\propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2)\mathrm{N}(\alpha_\phi, \beta_\phi^2)\infty_{(-1,+1)}(\phi),$$

$$\propto \exp\left(-\frac{1}{2}\left[\phi^2\left\{\underbrace{\frac{-(h_1-\mu)^2 + \sum_{t=1}^{T-1}(h_t-\mu)^2}{\sigma_\eta^2} + \frac{1}{\beta_\phi^2}}_{C}\right\}\right.\right.$$

$$\left.\left.-2\phi\left\{\underbrace{\frac{\sum_{t=1}^{T-1}(h_{t+1}-\mu)(h_t-\mu)}{\sigma_\eta^2} + \frac{\alpha_\phi}{\beta_\phi^2}}_{D}\right\}\right]\right)\infty_{(-1,+1)}(\phi),$$

$$\propto \mathrm{N}\left(\frac{D}{C}, \frac{1}{C}\right)\infty_{(-1,+1)}(\phi).$$

Bibliography

Andersen, T. G., Bollerslev, T., Diebold, F.X., and Labys, P. (2003). Modeling and Forecasting Realized Volatility, *Econometrica* **71**: 579–625.

Arteche, J. (2004). Gaussian semiparametric estimation in long memory in stochastic volatility and signal plus noise models, *Journal of Econometrics* **119**: 131–154.

Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics, *Journal of Econometrics* **73**: 5–59.

Bates, D. (2000). Post-'87 Crash fears in S&P 500 futures options, *Journal of Econometrics* **9**: 69–107.

Beran, J. (1994). *Statistics for Long-Memory Processes*, Chapman and Hall, Boca Raton.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society* **Series B 36**: 192–236.

Black, F. (1976). Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economical Statistics Section* 177–181.

Bollerslev, T. and Mikkelsen, H. O. (1996). Modeling and pricing long memory in stock market volatility, *Journal of Econometrics* **73**: 151–184.

Breidt, F. J., Carto, N., and de Lima, P. (1998). The detection and estimation of long memory in stochastic volatility models, *Journal of Econometrics* **83**: 325–348.

Brockwell, A. E. (2005). Likelihood-based analysis of a class generalized long-memory time series models, *Journal of Time Series Analysis* **28**: 386–407.

Bauwens, L., Lubrano, M., and Richard, J. F. (1999). *Bayesan Inference in Dynamic Econometric Models*, Oxford University Press & Sons, New York.

Campbell, J. Y. and Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns, *Journal of Financial Economics* **31**: 281–318.

Chib, S., Nardari, F., and Shephard, N. (2002). Markov Chain Monte Carlo methods for stochastic volatility models, *Journal of Econometrics* **108**: 281–316.

Cowles, M. K. and Carlin, B. P. (1996). Markov Chain Monte Carlo convergence diagnostics: A comparative review, *Journal of the American Statistical Association* **91**: 883–905.

Clark, P.K. (1973). A subordinated stochastic process model with finite variance for speculative prices,", *Econometrica* **41**: 135–156.

Danielsson, J. (1994). Stochastic volatility in asset prices: Estimation with simulated maximum likelihood, *Journal of Econometrics* **64**: 375–400.

Duffie, D., Pan, J., and Singleton, K. (2000). Transform analysis and asset pricing for affine jump-diffusions, *Econometrica* **68**: 1343–1376.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**: 987–1007.

Eraker, B., Johannes, M., and Polson, N. (2003). The impact of jumps in volatility and returns, *Journal of Finance* **58**: 1269–1300.

French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility, *Journal of Financial Economics* **19**: 3–29.

Frieze, A., Kannan, R., and Polson N. (1994). Sampling from log-concave distributions, *Annals of Applied Probability* **4**: 812–834.

Gallant, A.R., Hsie, D., and Tauchen G. (1997). Estimation of stochastic volatility models with diagnostics, *Journal of Econometrics* **81**: 159–192.

Geman, S. and Geman D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* J. M. Bernardo, J. O. Berger, A. P. Dawid, A. F. M. Smith (ed.), *Bayesian Statistics 4*, Oxford University Press, Amsterdam.

Ghysels, E., Harvey, A. C., and Renault E. (1996). Stochastic volatility, *in* G. S. Maddala and C. R. Rao (ed.), *Handbook of Statistics 14, Statistical Methods in Finance*, North-Holland, Amsterdam.

Granger, C. W. and Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis* **1**: 15–29.

Greenberg, E. (2008). *Introduction to Bayesian Econometrics*, Cambridge University Press, New York.

Hammersley, J. and Clifford, P. (1971). Markov fields on finite graphs and lattices, *Unpublished Manuscipt*.

Harvey, A. C. (1998). Long-memory in stochastic volatility, *in* J. Knight and S. E. Satchell (ed.), *Forecasting Volatility in Financial Markets*, Butterworth-Heinemann, London, pp. 203–226.

Harvey, A. C., Ruiz, E., and Shephard, N. (1994). Multivariate stochastic variance models, *Review of Economic Studies* **61**: 247–264.

Harvey, A. C. and Shephard, N. (1996). The estimation of an asymmetric stochastic volatility model for asset returns, *Journal of Business and Economics Statistics* **14**: 429–434.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994). Bayesian analysis of stochastic volatility models (with discussion), *Journal of Business and Economic Statistics* **12**: 371–417.

Jacquier, E., Polson, N. G., and Rossi, P. E. (2004). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors, *Journal of Econometrics* **122**: 185–212.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economic Studies* **65**: 361–393.

Koop, G. (2006). *Bayesian Econometrics*, Wiley-Interscience, London.

Liesenfeld, R. and Richard, J. F. (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics, *Journal of Empirical Finance* **10**: 505–531.

Liesenfeld, R. and Jung, R. C. (2000). Stochastic volatility models: Conditional normality versus heavy-tailed distributions, *Journal of Applied Econometrics* **15**: 137–160.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equations of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1092.

Melino, A. and Turnbull, S. M. (1990). Pricing foreign currency options with stochastic volatility, *Journal of Econometrics* **45**: 239–265.

Sandmann, G. and Koopman, S. J. (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood, *Journal of Econometrics* **87**: 271-301.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes, a study of daily sugar prices 1961-79, *in* O. D. Anderson (ed.), *Time Series Analysis: Theory and Practice 1*, North-Holland, Amsterdam, pp. 203–226.

Tierney, L. (1994). Markov Chains for exploring posterior distributions, *The Annals of Statistics* **22**: 1701–1762.

13 Measuring and Modeling Risk Using High-Frequency Data

Wolfgang Härdle, Nikolaus Hautsch and Uta Pigorsch

13.1 Introduction

Volatility modelling is the key to the theory and practice of pricing financial products. Asset allocation and portfolio as well as risk management depend heavily on a correct modelling of the underlying(s). This insight has spurred extensive research in financial econometrics and mathematical finance. Stochastic volatility models with separate dynamic structure for the volatility process have been in the focus of the mathematical finance literature, see Heston (1993) and Bates (2000), while parametric GARCH-type models for the returns of the underlying(s) have been intensively analyzed in financial econometrics.

The validity of these models in practice though depends upon specific distributional properties or the knowledge of the exact (parametric) form of the volatility dynamics. Moreover, the evaluation of the predictive ability of volatility models is quite important in empirical applications. However, the latent character of the volatility poses a problem. To what measure should the volatility forecasts be compared to? Conventionally, the forecasts of daily volatility models, such as GARCH-type or stochastic volatility models, have been evaluated with respect to absolute or squared daily returns. In view of the excellent in-sample performance of these models, the forecasting performance, however, seems to be disappointing.

The availability of ultra-high-frequency data opens the door for a refined measurement of volatility and model evaluation. An often used and very flexible model for logarithmic prices of speculative assets is the (continuous-time) stochastic volatility model:

$$dY_t = (\mu + \beta\sigma_t)dt + \sigma_t dW_t, \tag{13.1}$$

where σ_t^2 is the instantaneous (spot) variance, μ denotes the drift, β is the risk premium, and W_t defines the standard Wiener process. The object of interest is the amount of variation accumulated in a time interval Δ (e.g., a day, week, month etc.). If $n = 1, 2, \ldots$ denotes a counter for the time intervals of interest, then the term

$$\sigma_n^2 = \int_{(n-1)\Delta}^{n\Delta} \sigma_t^2 dt \tag{13.2}$$

is called the actual volatility, see Barndorff-Nielsen and Shephard (2002). The actual volatility is the quantity that reflects the market risk structure (scaled in Δ) and is the key element in pricing and portfolio allocation. Actual volatility (measured in scale Δ) is of course related to the integrated volatility:

$$V(t) = \int_0^t \sigma_s^2 ds. \tag{13.3}$$

It is worth noting that there is a small notational confusion here: the mathematical finance literature would denote σ_t as "volatility" and σ_t^2 as "variance", see Nelson and Foster (1994), for example.

An important result is that $V(t)$ can be estimated from Y_t via the quadratic variation:

$$[Y_t]_M = \sum (Y_{t_j} - Y_{t_{j-1}})^2, \tag{13.4}$$

where $t_0 = 0 < t_1 < \cdots < t_M = t$ is a sequence of partition points and $\sup_j |t_{j+1} - t_j| \to 0$. Andersen and Bollerslev (1998) have shown that

$$[Y_t]_M \xrightarrow{p} V(t), \ M \to \infty. \tag{13.5}$$

This observation leads us to consider in an interval Δ with M observations

$$RV_n = \sum_{j=1}^{M} (Y_{t_j} - Y_{t_{j-1}})^2 \tag{13.6}$$

with $t_j = \Delta\{(n-1) + j/M\}$. Note that RV_n is a consistent estimator of σ_n^2 and is called *realized volatility*. Barndorff-Nielsen and Shephard (2002) point out that $RV_n - \sigma_n^2$ is approximately mixed Gaussian and provide the asymptotic law of

$$\sqrt{M}(RV_n - \sigma_n^2). \tag{13.7}$$

The realized volatility turns out to be very useful in the assessment of the validity of volatility models. For instance, reconciling evidence in favor of the forecast accuracy of GARCH-type models is observed when using realized

volatility as a benchmark rather than daily squared returns. Moreover, the availability of the realized volatility measure initiated the development of a new and quite accurate class of volatility models. In particular, based on the ex-post observability of the realized volatility measure, volatility is now treated as an observed rather than a latent variable to which standard time series procedures can be applied.

The remainder of this chapter is structured as follows. We first discuss the practical problems encountered in the empirical construction of realized volatility which are due to the existence of market microstructure noise. Section 13.3 presents the stylized facts of realized volatility, while Section 13.4 reviews the most popular realized volatility models. Section 13.5 illustrates the usefulness of the realized volatility concept for measuring time-varying systematic risk within a conditional asset pricing model (CAPM).

13.2 Market Microstructure Effects

The consistency of the realized volatility estimator builds on the notion that prices are observed in continuous time and without measurement error. In practice, however, the sampling frequency is inevitably limited by the actual quotation or transaction frequency. Since high-frequency prices are subject to market microstructure noise, such as price-discreteness, bid-and-ask bounce effects, transaction costs etc., the true price is unobservable. Market microstructure effects induce a bias in the realized volatility measure, which can straightforwardly be illustrated in the following simple discrete-time setup. Assume that the logarithmic high-frequency prices are observed with noise, i.e.,

$$Y_{t_j} = Y_{t_j}^* + \varepsilon_{t_j}, \tag{13.8}$$

where $Y_{t_j}^*$ denotes the latent true price. Moreover, the microstructure noise ε_{t_j} is assumed to be iid distributed with mean zero and variance η^2, and is independent of the true return. Let $r_{t_j}^*$ denote the efficient return, then the high-frequency continuously compounded returns

$$r_{t_j} = r_{t_j}^* + \varepsilon_{t_j} - \varepsilon_{t_{j-1}} \tag{13.9}$$

follow an MA(1) process. Such a return specification is well established in the market microstructure literature and is usually justified by the existence of the bid-ask bounce effect, see, e.g., Roll (1984). In this model, the realized

volatility is given by

$$RV_n = \sum_{i=1}^{M} (r_{t_j}^*)^2 + 2 \sum_{j=1}^{M} r_{t_j}^* (\varepsilon_{t_j} - \varepsilon_{t_{j-1}}) + \sum_{j=1}^{M} (\varepsilon_{t_j} - \varepsilon_{t_{j-1}})^2. \qquad (13.10)$$

with

$$\mathsf{E}[RV_n] = \mathsf{E}[RV_n^*] + 2M\eta^2. \qquad (13.11)$$

If the sampling frequency goes to infinity, we know from the previous section that RV_n^* consistently estimates σ_n^2 and, thus, the realized volatility based on the observed price process is a biased estimator of the actual volatility with bias term $2M\eta^2$. Obviously, for $M \to \infty$, RV_n diverges.
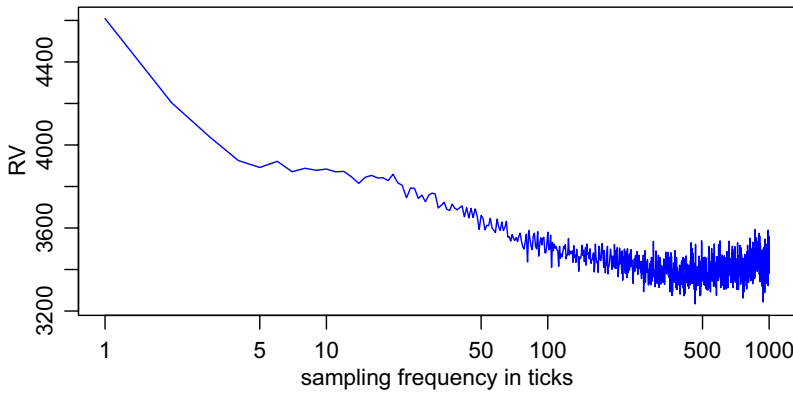
Figure 13.1.   Volatility signature plot for IBM, 2001-2006.
Average time between trades: 6.78 seconds. ◎ XFGsignature

This diverging behavior can also be observed empirically in so–called volatility signature plots. Figure 13.1 shows the volatility signature for one stock of the IBM incorporation over the period ranging from January 2, 2001 to December 29, 2006. The plot depicts the average annualized realized volatility over the full sample period constructed at different frequencies measured in number of ticks (depicted in log scale). Obviously, the realized volatility is large at the very high frequency, but decays for lower frequencies and stabilizes around a sampling frequency of 300 ticks, which corresponds approximately to a 30 minute sampling frequency, given that the average duration between two consecutive trades is around 6.78 seconds.

Thus, sampling at a lower frequency, such as every 10, 15 or 30 minutes, seems to alleviate the problem of market microstructure noise and has thus

frequently been applied in the literature. This so–called *sparse sampling*, however, comes at the cost of a less precise estimate of the actual volatility. Alternative methods have been proposed to solve this bias-variance trade-off for the above simple noise assumption as well as for more general noise processes, allowing also for serial dependence in the noise and/or for dependence between the noise and the true price process, which is sometimes referred to as endogenous noise. A natural approach to reduce the market microstructure noise effect is to construct the realized volatility measure based on prefiltered high-frequency returns, using, e.g., an MA(1) model.

In the following we briefly present two more elaborate and under specific noise assumptions consistent procedures for estimating actual volatility. Both have been theoretically considered in several papers. The subsampling approach originally suggested by Zhang et al. (2005) builds on the idea of averaging over various realized volatilities constructed from different high-frequency subsamples. For the ease of exposition we focus again on one time period, e.g., one day, and denote the full grid of time points at which the M intradaily prices are observed by $\mathcal{G}_t = \{t_0, \ldots, t_M\}$. The realized volatility that makes use of all observations in the full grid is denoted by $RV_n^{(all)}$. Moreover, the grid is partitioned into L nonoverlapping subgrids $\mathcal{G}^{(l)}$, $l = 1, \ldots, L$. A simple way for selecting such a subgrid may be the so–called regular allocation, in which the l-th subgrid is given by $\mathcal{G}^{(l)} = \{t_{l-1}, t_{l-1+L}, \ldots, t_{l-1+M_l L}\}$ for $l = 1, \ldots, L$, and M_l denoting the number of observations in each subgrid. E.g., consider 5-minute returns that can be measured at the time points 9:30, 9:35, 9:40, ..., and at the time points 9:31, 9:36, 9:41, ... and so forth. In analogy to the full grid, the realized volatility for subgrid l, denoted by $RV_n^{(l)}$, is constructed from all data points in subgrid l. Thus, $RV_n^{(l)}$ is based on sparsely sampled data.

The actual volatility is then estimated by:

$$RV_n^{(ZMA)} = \frac{1}{L} \sum_{l=1}^{L} RV_n^{(l)} - \frac{\bar{M}}{M} RV_n^{(all)}, \qquad (13.12)$$

where $\bar{M} = \frac{1}{L} \sum_{l=1}^{L} M_l$. The latter term on the right-hand side is included to bias-correct the averaging estimator $\frac{1}{L} \sum_{l=1}^{L} RV_n^{(l)}$. As the estimator (13.12) consists of a component based on sparsely sampled data and one based on the full grid of price observations, the estimator is also called the *two time scales estimator*.

Given the similarity to the problem of estimating the long-run variance of a stationary time series in the presence of autocorrelation, it is not surprising that kernel-based methods have been developed for estimating the realized

volatility. Most recently, Barndorff-Nielsen et al. (2008) proposed the flat-top realized kernel estimator

$$RV_n^{(BHLS)} = RV_n + \sum_{h=1}^{H^*} K\left(\frac{h-1}{H^*}\right)(\widehat{\gamma}_h + \widehat{\gamma}_{-h}) \qquad (13.13)$$

with

$$\widehat{\gamma}_h = \frac{M}{M-h}\sum_{j=1}^{M} r_{t_j} r_{t_{j-h}}, \qquad (13.14)$$

and $K(0) = 1$, $K(1) = 0$. Obviously, the summation term on the right-hand side is the realized kernel correction of the market microstructure noise. Zhou (1996), who was the first to consider realized kernels, proposed (13.13) with $H = 1$, while Hansen and Lunde (2006) allowed for general H but restricted $K(x) = 1$. Both of these estimators, however, have been shown to be inconsistent. Barndorff-Nielsen et al. (2008) instead propose several consistent realized kernel estimators with an optimally chosen H^*, such as the Tukey-Hanning kernel, i.e. $K(x) = \{1 - \cos\pi(1-x)^2\}/2$, which performs also very well in terms of efficiency as illustrated in a Monte Carlo analysis. They further show, that these realized kernel estimators are robust to market microstructure frictions that may induce endogenous and dependent noise terms.

13.3 Stylized Facts of Realized Volatility

Figure 13.2 shows kernel density estimates of the plain and logarithmic daily realized volatility in comparison to plots of a correspondingly fitted (log) normal distribution based on the IBM data, 2001-2006. The pictures in the top of Figure 13.2 show the unconditional distribution of the (plain) realized volatility in contrast to a fitted normal distribution. As also confirmed by the corresponding descriptive statistics displayed by Table 13.1, we observe that realized volatility reveals severe right-skewness and excess kurtosis. This result might be surprising given that the realized volatility consists of the sum of squared intra-day returns and thus central limit theorems should apply. However, it is a common finding that intra-day returns are strongly serially dependent requiring significantly higher intra-day sampling frequencies to observe convergence to normality. In contrast, the unconditional distribution of the logarithmic realized volatility is well approximated by a normal distribution. The sample kurtosis is strongly reduced and is close to 3. Though slight right-skewness and deviations from normality in the tails of the distribution
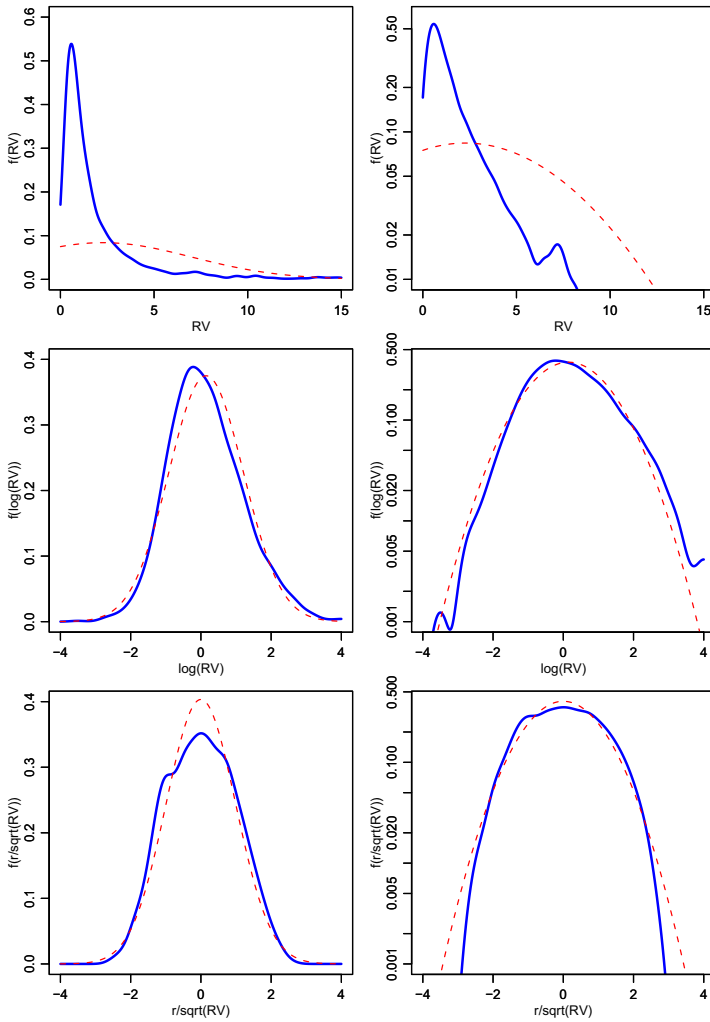
Figure 13.2. Kernel density estimates of the (logarithmic) realized volatility and of correspondingly standardized returns for IBM, 2001-2006. The dotted line depicts the density of the correspondingly fitted normal distribution. The left column depicts the kernel density estimates based on a log scale. ◨ XFGkernelcom

are still observed, the underlying distribution is remarkably close to that of a Gaussian distribution.

	RV_n	$\ln RV_n$	$r_n/\sqrt{RV_n}$
Mean	2.26	0.14	-.000
Median	1.05	0.05	-.013
Skewness	9.93	0.42	.035
Variance	22.57	1.13	.979
Kurtosis	150.47	3.43	2.349
1%-quantile	0.13	-2.03	-1.980
5%-quantile	0.24	-1.41	-1.558
95%-quantile	7.58	2.00	1.628
99%-quantile	17.66	2.87	2.141
LB(40)	2140.48	14213.07	39.780
p-value LB(40)	0.00	0.00	0.480
$\widehat{d}$	0.38	0.62	-

Table 13.1. Descriptive statistics of the realized volatility, log realized volatility and standardized returns, IBM stock, 2001-2006. LB (40) denotes the Ljung-Box statistic based on 40 lags. The last row gives an estimate of the order of fractional integration based on the Geweke and Porter-Hudak estimator. **Q** XFGIBm

A common finding is that financial returns have fatter tails than the normal distribution and reveal significant excess kurtosis. Though GARCH models can explain excess kurtosis, they cannot completely capture these properties in real data. Consequently, (daily) returns standardized by GARCH-induced volatility, typically still show clear deviations from normality. However, a striking result in recent literature is that return series standardized by the square root of realized volatility, $r_n/\sqrt{RV_n}$, are quite close to normality. This result is illustrated by the plots in the bottom of Figure 13.2 and the descriptive statistics in Table 13.1. Though we observe deviations from normality for returns close to zero resulting in a kurtosis which is even below 3, the fit in the tails of the distribution is significantly better than that for plain log returns. Summarizing the empirical findings from Figure 13.2, we can conclude that the unconditional distribution of daily returns is well described by a lognormal-normal mixture. This confirms the mixture-of-distribution hypothesis by Clark (1973) as well as the idea of the basic stochastic volatility model, where the log variance is modelled in terms of a Gaussian AR(1) process.

Figure 13.3 shows the evolvement of daily realized volatility over the analyzed sample period and the implied sample autocorrelation functions (ACFs). As also shown by the corresponding Ljung-Box statistics in Table 13.1, the realized volatility is strongly positively autocorrelated with high persistence. This is particularly true for the logarithmic realized volatility. The plot
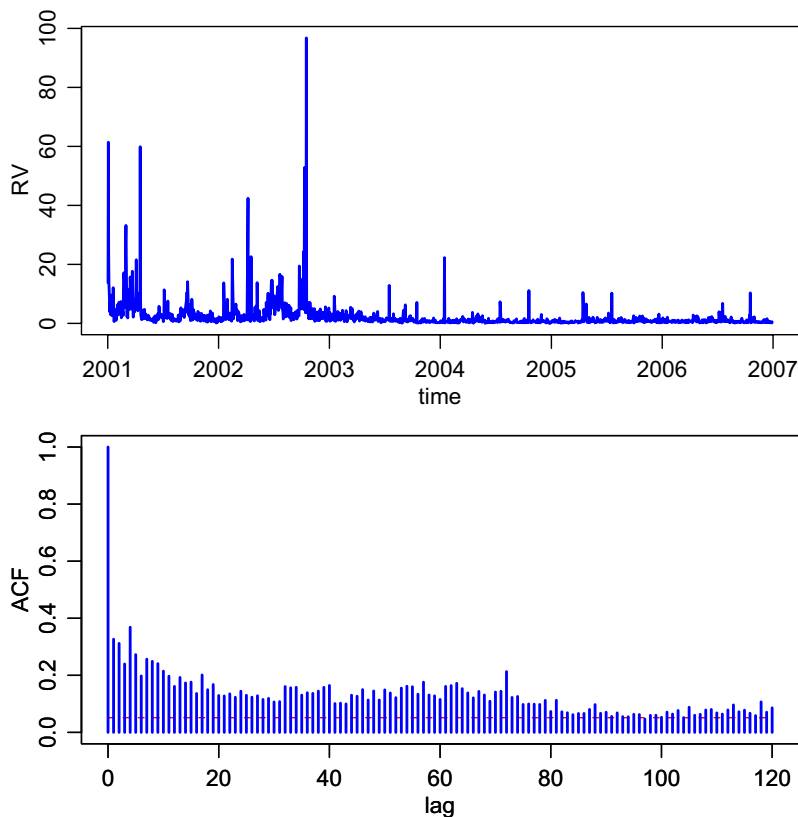
Figure 13.3.   Time evolvement and sample autocorrelation function of the realized volatility for IBM, 2001-2006.
Q XFGrvtsacf

shows that the ACF decays relatively slowly providing hints on the existence of long range dependence. Indeed, a common finding is that the realized volatility processes reveal long range dependence which is well captured by fractionally integrated processes. In particular, if RV_n is integrated of the order $d \in (0, 0.5)$, it can be shown that

$$\mathsf{Var}\left[\sum_{j=1}^{h} RV_{n+j}\right] \approx ch^{2d+1}, \tag{13.15}$$

with c denoting a constant. Then, plotting $\ln \mathsf{Var}\left[\sum_{j=1}^{h} RV_{n+j}\right]$ against $\ln h$ should result in a straight line with slope $2d + 1$. Most empirical studies

strongly confirm this relationship and find values for d between 0.35 and 0.4 providing clear evidence for long range dependence. Estimating d using the Geweke and Porter-Hudak estimator, we obtain $\widehat{d} = 0.38$ for the series of realized volatilities and $\widehat{d} = 0.62$ for its logarithmic counterpart. Hence, for both series we find clear evidence for long range dependence. However, the persistence in logarithmic realized volatilities is remarkably high providing even hints on non-stationarity of the process.

Summarizing the most important empirical findings, we can conclude that the unconditional distributions of logarithmic realized volatility and of correspondingly standardized log returns are well approximated by normal distributions and that realized volatility itself follows a long memory process. These results suggest (Gaussian) ARFIMA models as valuable tools to model and to predict (log) realized volatility.

13.4 Realized Volatility Models

As illustrated above, realized volatility models should be able to capture the strong persistence in the sample autocorrelation function. While this seemingly long-memory pattern is widely acknowledged, there is still no consensus on the mechanism generating it. One approach is to assume that the long memory is generated by a fractionally integrated process as originally introduced by Granger and Joyeux (1980) and Hosking (1981). In the GARCH literature this has lead to the development of the fractionally integrated GARCH model as, e.g., proposed by Baillie et al. (1996). For realized volatility the use of a fractionally integrated autoregressive moving average (ARFIMA) process was advocated, for example, by Andersen et al. (2003). The ARFIMA(p, q) model is given by

$$\phi(L)(1 - L)^d (y_n - \mu) = \psi(L)u_n, \tag{13.16}$$

with $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$, $\psi(L) = 1 + \psi_1 L + \ldots \psi_q L^q$, and d denoting the fractional difference parameter. Moreover, u_n is usually assumed to be a Gaussian white noise process, and y_n denotes either the realized volatility (see Koopman et al. (2005)) or its logarithmic transformation. Several extensions of the realized volatility ARFIMA model have been proposed, accounting, for example, for leverage effects (see Martens et al. (2004)), for non–Gaussianity of (log) realized volatility or for time-variation in the volatility of realized volatility (see Corsi et al. (2008)). Generally the empirical results show significant improvements in the point forecasts of volatility when using ARFIMA rather than GARCH-type models.

An alternative model for realized volatility has been suggested by Corsi (2004). The so-called heterogeneous autoregressive (HAR) model of realized volatility approximates the long-memory pattern by a sum of multi-period volatility components. The simulation results in Corsi (2004) show, that the HAR model can quite adequately reproduce the hyperbolic decay in the sample autocorrelation function of realized volatility even if the number of volatility components is small. For the HAR model, let the k–period realized volatility component be defined by the average of the single-period realized volatilities, i.e.,

$$RV_{n+1-k:n} = \frac{1}{k} \sum_{j=1}^{k} RV_{n-j}. \tag{13.17}$$

The HAR model with the so-defined daily, weekly and monthly realized-volatility components, is given by

$$
\begin{aligned}
\log RV_n &= \alpha_0 + \alpha_d \log RV_{n-1} + \alpha_w \log RV_{n-5:n-1} \\
&\quad + \alpha_m \log RV_{n-21:n-1} + u_n,
\end{aligned} \tag{13.18}
$$

with u_n typically being a Gaussian white noise. The HAR model has become very popular due to its simplicity in estimation and its excellent in-sample fit and predictive ability (see e.g. Andersen et al. (2004), Corsi et al. (2008)). Several extensions exist and deal, for example, with the inclusion of jump measures (see Andersen et al. (2004)) or non-linear specifications based on neural networks (see Hillebrand and Medeiros (2007)).

Alternative realized volatility models have been proposed in, e.g., Barndorff-Nielsen and Shephard (2002), who consider a superposition of Ornstein–Uhlenbeck processes, and in Deo et al. (2006), who specify a long-memory stochastic volatility model. A recent and comprehensive review on realized volatility models can also be found in McAleer and Medeiros (2008).

13.5 Time-Varying Betas

So far, our discussion focused on the measurement and modeling of the volatility of a financial asset using high-frequency transaction data. From a pricing perspective, however, systematic risk is most important. In this section, we therefore discuss, how high-frequency information can be used for the evaluation and modeling of systematic risk. A common measure for the systematic risk is given by the so-called (market) beta, which represents the sensitivity of a financial asset to movements of the overall market. As the beta plays a

crucial role in asset pricing, investment decisions, and the evaluation of the performance of asset managers, a precise estimate and forecast of betas is indispensable. While the unconditional capital asset pricing model implies a linear and stable relationship between the asset's return and the systematic risk factor, i.e., the return of the market, empirical results suggest that the beta is time-varying, see, for example, Bos and Newbold (1984), Hafner and Herwartz (1973), and Fabozzi and Francis (1978). Similar evidence has been found for multi-factor asset pricing models, where the factor loadings seem to be time-varying rather than constant. A large amount of research has therefore been devoted to conditional CAPM and APT models, which allow for time-varying factor loadings, see, for example, Dumas and Solnik (1995), Ferson and Harvey (1991), Ferson and Harvey (1993), and Ferson and Korajcyzyk (1995).

13.5.1 The Conditional CAPM

Below we consider the general form of the conditional CAPM. A similar discussion for multi-factor models can be found in Bollerslev and Zhang (2003). Assume that the continuously compounded return of a financial asset i from period n to $n+1$ is generated by the following process

$$r_{i;n+1} = \alpha_{i;n+1|n} + \beta_{i;n+1|n} r_{m;n+1} + u_{i;n+1}, \tag{13.19}$$

with $r_{m;n+1}$ denoting the excess market return and $\alpha_{n+1|n}$ denoting the intercept that may be time-varying conditional on the information set available at time n, as indicated by the subscript. The idiosyncratic risk u_{n+1} is serially uncorrelated, $\mathsf{E}_n(u_{n+1}) = 0$, but may exhibit conditionally time-varying variance. Note that $\mathsf{E}_n(\cdot)$ denotes the expectation conditional on the information set available at time n. Moreover, we assume that $\mathsf{E}(r_{m;n+1}u_{n+1}) = 0$ for all n. The conditional beta coefficient of the CAPM regression (13.19) is defined as

$$\beta_{i;n+1|n} = \frac{\mathsf{Cov}(r_{i;n+1}, r_{m;n+1})}{\mathsf{Var}(r_{i;n+1})}. \tag{13.20}$$

Now, assume that lending and borrowing at a one-period risk-free rate $r_{f;n}$ is possible. Then, the arbitrage-pricing theory implies that the conditional expectation of the next period's return at time n is given by

$$\mathsf{E}_\mathsf{n}(r_{i;n+1}) = r_{f;n} + \beta_{i;n+1|n}\mathsf{E}_\mathsf{n}(r_{m;n+1}). \tag{13.21}$$

Thus, the computation of the future return of asset i requires to specify how the beta coefficient evolves over time.

The most common approach to allow for time-varying betas is to re-run the CAPM regression in each period based on a sample of 3 or 5 years. We refer to this as the rolling regression (RR) method. More elaborate estimates of the beta can be obtained using the Kalman-filter, which builds on a state-space representation of the conditional CAPM or by specifying a dynamic model for the covariance matrix between the return of asset i and the market return.

13.5.2 Realized Betas

The evaluation of the in-sample fit and predictive ability of various beta models is also complicated by the unobservability of the true beta. Consequently, model comparisons are usually conducted in terms of implied pricing errors, i.e., $e_{i,n+1} = \widehat{r}_{i,n+1} - r_{i,n+1}$, with $\widehat{r}_{i,n+1} = r_{f;n} + \widehat{\beta}_{i;n+1|n} \, \mathsf{E}_n(r_{m;n+1})$. Owing to the discussion on the evaluation of volatility models, the question arises, whether high-frequency data may also be useful for the evaluation of competing beta estimates. The answer is a clear "yes". In fact, high-frequency based estimates of betas are quite informative for the dynamic behavior of systematic risk. The construction of so-called *realized betas* is straightforward and builds on realized covariance and realized volatility measures. In particular, denote the realized volatility of the market by $RV_{m;n}$ and the realized covariance between the market and asset i by $RCov_{m,i;n} = \sum_{j=1}^{M} r_{i,t_j} r_{m,t_j}$, where r_{i,t_j} and r_{m,t_j} denote the j-th high-frequency return of the asset and the market, respectively, during day n. The realized beta is then defined as

$$\widehat{\beta}_{HF;i;n} = \frac{RCov_{m,i;n}}{RV_{m;n}}. \tag{13.22}$$

Barndorff-Nielsen and Shephard (2004) show that the realized beta converges almost surely for all n to the integrated beta over the time period from $n-1$ to n, i.e., the daily systematic risk associated with the market index. Note that the realized beta can also be obtained from a simple regression of the high-frequency returns of asset i on the high-frequency returns of the market, see, e.g., Andersen et al. (2006). The preciseness of the realized beta estimator can easily be assessed by constructing the $(1-\alpha)$-percent confidence intervals, which have been derived in Barndorff-Nielsen and Shephard (2004) and are given by

$$\widehat{\beta}_{HF;i;n} \pm z_{\alpha/2} \sqrt{\left(\sum_{j=1}^{M} r_{m,t_j}^2 \right)^{-2} \widehat{g}_{i;n}}, \tag{13.23}$$

where $z_{\alpha/2}$ denotes the $(\alpha/2)$-quantile of the standard normal distribution,

$$\widehat{g}_{i;n} = \sum_{j=1}^{M} x_{i;j}^2 - \sum_{j=1}^{M-1} x_{i;j} x_{i;j+1}, \tag{13.24}$$

and

$$x_{i;j} = r_{i,t_j} r_{m,t_j} - \widehat{\beta}_{HF;i;n} r_{m,t_j}^2. \tag{13.25}$$

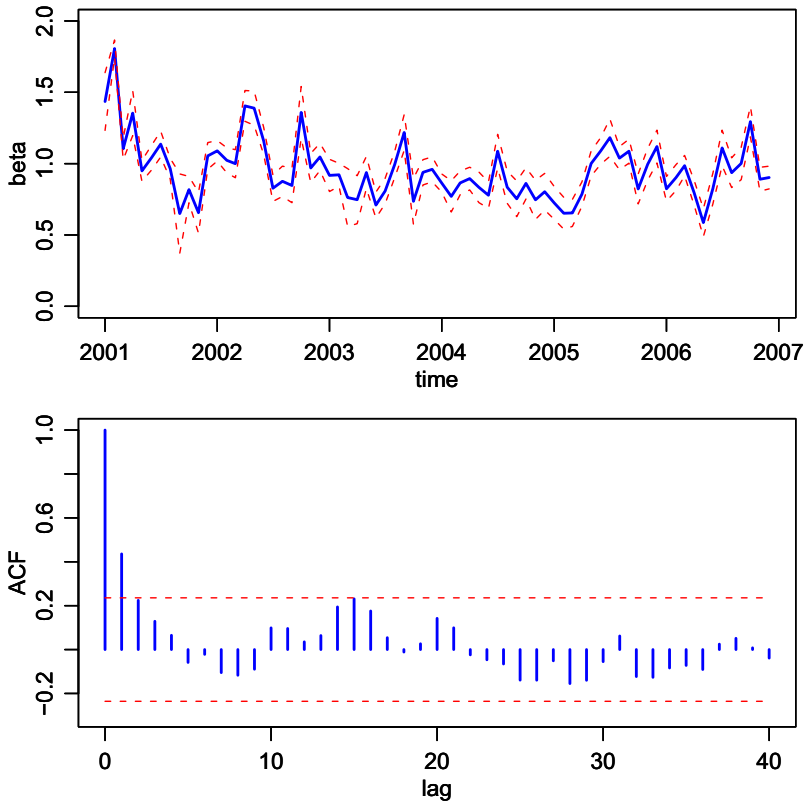The upper panel in Figure 13.4 presents the time-evolvement of the monthly

Figure 13.4. Time evolvement and sample autocorrelation function of monthly realized betas for IBM, 2001-2006. The dashed lines in the upper panel present the 95% confidence intervals of the realized beta estimator as given in (13.23). The dashed lines in the lower panel depict the 95% Bartlett confidence intervals. XFGbetatsacf

realized beta for IBM incorporation over the period ranging from 2001 to 2006. We use the Dow Jones Industrial Average Index as the market index and construct the realized betas using 30 minute returns. The graph also shows the 95%-confidence intervals of the realized beta estimator. The time-varying nature of systematic risk emerges strikingly from the figure and provides once more evidence for the relevance of its inclusion in asset pricing models.

Interestingly, the sample autocorrelation function of the realized betas depicted in the lower panel of Figure 13.4 indicates significant serial correlation over the short horizon. This dependency can be explored for the prediction of systematic risk. Bollerslev and Zhang (2003), for example, find that an autoregressive model for the realized betas outperforms the RR approach both in terms of forecast accuracy as well as in terms of pricing errors.

13.6 Summary

We review the usefulness of high-frequency data for measuring and modeling actual volatility at a lower frequency, such as a day. We present the realized volatility as an estimator of the actual volatility along with the practical problems arising in the implementation of this estimator. We show that market microstructure effects induce a bias to the realized volatility and we discuss several approaches for the alleviation of this problem. The realized volatility is a more precise estimator of the actual volatility than the conventionally used daily squared returns, and thus provides more accurate information on the distributional and dynamic properties of volatility. This is important for many financial applications, such as asset pricing, portfolio allocation or risk management. As a consequence, several modeling approaches for realized volatility exist and have been shown to usually outperform traditional GARCH or stochastic volatility models, both in terms of in-sample as well as out-of-sample performance. We further demonstrate the usefulness of the realized variance and covariance estimator for measuring and modeling systematic risk. For the empirical examples provided in this chapter we use tick-by-tick transaction data of one stock of the IBM incorporation and of the DJIA index.

Bibliography

P. R. Hansen and A. Lunde (2006). Realized Variance and Market Microstructure Noise, *Journal of Business & Economic Statistics* **24**: 127-161.

B. Zhou (1996). High-Frequency Data and Volatility in Foreign-Exchange Rates, *Journal of Business & Economic Statistics* **14**: 45-52.

Ole E. Barndorff-Nielsen, Peter R. Hansen, Asger Lunde and Neil Shephard (2008). Designing Realised Kernels to Measure the ex-post Variation of Equity Prices in the Presence of Noise, *Econometrica*, forthcoming.

L. Zhang, P. A. Mykland and Y. Ait-Sahalia(1005). A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data, *Journal of the American Statistical Association* **100**(472): 1394-1411.

R. Roll (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market, *Journal of Finance* **39**: 1127-1139.

O. E. Barndorff-Nielsen and N. Shephard (2002). Estimating Quadratic Variation Using Realized Variance, *Journal of Applied Econometrics* **4,5**: 457-477.

D. S. Bates (2000). Post-'87 Crash Fears in the S&P 500 Futures Option, *Journal of Econometrics* **94,1-2**: 181-238.

S. L. Heston, (1993). A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *The Review of Financial Studies* **6**(2): 327-343.

T. G. Andersen and T. Bollerslev, (1997). Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-Run in High Frequency Returns, *The Journal of Finance* **52**: 975-1005.

T. G. Andersen, T. Bollerslev, F. X. Diebold and P. Labys (2001). The Distribution of Realized Exchange Rate Volatility, *Journal of the American Statistical Association* **96**: 42-55.

T. G. Andersen, T. Bollerslev, F. X. Diebold and P. Labys (2003). Modeling and Forecasting Realized Volatility, *Econometrica* **71**: 579-625.

T. G. Andersen, T. Bollerslev, F. X. Diebold and J. Wu, (2006). Realized Beta: Persistence and Predictability, *Advances in Econometrics: Econometric Analysis of Economic and Financial Time Series, Volume B* **Editor: T.Fomby**: 1-40.

T. G. Andersen and T. Bollerslev (1998). Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts, *International Economic Review* **39**: 885-905.

T. G. Andersen, T. Bollerslev and F. X. Diebold, (2008). Roughing it Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility, *Review of Economics and Statistics*, forthcoming.

Torben G. Andersen, Tim Bollerslev and Xin Huang (2006). A Semiparametric Framework for Modeling and Forecasting Jumps and Volatility in Speculative Prices, *Duke University, Working Paper*.

R. T. Baillie, T. Bollerslev and H. O. Mikkelsen (1996). Fractionally Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* **74**: 3-30.

F. M. Bandi, J. R. Russell and J. Zhu (2008). Using High-Frequency Data in Dynamic Portfolio Choice, *Econometric Reviews*, forthcoming.

O. E. Barndorff-Nielsen and N. Shephard(2002). Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models, *Journal of the Royal Statistical Society, Series B* **64**: 253-280.

O. E. Barndorff-Nielsen and N.S hephard (2004). Econometric Analysis of Realized Co-variation: High Frequency Based Covariance, Regression, and Correlation in Financial Economics, *Econometrica* **72**: 885-925.

G. H. Bauer and K. Vorkink (2007). Multivariate Realized Stock Market Volatility, *Bank of Canada, Working Paper.*

T. Bollerslev, U. Kretschmer, C. Pigorsch, and G. Tauchen(2008). A Discrete-Time Model for Daily S&P500 Returns and Realized Variations: Jumps and Leverage Effects, *Journal of Econometrics*, forthcoming.

M. Blume (1975). Betas and Their Regression Tendencies, *Journal of Finance* **30**: 785-95.

T. Bollerslev, R. F. Engle, J. M. Wooldridge (1988). A Capital Asset Pricing Model with Time Varying Covariances, *Journal of Political Economy* **96**: 116-131.

T. Bollerslev and B. Y. B. Zhang (2003). Measuring and Modeling Systematic Risk in Factor Pricing Models Using High-frequency Data, *Journal of Empirical Finance* **10**: 533-558.

T. Bos and P. Newbold(1984). An Empirical Investigation of the Possibility of Stochastic Systematic Risk in the Market Model, *Journal of Business* **57**: 35-41.

P. A. Braun, D. B. Nelson and A. M. Sunier (1995). Good News, Bad News, Volatility, and Betas, *Journal of Business* **50**: 1575-1603.

R. Chiriac and V. Voev (2007). Long Memory Modelling of Realized Covariance Matrices, *University of Konstanz, Working Paper.*

F. Corsi, S. Mittnik, C. Pigorsch and U. Pigrosch (2008). The Volatility of Realized Volatility, *Econometric Reviews* **27**: 46-78.

D. W. Collins, J. Ledolter and J. Rayburn (1987). Some Further Evidence on the Stochastic Properties of Systematic Risk, *Journal of Business* **60**: 425-448.

F.Corsi (2004). A Simple Long Memory Model of Realized Volatility, *University of Southern Switzerland, Woking Paper.*

Z. Da and E. Schaumburg (2006). The Factor Structure of Realized Volatility and its Im-plications for Option Pricing, *University of Notre Dame, Working Paper.*

R. Deo, C. Hurvich and Y. Lu (2006). Forecasting Realized Volatility Using a Long-Memory Stochastic Volatility Model: Estimation, Prediction and Seasonal Adjustment, *Journal of Econometrics* **131**(1-2): 29-58.

F. X. Diebold and A. Inoue (2001). Long Memory and Regime Switching , *Journal of Econometrics* **2001**: 131-159.

B. Dumas and B. Solnik (1995). The World Price of Exchange Rate Risk, *Journal of Finance* **50**: 445-480.

R. F. Engle and G. G. J. Lee (1999). A Permanent and Transitory Component Model of Stock Return Volatility, *Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger*: 475-497.

F. J. Fabozzi and J. C. Francis (1978). Beta as a Random Coefficient, *Journal of Financial and Quantitative Analysis* **13**: 101-116.

R. W. Faff, D.Hillier and J. Hillier, (2000). Time Varying Beta Risk: An Analysis of Alter-native Modelling Techniques, *Journal of Business, Finance and Accounting* **27**: 523-554

W. E. Ferson and C. R. Harvey (1991). The Variation of Economic Risk Premiums, *Journal of Poltical Economy* **99**: 385-415.

W. E. Ferson and C. R. Harvey (1993). The Risk and Predictability of International Equity Returns, *Review of Financial Studies* **6**527-566.

W. E. Ferson and R. A. Korajczyk (1995). Do Arbitrage Pricing Models Explain the Predictability of Stock Returns?, *Journal of Business* **68**: 309-349.

C. Gourieroux and J. Jasiak (2001). Financial Econometrics: Problems, Models, and Methods, Princeton University Press, Princeton

C. W. J. Granger (1980). Long Memory Relationships and the Aggregation of Dynamic Models, *Journal of Econometrics* **14**: 227-238.

C. W. J. Granger and N. Hyung (2004). Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns, *Journal of Empirical Finance* **11**: 399-421.

C. W. J. Granger and R. Joyeux (1980). An Introduction to Long-Range Time Series Models and Fractional Differencing, *Journal of Time Series Analysis* **1**: 15-30.

C. W. J. Granger and T. Teräsvirta (1999). A Simple Nonlinear Time Series Model with Misleading Linear Properties, *Economic Letters* **62**: 161-165.

E. Hillebrand and M. Medeiros (2007). Forecasting Realized Volatility Models: the Benefits of Bagging and Non-Linear Specifications, *Louisana State University, Working Paper*.

J. R. M. Hosking (1981). Fractional Differencing, *Biometrika* **68**: 165-176.

N. L. Jacob (1971). The Measurement of Systematic Risk for Securities and Portfolios: Some Empirical Results, *Journal of Financial and Quantitative Analysis* **6**: 815-834.

S. J. Koopman, B. Jungbacker and E. Hall.(2005). Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised and Implied Volatility Measurements, *Journal of Empirical Finance* **12,3**: 445-475.

X. Li (2003). On Unstable Beta Risk and Its Modelling Techniques for New Zealand Industry Portfolios, *Massey University, Working Paper*.

T. Lux and M. Marchesi (1999). Scaling and Criticality in a Stochastic Multi-Agent Model of a Financial Market, *Nature* **397**: 498-500

M. Martens, D. van Dijk and M.dePooter (2004). Modeling and Forecasting S&P 500 Volatility: Long Memory, Structural Breaks and Nonlinearity, *Erasmus University Rotterdam, Working Paper*.

M. Martens and J. Zein (2004). Predicting Financial Volatility: High-Frequency Time-Series Forecasts vis-à-vis Implied Volatility, *Journal of Futures Markets* **11**: 1005-1028.

M. McAleer and M.Medeiros (2008). Realized Volatilty: A Review, *Econometric Reviews* **26**: 10-45.

M. McAleer and M.Medeiros (2006). A Multiple Regime Smooth Transition Heterogenous Autoregressive Model for Long Memory and Asymmetries, *Pontifical Catholic University of Rio de Janeiro, Working Paper*.

U. A. Müller, M. M. Dacorogna, R. D. Dav, R. B. Olsen, O. V. Pictet and J. E. von Weizsäcker(1997). Volatilities of Different Time Resolutions - Analyzing the Dynamics of Market Components, *Journal of Empirical Finance* **4**(2-3); 213-239.

T. Mikosch and C. Stărică (2004). Nonstationarities in Financial Time Series, the Long-range Dependence, and the IGARCH Effects, *The Review of Economics and Statistics* **86**: 378-390.

D. B. Nelson and D. P. Foster (1994). Asymptotic Filtering Theory for Univariate ARCH Models, *Econometrica* **62**: 1-41.

M. dePooter, M. Martens and D. van Dijk (). Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data - But Which Frequency to Use?, *Econoemtric Reviews*, forthcoming.

S. Pong, M. B. Shackleton, S. J. Taylor and X. Xu (). Forecasting Currency Volatility: A Comparison of Implied Volatilities and AR(FI)MA Models, *Journal of Banking & Finance* **28**: 2541-2563.

M. Scharth and M. Medeiros (2006). Asymmetric Effects and Long Memory in the Volatility of DJIA Stocks, *Pontifical Catholic University of Rio de Janeiro, Working Paper*.

S. Sunder (1980). Stationarity of Market Risk: Random Coefficients Tests for Individual Stocks, *Journal of Finance* **35**: 883-896.

D. D. Thomakos (2003). Realized Volatility in the Futures Markets, *Journal of Empirical Finance* **10**: 321-353.

V. Voev (2008). Dynamic Modelling of Large-Dimensional Covariance Matrices, in High Frequency Financial Econometrics, edited by L.Bauwens, W.Pohlmeier and D.Veredas, Physica-Verlag, Heidelberg: 293-312.

C. M. Hafner and H. Herwartz (1998). Time Varying Market Price of Risk in the CAPM-Approaches, Empirical Evidence and Implications, *Finance* **19**: 93-112.

P. K. Clark (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices, *Econometrica* **41**: 135-156.

14 Valuation of Multidimensional Bermudan Options

Shih-Feng Huang and Meihui Guo

14.1 Introduction

Multi-dimensional option pricing becomes an important topic in financial markets (Franke et al., 2008). Among which, the American-type derivative (e.g. the Bermudan option) pricing is a challenging problem. Unlike the European options which can only be exercised on the expiration date, the owner of a Bermudan option has the right to exercise early on a contractually specified finite set of dates. The dynamic programming approach is a practical and popular approach used to price the Bermudan option (Shreve, 2004, p.91). In that approach, the option value on each possible early exercise date is set to be the maximum of the payoff associated with immediate exercise, called the intrinsic value, and the discounted conditional expectation of the future option value, called the continuation value. The major problem of the approach lies in the computation of the continuation value.

In the literature, numerical methods, Barraquand and Martineau (1995) and Jeantheau (1998), and simulation based methods, Rust (1997), Tsitsiklis and Van Roy (1999), Longstaff and Schwartz (2001) and Broadie and Glasserman (2004), were proposed to solve this problem. Here we consider a dynamic semiparametric method to valuate multi-dimensional options. The proposed approach uses nonparametric step functions to approximate the option values on each possible early exercise date and evaluate the continuation values by parametric transition density. Unlike the simulation based method generating random sample paths, the proposed method selects the asset price points beforehand. And instead of numerically evaluating the multiple integral involved in computation of the continuation values, the proposed method provides closed form expressions for the integrals. Using this semiparametric technique, the proposed method provides a flexible and handy tool for multidimensional derivative pricing. Details of the dynamic semiparametric

method are given in Section 14.3. The computational effort of the method is linear in the number of exercise opportunities and quadratic in the number of partition points. In addition, it is easily implemented when the multivariate joint distributions of the underlying assets are modeled by copula functions (Nelsen, 2006), which are to be introduced in the next section.

Section 14.2 defines the model assumptions. The proposed approach for valuing multidimensional Bermudan option is introduced in Section 14.3. One dimensional Bermudan option pricing of Black-Scholes model, multidimensional Bermudan option of multivariate geometric Brownian processes and a real example are demonstrated in Section 14.4. Section 14.5 concludes.

14.2 Model Assumptions

Consider a Bermudan option on d-dimensional underlying assets. Assume the price of each underlying asset $S_{\ell,t}$ follows a risk-neutral geometric Brownian process:

$$\frac{dS_{\ell,t}}{S_{\ell,t}} = (r - q_\ell)dt + \sigma_\ell dW_{\ell,t}, \ \ell = 1, \cdots, d, \tag{14.1}$$

where q_ℓ and σ_ℓ are the continuously compounded dividend yield and the instantaneous volatility of the ℓth asset, respectively, $W_{\ell,t}$'s are Wiener processes and the dependence among the $W_{\ell,t}$'s will be modeled by copula function introduced below. Let $\mathbf{X}_t = (X_{1,t}, \cdots, X_{d,t})^\top$ be the standardized log price per strike price, that is, $X_{\ell,t} = \log(S_{\ell,t}/K)$, $\ell = 1, \cdots, d$. Thus the (conditional) marginal distribution of $X_{\ell,t}$ is $N(X_{\ell,0} + (r - q_\ell - \frac{1}{2}\sigma_\ell^2)t, \sigma_\ell^2 t)$. We will use copula functions to connect the asset marginals to their joint distribution. Since copula functions provide a flexible methodology for modeling of multivariate asset dependence, it has recently become a popular technique in financial markets, Sklar (1959), Cherubini et al. (2004), Nelsen (2006) and Giacomini et al. (2007). Let $F_\ell(x_\ell)$, $\ell = 1, \cdots, d$ denote the marginal distribution of X_ℓ, throughout we assume the joint distribution of $(X_1, \cdots, X_d)^\top$, $F(x_1, \cdots, x_d)$, is modeled by a copula function C, that is

$$F(x_1, \cdots, x_d) = C\{F_1(x_1), \cdots, F_d(x_d)\}. \tag{14.2}$$

For example, the Gaussian copula has the form $C(u_1, \cdots, u_d) = \Phi_R\{\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_d)\}$, where Φ is the distribution of $N(0,1)$ and Φ_R is the standardized multivariate normal distribution with correlation matrix R. When the univariate X_j's are normally distributed, the Gaussian copula is corresponding to the multivariate normal distribution. If the correlation matrix R is the identity matrix, then the Gaussian copula becomes the independence copula,

implying that the random variables are independent. For example, in case $R = I_2$ the 2×2 identity matrix, by (14.2) and the definition of Gaussian copula, we have the joint distribution

$$
\begin{aligned}
F(x_1, x_2) &= \Phi_{I_2}\Big[\Phi^{-1}\{F_1(x_1)\}, \Phi^{-1}\{F_2(x_2)\}\Big] \\
&= \Phi\Big[\Phi^{-1}\{F_1(x_1)\}\Big]\Phi\Big[\Phi^{-1}\{F_2(x_2)\}\Big] \\
&= F_1(x_1)F_2(x_2).
\end{aligned}
$$

Copulae also provide a natural perspective to study the dependence in the tail of a multivariate distribution. For bivariate case, the lower tail dependence of X_1 and X_2 is defined as $\lambda_L = \lim_{v \to 0^+} \mathrm{P}\{F_2(X_2) \le v \mid F_1(X_1) \le v\} = \lim_{v \to 0^+} \frac{C(v,v)}{v}$. If $\lambda_L > 0$, then the two variables X_1 and X_2 are said to have lower tail dependence. Similarly, the upper tail dependence is defined as $\lambda_U = \lim_{v \to 1^-} \mathrm{P}\{F_2(X_2) > v \mid F_1(X_1) > v\} = \lim_{v \to 1^-} \frac{1 - 2v + C(v,v)}{1 - v}$. If $\lambda_U > 0$, then there exists upper tail dependence. Archimedean copulae such as the Clayton and Gumbel copulae are two popular functions used to model the tail dependence of data. The Clayton copula has the form

$$
C(u_1, \cdots, u_p) = \Big(\sum_{j=1}^{p} u_j^{-\theta} - p + 1\Big)^{\frac{-1}{\theta}}, \ \theta > 0,
$$

and the Gumbel copula is

$$
C(u_1, \cdots, u_p) = \exp\Big[-\Big\{\sum_{j=1}^{p}(-\log u_j)^{\theta}\Big\}^{\frac{1}{\theta}}\Big], \ \theta \ge 1.
$$

The lower tail dependence of the bivariate Clayton copula is $\lambda_L = 2^{-1/\theta} > 0$, and the upper tail dependence of the Gumbel copula is $\lambda_U = 2 - 2^{1/\theta} > 0$, for $\theta > 1$. Thus the Clayton and Gumbel copulae are usable to model assets with lower and upper tail dependence, respectively. On the contrary, the Gaussian copula has neither upper nor lower dependence, unless the correlation coefficient $\rho = 1$. In Figure 21.1, we plot the random samples of four bivariate copulae, independent and correlated Gaussian, Clayton and Gumbel copulae with $\mathrm{N}(0, 1)$ marginals. Although the marginals are the same in the four cases, the plots display different tail dependence. The top-left is the independent Gaussian copula, denoted by Gaussian(0). The top-right is the Gaussian copula with correlation 0.5, which is the same as the bivariate normal distribution with zero mean, unit variance and correlation 0.5. The bottom-left and bottom-right are the Clayton and Gumbel copulae with parameter $\theta = 2$, which show a lower tail dependence and a upper tail dependence, respectively.
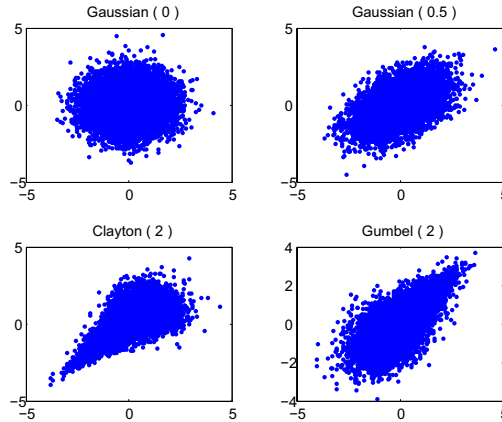
Figure 14.1. Bivariate copula plots with $N(0,1)$ marginals.
Output of 🔲XFGbicopula .

14.3 Methodology

In this section, we introduce the proposed semiparametric method to valuate a d-dimensional Bermudan option with expiration date T. Assume the Bermudan option can only be exercised at time t_i, $i = 1, \cdots, n$, where $0 = t_0 < t_1 < \cdots < t_n = T$ and for simplicity we assume t_i's are equidistant with constant interval length $\Delta = t_i - t_{i-1}$. Let V_i denote the time t_i value of the Bermudan option, $\mathbf{S}_i = (S_{1,i}, \cdots, S_{d,i})^\top$ be the corresponding d underlying asset values, and $g(\mathbf{S}_i)$ be the option payoff function. Then the no arbitrage option values on possible early exercise dates are

$$\begin{cases} V_n(\mathbf{S}_n) = g(\mathbf{S}_n) \text{ and} \\ V_i(\mathbf{S}_i) = \max\{g(\mathbf{S}_i), e^{-r\Delta}\,\mathsf{E}(V_{i+1} \mid \mathbf{S}_i)\}, \text{ if } i < n \end{cases}, \qquad (14.3)$$

where $r > 0$ is the riskless interest rate and $\mathsf{E}(\cdot|\mathbf{S}_i)$ is the conditional expectation under a risk-neutral probability measure given the information up to time t_i, Shreve (2004, p.91). In (14.3), the term $g(\mathbf{S}_i)$ is also called the *early exercise value* and $e^{-r\Delta}\,\mathsf{E}(V_{i+1}|\mathbf{S}_i)$ is the *continuation value* at time t_i. The Bermudan option will be exercised at time t_i if $g(\mathbf{S}_i) \geq e^{-r\Delta}\,\mathsf{E}(V_{i+1}|\mathbf{S}_i)$, and will be held continuously if $g(\mathbf{S}_i) < e^{-r\Delta}\,\mathsf{E}(V_{i+1}|\mathbf{S}_i)$.

The objective is to derive the initial option value $V_0(S_0)$, the main difficulty arises from evaluation of the continuation value. For instance, consider a univariate Bermudan put option on an underlying asset without paying dividend, $q = 0$, with payoff function $(K - S_n)^+$. Under geometric Brownian motion

model assumption, the continuation value at time t_{n-1} can be obtained by the following Black and Scholes (1973) formula,

$$e^{-r\Delta}\,\mathsf{E}[(K - S_n)^+ \mid S_{n-1}] = Ke^{-r\Delta}\Phi(-d_2) - S_{n-1}\Phi(-d_1),$$

where $d_1 = \frac{\log(S_{n-1}/K)+(r+\frac{\sigma^2}{2})\Delta}{\sigma\sqrt{\Delta}}$ and $d_2 = d_1 - \sigma\sqrt{\Delta}$. Thus the continuation value at time t_{n-2} is

$$e^{-r\Delta}\,\mathsf{E}(V_{n-1} \mid S_{n-2})$$
$$= e^{-r\Delta}\,\mathsf{E}\left(\max\{(K - S_{n-1})^+, Ke^{-r\Delta}\Phi(-d_2) - S_{n-1}\Phi(-d_1)\} \mid S_{n-2}\right),$$

which is difficult to evaluate and has no closed-form solution. As the time move backwards to time t_0, the problem becomes more knotty. To handle the problem, we use step functions to approximate the option value at time t_n, $V_n(\mathbf{S}_n)$ defined in (14.3). Since the conditional joint distribution of $\mathbf{X}_t$ given $\mathbf{X}_{t-1}$ is modeled by the copula function $C\{F_1(X_{1,t}|X_{1,t-1}), \cdots, F_d(X_{d,t}|X_{d,t-1})\}$, it is relatively easy to evaluate the continuation value at time t_{n-1}. Accordingly, we define the approximate option value at time t_{n-1} to be the maximum of the intrinsic value and this continuation value. Continue the procedure backwards to t_0, we can obtain the initial option value. The proposed procedure uses a dynamic semiparametric approach, which incorporates nonparametric step function approximation and parametric model assumption, to tackle the difficult multiple integral computation involved in the high-dimensional derivative pricing problem. The details of the procedure is given below.

First, we confine the space of $\mathbf{X}_T$ to a proper finite region, say ± 5 standard deviation region of a given initial value $\mathbf{X}_0$, and then partition the region with equidistant grid points, denoted by $\mathbf{x}^{(j)} = (x_1^{(j)}, \cdots, x_d^{(j)})^\top$, $j = 1, \cdots, N$. The distance between two adjacent points in each dimension is denoted by Δ_x (see Figure 21.2 for the two-dimensional case). We keep the partition length Δ_x constant throughout the time. Start from the time point $i = n$, we use $\widetilde{V}_i(\cdot)$ to denote the approximate option function at time t_i, and set $\widetilde{V}_n(\mathbf{x}^{(j)}) = g(\mathbf{x}^{(j)})$ on the expiration date. The proposed steps to compute the d-dimensional Bermudan option are:

(1) Set the grid $\mathbf{A}^{(j)} = \prod_{\ell=1}^d [x_\ell^{(j)} - (1 - c)\Delta_x, x_\ell^{(j)} + c\Delta_x], j = 1, \cdots, N$ (see Figure 21.2 for the two-dimensional case) and c is a pre-chosen constant. Based on the grids $\{\mathbf{A}^{(j)}\}_{j=1}^N$, define the step function

$$\widehat{V}_i = \sum_{j=1}^N \widetilde{V}_i(\mathbf{x}^{(j)})\mathbf{1}\{\mathbf{X}_i \in \mathbf{A}^{(j)}\}.$$
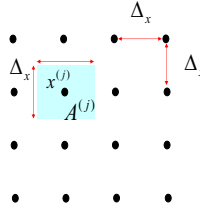
Figure 14.2. Two-dimensional grid points.

The constant c is chosen to meet the criterion that the European option values derived from this scheme are close to the benchmarks. In which the European benchmark option values can either be obtained analytically or by Monte Carlo simulation. For instance, the option on a geometric average for multivariate normal distributed underlying assets, the benchmark can be obtained by Black-Scholes formula since it can be reduced to a one-dimensional problem (for details see example 14.2).

(2) Compute the continuation value at time t_{i-1} given $\mathbf{X}_{i-1} = \mathbf{x}^{(h)}$ by

$$\mathsf{E}(\widehat{V}_i|\mathbf{X}_{i-1}) = \sum_{j=1}^{N} \widetilde{V}_i(\mathbf{x}^{(j)})\mathsf{P}(\mathbf{X}_i \in \mathbf{A}^{(j)} \mid \mathbf{X}_{i-1} = \mathbf{x}^{(h)}) = \mathbf{P}_h \widetilde{\mathbf{V}}_i,$$

where $\mathbf{P}_h$ is the hth row of the transition matrix $\mathbf{P} = (p_{hj})_{N \times N}$ with

$$
\begin{aligned}
p_{hj} &= \mathsf{P}(\mathbf{X}_i \in \mathbf{A}^{(j)} \mid \mathbf{X}_{i-1} = \mathbf{x}^{(h)}) \\
&= \sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1 + \cdots + i_d} C(u_{1i_1}, \cdots, u_{di_d}),
\end{aligned}
$$

where C is the copula, $u_{\ell 1} = F_\ell(x_\ell^{(j)} - (1-c)\Delta_x \mid x_\ell^{(h)})$ and $u_{\ell 2} = F_\ell(x_\ell^{(j)} + c\Delta_x \mid x_\ell^{(h)})$ for all $\ell = 1, \cdots, d$, and $\widetilde{\mathbf{V}}_i = \left(\widetilde{V}_i(\mathbf{x}^{(1)}), \cdots, \widetilde{V}_i(\mathbf{x}^{(N)})\right)^\top$ is the approximate option value at time t_i, McNeil et al. (2005). Note that the transition matrix $\mathbf{P}$ is the same for $i = 1, \cdots, n$.

(3) The approximate option value at time t_{i-1} given $\mathbf{X}_{i-1} = \mathbf{x}^{(h)}$ is obtained by $\widetilde{V}_{i-1}(\mathbf{x}^{(h)}) = \max\{g(\mathbf{x}^{(h)}), e^{-r\Delta}\mathbf{P}_h\widetilde{\mathbf{V}}_i\}$. Note that if the interest is to valuate a European option, then just set $\widetilde{V}_{i-1}(\mathbf{x}^{(h)}) = e^{-r\Delta}\mathbf{P}_h\widetilde{\mathbf{V}}_i$.

(4) If $i - 1 = 0$, then stop; otherwise set $i = i - 1$ and return to (1).

Since the proposed method performs iterative matrix vector multiplication at each time t_i, its computational effort is linear in the number of exercise
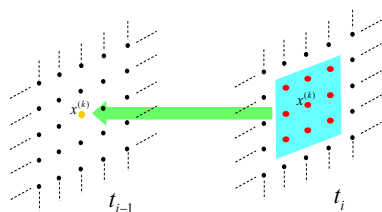
Figure 14.3. Non-zero elements in the two-dimensional case.

opportunities n. At each time t_i, on account of the matrix size $(N \times N)$, the computational work of matrix multiplication is quadratic in the total number of grid points N. Although the size $(N \times N)$ of the transition matrix $\mathbf{P}$ gets large as either the maturity time T or the dimension d of the underlying assets increases, lots (most) of it elements are zeros. This is due to the reason that the transition probabilities are negligible for far apart grid points, say more greater than five standard deviations (see Figure 21.3 for the two-dimensional example). Specifically, the row length (N) of the transition matrix $\mathbf{P}$ is of order $\mathcal{O}(T^{d/2})$ and the number of nonzero entries of each row is of order $\mathcal{O}(\Delta^{d/2}) = \mathcal{O}((\frac{T}{n})^{d/2})$, as a result the ratio of non-zero elements of $\mathbf{P}$ is of order $\mathcal{O}(n^{-d/2})$. In another word, the transition matrix $\mathbf{P}$ is a sparse matrix populated primarily with zeros.

When storing and manipulating sparse matrices on a computer, we can utilize specialized algorithms and data structures, eg. the SPARSE routine of MATLAB, to save the computation time and to consume less memory. Furthermore, since the partition grid points of the d-dimensional asset price space are determined in advance and kept fixed, the transition matrix remains unchanged throughout the time, which contrasts sharply with the time varying transition matrix used in simulation based approach. In the simulation based method, e.g. see Rust (1997) and Broadie and Glasserman (2004), random samples are generated by Monte Carlo method at each time t_i, and the continuation value at time t_{i-1} given the kth random sample, $\mathbf{S}_{i-1}^{(k)}$, is approximated by $\sum_{j=1}^{N} w_i^{(k,j)} V_i(\mathbf{S}_i^{(j)})$, where $w_i^{(k,j)}$ determines the stochastic weights of the sample at time t_i. In Figure 21.4 and 21.5, we illustrate the design grid points of the proposed scheme and the random samples of the simulation based method, respectively. Let $\mathbf{P}^{(i)} = (w_i^{(k,j)})_{k,j}$, the matrix of stochastic weights, then the continuation value at time t_{i-1} can be viewed as a matrix multiplication of the option value at time t_i, and the matrix $\mathbf{P}^{(i)}$ varies as time changes.
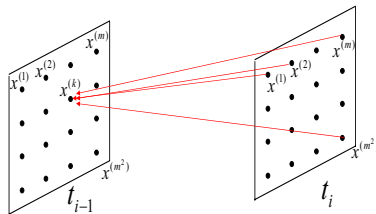
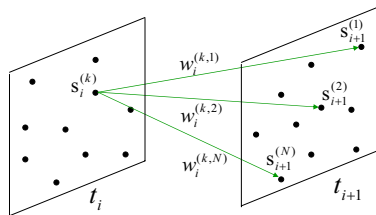Figure 14.4. The designed points of the proposed scheme.

Figure 14.5. The random samples of the simulation based method.

14.4 Examples

In this section, we demonstrate three simulated examples (example 14.1-14.3) and one real application (example 14.4) to valuate Bermudan options by the proposed scheme.

EXAMPLE 14.1 *Suppose the underlying asset satisfies the following risk-neutral geometric Brownian motion*

$$dS_t = (r - q)S_t dt + \sigma S_t dW_t, \tag{14.4}$$

where $r = 0.08$, $\sigma = 0.2$ and $q = 0$, 0.04, 0.08 or 0.12. Consider a one dimensional Bermudan put option with strike price $K = 100$, time to maturity $T = 3$, length of time interval $\Delta = \frac{1}{52}$ (i.e. $n = 156$) and payoff function $g(S_t) = (K - S_t)^+$. In Figure 21.6, we plot a simulated path of $\{S_t\}$ satisfying (14.4) with $r = 0.08$, $\sigma = 0.2$, $q = 0$, $T = 3$, $\Delta = \frac{1}{52}$ and the initial stock price $S_0 = K = 100$. The stock price at the maturity date is 117. Thus the payoff is 17 at time T in this realization. At each time $t < T$, the owner of this option would exercise early only when the payoff is positive, i.e. $S_t > K$, and would hold the option continuously when $S_t < K$. If $S_t > K$, then she needs to compute the continuation value of her option in order to decide whether exercising immediately or not.
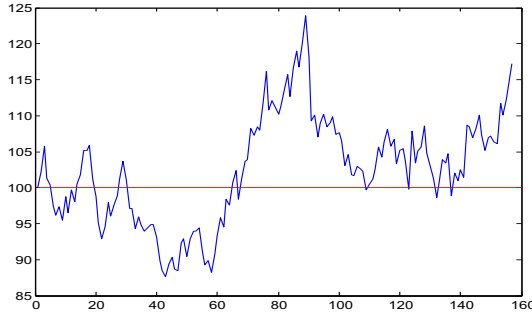
Figure 14.6. A simulated path of the stock price process satis-
fying (14.4) with $r = 0.08$, $\sigma = 0.2$, $q = 0$, $T = 3$, $\Delta = \frac{1}{52}$ and
the initial stock price $S_0 = K = 100$. Output of 🔍 XFGstock
.

Let $X_t = \log(S_t/K)$ denote the standardized log price per strike price and
let $\{x^{(j)}\}_{j=1}^{401}$ denote the 401 pre-chosen equidistant grid points of X_t, where
$x^{(1)} = X_0 - 5\sigma\sqrt{T}$ and $x^{(401)} = X_0 + 5\sigma\sqrt{T}$, that is the distance between
two adjacent points is $\Delta_x = 0.0087$ and $X_0 = x^{(201)}$. In the following, we
illustrate the procedure to compute the approximate option values backwards
from time t_{155} to t_{154}. First the continuation values of $x^{(j)}$ at time t_{155},
$e^{-0.0015}\,\mathsf{E}(V_{155} \mid x^{(j)})$, are derived by the Black-Scholes formula, and the option
values of $x^{(j)}$'s are obtained by

$$\widetilde{V}_{155}(x^{(j)}) = \max\{100 - 100\exp(x^{(j)}),\, e^{-0.0015}\,\mathsf{E}(V_{156} \mid x^{(j)})\},$$

$j = 1, \cdots, 401$. In Figure 21.7, we show the evolution of the intrinsic and
continuation values at t_i in (a) to the approximate option value at t_{i-1} in
(b). In Figure 21.7 (a), the green line is the intrinsic value, the red dash
curve is the continuation value and the intersection of the green line and the
red curve represents the early exercised boundary of the Bermuda option. In
Figure 21.7 (b), the blue curve is the approximate option value, $\widetilde{V}_{i-1}$. Define
the following step function

$$\widehat{V}_{155} = \sum_{j=1}^{401} \widetilde{V}_{155}(x^{(j)})\mathbf{1}_{\{X_{155} \in A^{(j)}\}},$$

where $A^{(j)} = [x^{(j)} - (1-c)\Delta_x, x^{(j)} + c\Delta_x]$ and c is chosen to let the European
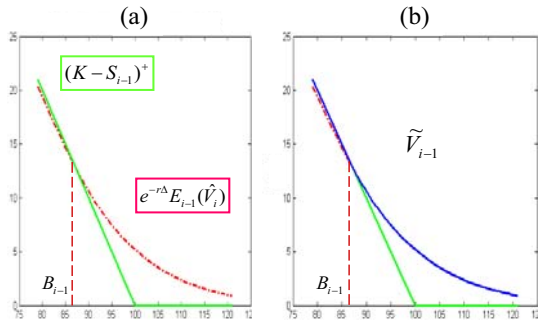option price of X_0 computed by proposed method meets that of the Black-

Figure 14.7. (a) The intrinsic and continuation values at t_i
(b) The approximate option value, $\widetilde{V}_{i-1}$, at t_{i-1}.

Scholes formula. The continuation value of $x^{(h)}$ at time t_{154} is given by

$$
e^{-0.0015}\, \mathsf{E}(\widehat{V}_{155} \mid x^{(h)}) \;=\; e^{-0.0015} \sum_{j=1}^{401} \widetilde{V}_{155}(x^{(j)}) \mathrm{P}(X_{155} \in A^{(j)} \mid x^{(h)})
$$
$$
=\; e^{-0.0015} \mathbf{P}_h \widetilde{\mathbf{V}}_{155},
$$

where $\mathbf{P}_h$ is the hth row of the transition matrix $\mathbf{P} = (p_{hj})_{401 \times 401}$ with

$$
p_{hj} \;=\; \mathrm{P}(X_{155} \in A^{(j)} \mid x^{(h)})
$$
$$
=\; \Phi\big(\tfrac{x^{(j)}+c\Delta_x-x^{(h)}-(r-q-0.5\sigma^2)\Delta}{\sigma\sqrt{\Delta}}\big) - \Phi\big(\tfrac{x^{(j)}-(1-c)\Delta_x-x^{(h)}-(r-q-0.5\sigma^2)\Delta}{\sigma\sqrt{\Delta}}\big),
$$

$\Phi(\cdot)$ is the standard normal cumulative distribution function and $\widetilde{\mathbf{V}}_{155} = \left(\widetilde{V}_{155}(x^{(1)}), \cdots, \widetilde{V}_{155}(x^{(401)})\right)^{\top}$ are the approximate option values at time t_{155}. Therefore, the approximate option values of $x^{(j)}$'s at time t_{154} are $\widetilde{V}_{154}(x^{(j)}) = \max\{100 - 100\exp(x^{(j)}), e^{-0.0015}\, \mathsf{E}(\widehat{V}_{155} \mid x^{(j)})\}$, $j = 1, \cdots, 401$. Note that the transition matrix $\mathbf{P}$ remains unchanged throughout the time. Proceeding the above procedure backwards to time zero, one obtains the desired option value.

Table 14.1 presents the simulation results for different initial stock prices, $S_0 = 90, 100, 110$. In the table, we give the option prices obtained by the proposed method and the methods by Ju (1998), denoted as EXP3, by and Lai and AitSahalia (2001), denoted as LSP4. In approximating the early exercise boundary of the Bermuda option, Ju (1998) adopts multipiece exponential function and Lai and AitSahalia (2001) adopt a linear spline method. The

S_0		Bin.	LSP4	EXP3	Alg.
90	(1)	20.08	20.08	20.08	20.09
100	$q = 0.12$	15.50	15.51	15.50	15.50
110		11.80	11.81	11.80	11.81
90	(2)	16.21	16.20	16.20	16.21
100	$q = 0.08$	11.70	11.70	11.70	11.71
110		8.37	8.37	8.36	8.37
90	(3)	13.50	13.49	13.49	13.50
100	$q = 0.04$	8.94	8.94	8.93	8.95
110		5.91	5.91	5.90	5.92
90	(4)	11.70	11.70	11.69	11.69
100	$q = 0.00$	6.93	6.93	6.92	6.93
110		4.16	4.15	4.15	4.16

Table 14.1. Bermudan put values of example 14.1 with parameters $r = 0.08$, $\sigma = 0.20$, $K = 100$, $T = 3$ and $\Delta = 1/52$. Output of XFGBP1 .

values based on 10,000 steps of the binomial method are taken as the bench-mark option prices. The results show that our approach is competitive and comparable with the LSP4 and EXP3 methods.

EXAMPLE 14.2 *Assume now two underlying assets satisfying (14.1), i.e. $d = 2$, with parameters $r = 0.05$, $q_1 = q_2 = 0$, $\sigma_1 = \sigma_2 = 0.2$ and the joint distribution of the two log stock price processes is bivariate normal with correlation coefficient $\rho = 0.3$. Consider a Bermudan put option on a geometric average with $K = 100$, $T = 1$, $\mathbf{S}_0 = 100$, $\Delta = 1/12$ (i.e. $n = 12$) and payoff function $g(\mathbf{S}_t) = (K - \sqrt{S_{1,t}S_{2,t}})^+$. First, we confine the space of $\mathbf{X}_{12}$ to $[-0.57, 0.63]^2$, and partition the region with 25 equidistance partition points in each dimension, that is we have 625 two-dimensional grid points, denoted by $\mathbf{x}^{(1)}, \cdots, \mathbf{x}^{(625)}$. The transition matrix $\mathbf{P} = (p_{hj})_{625 \times 625}$ has entries*

$$p_{hj} = \mathrm{P}(\mathbf{X}_i \in \mathbf{A}^{(j)} \mid \mathbf{x}^{(h)}) = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}(-1)^{i_1+i_2}C_{0.3}^{Ga}(u_{1i_1}, u_{2i_2}),$$

where $u_{\ell 1} = \Phi(\frac{x_\ell^{(j)}+c\Delta_x-x_\ell^{(h)}-0.03\Delta}{0.2\sqrt{\Delta}})$, $u_{\ell 2} = \Phi(\frac{x_\ell^{(j)}+c\Delta_x-x_\ell^{(h)}-0.03\Delta}{0.2\sqrt{\Delta}})$, for $\ell = 1, 2$, and $C_{0.3}^{Ga}$ is the Gaussian copula with correlation coefficient 0.3. Obviously, the the entries of p_{hj} of the transition matrix are independent of the time index. To decide the adjusting coefficient c of the grids, we demonstrate the

	Copula	T (year)	European		Bermudan	
			Ben. (std.)	Alg.	Ben.	Alg.
2-dim.	Gaussian(0)	0.25	2.33	2.34	2.38	2.39
		0.5	3.02	3.02	3.18	3.18
		1	3.75	3.75	4.13	4.13
	Gaussian(0.3)	0.25	2.69	2.69	2.74	2.75
		0.5	3.50	3.50	3.67	3.67
		1	4.38	4.37	4.79	4.79
	Clayton(5)	0.25	3.30 (0.004)	3.30		3.37
		0.5	4.33 (0.006)	4.33		4.52
		1	5.46 (0.006)	5.48		5.94
	Gumbel(5)	0.25	3.33 (0.004)	3.33		3.39
		0.5	4.36 (0.005)	4.36		4.55
		1	5.50 (0.008)	5.49		5.96
3-dim.	Gaussian(0)	0.25	1.86	1.86	1.91	1.92
		0.5	2.38	2.37	2.53	2.53
	Gaussian(0.3)	0.25	2.41	2.41	2.47	2.47
		0.5	3.13	3.11	3.29	3.28
	Clayton(5)	0.25	3.27 (0.002)	3.27		3.35
		0.5	4.29 (0.004)	4.29		4.49

Table 14.2. Multi-dimensional put option prices on a geometric average with parameters $r = 0.05$, $\sigma = 0.2$, $\mathbf{S_0}$=K=100 and $\Delta = 1/12$ (year). Gaussian(ρ): ρ denotes the equi-correlation among securities. Clayton(α) and Gumbel(α): α is the parameter of Clayton and Gumbel copulae. The Ben. values of the Gaussian cases are computed by **Q** XFGBPgmeanR1 , while the Ben. values of the Clayton and Gumbel cases are from **Q** XFGEPmean2MC (2 dimensional case) and **Q** XFGEPmean3MC (3 dimensional case). The 2 and 3 dimensional Alg. values are obtained by **Q** XFGBPgmean2 and **Q** XFGBPgmean3 , respectively.

Gaussian copula case. In the case of Gaussian copula, this problem can also be considered as a one-dimensional option pricing problem. Let $\bar{S}_t$ denote the geometric mean of $S_{1,t}$ and $S_{2,t}$, that is $\bar{S}_t = \sqrt{S_{1,t}S_{2,t}}$. Since $S_{1,t}$ and $S_{2,t}$ both are geometric Brownian motions, thus by Ito's lemma we have

$$d \log \bar{S}_t = (\widetilde{r} - \frac{1}{2}\widetilde{\sigma}^2)dt + \widetilde{\sigma}dW_t,$$

where W_t is a Wiener process, $\widetilde{\sigma}^2 = \frac{1}{4}(\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2)$, which is due to the bivariate normal distributed assumption, and $\widetilde{r} = r + \frac{1}{2}\widetilde{\sigma}^2 - \frac{1}{4}(\sigma_1^2 + \sigma_2^2)$. Consequently, the European put option values can be obtained by the following

		European		Bermudan
Copula	T (year)	Ben. (std.)	Alg.	Alg.
Gaussian(0)	2/3	8.46 (0.006)	8.45	8.65
	1	9.55 (0.009)	9.55	10.04
Gaussian(0.3)	2/3	7.90 (0.011)	7.89	8.07
	1	8.93 (0.013)	8.93	9.36
Clayton(5)	2/3	6.73 (0.014)	6.74	6.92
	1	7.66 (0.013)	7.61	8.02

Table 14.3. Multi-dimensional max call option prices with parameters $r = 0.05$, $q = 0.1$, $\sigma = 0.2$, $\mathbf{S_0}$=K=100 and $\Delta = 1/3$ (year). The Ben. values are computed by **Q** XFGECmax2MC and the Alg. values are from **Q** XFGBCmax2 .

formula

$$V_0(\mathbf{S}_0) = e^{-rT}\,\mathsf{E}[(K - \bar{S}_T)^+ \mid \mathbf{S}_0] = e^{-(r-\tilde{r})T}\{Ke^{-\tilde{r}T}\Phi(-d_2) - \bar{S}_0\Phi(-d_1)\},$$
(14.5)

where $d_1 = \frac{\log(\bar{S}_0/K)+(\tilde{r}+0.5\tilde{\sigma}^2)T}{\tilde{\sigma}\sqrt{T}}$, $d_2 = d_1 - \tilde{\sigma}\sqrt{T}$ *and the second equality is due to the Black-Scholes formula. The above result can also be extended to d-dimensional European option on geometric average.*

Assume that the random vector $(\log S_{1,t}, \cdots, \log S_{d,t})^\top$ *has a multivariate normal distribution with covariance matrix* $t{\cdot}\Sigma = t{\cdot}(\sigma_{jk})$. *Let* $\bar{S}_t = (S_{1,t}\cdots S_{d,t})^{1/d}$, *thus* $\log\bar{S}_t$ *given* $\mathbf{S}_0$ *is normally distributed with mean* $\log\bar{S}_0 + (\tilde{r} - \frac{1}{2}\tilde{\sigma}^2)t$ *and variance* $\tilde{\sigma}^2 t$, *where* $\tilde{\sigma} = \frac{1}{d}\sqrt{\sum_{j,k}\sigma_{jk}}$ *and* $\tilde{r} = r + \frac{1}{2}\tilde{\sigma}^2 - \frac{1}{2d}\sum_j\sigma_{jj}$. *Thus the European put option values on a d-dimensional geometric average can be obtained by (14.5) analogously. And the corresponding Bermudan option can also be valuated using this reduced one-dimensional version. Thus for Gaussian copula, we can use (14.5) to obtain the benchmarks of the European and Bermudan geometric option prices and the adjusting coefficient c can then be determined.*

Table 14.2 presents the results of several expiration dates T for Gaussian, Clayton and Gumbel copulae. For Clayton and Gumbel copulae, since no closed-form solutions exit, thus the benchmarks of European options are obtained by Monte Carlo simulation. For the Gaussian cases, the estimated option values are all close to the benchmarks, which shows the proposed scheme provides a promising approach for multi-dimensional options on a geometric average.

EXAMPLE 14.3 *Suppose two underlying assets satisfying (14.1) with* $r = 0.05$, $q_1 = q_2 = 0.1$ *and* $\sigma_1 = \sigma_2 = 0.2$. *Consider a Bermudan max-call*

S_0/K	$T = 1/4$		$T = 1/2$		$T = 1$	
	Euro.	Berm.	Euro.	Berm.	Euro.	Berm.
0.9	7.86	20.49	4.01	20.49	1.18	20.49
1.0	2.41	3.90	1.45	4.27	0.48	4.42
1.1	0.62	0.84	0.49	1.11	0.19	1.24

Table 14.4. European and Bermudan put values of example 14.4 with parameters $r = 0.5736$, $\sigma = 0.304$, $\Delta = 1/12$ and $S_0 = 184.375$. Output of **Q** XFGBP1 .

option with $K = 100$, $T = 2/3, 1$, $\mathbf{S}_0 = 100$, $\Delta = 1/3$ (i.e. $n = 3$) and payoff function $g(\mathbf{S}_t) = \{\max(S_{1,t}, S_{2,t}) - K\}^+$. Table 14.3 gives the results of Bermudan max call option prices for Gaussian copula with $\rho = 0, 0.3$, and Clayton copula with $\alpha = 5$. Since no closed-form solutions of European max-call option exit for Gaussian and Clayton copulae, the European benchmarks are obtained by Monte Carlo simulation. The simulation results show that all the Bermudan options are more valuable than their European counterparts.

EXAMPLE 14.4 *Consider a standard Bermuda put option on the IBM shares. In (Tsay, 2005, p.259 − 260) a geometric Brownian motion process (14.4) is fitted to the 252 daily IBM stock prices of 1998. The parameters' estimated values are $r = 0.5732$, $q = 0$ and $\sigma = 0.304$. The stock price of IBM on Dec. 31, 1998 is $S_0 = 184.375$. Assume the possible early exercise dates are at the end of each month, that is the length of the time interval is $\Delta = 1/12$. Table 14.4 presents the European and Bermudan put option values on Dec. 31, 1998, for different $S_0/K = 0.9, 1, 1.1$, where K is the strike price, and maturity time $T = 1/4, 1/2, 1$. The European put values are computed by the Black-Scholes formula and the Bermudan put values are computed by the proposed method with 401 pre-chosen equidistant grid points as in example 14.1. The results show the Bermudan options are all more valuable than their European counterparts.*

14.5 Conclusion

The proposed method gives an innovative semiparametric approach to multidimensional Bermudan option pricing. The method is applicable to use copula functions modeling multivariate asset dependence. The simulation results show that the proposed approach is very tractable for numerical implementation and provides an accurate method for pricing Bermudan options. Although the transition matrix of the proposed method is a sparse matrix

containing lots of zeros, the geometrically increasing rate (in time) of the matrix size still impedes its application. To tackle this problem, Huang and Guo (2007) apply important sampling idea to re-weight the grid probabilities and keep the matrix size constant throughout the time.

Bibliography

Barraquand, J. and Martineau, D. (1995). Numerical valuation of high dimensional multivariate American securities, *J. Fin. Quant. Anal.* **30**: 383–405.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *J. Political Economy* **81**: 637–654.

Broadie, M. and Glasserman, P. (2004). A stochastic mesh method for pricing high-dimensional American options, *J. Comput. Finance* **7**: 35–72.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula Methods in Finance*, Wiley, West Sussex.

Franke, J., Härdle, W. and Hafner, C. M. (2008). *Statistics of Financial Markets*, 2nd edn, Springer, Berlin.

Giacomini, E., Härdle, W., Ignatieva, E. and Spokoiny, V. (2007). Inhomogeneous dependency modelling with time varying copulae, *SFB 649 Economic Risk Berlin*.

Huang, S. F. and Guo, M. H. (2007). Financial Derivative Valuation - A Dynamic Semiparametric Approach, *Submitted for publishing*.

Ju, N. (1998). Pricing an American option by approximating its early exercise boundary as multipiece exponential function, *Rev. Financial Stud.* **11**: 627–646.

Judd, K. (1998). *Numerical Methods in Economics*, MIT Press, Cambridge, Mass.

Lai, T. L. and AitSahalia, F. (2001). Exercise boundaries and efficient approximations to American option prices and hedge parameters, *J. Computational Finance* **4**: 85–103.

Longstaff, F. A. and Schwartz, E. S. (2001). Valuing American options by simulation: a simple least-squares approach, *Rev. Financial Stud.* **14**: 113–147.

McNeil, A. J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management*, Princeton University Press, New Jersey.

Nelsen, R. B. (2006). *An Introduction to Copulas*, Springer, Berlin.

Rust, J. (1997). Using randomization to break the curse of dimensionality, *Econometrica* **65**: 487–516.

Shreve, S. E. (2004). *Stochastic Calculus for Finance I*, Springer, New York.

Sklar, A. (1959). Fonctions de repartition á n dimensions et leurs marges, *Pub. Inst. Statist. Univ. Paris* **8**: 229–231.

Tsay, R. S. (2006). *Analysis of Financial Time Series*, 2nd edn, Wiley, New York.

Tsitsiklis, J. and Van Roy, B. (1999). Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives, *IEEE Transactions on Automatic Control* **44**: 1840–1851.

Part IV

Econometrics

15 Multivariate Volatility Models

Matthias R. Fengler and Helmut Herwartz

Multivariate volatility models are widely used in Finance to capture both volatility clustering and contemporaneous correlation of asset return vectors. Here we focus on multivariate GARCH models. In this common model class it is assumed that the covariance of the error distribution follows a time dependent process conditional on information which is generated by the history of the process. To provide a particular example, we consider a system of exchange rates of two currencies measured against the US Dollar (USD), namely the Deutsche Mark (DEM) and the British Pound Sterling (GBP). For this process we compare the dynamic properties of the bivariate model with univariate GARCH specifications where cross sectional dependencies are ignored. Moreover, we illustrate the scope of the bivariate model by ex-ante forecasts of bivariate exchange rate densities.

15.1 Introduction

Volatility clustering, i.e. positive correlation of price variations observed on speculative markets, motivated the introduction of autoregressive conditionally heteroskedastic (ARCH) processes by Engle (1982) and its popular generalizations by Bollerslev (1986) (Generalized ARCH, GARCH) and Nelson (1991) (exponential GARCH, EGARCH). Being univariate in nature, however, such models neglect a further stylized fact of empirical price variations, namely contemporaneous cross correlation e.g. over a set of assets, stock market indices, or exchange rates.

Cross section relationships are often implied by economic theory. Interest rate parities, for instance, provide a close relation between domestic and foreign bond rates. Assuming absence of arbitrage, the so-called triangular equation formalizes the equality of an exchange rate between two currencies on the one hand and an implied rate constructed via exchange rates measured towards a third currency. Furthermore, stock prices of firms acting on the same market often show similar patterns in the sequel of news that are important for the entire market (Hafner and Herwartz, 1998). Similarly, analyzing global

volatility transmission Engle, Ito and Lin (1990) and Hamao, Masulis and
Ng (1990) found evidence in favor of volatility spillovers between the world's
major trading areas occurring in the sequel of floor trading hours. From this
point of view, when modeling time varying volatilities, a multivariate model
appears to be a natural framework to take cross sectional information into
account. Moreover, the covariance between financial assets is of essential
importance in finance. Effectively, many problems in financial practice like
portfolio optimization, hedging strategies, or Value-at-Risk evaluation require
multivariate volatility measures (Bollerslev et al., 1988; Cecchetti, Cumby
and Figlewski, 1988).

15.1.1 Model Specifications

Let $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})^\top$ denote an N-dimensional error process, which is
either directly observed or estimated from a multivariate regression model.
The process ε_t follows a multivariate GARCH process if it has the representation

$$\varepsilon_t = \Sigma_t^{1/2} \xi_t, \tag{15.1}$$

where Σ_t is measurable with respect to information generated up to time
$t - 1$, denoted by the filtration $\mathcal{F}_{t-1}$. By assumption the N components of
ξ_t follow a multivariate Gaussian distribution with mean zero and covariance
matrix equal to the identity matrix.

The conditional covariance matrix, $\Sigma_t = \mathsf{E}[\varepsilon_t \varepsilon_t^\top | \mathcal{F}_{t-1}]$, has typical elements
σ_{ij} with $\sigma_{ii}, i = 1, \ldots, N$, denoting conditional variances and off-diagonal
elements $\sigma_{ij}, i, j = 1, \ldots, N, i \neq j$, denoting conditional covariances. To
make the specification in (15.1) feasible a parametric description relating
Σ_t to $\mathcal{F}_{t-1}$ is necessary. In a multivariate setting, however, dependencies
of the second order moments in Σ_t on $\mathcal{F}_{t-1}$ become easily computationally
intractable for practical purposes.

Let vech(A) denote the half-vectorization operator stacking the elements of a
quadratic ($N \times N$)-matrix A from the main diagonal downwards in a $\frac{1}{2}N(N+1)$ dimensional column vector. Within the so-called vec-representation of the
GARCH(p, q) model Σ_t is specified as follows:

$$\text{vech}(\Sigma_t) = c + \sum_{i=1}^{q} \widetilde{A}_i \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}^\top) + \sum_{i=1}^{p} \widetilde{G}_i \text{vech}(\Sigma_{t-i}). \tag{15.2}$$

In (15.2) the matrices $\widetilde{A}_i$ and $\widetilde{G}_i$ each contain $\{N(N + 1)/2\}^2$ elements.
Deterministic covariance components are collected in c, a column vector of

dimension $N(N+1)/2$. We consider in the following the case $p = q = 1$ since in applied work the GARCH(1,1) model has turned out to be particularly useful to describe a wide variety of financial market data (Bollerslev, Engle and Nelson, 1994).

On the one hand the vec–model in (15.2) allows for a very general dynamic structure of the multivariate volatility process. On the other hand this specification suffers from high dimensionality of the relevant parameter space, which makes it almost intractable for empirical work. In addition, it might be cumbersome in applied work to restrict the admissible parameter space such that the implied matrices Σ_t, $t = 1, \ldots, T$, are positive definite. These issues motivated a considerable variety of competing multivariate GARCH specifications.

Prominent proposals reducing the dimensionality of (15.2) are the constant correlation model (Bollerslev, 1990) and the diagonal model (Bollerslev et al., 1988). Specifying diagonal elements of Σ_t both of these approaches assume the absence of cross equation dynamics, i.e. the only dynamics are

$$\sigma_{ii,t} = c_{ii} + a_i \varepsilon_{i,t-1}^2 + g_i \sigma_{ii,t-1}, \ i = 1, \ldots, N. \qquad (15.3)$$

To determine off-diagonal elements of Σ_t Bollerslev (1990) proposes a constant contemporaneous correlation,

$$\sigma_{ij,t} = \rho_{ij} \sqrt{\sigma_{ii}\sigma_{jj}}, \ i, j = 1, \ldots, N, \qquad (15.4)$$

whereas Bollerslev et al. (1988) introduce an ARMA-type dynamic structure as in (15.3) for $\sigma_{ij,t}$ as well, i.e.

$$\sigma_{ij,t} = c_{ij} + a_{ij}\varepsilon_{i,t-1}\varepsilon_{j,t-1} + g_{ij}\sigma_{ij,t-1}, i, j = 1, \ldots, N. \qquad (15.5)$$

For the bivariate case ($N = 2$) with $p = q = 1$ the constant correlation model contains only 7 parameters compared to 21 parameters encountered in the full model (15.2). The diagonal model is specified with 9 parameters. The price that both models pay for parsimony is in ruling out cross equation dynamics as allowed in the general vec-model. Positive definiteness of Σ_t is easily guaranteed for the constant correlation model ($|\rho_{ij}| < 1$), whereas the diagonal model requires more complicated restrictions to provide positive definite covariance matrices.

The so-called BEKK-model (named after Baba, Engle, Kraft and Kroner, 1990) provides a richer dynamic structure compared to both restricted processes mentioned before. Defining $N \times N$ matrices A_{ik} and G_{ik} and an upper triangular matrix C_0 the BEKK–model reads in a general version as follows:

$$\Sigma_t = C_0^\top C_0 + \sum_{k=1}^{K}\sum_{i=1}^{q} A_{ik}^\top \varepsilon_{t-i}\varepsilon_{t-i}^\top A_{ik} + \sum_{k=1}^{K}\sum_{i=1}^{p} G_{ik}^\top \Sigma_{t-i} G_{ik}. \qquad (15.6)$$

If $K = q = p = 1$ and $N = 2$, the model in (15.6) contains 11 parameters and implies the following dynamic model for typical elements of Σ_t:

$$
\begin{aligned}
\sigma_{11,t} &= c_{11} + a_{11}^2\varepsilon_{1,t-1}^2 + 2a_{11}a_{21}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{21}^2\varepsilon_{2,t-1}^2 \\
&\quad + g_{11}^2\sigma_{11,t-1} + 2g_{11}g_{21}\sigma_{21,t-1} + g_{21}^2\sigma_{22,t-1}, \\
\sigma_{21,t} &= c_{21} + a_{11}a_{22}\varepsilon_{1,t-1}^2 + (a_{21}a_{12} + a_{11}a_{22})\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{21}a_{22}\varepsilon_{2,t-1}^2 \\
&\quad + g_{11}g_{22}\sigma_{11,t-1} + (g_{21}g_{12} + g_{11}g_{22})\sigma_{12,t-1} + g_{21}g_{22}\sigma_{22,t-1}, \\
\sigma_{22,t} &= c_{22} + a_{12}^2\varepsilon_{1,t-1}^2 + 2a_{12}a_{22}\varepsilon_{1,t-1}\varepsilon_{2,t-1} + a_{22}^2\varepsilon_{2,t-1}^2 \\
&\quad + g_{12}^2\sigma_{11,t-1} + 2g_{12}g_{22}\sigma_{21,t-1} + g_{22}^2\sigma_{22,t-1}.
\end{aligned}
$$

Compared to the diagonal model the BEKK–specification economizes on the number of parameters by restricting the vec–model within and across equations. Since A_{ik} and G_{ik} are not required to be diagonal, the BEKK-model is convenient to allow for cross dynamics of conditional covariances. The parameter K governs to which extent the general representation in (15.2) can be approximated by a BEKK-type model. In the following we assume $K = 1$. Note that in the bivariate case with $K = p = q = 1$ the BEKK-model contains 11 parameters. If $K = 1$ the matrices A_{11} and $-A_{11}$, imply the same conditional covariances. Thus, for uniqueness of the BEKK-representation $a_{11} > 0$ and $g_{11} > 0$ is assumed. Note that the right hand side of (15.6) involves only quadratic terms and, hence, given convenient initial conditions, Σ_t is positive definite under the weak (sufficient) condition that at least one of the matrices C_0 or G_{ik} has full rank (Engle and Kroner, 1995).

15.1.2 Estimation of the BEKK-Model

As in the univariate case the parameters of a multivariate GARCH model are estimated by maximum likelihood (ML) optimizing numerically the Gaussian log-likelihood function.

With f denoting the multivariate normal density, the contribution of a single observation, l_t, to the log-likelihood of a sample is given as:

$$
\begin{aligned}
l_t &= \ln\{f(\varepsilon_t|\mathcal{F}_{t-1})\} \\
&= -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln(|\Sigma_t|) - \frac{1}{2}\varepsilon_t^\top \Sigma_t^{-1}\varepsilon_t.
\end{aligned}
$$

Maximizing the log-likelihood, $l = \sum_{t=1}^{T} l_t$, requires nonlinear maximization methods. Involving only first order derivatives the algorithm introduced by

Berndt, Hall, Hall, and Hausman (1974) is easily implemented and particularly useful for the estimation of multivariate GARCH processes.

If the actual error distribution differs from the multivariate normal, maximizing the Gaussian log-likelihood has become popular as Quasi ML (QML) estimation. In the multivariate framework, results for the asymptotic properties of the (Q)ML-estimator have been derived recently. Jeantheau (1998) proves the QML-estimator to be consistent under the main assumption that the considered multivariate process is strictly stationary and ergodic. Further assuming finiteness of moments of ε_t up to order eight, Comte and Lieberman (2000) derive asymptotic normality of the QML-estimator. The asymptotic distribution of the rescaled QML-estimator is analogous to the univariate case and discussed in Bollerslev and Wooldridge (1992).

15.2 An Empirical Illustration

15.2.1 Data Description

We analyze daily quotes of two European currencies measured against the USD, namely the DEM and the GBP. The sample period is December 31, 1979 to April 1, 1994, covering $T = 3720$ observations. Note that a subperiod of our sample has already been investigated by Bollerslev and Engle (1993) discussing common features of volatility processes.

The data is provided in `fx`. The first column contains DEM/USD and the second GBP/USD. In XploRe a preliminary statistical analysis is easily done by the **Q** `summarize` command. Before inspecting the summary statistics, we load the data, R_t, and take log differences, $\varepsilon_t = \ln(R_t) - \ln(R_{t-1})$. **Q** `XFGmvol01` produces the following table:

```
[2,] "          Minimum    Maximum     Mean    Median Std.Error"
[3,] "---------------------------------------------------------"
[4,] "DEM/USD  -0.040125   0.031874 -4.7184e-06    0   0.0070936"
[5,] "GBP/USD  -0.046682   0.038665  0.00011003    0   0.0069721"
```

<div align="right">**Q** XFGmvol01</div>

Evidently, the empirical means of both processes are very close to zero (-4.72e-06 and 1.10e-04, respectively). Also minimum, maximum and standard errors are of similar size. First differences of the respective log exchange rates are shown in Figure 15.1. As is apparent from Figure 15.1, variations
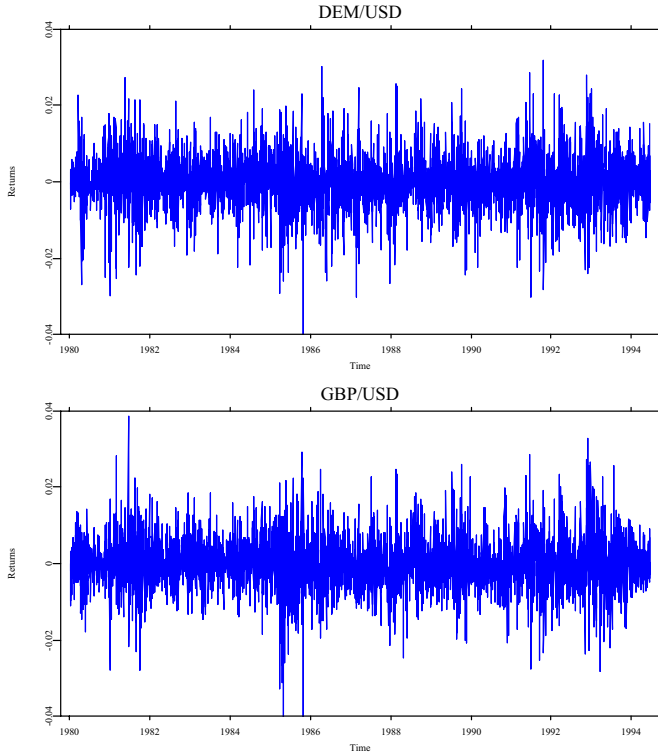
Figure 15.1. Foreign exchange rate data: returns.

of exchange rate returns exhibit an autoregressive pattern: Large returns in foreign exchange markets are followed by large returns of either sign. This is most obvious in periods of excessive returns. Note that these volatility clusters tend to coincide in both series. It is precisely this observation that justifies a multivariate GARCH specification.

15.2.2 Estimating Bivariate GARCH

The quantlet ◘ bigarch provides a fast algorithm to estimate the BEKK representation of a bivariate GARCH(1,1) model. QML-estimation is implemented by means of the BHHH-algorithm which minimizes the negative Gaussian log-likelihood function. The algorithm employs analytical first order derivatives of the log-likelihood function (Lütkepohl, 1996) with respect

to the 11-dimensional vector of parameters containing the elements of C_0, A_{11} and G_{11} as given in (15.6).

The standard call is

```
{coeff, likest}=bigarch(theta, et),
```

where as input parameters we have initial values `theta` for the iteration algorithm and the data set, e.g. financial returns, stored in `et`. The estimation output is the vector `coeff` containing the stacked elements of the parameter matrices C_0, A_{11} and G_{11} in (15.6) after numerical optimization of the Gaussian log-likelihood function. Being an iterative procedure the algorithm requires to determine suitable initial parameters `theta`. For the diagonal elements of the matrices A_{11} and G_{11} values around 0.3 and 0.9 appear reasonable, since in univariate GARCH(1,1) models parameter estimates for a_1 and g_1 in (15.3) often take values around $0.3^2 = 0.09$ and $0.81 = 0.9^2$. There is no clear guidance how to determine initial values for off diagonal elements of A_{11} or G_{11}. Therefore it might be reasonable to try alternative initializations of these parameters. Given an initialization of A_{11} and G_{11} the starting values for the elements in C_0 are immediately determined by the algorithm assuming the unconditional covariance of ε_t to exist, Engle and Kroner (1995).

Given our example under investigation the bivariate GARCH estimation yields as output:

```
Contents of coeff

[ 1,]   0.0011516
[ 2,]   0.00031009
[ 3,]   0.00075685
[ 4,]   0.28185
[ 5,]  -0.057194
[ 6,]  -0.050449
[ 7,]   0.29344
[ 8,]   0.93878
[ 9,]   0.025117
[10,]   0.027503
[11,]   0.9391

Contents of likest

[1,]    -28599
```

 XFGmvol02

The last number is the obtained minimum of the negative log-likelihood func-

tion. The vector `coeff` given first contains as first three elements the parameters of the upper triangular matrix C_0, the following four belong to the ARCH (A_{11}) and the last four to the GARCH parameters (G_{11}), i.e. for our model

$$\Sigma_t = C_0^\top C_0 + A_{11}^\top \varepsilon_{t-1} \varepsilon_{t-1}^\top A_{11} + G_{11}^\top \Sigma_{t-1} G_{11} \tag{15.7}$$

stated again for convenience, we find the matrices C_0, A, G to be:

$$C_0 = 10^{-3} \begin{pmatrix} 1.15 & .31 \\ 0 & .76 \end{pmatrix},$$

$$A_{11} = \begin{pmatrix} .282 & -.050 \\ -.057 & .293 \end{pmatrix}, \ G_{11} = \begin{pmatrix} .939 & .028 \\ .025 & .939 \end{pmatrix}. \tag{15.8}$$

15.2.3 Estimating the (Co)Variance Processes

The (co)variance is obtained by sequentially calculating the difference equation (15.7) where we use the estimator for the unconditional covariance matrix as initial value ($\Sigma_0 = \frac{E^\top E}{T}$). Here, the $T \times 2$ vector E contains log-differences of our foreign exchange rate data. Estimating the covariance process is also accomplished in the quantlet **Q**`XFGmvol02` and additionally provided in `sigmaprocess`.

We display the estimated variance and covariance processes in Figure 15.2. The upper and the lower panel of Figure 15.2 show the variances of the DEM/USD and GBP/USD returns respectively, whereas in the middle panel we see the covariance process. Except for a very short period in the beginning of our sample the covariance is positive and of non-negligible size throughout. This is evidence for cross sectional dependencies in currency markets which we mentioned earlier to motivate multivariate GARCH models.

Instead of estimating the realized path of variances as shown above, we could also use the estimated parameters to *simulate* volatility paths (**Q**`XFGmvol03`).

For this at each point in time an observation ε_t is drawn from a multivariate normal distribution with variance Σ_t. Given these observations, Σ_t is updated according to (15.7). Then, a new residual is drawn with covariance Σ_{t+1}. We apply this procedure for $T = 3000$. The results, displayed in the upper three panels of Figure 15.3, show a similar pattern as the original process given in Figure 15.2. For the lower two panels we generate two
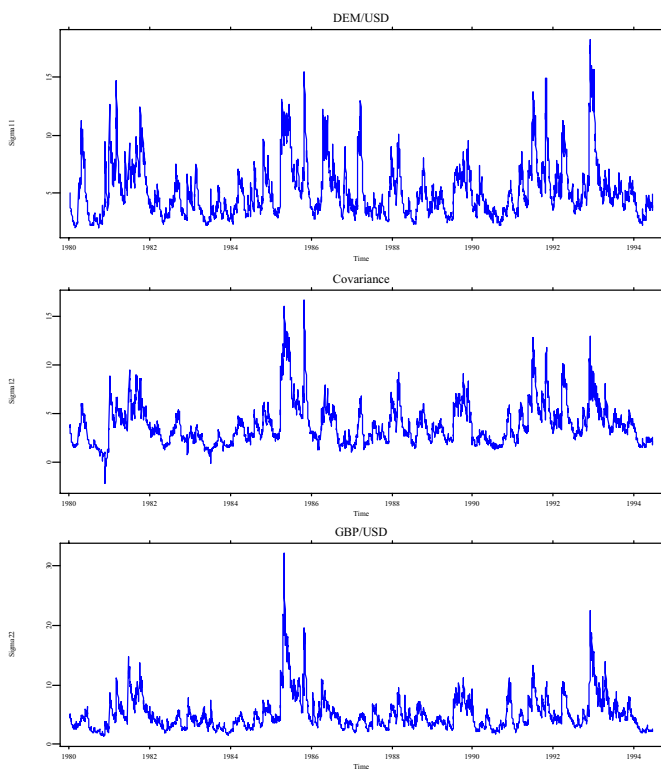
Figure 15.2. Estimated variance and covariance processes, $10^5\widehat{\Sigma}_t$.

 XFGmvol02

variance processes from the *same* residuals ξ_t. In this case, however, we set off-diagonal parameters in A_{11} and G_{11} to zero to illustrate how the unrestricted BEKK model incorporates cross equation dynamics. As can be seen, both approaches are convenient to capture volatility clustering. Depending on the particular state of the system, spillover effects operating through conditional covariances, however, have a considerable impact on the magnitude of conditional volatility.
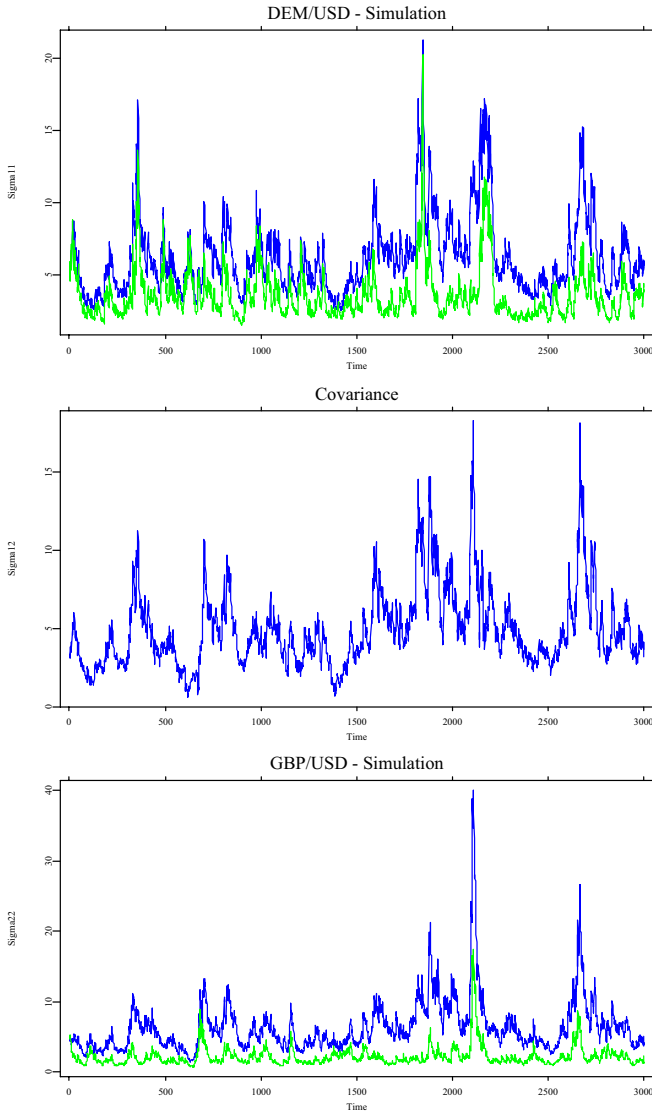
Figure 15.3. Simulated variance and covariance processes, both bivariate (blue) and univariate case (green), $10^5\widehat{\Sigma}_t$.

XFGmvol03

15.3 Forecasting Exchange Rate Densities

The preceding section illustrated how the GARCH model may be employed effectively to describe empirical price variations of foreign exchange rates. For practical purposes, as for instance scenario analysis, VaR estimation (Chapter **??**), option pricing (Chapter **??**), one is often interested in the future joint density of a set of asset prices. Continuing the comparison of the univariate and bivariate approach to model volatility dynamics of exchange rates it is thus natural to investigate the properties of these specifications in terms of forecasting performance.

We implement an iterative forecasting scheme along the following lines: Given the estimated univariate and bivariate volatility models and the corresponding information sets $\mathcal{F}_{t-1}, t = 1, \ldots, T - 5$ (Figure 15.2), we employ the identified data generating processes to simulate one-week-ahead forecasts of both exchange rates. To get a reliable estimate of the future density we set the number of simulations to 50000 for each initial scenario. This procedure yields two bivariate samples of future exchange rates, one simulated under bivariate, the other one simulated under univariate GARCH assumptions.

A review on the current state of evaluating competing density forecasts is offered by Tay and Wallis (1990). Adopting a Bayesian perspective the common approach is to compare the expected loss of actions evaluated under alternative density forecasts. In our pure time series framework, however, a particular action is hardly available for forecast density comparisons. Alternatively one could concentrate on statistics directly derived from the simulated densities, such as first and second order moments or even quantiles. Due to the multivariate nature of the time series under consideration it is a nontrivial issue to rank alternative density forecasts in terms of these statistics. Therefore, we regard a particular volatility model to be superior to another if it provides a higher simulated density estimate of the actual bivariate future exchange rate. This is accomplished by evaluating both densities at the actually realized exchange rate obtained from a bivariate kernel estimation. Since the latter comparison might suffer from different unconditional variances under univariate and multivariate volatility, the two simulated densities were rescaled to have identical variance. Performing the latter forecasting exercises iteratively over 3714 time points we can test if the bivariate volatility model outperforms the univariate one.

To formalize the latter ideas we define a success ratio SR_J as

$$SR_J = \frac{1}{|J|} \sum_{t \in J} \mathbf{1}\{\widehat{f}_{biv}(R_{t+5}) > \widehat{f}_{uni}(R_{t+5})\}, \tag{15.9}$$

Time window J		Success ratio SR_J
1980	1981	0.744
1982	1983	0.757
1984	1985	0.793
1986	1987	0.788
1988	1989	0.806
1990	1991	0.807
1992	1994/4	0.856

Table 15.1. Time varying frequencies of the bivariate GARCH model outperforming the univariate one in terms of one-week-ahead forecasts (success ratio)

where J denotes a time window containing $|J|$ observations and **1** an indicator function. $\widehat{f}_{biv}(R_{t+5})$ and $\widehat{f}_{uni}(R_{t+5})$ are the estimated densities of future exchange rates, which are simulated by the bivariate and univariate GARCH processes, respectively, and which are evaluated at the actual exchange rate levels R_{t+5}. The simulations are performed in **Q** XFGmvol04 .

Our results show that the bivariate model indeed outperforms the univariate one when both likelihoods are compared under the actual realizations of the exchange rate process. In 81.6% of all cases across the sample period, $SR_J = 0.816, \quad J = \{t : t = 1, ..., T - 5\}$, the bivariate model provides a better forecast. This is highly significant. In Table 15.1 we show that the overall superiority of the bivariate volatility approach is confirmed when considering subsamples of two-years length. A-priori one may expect the bivariate model to outperform the univariate one the larger (in absolute value) the covariance between both return processes is. To verify this argument we display in Figure 15.4 the empirical covariance estimates from Figure 15.2 jointly with the success ratio evaluated over overlapping time intervals of length $|J| = 80$.

As is apparent from Figure 15.4 there is a close co-movement between the success ratio and the general trend of the covariance process, which confirms our expectations: the forecasting power of the bivariate GARCH model is particularly strong in periods where the DEM/USD and GBP/USD exchange rate returns exhibit a high covariance. For completeness it is worthwhile to mention that similar results are obtained if the window width is varied over reasonable choices of $|J|$ ranging from 40 to 150.

With respect to financial practice and research we take our results as strong support for a multivariate approach towards asset price modeling. Whenever contemporaneous correlation across markets matters, the system approach offers essential advantages. To name a few areas of interest multivariate
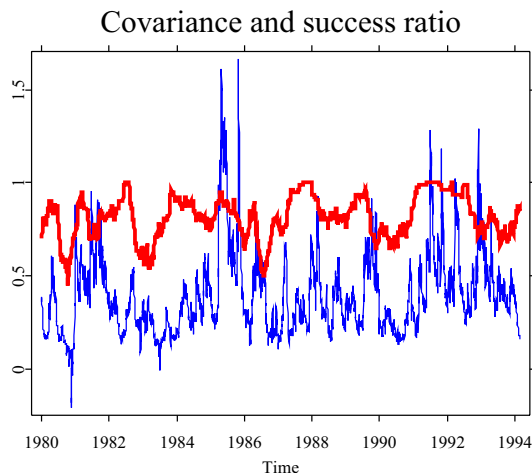
Figure 15.4. Estimated covariance process from the bivariate GARCH model ($10^4\hat{\sigma}_{12}$, blue) and success ratio over overlapping time intervals with window length 80 days (red).

volatility models are supposed to yield useful insights for risk management, scenario analysis and option pricing.

Bibliography

Baba, Y., Engle, R.F., Kraft, D.F., and Kroner, K.F. (1990). Multivariate Simultaneous Generalized ARCH, *mimeo*, Department of Economics, University of California, San Diego.

Berndt, E.K., Hall B.H., Hall, R.E., and Hausman, J.A. (1974). Estimation and Inference in Nonlinear Structural Models, *Annals of Economic and Social Measurement* **3/4**: 653–665.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroscedasticity, *Journal of Econometrics* **31**: 307-327.

Bollerslev, T. (1990). Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Approach, *Review of Economics and Statistics* **72**: 498–505.

Bollerslev, T. and Engle, R.F. (1993). Common Persistence in Conditional Variances, *Econometrica* **61**: 167–186.

Bollerslev, T., Engle, R.F. and Nelson, D.B. (1994). GARCH Models, in: Engle, R.F., and McFadden, D.L. (eds.) Handbook of Econometrics, Vol. 4, Elsevier, Amsterdam, 2961–3038.

Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988). A Capital Asset Pricing Model with Time-Varying Covariances, *Journal of Political Economy* **96**: 116–131.

Bollerslev, T. and Wooldridge, J.M. (1992). Quasi–Maximum Likelihood Estimation and Inference in Dynamic Models with Time–Varying Covariances, *Econometric Reviews*, **11**: 143–172.

Cecchetti, S.G., Cumby, R.E. and Figlewski, S. (1988). Estimation of the Optimal Futures Hedge, *Review of Economics and Statistics* **70**: 623-630.

Comte, F. and Lieberman, O. (2000). Asymptotic Theory for Multivariate GARCH Processes, *Manuscript, Universities Paris 6 and Paris 7*.

Engle, R.F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation. *Econometrica* **50**: 987-1008.

Engle, R.F., Ito, T. and Lin, W.L. (1990). Meteor Showers or Heat Waves? Heteroskedastic Intra-Daily Volatility in the Foreign Exchange Market, *Econometrica* **58**: 525–542.

Engle, R.F. and Kroner, K.F. (1995). Multivariate Simultaneous Generalized ARCH, *Econometric Theory* **11**: 122–150.

Hafner, C.M. and Herwartz, H. (1998). Structural Analysis of Portfolio Risk using Beta Impulse Response Functions, *Statistica Neerlandica* **52**: 336-355.

Hamao, Y., Masulis, R.W. and Ng, V.K. (1990). Correlations in Price Changes and Volatility across International Stock Markets, *Review of Financial Studies* **3**: 281–307.

Jeantheau, T. (1998). Strong Consistency of Estimators for Multivariate ARCH Models, *Econometric Theory* **14**: 70-86.

Lütkepohl, H. (1996). *Handbook of Matrices*, Wiley, Chichester.

Nelson, D.B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach, *Econometrica* **59**: 347–370.

Tay, A. and Wallis, K. (2000). Density forecasting: A Survey, *Journal of Forecasting* **19**: 235–254.

16 The Accuracy of Long-term Real Estate Valuations

Rainer Schulz, Markus Staiber, Martin Wersing and Axel Werwatz

16.1 Introduction

Real estate valuations are important for financial institutions, especially banks, for at least two reasons. First, valuations are often needed during the underwriting or refinancing of mortgage loans, where valuations should provide a fair assessment of the (future) market value of the property that will serve as collateral for the loan. Second, valuations are needed if the institution or bank wants an updated assessment of collateral values for outstanding loans it holds on its balance sheet. Such reassessments might be necessary and required by Basel II if new information arrives or market sentiments change.

The two most common approaches for the valuation of single-family houses are the sales comparison approach and the cost approach. Focussing on a short-term horizon, the studies of Dotzour (1990) and Schulz and Werwatz (2008) have shown that sales comparison values are more accurate than cost values when used as forecasts of current house prices. Further, the latter study finds that a weighted average of sales comparison values and cost values performs best.

In this study, we complement the above results by focussing on a long-term horizon and examine the accuracy of single-family house valuations when used as forecasts of future house prices. Here, the future could refer to the date when the borrower is most likely to default. The long-term valuation would then be a forecast of collateral recovery value given default. Informal evidence indicates that the default probabilities are highest in the early years of a mortgage loan, so that a long-term horizon of up to five years seems to be a reasonable choice.

It should be noted that mortgage banks in several countries are required to compute so-called mortgage lending values for the underwriting process. The

rules for the computation of mortgage lending values are binding and defined in detail by the financial market supervisory authorities. This applies to Germany, the country our data comes from. According to the German rules, sales and cost values form the basis for the computation of single-family house mortgage lending values, but further adjustments and deductions are required. Deductions are reasonable if the economic loss function of valuation errors is asymmetric. The long-term valuations we examine and the mortgage lending values are thus not identical, but related. Evaluating the accuracy of long-term valuations might thus also be useful for an understanding of the accuracy of mortgage lending values.

The results of our study show that the sales comparison values provide better long-term forecasts than cost values if the economic loss function is symmetric, but a weighted average of sales comparison and cost values performs best. If the economic loss function is asymmetric, however, then—as kernel density estimates of the valuation error distributions reveal—cost values might provide better long-term forecasts. In summary, the study proves that it is possible and useful to assess the long-term performance of different valuation approaches empirically. Future work has to provide better understanding of the economic loss function. Moreover, a discussion of the accuracy of the different valuation approaches in a portfolio context seems to be worthwhile (Shiller and Weiss, 1999).

The study is organized as follows. Section 16.2 discusses the sales comparison and the cost approach in detail and explains our data set and how we compute the different valuations. Section 16.3 presents the empirical results. Section 16.4 concludes.

16.2 Implementation

In this study, the accuracy of long-term valuations is explored with single-family house data from Berlin. Our data set allows the computation of sales comparison and cost values over a period of 30 quarters. These valuations are computed for different forecast horizons and are then compared to actual transaction prices. More precisely, we compute valuations for every transaction backdated up to five years, taking into account only the information that was available that time. These valuations are adjusted for the expected future levels of house prices and replacement cost, respectively, and also for depreciation when necessary. In addition to a direct comparison of sales comparison and cost values, we also compute an equally-weighted combination of

both. In practice, appraisers sometimes compute such weighted combinations if two or more valuations of the same property are available.

16.2.1 Computation of the Valuations

Sales comparison approach: This approach uses transaction prices of comparable houses to estimate the value of the subject house. Several adjustments might be necessary when this approach is applied, either because the recent transactions are not completely comparable to the subject house or because house prices in the aggregate have changed.

We use hedonic regression techniques to compute sales comparison values. According to the technique, the observed transaction price of a house is a function of an aggregate price level, the house's characteristics and an unexplained part, assumed to be random. In particular, we employ the following specification

$$p_t = \beta_{0t} + \sum_{c=1}^{C} \left\{ \beta_{c1} T_c(x_{ct}) + \beta_{c2} T_c(x_{ct})^2 \right\} + \sum_{d=1}^{D} \gamma_d x_{dt} + \varepsilon_t . \qquad (16.1)$$

The dependent variable p_t is the log price for a house transacted in period t. β_{0t} captures the price level in period t. $T_c(\cdot)$ is a Box-Cox type transformation function for the cth continuous characteristic. Examples of continuous characteristics x_c are size of the lot, amount of floor space, and age of the building. β_{c1} and β_{c2} are the implicit prices for the respective—possibly transformed—characteristic. x_d is an indicator for the dth discrete characteristic, which could be a location indicator or the type of cellar. γ_d is the implicit price of the discrete characteristic. ε_t is a random noise term.

Fitting equation (16.1) to transaction data requires the choice of a specific transformation function $T_c(\cdot)$ for each of the continuous characteristics. In principle, these transformations might depend on the sample period used to fit the model. To simplify our analysis, we choose the transformations based on the entire sample and use these transformations throughout. As a by-product of our hedonic regressions, we also obtain constant-quality house price indices, which we use later for forecasting the expected future house price level. We start with a regression using the data over the period 1980Q1-1991Q2 to obtain estimates of the price levels β_{0t}. The second regression covers the period 1991Q2-1995Q1 and is used to make valuations based on information up to 1995Q1. The estimated coefficients of the price levels are used to construct the price index series from 1980Q1-1995Q1, which is used to forecast the future trend of the price level. The procedure continues by shifting the sample by one quarter and fitting a new regression. The last

regression is for the period 2002Q3-2006Q2 and we fit 47 regressions in total. For further details on the hedonic regression model and, in particular, the choice of functional form, see Schulz and Werwatz (2004).

The individual long-term sales comparison value of a house transacted in $t + h$ is computed in two steps. In the first step, we use hedonic regression fitted with data up to quarter t to compute the market value of the subject house in the valuation period t. Since the dependent variable in our hedonic regression is measured in logs, a re-transformation of the computed value is necessary. The re-transformation also corrects for any potential bias by using an 'optimal linear correction' factor (Theil, 1966, pp. 34). In the second step, we adjust the computed period t sales comparison value for the expected future price level over the forecast horizon h, see Section 16.3.1.

As stated above, the hedonic regression technique is only one of many possible ways of implementing the sales comparison approach. A great advantage of the hedonic regression technique is that it copes easily with large data sets and is suitable for mass appraisals (automated valuation). Once the regression is fitted, the value of a house—its expected price—is readily computed. The disadvantage of the hedonic regression technique is that it cannot take into account information that is not systematically recorded in the data set being used to fit the model. Such missing information is often of 'soft' nature, i.e., hard to quantify exactly. Examples are the style of decoration or the appearance of the immediate neighborhood. A valuer visiting the subject house would take such soft factors into account when forming his appraisal. The results presented below on the performance of the sales comparison values might thus be seen as conservative, because the performance could be improved if soft factors were taken into account.

Cost approach: This approach uses the replacement cost of the subject house as valuation, i.e., the sum of building cost and land cost. In case where the building of the subject property is not new, building cost needs to be adjusted for depreciation. The cost value C for a property is given by

$$C = L + \{1 - \delta(a)\}B \; ,$$

where L is land cost, B is the construction cost of a new building, and $\delta(a)$ is the depreciation due to age a. Obviously, $\delta(0) = 0$ and $\delta(a)$ approaches 1 as age a becomes large. Both building cost and land cost are computed by our data provider for the transaction period $t + h$, for details see Section 16.2.2.

We compute the cost values in two steps. First, we discount the land cost of the subject house to the valuation period t by using a land cost index. This

land cost index is derived from estimating a hedonic regression over the full sample period. The land cost in period t is then adjusted for the expected future growth of land cost over the forecast horizon h by using a time series model fitted to the land cost index estimated with information up to period t. In the second step, the observed building cost in period $t+h$ is discounted to the valuation period t by using the official construction cost index, see Section 16.3.1. The building cost is then adjusted for the expected future growth over the forecast horizon h by using a time series model fitted to the construction cost index up to period t. The building cost for the subject house is finally adjusted for depreciation accrued in period $t+h$. We employ the following depreciation function

$$\delta(a) = 1 - \left(1 - \frac{a}{l}\right)^{0.65} \quad \text{with} \quad l = \begin{cases} 98 & \text{if } a \leqslant 66 \\ 98 + (a - 66) & \text{if } a > 66 \end{cases}, \quad (16.2)$$

where l is the conditional life span of a new building and a is the age of the building. A simpler version of this function was first introduced by Cannaday and Sunderman (1986). Observe that for $a \leqslant 66$ the depreciation accelerates with age. Once a building has reached the age of 66, however, depreciation slows, reflecting superior quality of long-lived buildings.

The building cost adjusted for depreciation and the land cost are then added together to form replacement cost, i.e., the cost value C. If a valuation is for the short term, it might be advisable to further adjust C for current deviations of prices from cost. Such an adjustment is not necessary for long-term valuations, however, if prices and replacement cost realign quickly over time, as it is the case for the test market (Schulz and Werwatz, 2008).

16.2.2 Data

The data used in the study consists of transactions of single-family houses in Berlin between 1980Q1 and 2007Q2. Data are provided by Berlin's local real estate surveyor commission (Gutachterausschuss für Grundstückswerte, GAA) out of its transaction database (Automatisierte Kaufpreissammlung, AKS). This transaction database covers information on all real estate transactions in Berlin. All observations in our data set have information on the price, appraised land cost, and many different characteristics of the house. Only transactions from 2000Q1 onwards, however, have current information on new building cost. Between 2000Q1 and 2007Q2, we have 9088 observations, with at least 135, at most 628, and on average 303 transactions per quarter. Table 16.1 reports summary statistics for the main characteristics

of the houses. Obviously, all the characteristics that a valuer would use for computing a sales comparison value are observed.

Panel A: Continuous Characteristics, Prices, and Cost				
	Mean	Median	Std. Dev.	Units
Lot size	566.8	514.0	308.3	Sqm
Floor space	147.7	137.0	53.3	Sqm
Gross volume	657.2	599.0	253.1	Cm
Gross base	247.4	232.0	90.0	Sqm
Year of construction	1961	1962	29.0	Year
Price	228.7	198.5	14.0	(000)
Building cost	185.8	173.4	82.4	(000)
Land cost	120.7	91.1	117.2	(000)
Panel B: House Type				
Detached	52.7%	Semi-detached		22.2%
End-row	16.9%	Mid-row		15.8%
Panel C: Location and Lake Side				
Simple	32.1%	Average		46.5%
Good	18.9%	Excellent		2.0%
Lake side	0.9%			
Panel D: Number of Storeys and Attic				
One	54.3%	Two		43.6%
Three	2.1%	Attic		55.0%
Panel E: Cellar				
Full	77.4%	Part		11.6%
No	10.9%			

Notes: 9088 observations. Gross base is the sum of all base areas in all storeys, gross volume is the corresponding volume. 4017 objects have information on the gross volume and 9063 on the gross base. Prices and cost are in year 2000 Euros. Building cost are cost of constructing a new building. Attic in Panel D means that the attic is upgraded for living.

Table 16.1. Summary statistics for transacted single-family houses in Berlin between 2000Q1 to 2007Q2.

The building cost in our data set were computed by GAA surveyors based on information gathered and published by the German government (Bundesministerium für Raumordnung, Bauwesen und Städtebau, 1997; Bundesministerium für Verkehr, Bau- und Wohnungswesen, 2001). The published infor-

mation gives the average building cost for many different house specifications in Germany. The land cost in our data set are the value of land if the site of the subject house were undeveloped. GAA surveyors appraised these land cost using the sales comparison approach and their database of all land transactions.

16.3 Empirical Results

16.3.1 Characterization of the Test Market

Figure 16.1 shows the trend of house price, land cost, and construction cost for a constant-quality single-family house in Berlin over the period 1980Q1 to 2007Q2. The index values are computed as

$$100 \exp\{\widehat{\beta}_{0t} - 0.5\widehat{\sigma}_t^2\} ,$$

which corrects for small-sample bias (Kennedy, 1998, p. 37). $\widehat{\beta}_{0t}$ is the estimated coefficient of the period-t dummy variable in a hedonic regression with either house price or land cost as the dependent variable and $\widehat{\sigma}_t^2$ is the corresponding estimated robust variance of the coefficient estimator. The quarterly construction cost index is provided by the Statistical Office Berlin in its Statistical Report M I 4. It measures the change of the construction cost for a new single-family building.

The movement of prices for existing houses and the cost of constructing new houses are closely related. This is in line with economic reasoning because if house prices are above replacement cost (i.e., the *sum* of land cost and building cost) then it is profitable for developers to construct new houses. The additional supply will increase the housing stock and, given unchanged demand, dampen house price growth. Developers will provide additional supply until prices of existing houses are realigned with replacement cost and no extra profits can be made. In the case that house prices fall below replacement cost, developers will provide no new supply at all and the housing stock will shrink until equilibrium is reached again. This reasoning motivates the use of the cost approach for forecasting long-term house values, because even if prices and replacement cost deviate at the date of valuation, they ought to move closer to each other in the near future. If replacement cost is a better predictor of the future price of a home than any function of past prices, then this could put the cost approach at advantage even if the sales comparison approach has been found in previous studies to perform better with respect to short-term valuations.
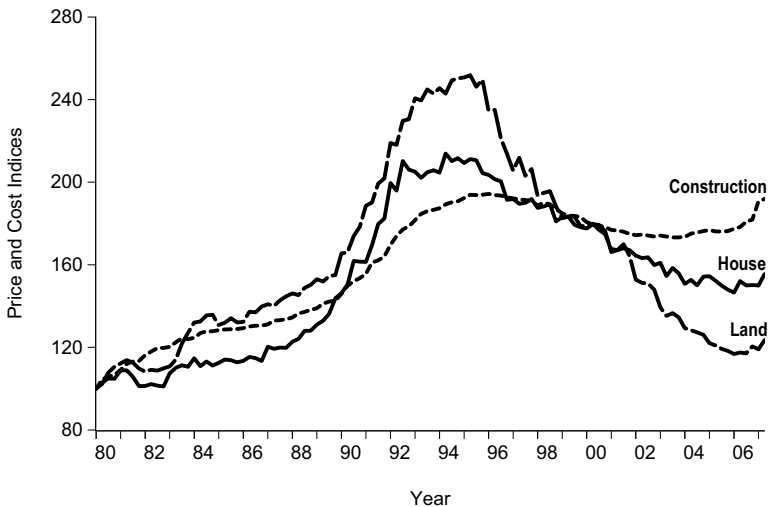
Figure 16.1. Constant-quality house price and land cost indices, and construction cost indices for single-family houses in Berlin, 1980Q1-2007Q2. Series are normalized to 100 in 1980Q1.

Variable y	Model specification	$\widehat{\sigma}_{\Delta \ln y}$	R^2
House price	$\Delta \ln y_t = c_t + \varepsilon_t$	2.6	13.2
Land cost	$\Delta \ln y_t = c_t + \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} + \theta_4 \varepsilon_{t-4} + \varepsilon_t$	2.8	44.1
Construction cost	$\Delta \ln y_t = c_t + \phi_4 \Delta \ln y_{t-4} + \theta_3 \varepsilon_{t-3} + \varepsilon_t$	0.9	50.6

Notes: The constant is $c_t = c_0 + c_1 I_{1993Q2}(t)$, where $I_{1993Q2}(t)$ is an indicator function, which is 1 if $t \geqslant 1993Q2$ and 0 otherwise. ε_t is the residual. The estimated volatility $\widehat{\sigma}_{\Delta \ln y}$ and the coefficient of determination R^2 are expressed in percent.

Table 16.2. Time series model specifications fitted to the three different index series. Volatility and coefficient of determination are for the full sample fit with data from 1980Q1 to 2007Q2.

For the forecast experiment, all three series are treated as difference-stationary time series and ARMA models are fitted to their growth rates. Table 16.2 presents the ARMA specifications for the three different series, the volatility of the growth rates over the full sample and the respective regression fit. In the case of the two estimated constant-quality series we take the log indices directly from the hedonic regressions (instead of re-transforming the

indices again). The regression constant c_t of the specifications in Table 16.2 allows for a shift in the respective growth rate after the introduction of the European single market in 1993. The specifications have a parsimonious parametrization and the fitted models have uncorrelated residuals according to the standard tests for autocorrelation (Q-Statistic and LM test). To simplify the forecast experiment, we fit the same specifications to all sample periods, regardless of their length. In most cases the residuals of the specifications fitted over shorter sample periods rather than the full sample period are uncorrelated and all coefficients are statistically significant.

It follows from the specification for the house price growth rate in Table 16.2 that the house price index follows a random walk. If we were to assume that the required return rate for a housing investment is constant and the unobserved imputed rent is proportional to the house price, then a random walk would indicate that prices are set in an informational efficient manner. Without the lagged MA terms, the land cost index would follow a random walk, too. It seems reasonable to attribute the moving average terms to the valuation process with which land cost are computed (appraisal smoothing). The growth of construction cost exhibits a strong seasonal component.

As is obvious from Figure 16.1, the construction cost series has a much smaller volatility than the other two series. Moreover, because of the strong seasonal component, the in-sample predictability of construction cost growth is higher than for the other two series as indicated by the R^2s. Thus, it might be possible to forecast construction cost with greater accuracy. Compared to the price regression, the land cost regression provides a much better fit of the data, which might indicate that land cost can be forecasted more accurately as well, making a combination of construction cost and land cost superior to direct forecasting of the house price index.

Figure 16.2 compares two different price forecasts for the last five years of the full sample period with the full sample house price index. The first forecast is based on the house price specification fitted to the data up to 2002Q2. This is a forecast of the house price index itself and corresponds to the very idea of the sales comparison approach. The second forecast is based on a weighted average of the land and construction cost indices, both forecasted in 2002Q2 based on the information available at that time. We assume that building cost account for 55% of replacement cost while land cost account for the remaining 45%. Using the replacement cost index as a forecast of the future level of house prices corresponds to the very idea of cost approach. Figure 16.2 reveals that both forecasts seem to perform well.

Although the house price index estimated with the data up to 2002Q2 and the index estimated with the full data sample show a very similar behavior
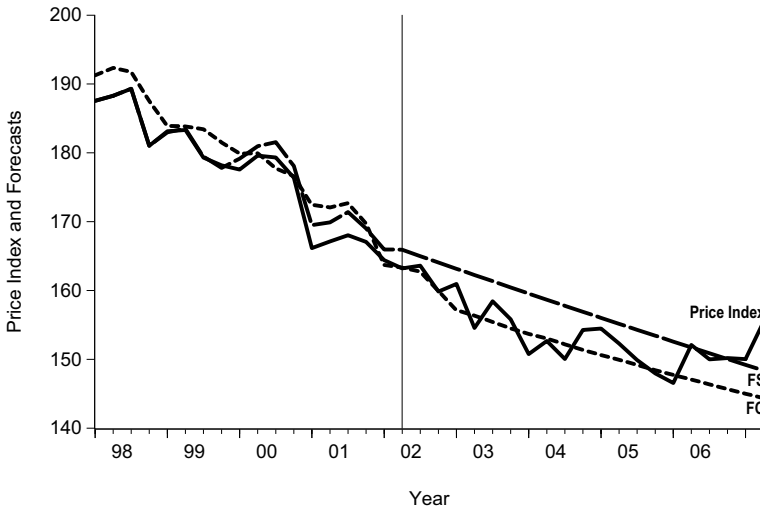
Figure 16.2. Full sample house price index and forecasts for the period 2002Q3-2007Q2 (right from vertical line) based on information up to 2002Q2. The sales comparison approach forecast (FS) is based on the time series model for the price index, the cost approach forecast (FC) is a weighted average of the forecasted land cost and construction cost indices.

before 2002Q3, they are not identical. This is the results of the rolling window estimation technique we apply. New information due to the extension of the estimation sample can influence the estimated index coefficients in preceding quarters. The difference of the two house price indices in Figure 16.2 before 2002Q2 are not statistically significant, but the point estimates differ. The index revision problem is not specific to the constant-quality indices, but applies also to official indices like the construction cost index. Consequently, the forecaster often has to work with provisional time series and there is no solution to this problem.

There are two additional aspects that have to be considered. First, the full sample house price index itself might not be the best benchmark for assessing forecast accuracy. Second, and closely related, because the time series are normalized indices, the seemingly good forecasting behavior of the replacement cost in Figure 16.2 should not be misinterpreted: the near equality of the full sample house price index and the replacement cost index in period 2002Q2 might simply be the result of the arbitrary index normalization. House prices in that period might be larger than replacement cost, in which case forecasted long-term cost values will be below prices during the whole forecasting hori-

zon. If, on the other hand, replacement cost are slightly above house prices in period 2002Q2, then the forecasted long-term cost values might be even closer to prices over the forecasting horizon than Figure 16.2 indicates.

Because of these possible estimation and normalization effects, a pure comparison of index series is no substitute for the evaluation of individual house-specific forecast errors. Only a direct comparison of valuations and transaction prices can reveal the accuracy of a valuation technique. The results of such a direct comparison are presented in the next section.

16.3.2 Horse Race

To measure forecasting accuracy at the individual level we use the valuation error defined as

$$e_{t+h} = \log P_{t+h} - \log V_t \;,$$

where P_{t+h} is the observed transaction price of a house in period $t + h$ and V_t is the valuation made for this house based on information in period t. We focus on the five quarterly forecast horizons $h \in \{4, 8, 12, 16, 20\}$, which correspond to 1, 2, 3, 4, and 5 years, respectively. We use log errors, because they treat over- and undervaluations symmetrically. If the errors are small, then e_{t+h} is a close approximation of the error relative to the valuation

$$\frac{P_{t+h} - V_t}{V_t}$$

and $-e_{t+h}$ is a close approximation of the valuation error relative to the price

$$\frac{V_t - P_{t+h}}{P_{t+h}} \;.$$

Clearly, a valuation technique is the better the smaller the valuation errors are on average and the less dispersed they are. To save on notation, we use N_h to denote the number of transactions for which we make valuations with a horizon of h and we use $e_{h,n}$ to denote the valuation error for house n. The mean error of a valuation technique for forecast horizon h is then

$$\mathrm{ME}_h = \frac{1}{N_h} \sum_{n=1}^{N_h} e_{h,n} \;,$$

i.e., the arithmetic average over all errors with the same forecast horizon h. The mean error does not take the dispersion of the errors into account. A valuation technique might have a small mean error while individual valuations

are never on the mark but either far too large or far too small. The following two measures take the dispersion into account. The first is the mean absolute error

$$\mathrm{MAE}_h = \frac{1}{N_h} \sum_{n=1}^{N_h} |e_{h,n}|$$

and the second is the mean squared error

$$\mathrm{MSE}_h = \frac{1}{N_h} \sum_{n=1}^{N_h} e_{h,n}^2 \ .$$

Both measures are symmetric and give the same weight to positive and

Valuation approach	Horizon	ME	MDE	MSE	MAE	PE25
	1	0.9	1.8	8.8	22.6	65.0
	2	-0.3	0.8	8.8	22.6	65.2
Cost value	3	-2.2	-0.9	9.1	22.8	64.7
	4	-5.6	-4.3	9.5	23.4	63.4
	5	-11.4	-10.3	11.0	25.5	59.5
	1	-3.3	-2.3	6.2	18.7	73.4
	2	-3.3	-2.4	6.7	19.5	71.6
Sales comparison value	3	-4.6	-4.3	7.2	20.2	69.7
	4	-6.6	-5.7	7.9	21.3	67.2
	5	-7.9	-7.3	8.6	22.5	64.3
	1	-1.9	-0.6	6.2	18.7	72.9
	2	-2.5	-1.4	6.4	19.1	72.0
Combination	3	-4.1	-3.0	6.7	19.5	71.2
	4	-6.9	-5.8	7.3	20.5	69.3
	5	-10.6	-9.4	8.2	21.8	66.2

Notes: All reported measures are in percent. Number of observations is 9088 per valuation method and forecast horizon. ME is the mean error, MDE the median error, MSE the mean squared error, MAE the mean absolute error, and PE25 is the relative frequency of valuation errors within the ±25% range. Combination is an equally-weighted average of the cost and the sales comparison values.

Table 16.3. Performance of sales comparison and cost values over different yearly forecast horizons. Summary statistics of valuation errors for transactions between 2000:1 to 2007:2.

negative errors of equal absolute magnitude. In many situations where the economic loss due to under- or overvaluations is unknown, this might be a good compromise. A negative valuation error corresponds to a forecasted value above the realized transaction price. In the context of the mortgage

underwriting process, such overestimation could lead to underwriting based on a false perception of collateral value in the case of default. Overestimation does not necessarily need to lead to an actual loss in the case of default, because the loss also depends on the outstanding loan balance. The sale of the collateral may still be enough to cover borrower's outstanding liabilities. However, from a risk management perspective, it is desirable that loan underwriting is based on a correct assessment of the recovery value of the collateral. Moreover, the initial loan might be directly related to the collateral value and overestimation could lead to larger and more risky loans than are perceived during the underwriting process. A positive valuation error corresponds to a forecasted value below the realized price. In this case the collateral will always be sufficient to cover any outstanding loan balance. The economic loss due to underestimated collateral values stems from the fact that loan applications may get declined during the underwriting process. This is foregone business for the mortgage underwriter, because the true value of the collateral could have been more than sufficient to fulfill the underwriting criteria. Using the MSE and the MAE as accuracy measures thus corresponds to the assumption that the economic loss of over- and undervaluation is the same.

Table 16.3 presents the forecast evaluation measures for cost and sales comparison values and an equally-weighted combination of both. In addition to the measures already discussed above, Table 16.3 also reports the median valuation error and the percentage of observations for which the valuation lies within $\pm 25\%$ of the observed transaction price. The first two panels of Table 16.3 show that the sales comparison values perform better than the cost values for each of the five forecast horizons. Although the cost values have smaller mean errors than the sales comparison values for all but the five year horizon, the variation of these errors is larger, as the MSE and the MAE clearly indicate. Moreover, the percentage of valuations that lie within $\pm 25\%$ of the transaction price is larger for the sales comparison approach than for the cost approach.

One may object that the above comparison is based on a sample of transaction prices only and that transaction prices in general may deviate from unobserved market values, i.e., the expected price. It could be that cost values forecast market values perfectly well, but this goes undetected, because observed prices can and will deviate from market values. Diebold and Mariano (1995) proposed several tests for the comparison of different forecast methods that take such uncertainty into account. The test on the MSE uses the $N = 9088$ differences of the squared errors

$$e^2_{C,h,n} - e^2_{S,h,n} \,,$$

where C stands for the cost valuation error and S for the sales comparison

valuation error, and tests if the difference is equal to zero on average (same MSE) or if the difference is at most as large as zero (cost values are at least as good as sales comparison values, possibly even better). The test on the MAE uses

$$|e_{C,h,n}| - |e_{S,h,n}| \,,$$

but is otherwise identical to the test on the MSE. Applied to our data, we can reject the hypothesis that the cost values have a MSE at most as large as the sales comparison values at the 1% significance level, i.e., we reject $\mathrm{MSE}_C \leqslant \mathrm{MSE}_S$. We can reject the equivalent hypothesis for the MAE at the same level of significance, i.e., we reject $\mathrm{MAE}_C \leqslant \mathrm{MAE}_S$. Another important test is the Sign test, which counts the number of observations where the cost value is closer to the price than the respective sales comparison value, i.e., how often it is true that

$$|e_{C,h,n}| \leqslant |e_{S,h,n}| \,.$$

If both valuation approaches were of equal accuracy, then the probability of one being better than the other would be 0.5. If we have N pairwise observations of valuation errors, then we expect under the assumption of equal accuracy that the cost values are better in 50% of the observations and the sales comparison values in the remaining 50%. For our data, however, the cost values are better for only 44.2% of the pairwise observations over all forecast horizons, whereas the sales comparison values are better for 55.8% of the observations. Given the total number of observations, $N = 9088$, these frequencies are unlikely to have been generated by valuation approaches with equal accuracy. We can reject the hypothesis that the cost values are at least as accurate as the sales comparison values for each of the forecast horizons at the 1% significance level.

Taking the first two panels of Table 16.3 and the test results together, we conclude that the sales comparison approach performs better than the cost approach based on the MSE and MAE criteria. The third panel of Table 16.3 shows that an equally-weighted average of both approaches delivers an even better performance than stand-alone sales comparison values. Other than equal weights for the two values are possible, which might enhance the performance even further. The performance results on the long-term valuation accuracy of sale comparison and cost values are thus identical to the results obtained in previous studies for valuations with a short-term horizon.

Both the MSE and the MAE weigh positive and negative valuation errors symmetrically. In the context of mortgage underwriting, however, it is open to debate if the cost of foregone business due to underestimating the collateral value is the same as the cost of a loan that is collateralized with a property that has a much lower market value than indicated by the forecasted long-

term valuation. One could therefore argue that positive valuation errors are less costly than negative valuation errors. The true economic loss function would be then asymmetric, putting more weight on negative valuation errors. The main problem with this reasoning is that the true economic loss function is unknown and might be complicated to establish. Because of this, Shiller and Weiss (1999) have proposed to investigate the asymmetry issue by looking at estimates of the distributions of the valuation errors.

Figures 16.3 and 16.4 show kernel density estimates for the valuation error distributions with a horizon of two and five years. We select the bandwidth according to Silverman's rule of thumb; asymptotic confidence bands are estimated at the 95% level, see Härdle, Müller, Sperlich and Werwatz (2004, Chapter 3). The density estimates for the horizons of one, three, and four
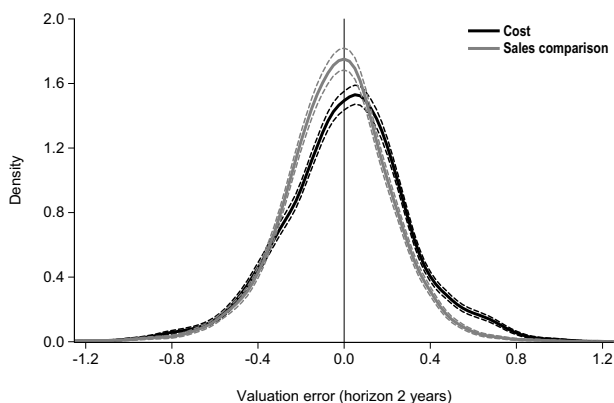
Figure 16.3. Kernel density estimates for the valuation error distributions of the cost and the sales comparison values. The forecast horizon is two years. The dashed lines are 95% confidence intervals.

years are very similar in shape to the density for the two year horizon in Figure 16.3. It emerges from these density estimates that the valuation error distribution of the sales comparison values is quite symmetric around its mean error, which is -3.3%, but shifted to the right if an expected error of zero is taken as reference. The distribution of the valuation errors of the cost values, on the other hand, is less symmetric around its mean error of -0.3%. Furthermore, the cost values have a larger probability (51.3%) for producing non-negative errors than the sales comparison values (45.6%). Compared to the sales comparison values, it is more likely that a cost value underestimates the future price. Severe underestimations, where the valuation is only 20-40% of the transaction price, are much more likely to occur with cost values

compared than with sales comparison values. This is shown by the dent in
the density function on the right side. If underestimation leads to a lower
economic loss than overestimation, then this might indicate an advantage
of the cost approach. Without an explicitly specified asymmetric economic
loss function, however, it is not possible to compute the magnitude of this
possible advantage.

A different picture emerges for the distribution of the valuation errors at
the five year forecast horizon. Both distributions are shifted to the left and
only 35.6% of the cost values produce a positive valuation error compared to
38.5% of the sales comparison values. The dent in the density function for
large underestimations of the transaction price is visible again.
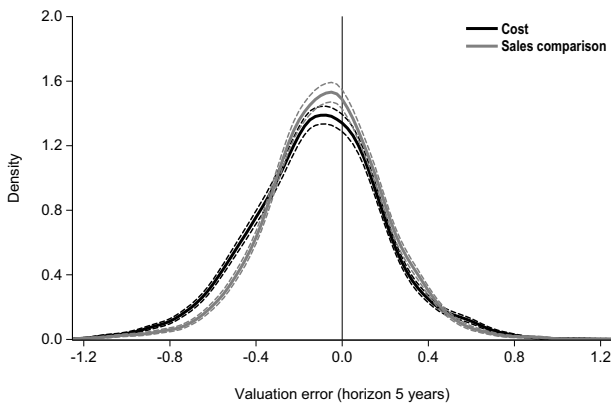
Figure 16.4. Kernel density estimates for the valuation error distributions of the cost and the sales comparison values. The forecast horizon is five years. The dashed lines are 95% confidence intervals.

Figures 16.3 and 16.4 are also useful to assess the effect of proportional deductions on valuation errors. Such deductions are required for the computation of mortgage lending values. Let γ denote the proportional deduction, say 20%, then the resulting mortgage lending value is $(1-\gamma)V$. The corresponding lending valuation error distribution would then simply correspond to the plotted valuation error distributions shifted to the right by approximately γ.

16.4 Conclusion

The direct comparison has shown that sales comparison values perform better than cost values if the economic loss function is symmetric. If both values are available, however, then an equally-weighted average of both cost and sales comparison values produces smaller losses on average than each of the values alone. Pooling the valuations is thus advisable and the cost value, although inferior to the sales comparison values in a direct comparison, still provides information for better valuations. If the loss function is asymmetric, penalizing overvaluations more than undervaluations, then it might be possible that cost values are better in a direct comparison than the sales comparison values. It is more likely for a cost value to underestimate the transaction price of a house than it is for a sales comparison value.

Without further knowledge on the proper economic loss function to be applied to valuation errors it is not possible to arrive at a final assessment. Further work needs to explore and incorporate a specific form of the economic loss function. Given the deductions required for the computation of mortgage lending values, it seems plausible that losses from overestimation are more problematic in practice than losses from underestimation.

A shortcoming of our study is that from the first quarter of 2000 onwards prices were steadily falling – only in the last quarter do prices seem to have gained some upward momentum. This may explain why the mean valuation errors are negative in all but one case. Moreover, our data are for only one region with a large number of comparable sales. The performance of the sales comparison approach might be worse in regions with less active markets. Future studies have to make use of longer time periods and should also extend the horizons over which forecasts are being made.

Bibliography

Bundesministerium für Raumordnung, Bauwesen und Städtebau (1997). Normalherstellungskosten 1995. NHK 95, Bonn.

Bundesministerium für Verkehr, Bau- und Wohnungswesen (2001). Normalherstellungskosten 2000. NHK 2000, Berlin.

Cannaday, R. E. and Sunderman, M. A. (1986). Estimation of depreciation for single-family appraisals, *AREUEA Journal* **14**: 255–273.

Diebold, F. X. and Mariano, R. S. (1995). Comparing predictive accuracy, *Journal of Business & Economic Statistics* **13**: 253–263.

Dotzour, M. G. (1990). An empirical analysis of the reliability and precision of the cost approach in residential appraisal, *Journal of Real Estate Research* **5**: 67–74.

Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). *Nonparametric and Semiparametric Models*, Springer Series in Statistics, Springer, Berlin, New York.

Kennedy, P. (1998). *A Guide to Econometrics*, fourth edn, Blackwell, Oxford.

Schulz, R. and Werwatz, A. (2004). A state space model for Berlin house prices: Estimation and economic interpretation, *Journal of Real Estate Finance and Economics* **28**: 37–57.

Schulz, R. and Werwatz, A. (2008). House prices and replacement cost: A micro-level analysis, *SFB* 649 *Discussion Paper* 2008-013, Humboldt-Universität zu Berlin, SFB 649.

Shiller, R. J. and Weiss, A. N. (1999). Evaluating real estate valuation systems, *Journal of Real Estate Finance and Economics* **18**: 147–161.

Theil, H. (1966). *Applied Economic Forecasting*, North-Holland, Amsterdam. Assisted by G.A.C. Beerens and C.G. De Leeuw and C.B. Tilanus.

17 Locally Time Homogeneous Time Series Modelling

Mstislav Elagin and Vladimir Spokoiny

17.1 Introduction

Modelling particular features ("stylized facts") of financial time series such as volatility clustering, heavy tails, asymmetry, etc. is an important task arising in financial engineering. For instance, attempts to model volatility clustering, i.e. the tendency of volatility jumps to appear in groups followed by periods of stability, led to the development of conditional heteroskedastic (CH) models including ARCH by Engle (1982) and GARCH by Bollerslev (1986) as well as their derivatives. The main idea underlying the mentioned methods is that volatility clustering can be modelled globally by a stationary process.

However, the assumption of stationarity is often compromised by the shape of the autocorrelation function (ACF) of squared log returns that for a typical financial time series decays slower than exponentially. Furthermore, Mikosch and Stărică (2004) showed that long range memory effects in financial time series may be caused by structural breaks rather than that constitutes an essential feature of stationary processes to be modeled by global methods. Diebold and Inoue (2001) and Hillebrand (2005) argue that one can easily overlook structural breaks with negative impact on the quality of modelling, estimation and forecasting. This circumstance motivates the development of methods involving processes that are stationary only locally. Local methods consider just the most recent data and imply subsetting of data using some localization scheme that can itself be either global or local and adaptive. Methods of this kind have been presented e.g. in Fan and Gu (2003) for adaptive selection of the decay factor used to weight components of the pseudo-likelihood function, in Dahlhaus and Subba Rao (2006) for the formulation of the locally stationary ARCH(∞) processes, in Cheng, Fan and Spokoiny (2003) for locally choosing parameters of a filter. In a recent paper

by Giacomini, Härdle and Spokoiny (2008) a local adaptive method has been applied to the problem of copulae estimation.

Below we compare three methods for estimation of parameters in the context of univariate time series: the local change point (LCP) procedure by Mercurio and Spokoiny (2004), the local model selection (LMS), also known as the intersection of confidence intervals (ICI) by Katkovnik and Spokoiny (2008), and the stagewise aggregation (SA) by Belomestny and Spokoiny (2007). A universal procedure for the choice of parameters (*critical values*) is given. The performance of the procedures is compared using genuine financial data. It is shown that adaptive methods often outperform the standard GARCH(1,1) method.

The chapter is organized as follows. Section 17.2 is devoted to the formulation of the problem and theoretical introduction. Section 17.3 describes the methods under comparison. In Section 17.4 the procedure for obtaining *critical values*, essential parameters of the procedures, is given. Section 17.5 shows the application of the adaptive methods to the computation of the value-at-risk.

17.2 Model and Setup

17.2.1 Conditional Heteroskedastic Model

Let S_t be a one-dimensional stochastic asset price process in discrete time $t \in \mathbb{N}$ and $R_t = \log S_t / S_{t-1}$ be the corresponding log returns process. The latter is typically described using the *conditional heteroskedastic* model

$$R_t = \sigma_t \varepsilon_t, \tag{17.1}$$

where ε_t are independent and identically (standard Gaussian) distributed innovations, and σ_t is the *volatility* process progressively measurable w.r.t. the filtration $(\mathcal{F}_{t-1}) = \mathcal{F}(R_1, \ldots, R_{t-1})$ generated by past returns. Equivalently,

$$Y_t = \theta_t \varepsilon_t^2 \tag{17.2}$$

where $Y_t = R_t^2$ are the squared log returns and $\theta_t = \sigma_t^2$. We aim to estimate θ_t from the past observations $Y_1, \ldots, Y_{t-1}$. This problem commonly arises in financial applications such as value-at-risk determination and portfolio optimisation.

17.2.2 Parametric and Local Parametric Estimation and Inference

If $\theta_t = \theta$ one can apply the method of maximum likelihood to obtain the estimate $\widehat{\theta}$. The model (17.2) leads to the log-likelihood function

$$L(\theta) = \sum_t \ell(Y_t, \theta)$$

where $\ell(y, \theta) = -\frac{1}{2}\log(2\pi\theta) - y/(2\theta)$ is the log density of the normal distribution with zero mean. The estimate $\widehat{\theta}$ is then obtained by maximizing the log-likelihood function w.r.t. to θ:

$$\widehat{\theta} = \arg\max_\theta L(\theta) = \frac{\sum_t Y_t}{N},$$

where N is the sample size. When the volatility does depend on time, $\theta_t = \theta(t) \neq \text{const.}$, the method of maximum likelihood is not directly applicable, since the joint distribution of the observations and therefore the log likelihood function are not available. Hence, we take the *local parametric approach* by supposing that for the time point of estimation T there exists some interval $\mathcal{I} = [T - N_\mathcal{I}, T]$ of length $N_\mathcal{I}$, to be estimated from the data, within which the model (17.2) describes the process adequately. If the interval $\mathcal{I}$ has been found, then the log likelihood function assumes the form

$$L_\mathcal{I}(\theta) = \sum_{t \in \mathcal{I}} \ell(Y_t, \theta)$$

and the maximum likelihood estimate corresponding to the interval $\mathcal{I}$ is

$$\widetilde{\theta}_\mathcal{I} = \arg\max_\theta L_\mathcal{I}(\theta) = \sum_{t \in \mathcal{I}} Y_t / N_\mathcal{I}.$$

For the purpose of describing the quality of estimation we use the fitted likelihood $L(\widetilde{\theta}, \theta)$ defined as the difference between the likelihood corresponding to the ML estimate $\widetilde{\theta}$ and the likelihood corresponding to a different parameter value:

$$L(\widetilde{\theta}, \theta) = L(\widetilde{\theta}) - L(\theta).$$

For the model considered here the fitted likelihood can be represented in the form

$$L(\widetilde{\theta}_\mathcal{I}, \theta) = N_\mathcal{I} \mathcal{K}(\widetilde{\theta}_\mathcal{I}, \theta), \tag{17.3}$$

where

$$\mathcal{K}(\theta_1, \theta_2) = \frac{1}{2}(\theta_1/\theta_2 - 1) - \frac{1}{2}\log(\theta_1/\theta_2)$$

denotes the Kullback – Leibler divergence that measures the "distance" between distributions indexed by θ_1 and θ_2.

17.2.3 Nearly Parametric Case

In practice the parametric assumption may be overly stringent and not hold even within an arbitrarily small interval. We describe the deviation from the parametric situation within an interval $\mathcal{I}$ by a magnitude:

$$\Delta_{\mathcal{I}}(\theta) = \sum_{t \in \mathcal{I}} \mathcal{K}(\theta_t, \theta),$$

that we shall call *divergence*. The following *small modelling bias (SMB)* condition imposes a limit on the deviation from the parametric case which provides the applicability of the local parametric approach.

Condition 1 *There exists some parameter value $\theta \in \Theta$ and some interval $\mathcal{I}$ such that the expectation under the true measure of the divergence $\Delta_{\mathcal{I}}(\theta)$ over the interval $\mathcal{I}$ is bounded by some $\Delta \geq 0$:*

$$\mathsf{E}\,\Delta_{\mathcal{I}}(\theta) \leq \Delta. \tag{17.4}$$

If the SMB condition 17.4 holds, then for any $r > 0$ the risk of the local maximum likelihood estimate in the nearly parametric case satisfies:

$$\mathsf{E}\log\left(1 + \frac{\left|N_{\mathcal{I}}\mathcal{K}(\widetilde{\theta}_{\mathcal{I}}, \theta)\right|^r}{\mathcal{R}_{r,\theta}}\right) \leq \Delta + 1,$$

where

$$\mathcal{R}_{r,\theta} = \mathsf{E}_\theta\left|N_{\mathcal{I}}\mathcal{K}(\widetilde{\theta}_{\mathcal{I}}, \theta)\right|^r \tag{17.5}$$

is the risk of the local maximum likelihood estimate in the parametric case. Here the logarithm under the expectation comes from the Cramér – Rao inequality, and the additional term Δ on the right-hand side can be interpreted as payment for the violation of the parametric assumption.

The last result leads to the notion of the *oracle* estimate as the "largest" one under the small modelling bias condition. In the next section we present three methods suitable for construction of estimates performing almost as well as the oracle estimate.
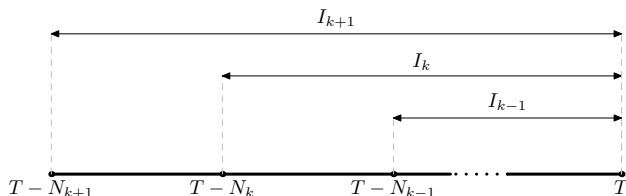
Figure 17.1. Nested intervals.

17.3 Methods for the Estimation of Parameters

17.3.1 Sequence of Intervals

Local methods imply subsetting of data. A localization scheme that we use is a growing sequence of intervals. Let T denote the time point at which the value of interest is to be estimated. We define an ordered sequence of intervals $\{\mathcal{I}_k\}_{k=1}^K$ of length N_k with the common right edge at T (Figure 17.1), so $\mathcal{I}_k = [T - N_k, T[$. We associate with each interval $\mathcal{I}_k$ from this sequence the corresponding maximum likelihood estimate $\widetilde{\theta}_k \equiv \widetilde{\theta}_{\mathcal{I}_k}$, which we shall call *weak estimate.* We aim to select or construct the "largest" one still satisfying the small modelling bias condition. The LCP and LMS procedures obtain the best estimate by choosing one from the sequence, whereas SSA builds the estimate by taking convex combinations of previously found estimates. Below we describe each of the methods.

17.3.2 Local Change Point Selection

The LCP method introduced in Mercurio and Spokoiny (2004) is a procedure that detects the largest interval of homogeneity and provides an adaptive estimate as the one associated with the interval found. The idea of the method consists in the testing of the null hypothesis of an interval containing no change points against the alternative hypothesis of a change point being present, whereas the interval under testing is taken from the growing sequence.

Consider a *tested interval* $\mathcal{I}$ that possibly contains a change point, and an enclosing *testing interval* I (Figure 17.2). The statistic to test the hypothesis about the parameter change in some internal point τ of the candidate interval can be expressed as the difference between the sum of log likelihoods corresponding to the intervals I', I'' into which the change point splits the testing interval, and the log likelihood corresponding to the testing interval
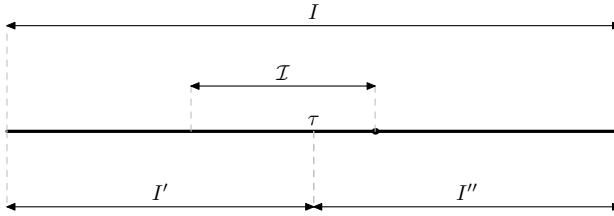
Figure 17.2. Intervals involved in the change point detection procedure.

containing no change points:

$$T_{\mathcal{I},\tau} = \max_{\theta',\theta''} \left\{ L_{I''}(\theta'') + L_{I'}(\theta') \right\} - \max_{\theta} L_I(\theta) = L_{I'}(\widetilde{\theta}_{I'}) + L_{I''}(\widetilde{\theta}_{I''}) - L_I(\widetilde{\theta}_I),$$

where $L(\cdot)$ denotes the log likelihood function. For the volatility distribution the test statistic can be represented in the form

$$T_{\mathcal{I},\tau} = \min_{\theta} \left\{ N_{I''}\mathcal{K}(\widetilde{\theta}_{I''},\theta) + N_{I'}\mathcal{K}(\widetilde{\theta}_{I'},\theta) \right\} = N_{I''}\mathcal{K}(\widetilde{\theta}_{I''},\widetilde{\theta}_I) + N_{I'}\mathcal{K}(\widetilde{\theta}_{I'},\widetilde{\theta}_I) \tag{17.6}$$

due to (17.3). The test statistic for the whole candidate interval is the maximum of the pointwise statistics over all internal points:

$$T_{\mathcal{I}} = \max_{\tau \in \mathcal{I}} T_{\mathcal{I},\tau}$$

The hypothesis is rejected if the test statistic exceeds some *critical value* $\mathfrak{z}$, which is a parameter of the procedure specific to the problem design.

We let $\mathcal{I} = I_k \backslash I_{k-1}$ and $I = \mathcal{I}_{k+1}$ and take the adaptive estimate $\widehat{\theta}$ to be equal to the $\widehat{k}$-th weak estimate, where $\widehat{k}$ is the largest interval number such that all test statistics corresponding to the intervals $\mathcal{I}_1, \ldots, \mathcal{I}_{\widehat{k}}$ do not exceed their critical values with the opposite holding for $\widehat{k} + 1$:

$$\widehat{\theta} = \widetilde{\theta}_{\widehat{k}}, \quad \text{where} \quad \widehat{k} = \max k \quad \text{such that} \quad T_l \leq \mathfrak{z}_l \quad \text{for all} \quad l \leq \widehat{k}.$$

The initial condition is that the smallest interval is always considered to be homogeneous. Since it is not feasible to test the largest interval, the greatest possible value of $\widehat{k}$ is $K - 1$.

17.3.3 Local Model Selection

The idea of the local model selection procedure introduced in Katkovnik and Spokoiny (2008) consists in the choice of the "largest" weak estimate among
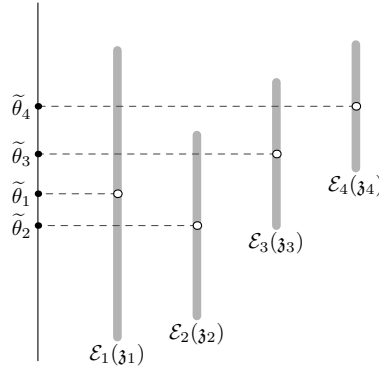
Figure 17.3. Principle of the local model selection. $\widehat{\theta} = \widetilde{\theta}_3$.

$\widetilde{\theta}_1 \ldots \widetilde{\theta}_K$ as the adaptive estimate $\widehat{\theta}$ in such a way that the adaptive estimate belongs to the confidence interval $\mathcal{E}$ of each of the previous weak estimates (Figure 17.3). Formally, $\widehat{\theta} = \widetilde{\theta}_{\widehat{k}}$, where

$$\widehat{k} \text{ is such that } \begin{cases} \widetilde{\theta}_{\widehat{k}} \in \mathcal{E}_l & \text{for all } l < \widehat{k} \\ \widetilde{\theta}_{\widehat{k}+1} \notin \mathcal{E}_l & \text{for some } l < \widehat{k} + 1 \end{cases}$$

Confidence interval of level α for a weak estimate $\widetilde{\theta}$ is provided by

$$\mathcal{E}(\mathfrak{z}_\alpha) = \left\{ \theta : L(\widetilde{\theta}, \theta) \leq \mathfrak{z}_\alpha \right\}.$$

As with the LCP procedure, the first weak estimate is always accepted. However, the LMS procedure checks all estimates including the one corresponding to the last interval.

17.3.4 Stagewise Aggregation

The SA procedure introduced in Belomestny and Spokoiny (2007) differs from the two methods described above in that it does not choose the adaptive estimate $\widehat{\theta}$ from the weak estimates $\widetilde{\theta}_1 \ldots \widetilde{\theta}_K$. Instead, based on the weak estimates, it sequentially constructs *aggregated estimates* $\widehat{\theta}_1 \ldots \widehat{\theta}_K$ possessing the property that any aggregated estimate $\widehat{\theta}_k$ has smaller variance than the corresponding weak estimate $\widetilde{\theta}_k$, while keeping "close" to it in terms of the statistical difference, the latter being measured through the likelihood ratio $L(\widehat{\theta}_k, \widehat{\theta}_{k-1}) = L(\widehat{\theta}_k) - L(\widehat{\theta}_{k-1})$. The adaptive estimate is finally taken equal to the last aggregated estimate: $\widehat{\theta} = \widehat{\theta}_K$ (unless an *early stopping* occurs).

Formally, the first aggregated estimate is equal to the first weak estimate and every next aggregated estimate is a convex combination of the previous aggregated estimate and the current weak estimate:

$$\widehat{\theta}_k = \begin{cases} \widetilde{\theta}_1, & k = 1 \\ \gamma_k \widetilde{\theta}_k + (1 - \gamma_k)\widehat{\theta}_{k-1}, & k = 2, \ldots, K \end{cases}$$

Here γ_k is the *mixing coefficient* that reflects the statistical difference between the previous aggregated estimate $\widehat{\theta}_{k-1}$ and the current weak estimate $\widetilde{\theta}_k$, and is obtained by applying an *aggregation kernel* K_{ag} to the likelihood ratio $L(\widetilde{\theta}_k, \widehat{\theta}_{k-1})$ scaled by the *critical value* $\mathfrak{z}_k$:

$$\gamma_k = K_{\mathrm{ag}} \left(\frac{L(\widetilde{\theta}_k, \widehat{\theta}_{k-1})}{\mathfrak{z}_k} \right).$$

The aggregation kernel acts as a link between the likelihood ratio and the mixing coefficient. The principle behind its selection is that a smaller statistical difference between $\widetilde{\theta}_k$ and $\widehat{\theta}_{k-1}$ should lead to the mixing coefficient close to 1 and thus to the aggregated estimate $\widehat{\theta}_k$ close to $\widetilde{\theta}_k$, whereas a larger difference should provide the mixing coefficient close to zero and thus keep $\widehat{\theta}_k$ close to $\widehat{\theta}_{k-1}$. Whenever the difference is very large, the mixing coefficient is zero, and the procedure stops prematurely by setting $\widehat{\theta} = \widehat{\theta}_{k-1}$. We call this situation *early stopping*.

To satisfy the stated requirements, the kernel must be supported on the closed interval $[0, 1]$ and monotonously decrease from 1 on the left edge to 0 on the right edge. It is also recommended that the kernels have a plateau of size b starting with zero. Thus, the aggregation kernel assumes the form:

$$K_{\mathrm{ag}}(u) = \begin{cases} 1, & 0 \le u < b \\ 1 - \bar{K}_{\mathrm{ag}}(u), & b \le u \le 1 \end{cases}$$

Examples of $\bar{K}_{\mathrm{ag}}(u)$ include $\frac{u-b}{1-b}$ (triangular kernel), $\left(\frac{u-b}{1-b}\right)^2$ (Epanechnikov kernel) etc.

17.4 Critical Values and Other Parameters

All procedures described above depend on the set of parameters $\mathfrak{z}_1 \ldots \mathfrak{z}_K$ known as critical values. The critical values reflect the problem design (interval length, model, method etc.). They are selected based on the following *propagation condition*:

Condition 2 (Propagation condition) *For any* $\theta^* \in \Theta$

$$\frac{\mathsf{E}_{\theta^*} |L(\widetilde{\theta}_k, \widehat{\theta}_k)|^r}{\mathcal{R}_{r,\theta^*}} \leq \alpha \frac{k}{K} \quad \textit{for} \ \ k = 1, \dots, K, \tag{17.7}$$

where $\widehat{\theta}_k$ is the adaptive estimate obtained on the k-th step and $\mathcal{R}_{r,\theta^*}$ is the risk delivered by the local maximum likelihood estimate in the parametric case (see (17.5)).

This condition means that in the homogeneous case the risk associated with the k-th adaptive estimate must not exceed a certain fraction of the risk in the parametric case.

Critical values constructed this way provide with high probability the prescribed performance of the procedures in the parametric situation (under the null hypothesis). Namely, under the parametric hypothesis on every step k the adaptive estimate $\widehat{\theta}_k$ should be close enough to the oracle estimate $\widetilde{\theta}_k$. However, the propagation condition is not explicit. For the computation of critical values we use the following sequential method based on Monte-Carlo simulations. Denote as $\widehat{\theta}_l(\mathfrak{z}_k)$ for $l \geq k$ the adaptive estimate obtained after the l-th step of the procedure run with he critical values $\mathfrak{z}_1, \dots, \mathfrak{z}_{k-1}$ known and $\mathfrak{z}_{k+1}, \dots, \mathfrak{z}_K$ set to infinity:

$$\widehat{\theta}_l(\mathfrak{z}_k) = \widehat{\theta}_l(\mathfrak{z}_1, \dots, \mathfrak{z}_k, \mathfrak{z}_{k+1} = \infty, \dots, \mathfrak{z}_K = \infty).$$

The first critical value can be selected to satisfy the conditions

$$\frac{\mathsf{E}_{\theta^*} \left| L(\widetilde{\theta}_l, \widehat{\theta}_l(\mathfrak{z}_1)) \right|^r}{\mathcal{R}_{r,\theta^*}} \leq \frac{\alpha}{K}, \quad l = 2, \dots, K.$$

Such a value exists, since for $\mathfrak{z}_1$ taken sufficiently large the weak and adaptive estimates coincide for any l and all Monte-Carlo paths, thus leading to the zero risk. With the first $k - 1$ critical values fixed the procedure is carried out sequentially for the remaining critical values. The k-th critical value is selected using the condition

$$\frac{\mathsf{E}_{\theta^*} \left| L(\widetilde{\theta}_l, \widehat{\theta}_l(\mathfrak{z}_k)) \right|^r}{\mathcal{R}_{r,\theta^*}} \leq k\frac{\alpha}{K}, \quad l = k+1, \dots, K.$$

Obviously, the critical values depend on the specific form of the likelihood function and hence of the Kullback-Leibler distance. Further, the critical values depend on the global parameters α and r.
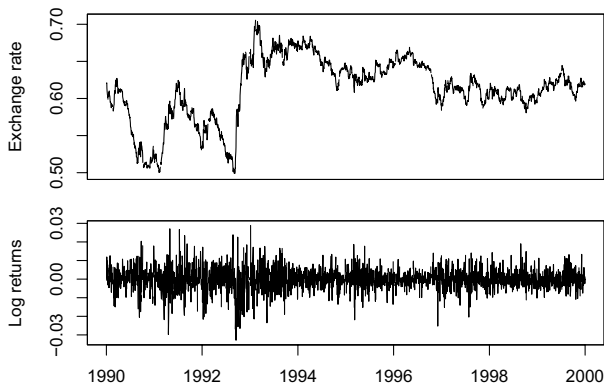
Figure 17.4. Exchange rate of the British pound to the US dollar 19900101-19991231 (above) and corresponding log returns (below).

17.5 Applications

We illustrate the performance of the methods introduced in the section 17.3 by analyzing daily exchange rates of six currencies (GBP, AUD, NZD, JPY, CAD, DKR) to the US dollar available from the site of the US Federal Reserve. We use the data for the period from Januar 1, 1990 till December 31, 1999. Unless indicated otherwise, we use the GBP/USD exchange rate. Observed GBP/USD exchange rates along with the log returns are shown on the Figure 17.4, while Figure 17.5 presents the volatility estimates obtained by three adaptive methods.

A well known feature of financial time series is the uncorrelatedness of the log returns. However, in spite of the uncorrelatedness, the log returns are not independent, as one can see by plotting the autocorrelation of a non-linear transformation. For instance, absolute log returns show significant autocorrelation (Figure 17.6, upper plot). We obtain standardized absolute log returns by dividing the absolute log returns by the volatility estimated using the LCP method. The ACF plot (Figure 17.6, lower plot) shows that nearly all autocorrelation has been removed by standardizing. This result indicates the reasonable quality of volatility estimation.
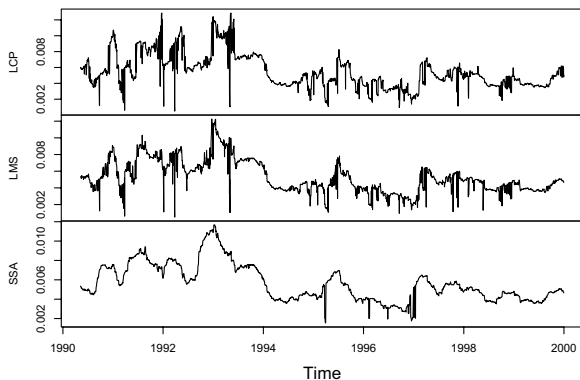
Figure 17.5. Volatility estimates for the British pound obtained by LCP, LMS and SA. **Q** XFGgbpvolaest
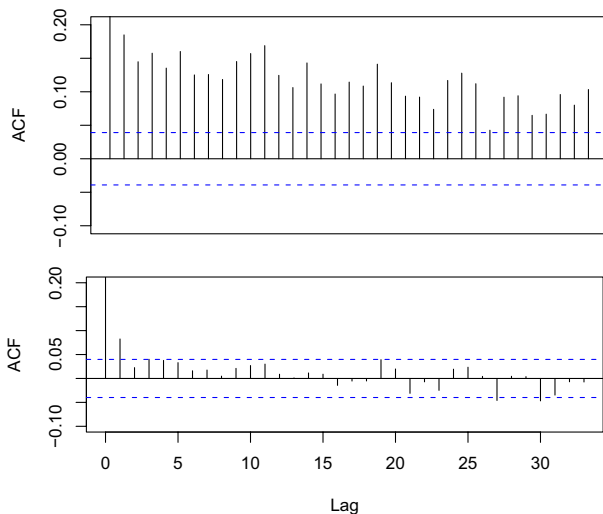
Figure 17.6. Autocorrelation functions of the absolute log returns (above) and of the absolute standardized log returns (below). **Q** XFGacfgbplog

17.5.1 Forecasting Performance for One and Multiple Steps

In order to assess the performance of the adaptive procedures we compare their ability to forecast the conditional variance of the aggregated returns

with that of the GARCH(1,1) model, one of the most popular parameterizations of the volatility process of financial time series. Namely, for a sequence of intervals and forecasting horizons we use the mean square root error (MSqE) criterion

$$\mathrm{MSqE}_{\mathcal{I}} = \sum_{t \in \mathcal{I}} |V_{t,h}^{\heartsuit} - V_{t,h}^{\circ}|^{1/2} \Big/ \sum_{t \in \mathcal{I}} |V_{t,h}^{\spadesuit} - V_{t,h}^{\circ}|^{1/2}, \qquad (17.8)$$

where

$$V_{t,h}^{\circ} = R_{t+1}^2 + \ldots + R_{t+h}^2 \qquad (17.9)$$

is the realized variance of h aggregated returns starting at time t, and $V_{t,h}^{\heartsuit}, V_{t,h}^{\spadesuit}$ denote the conditional variance forecast of the aggregated returns by an adaptive procedure and GARCH(1,1), respectively.

The h-step ahead conditional variance forecast originating at time t is defined as

$$V_{t,h} \stackrel{\mathrm{def}}{=} \mathsf{Var}\left(\sum_{k=1}^h R_{t+k} \,\Big|\, \mathcal{F}_t\right).$$

By definition of the conditional variance

$$\mathsf{Var}\left(\sum_{k=1}^h R_{t+k} \,\Big|\, \mathcal{F}_t\right) = \mathsf{E}\left[\left\{\sum_{k=1}^h R_{t+k} - \mathsf{E}\left(\sum_{k=1}^h R_{t+k} \,\Big|\, \mathcal{F}_t\right)\right\}^2 \,\Big|\, \mathcal{F}_t\right],$$

but since

$$\mathsf{E}\left(R_{t+k} \,\big|\, \mathcal{F}_t\right) = 0 \qquad (17.10)$$

the conditional variance simplifies to

$$\mathsf{Var}\left(\sum_{k=1}^h R_{t+k} \,\Big|\, \mathcal{F}_t\right) = \mathsf{E}\left\{\left(\sum_{k=1}^h R_{t+k}\right)^2 \,\Big|\, \mathcal{F}_t\right\}.$$

As the log returns are conditionally uncorrelated, conditional expectation of the squared sum is equal to the conditional expectation of the sum of squares:

$$\mathsf{Var}\left(\sum_{k=1}^h R_{t+k} \,\Big|\, \mathcal{F}_t\right) = \mathsf{E}\left(\sum_{k=1}^h R_{t+k}^2 \,\Big|\, \mathcal{F}_t\right).$$

Using the linearity of the expectation and equation (17.10), one finally obtains

$$\mathsf{Var}\left(\sum_{k=1}^h R_{t+k} \,\Big|\, \mathcal{F}_t\right) = \sum_{k=1}^h \mathsf{E}\left(R_{t+k}^2 \,\big|\, \mathcal{F}_t\right) = \sum_{k=1}^h \mathsf{Var}\left(R_{t+k} \,\big|\, \mathcal{F}_t\right).$$

By definition of the local constant approach the conditional variance of the log returns is constant for a certain horizon h:

$$\mathsf{Var}\left(R_{t+k}|\,\mathcal{F}_t\right) = \widehat{\sigma}_t^2, \quad k = 1, \ldots, h. \tag{17.11}$$

Therefore the estimated conditional variance of the aggregated returns $R_t + R_{t+1} + \ldots + R_{t+h}$ is simply

$$V_{t,h}^{\heartsuit} = h\widehat{\sigma}_t^2. \tag{17.12}$$

The GARCH(1,1) model describes the volatility dynamics by the relation

$$\sigma_t^2 = \omega + \alpha R_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where the requirement of the stationarity implies the following conditions on the coefficients:

$$\alpha > 0, \quad \beta > 0, \quad \alpha + \beta < 1.$$

The h-step ahead variance forecast of the GARCH(1,1) model is given by:

$$\sigma_{t+h|t}^{2,\spadesuit} \overset{\text{def}}{=} \sum_{k=1}^{h} \mathsf{E}\left(R_{t+h}^2 \,\middle|\, \mathcal{F}_t\right) = \bar{\sigma}^2 + (\alpha + \beta)^h(\sigma_t^2 - \bar{\sigma}^2),$$

where $\bar{\sigma}$ is the unconditional volatility. Thus, the conditional variance forecast of the aggregated returns is

$$V_{t,h}^{\spadesuit} = \sum_{k=1}^{h} \sigma_{t+k|t}^{2,\spadesuit}. \tag{17.13}$$

Substituting the expressions (17.9), (17.12) and (17.13) for $V_{t,h}^{\circ}$, $V_{t,h}^{\heartsuit}$ and $V_{t,h}^{\spadesuit}$ respectively in (17.8), one obtains the performance data shown in the Figure 17.7. The results are presented for various years and forecasting horizons. As seen from the figure, adaptive methods outperform the GARCH(1,1) in many cases.

17.5.2 Value-at-Risk

In the present section we apply the adaptive procedures to the computation of value at risk, an important problem in financial engineering. The value at risk (VaR) is defined as "the maximum loss not exceeded with a given probability defined as the confidence level, over a given period of time". The problem of the VaR estimation can be represented as the problem of quantile estimation
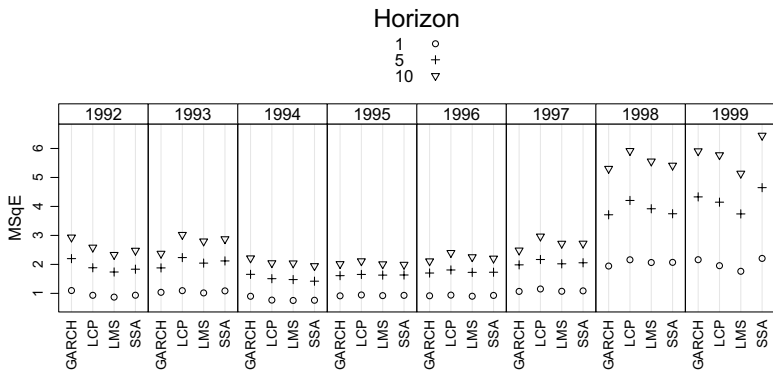
Figure 17.7.   Peformance of adaptive methods and GARCH(1,1) in terms of MSqE. ▣ `XFGadamethperf`

for the distribution of aggregated returns. We consider three distributions of innovations: standard Gaussian distributions, Student's scaled distribution with 5 degrees of freedom and the empirical distribution:

$$R_{t+h} = \widehat{\sigma}_t \xi_{t+h}, \quad \text{with} \quad \xi_{t+h} \sim N(0,1), \text{ or } \sqrt{5/3}\xi_{t+h} \sim t_5, \text{ or } \xi_{t+h} \sim \widehat{F}_t.$$

We aim to describe the quality of VaR computation in terms of the frequency of exceptions, where an "exception" is the event of the predicted value at risk exceeding the aggregated returns. According to the prescribed assessment rule, we examine the particular case of the value at risk predicted at 1% level for 10 steps ahead on 250 observations. Under the assumption that the exceptions follow the binomial distribution, we conduct a test with the null hypothesis about the probability of exception being equal to 0.01, and one-sided alternative hypothesis about the probability of exception exceeding 0.01. A procedure predicting the value at risk belongs in one of the three "zones": "green" zone if the null hypothesis can not be rejected with 95% confidence (corresponding to not more than 5 exceptions on 250 observations, or 2% frequency), "yellow" zone if the null should be rejected with 95% confidence (from 6 to 10 exceptions, or not more than 4% frequency), and "red zone" if the null should be rejected with 99.99% confidence (11 or more exceptions, or more than 4% frequency).

Figure 17.8 shows the percentage of time points at which the loss within a certain horizon overshoots the value at risk predicted with the corresponding confidence level. The results were obtained for three distributions of innovations. One observes that none of the adaptive methods falls in the red zone. Stagewise aggregation always belongs to the green zone. LCP and
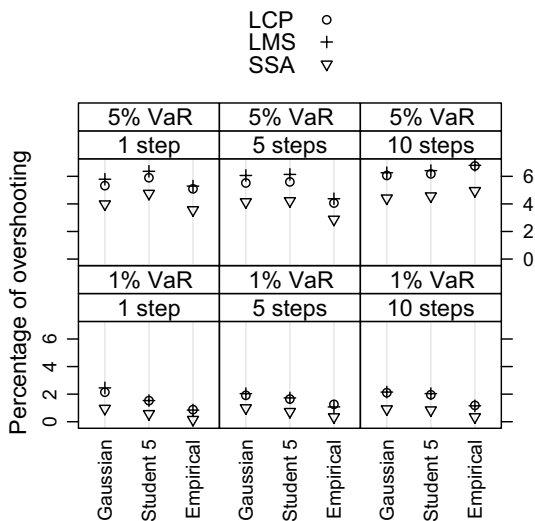
Figure 17.8. Percentage of overshooting the value-at-risk estimated by three methods for various distributions of innovations, number of forecasting steps and value-at-risk levels. Currency: Australian dollar. ⍰ `XFGvarAUD`

LMS combined with the Gaussian innovations sometimes fall into the yellow zone. Use of Student's innovations slightly, and of the empirically distributed innovations considerably improves the performance. Overall performance of the adaptive methods is rather good.

17.5.3 A Multiple Time Series Example

The local parametric approach can be extended to multiple time series. In this case one observes a vector of exchange rate processes $S_t \in \mathbb{R}^d$, $t = 1, 2, \ldots$ and $R_{t,m}$ is the vector of the corresponding log returns:

$$R_{t,m} = \log(S_{t,m}/S_{t-1,m}), \quad m = 1, \ldots, d.$$

The *conditional heteroskedasticity* model reads in this case as

$$R_t = \Sigma_t^{1/2} \varepsilon_t,$$

where ε_t, $t \geq 1$, is a sequence of independent standard Gaussian random innovations and Σ_t is a symmetric $d \times d$ *volatility* matrix, which is to be estimated. As an example, Figure 17.9 shows annualized volatility estimated
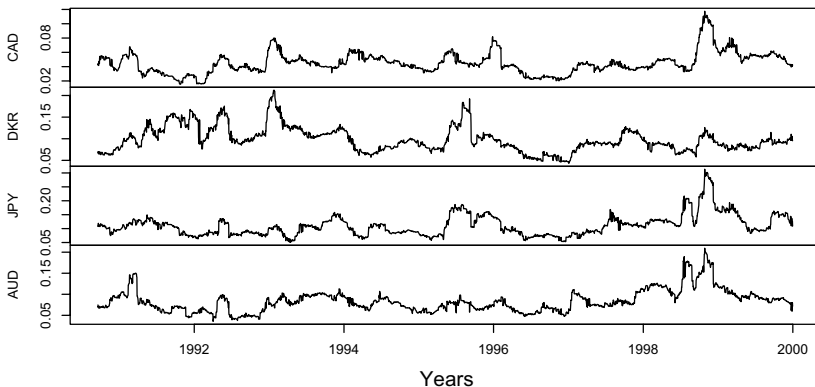
Figure 17.9. Adaptive estimation of the annualized volatility of four exchange rates. **Q** XFGcovmatexch

for exchange rates of several currencies to the US dollar. Annualized volatility is defined as $\sqrt{250\widehat{\Sigma}_{ii}}$, where $\widehat{\Sigma}_{ii}$ represent diagonal elements of the volatility matrix, Similar evolution of the estimates indicates a possible common low-order component.
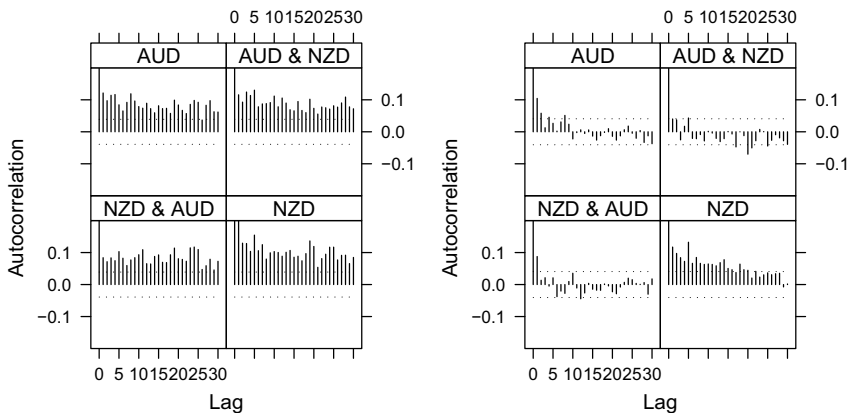
Figure 17.10. ACF for the NZD and AUD time series. Left: absolute log returns, right: absolute standardized log returns. **Q** XFGacfabsaud

As in one-dimensional case, we observe significant correlation and autocor-

relation of the absolute log returns (Figure 17.10, left) as a non-linear transformation of the log returns, indicating lack of independence in spite of the log returns being uncorrelated. We estimate the volatility matrix using the LCP method and obtain the standardized absolute log returns by solving the equation

$$R_t = \widehat{\Sigma}_t^{1/2} \xi_t$$

for ξ_t. The multivariate ACF plot of the standardized absolute log returns is shown in the right part of Figure 17.10. Although some autocorrelation still remains in the NZD series, the remaining three ACF plots show almost no significant correlation.

Bibliography

Belomestny, D. and Spokoiny, V. (2007). Spatial aggregation of local likelihood estimates with applications to classification, *Ann. Statist.* **25**: 2287–2311.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics* **31**(3): 307–327.

Cheng, M.-Y., Fan, J. and Spokoiny, V. (2003). Dynamic nonparametric filtering with application to volatility estimation, *in* M. G. Akritas and D. N. Politis (eds), *Recent Advances and Trends in Nonparametric Statistics*, Elsevier.

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes, *Ann. Statist.* **34**(3): 1075–1114.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching, *J. Econometrics* **105**(1): 131–159.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica* **50**(4): 987–1007.

Fan, J. and Gu, J. (2003). Semiparametric estimation of value at risk, *Econom. J.* **6**(2): 261–290.

Giacomini, E., Härdle, W. and Spokoiny, V. (2008). Inhomogeneous dependency modelling with time varying copulae, *Journal of Business and Economic Statistics.* Forthcoming.

Hillebrand, E. (2005). Neglecting parameter changes in GARCH models, *J. Econometrics* **129**(1-2): 121–138.

Katkovnik, V. and Spokoiny, V. (2008). Spatially adaptive estimation via fitted local likelihood techniques, *ForthcomingIEEE Transactions on Signal Processing.* Forthcoming.

Mercurio, D. and Spokoiny, V. (2004). Statistical inference for time-inhomogeneous volatility models, *Ann. Statist.* **32**(2): 577–602.

Mikosch, T. and Stărică, C. (2004). Changes of structure in financial time series and the GARCH model, *REVSTAT* **2**(1): 41–73.

18 Simulation Based Option Pricing

Denis Belomestny and Grigori N. Milstein

Here we develop an approach for efficient pricing discrete-time American and Bermudan options which employs the fact that such options are equivalent to the European ones with a consumption, combined with analysis of the market model over a small number of steps ahead. This approach allows constructing both upper and lower bounds for the true price by Monte Carlo simulations. An adaptive choice of local lower bounds and use of the kernel interpolation technique enhance efficiency of the whole procedure, which is supported by numerical experiments.

18.1 Introduction

The valuation of high-dimensional American and Bermudan options is one of the most difficult numerical problems in financial engineering. Several approaches have recently been proposed for pricing such options using Monte Carlo simulation technique (see, e.g. Andersen and Broadie (2004), Bally, Pagès, and Printems (2005), Belomestny and Milstein (2004), Boyle, Broadie, and Glasserman (1997), Broadie and Glasserman (1997), Clément, Lamberton and Protter (2002), Glasserman (2004), Haugh and Kogan (2004), Jamshidian (2003), Kolodko and J. Schoenmakers (2004), Longstaff and Schwartz (2001), Rogers (2001) and references therein). In some papers, procedures are proposed that are able to produce upper and lower bounds for the true price and hence allow for evaluating the accuracy of price estimates.

In Belomestny and Milstein (2004) we develop the approach for pricing American options both for discrete-time and continuous-time models. The approach is based on the fact that any American option is equivalent to the European one with a consumption process involved. This approach allows us, in principle, to construct iteratively a sequence v^1, V^1, v^2, V^2, v^3, ..., where v^1, v^2, v^3, ... is an increasing (at any point) sequence of lower bounds and V^1, V^2, ..., is a decreasing sequence of upper bounds. Unfortunately, the complexity of the procedure increases dramatically with any new iteration step. Even V^2 is too expensive for the real construction.

Let us consider a discrete-time financial model and let

$$(B_{t_i}, X_{t_i}) = (B_{t_i}, X_{t_i}^1, ..., X_{t_i}^d), \ i = 0, 1, ..., L,$$

be the vector of prices at time t_i, where B_{t_i} is the price of a scalar riskless asset (we assume that B_{t_i} is deterministic and $B_{t_0} = 1$) and $X_{t_i} = (X_{t_i}^1, ..., X_{t_i}^d)^\top$ is the price vector process of risky assets (along with index t_i we shall use below the index i and instead of (t_i, X_{t_i}) we will write (t_i, X_i)). Let $f_i(x)$ be the profit made by exercising an American option at time t_i if $X_{t_i} = X_i = x$.

Here we propose to use an increasing sequence of lower bounds for constructing an upper bound and lower bound for the initial position (t_0, X_0). It is supposed that the above sequence is not too expensive from the computational point of view. This is achieved by using local lower bounds which take into account a small number of exercise dates ahead.

Let $(t_i, X_{i,m})$, $i = 0, 1, ..., L$; $m = 1, ..., M$, be M independent trajectories all starting from the point (t_0, X_0) and let $v^1 \leq v^2 \leq ... \leq v^l$ be a finite sequence of lower bounds which can be calculated at any position (t_i, x). Clearly, these lower bounds are also ordered according to their numerical complexities and a natural number l indicates the maximal such complexity as well as the quality of the lower bound v^l. Any lower bound gives a lower bound for the corresponding continuation value (lower continuation value) and an upper bound for the consumption process (upper consumption process). If the payoff at $(t_i, X_{i,m})$ is less or equal to the lower continuation value, then the position $(t_i, X_{i,m})$ belongs to the continuation region and the consumption at $(t_i, X_{i,m})$ is equal to zero. Otherwise the position $(t_i, X_{i,m})$ can belong either to the exercise region or to the continuation region. In the latter cases we compute the upper consumption at $(t_i, X_{i,m})$ as a difference between the payoff and the lower continuation value.

It is important to emphasize that the lower bounds are applied adaptively. It means that if, for instance, using the lower bound v^1 (which is the cheapest one among $v^1, v^2, ..., v^l$) at the position $(t_i, X_{i,m})$, we have found that this position belongs to the continuation region (i.e., the corresponding upper consumption process is equal to zero), we do not calculate any further bounds. Similarly, if the upper consumption process is positive but comparatively small, we can stop applying further bounds at $(t_i, X_{i,m})$ because a possible error will not be large. Finally, if the upper consumption process is not small enough after applying lower bounds $v^1, ..., v^j$ but changes not significantly after applying v^{j+1}, we can stop applying further bounds as well. The lower bounds are prescribed to every position $(t_i, X_{i,m})$ and are, as a rule, local. Applying them means, in some sense, a local analysis of the considered financial market at any position. Such a local analysis for all

positions $(t_i, X_{i,m})$, $i = 0, 1, ..., L$; $m = 1, ..., M$, yields some global lower bound and upper bound at the original position (t_0, X_0). If we detect that the difference between the global upper and lower bounds is large, we can return to the deeper local analysis. It is clear that, in principle, this analysis can give exhaustive results in a finite number of steps (it suffices to take the following sequence of American options at $(t_i, X_{i,m})$: v^1 is the price of the American option on the time interval $[t_i, t_{i+1}]$, v^2 is the price on $[t_i, t_{i+2}]$ and so on, in a way that v^{L-i} is the price on $[t_i, t_L]$). Thus, we have no problems with convergence of the algorithms based on the approach considered.

In Subsection 18.2 we recall the basic notions related to the pricing of American and Bermudan options and sketch the approach developed in Belomestny and Milstein (2004). The developed method is presented in Subsection 18.3 . Two numerical examples are given in Subsection 18.4 .

18.2 The Consumption Based Processes

To be self-contained, let us briefly recall the approach to pricing American options that has been developed in Belomestny and Milstein (2004).

18.2.1 The Snell Envelope

We assume that the modelling is based on the filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_i)_{0 \leq i \leq L}, Q)$, where the probability measure Q is the risk-neutral pricing measure for the problem under consideration, and X_i is a Markov chain with respect to the filtration $(\mathcal{F}_i)_{0 \leq i \leq L}$.

The discounted process $\widetilde{X}_i \stackrel{\text{def}}{=} X_i/B_i$ is a martingale with respect to the Q and the price of the corresponding discrete American option at (t_i, X_i) is given by

$$u_i(X_i) = \sup_{\tau \in \mathcal{T}_{i,L}} B_i \, \mathsf{E} \left\{ \frac{f_\tau(X_\tau)}{B_\tau} | \mathcal{F}_i \right\}, \tag{18.1}$$

where $\mathcal{T}_{i,L}$ is the set of stopping times τ taking values in $\{i, i+1, ..., L\}$. The value process u_i (Snell envelope) can be determined by the dynamic programming principle:

$$u_N(x) = f_N(x), \tag{18.2}$$
$$u_i(x) = \max \left\{ f_i(x), B_i \, \mathsf{E} \left\{ \frac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right\}, \ i = L - 1, ..., 0.$$

We see that theoretically the problem of evaluating $u_0(x)$, the price of the discrete-time American option, is easily solved using iteration procedure (18.2). However, if X is high dimensional and/or L is large, the above iteration procedure is not practical.

18.2.2 The Continuation Value, the Continuation and Exercise Regions

For the considered American option, let us introduce the continuation value

$$C_i(x) = B_i \, \mathsf{E} \left\{ \frac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\}, \tag{18.3}$$

the continuation region $\mathcal{C}$ and the exercise (stopping) region $\mathcal{E}$:

$$\begin{aligned}
\mathcal{C} &= \{(t_i, x) : f_i(x) < C_i(x)\}, \\
\mathcal{E} &= \{(t_i, x) : f_i(x) \geq C_i(x)\}.
\end{aligned} \tag{18.4}$$

Let $X_j^{i,x}$, $j = i, i+1, ..., L$, be the Markov chain starting at time t_i from the point $x : X_i^{i,x} = x$, and $X_{j,m}^{i,x}$, $m = 1, ..., M$, be independent trajectories of the Markov chain. The Monte Carlo estimator $\widehat{u}_i(x)$ of $u_i(x)$ (in the case when $\mathcal{E}$ is known) has the form

$$\widehat{u}_i(x) = \frac{1}{M} \sum_{m=1}^{M} \frac{B_i}{B_\tau} f(X_{\tau,m}^{i,x}), \tag{18.5}$$

where τ is the first time at which $X_j^{i,x}$ gets into $\mathcal{E}$ (of course, τ in (18.5) depends on i, x, and $m : \tau = \tau_m^{i,x}$). Thus, for estimating $u_i(x)$, it is sufficient to examine sequentially the position $(t_j, X_{j,m}^{i,x})$ for $j = i, i+1, ..., L$, whether it belongs to $\mathcal{E}$ or not. If $(t_j, X_{j,m}^{i,x}) \in \mathcal{E}$, then we stop at the instant $\tau = t_j$ on the trajectory considered. If $(t_j, X_{j,m}^{i,x}) \in \mathcal{C}$, we move one step more along the trajectory.

Let v be any lower bound, i.e. $u_i(x) \geq v_i(x)$, $i = 0, 1, ..., L$. Clearly, $f_i(x)$ is a lower bound. If $v_i^1, ..., v_i^l$ are some lower bounds then the function $v_i(x) = \max_{1 \leq k \leq l} v_i^k(x)$ is also a lower bound. Henceforth we consider lower bounds satisfying the inequality $v_i(x) \geq f_i(x)$. Introduce the set

$$\mathcal{C}_v = \left\{ (t_i, x) : f_i(x) \leq B_i \, \mathsf{E} \left\{ \frac{v_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right\}.$$

Since $\mathcal{C}_v \subset \mathcal{C}$, any lower bound provides us with a sufficient condition for moving along the trajectory: if $(t_j, X_{j,m}^{i,x}) \in \mathcal{C}_v$, we do one step ahead.

18.2.3 Equivalence of American Options to European Ones with Consumption Processes

For $0 \leq i \leq L - 1$ the equation (18.2) can be rewritten in the form

$$u_i(x) = B_i E\left\{\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right\} + \left[f_i(x) - B_i E\left\{\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right\}\right]^+.$$
(18.6)

Introduce the functions

$$\gamma_i(x) = \left[f_i(x) - B_i \, \mathsf{E}\left\{\frac{u_{i+1}(X_{i+1})}{B_{i+1}}|X_i = x\right\}\right]^+, \ i = L - 1, ..., 0.$$
(18.7)

Due to (18.6), we have

$$u_{L-1}(X_{L-1}) = B_{L-1} \, \mathsf{E}\left\{\frac{f_L(X_L)}{B_L}|\mathcal{F}_{L-1}\right\} + \gamma_{L-1}(X_{L-1}),$$

$$u_{L-2}(X_{L-2}) = B_{L-2} \, \mathsf{E}\left\{\frac{u_{L-1}(X_{L-1})}{B_{L-1}}|\mathcal{F}_{L-2}\right\} + \gamma_{L-2}(X_{L-2})$$

$$= B_{L-2} \, \mathsf{E}\left\{\frac{f_L(X_L)}{B_L}|\mathcal{F}_{L-2}\right\} + B_{L-2} \, \mathsf{E}\left\{\frac{\gamma_{L-1}(X_{L-1})}{B_{L-1}}|\mathcal{F}_{L-2}\right\} + \gamma_{L-2}(X_{L-2}).$$

Analogously, one gets

$$u_i(X_i) = B_i \, \mathsf{E}\left\{\frac{f_L(X_L)}{B_L}|\mathcal{F}_i\right\} + B_i \sum_{k=1}^{L-(i+1)} \mathsf{E}\left\{\frac{\gamma_{L-k}(X_{L-k})}{B_{L-k}}|\mathcal{F}_i\right\}$$
(18.8)

$$+\gamma_i(X_i), \ i = 0, ..., L - 1.$$

Putting $X_0 = x$ and recalling that $B_0 = 1$, we obtain

$$u_0(x) = \mathsf{E}\left\{\frac{f_L(X_L)}{B_L}\right\} + \gamma_0(x) + \sum_{i=1}^{L-1} \mathsf{E}\left\{\frac{\gamma_i(X_i)}{B_i}\right\}.$$
(18.9)

Formula (18.9) gives us the price of the European option with the payoff function $f_i(x)$ in the case when the underlying price process is equipped with the consumption γ_i defined in (18.7).

18.2.4 Upper and Lower Bounds Using Consumption Processes

The results about the equivalence of the discrete-time American option to the European one with the consumption process cannot be used directly because

$u_i(x)$ and consequently $\gamma_i(x)$ are unknown. We take the advantage of this connection in the following way (see Belomestny and Milstein (2004)).

Let $v_i(x)$ be a lower bound on the true option price $u_i(x)$. Introduce the function (upper consumption process)

$$\gamma_{i,v}(x) = \left[f_i(x) - B_i \, \mathsf{E} \left\{ \frac{v_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right]^+ , \; i = 0, ..., L-1. \quad (18.10)$$

Clearly,

$$\gamma_{i,v}(x) \geq \gamma_i(x).$$

Hence the price $V_i(x)$ of the European option with payoff function $f_i(x)$ and upper consumption process $\gamma_{i,v}(x)$ is an upper bound: $V_i(x) \geq u_i(x)$.

Conversely, if $V_i(x)$ is an upper bound on the true option price $u_i(x)$ and

$$\gamma_{i,V}(x) = \left[f_i(x) - B_i \, \mathsf{E} \left\{ \frac{V_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right]^+ , \; i = 0, ..., L-1, \quad (18.11)$$

then the price $v_i(x)$ of the European option with lower consumption process $\gamma_{i,V}(x)$ is a lower bound.

Thus, starting from a lower bound $v_i^1(x)$, one can construct the sequence of lower bounds $v_i^1(x) \leq v_i^2(x) \leq v_i^3(x) \leq ... \leq u_i(x)$, and the sequence of upper bounds $V_i^1(x) \geq V_i^2(x) \geq ... \geq u_i(x)$. All these bounds can be, in principle, evaluated by the Monte Carlo simulations. However, each further step of the procedure requires labor-consuming calculations and in practice it is possible to realize only a few steps of this procedure. In this connection, much attention in Belomestny and Milstein (2004) is given to variance reduction technique and some constructive methods for reducing statistical errors are proposed there.

18.2.5 Bermudan Options

As before, let us consider the discrete-time model

$$(B_i, X_i) = (B_i, X_i^1, ..., X_i^d), \quad i = 0, 1, ..., L.$$

Suppose that an investor can exercise only at an instant from the set of stopping times $S = \{s_1, ..., s_l\}$ within $\{0, 1, ..., L\}$, where $s_l = L$. The price $u_i(X_i)$ of the so called Bermudan option is given by

$$u_i(X_i) = \sup_{\tau \in \mathcal{T}_{S \cap [i,L]}} B_i \, \mathsf{E} \left\{ \frac{f_\tau(X_\tau)}{B_\tau} | \mathcal{F}_i \right\},$$

where $\mathcal{T}_{S \cap [i,L]}$ is the set of stopping times τ taking values in $\{s_1, ..., s_l\} \cap \{i, i+1, ..., L\}$ with $s_l = L$.

The value process u_i is determined as follows:

$$u_L(x) = f_L(x),$$

$$u_i(x) = \begin{cases} \max\left\{ f_i(x), \ B_i \, \mathsf{E}\left\{ \dfrac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right\}, \ i \in S, \\ B_i \left\{ \dfrac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\}, \ i \notin S. \end{cases}$$

Similarly to American options, any Bermudan option is equivalent to the European one with the payoff function $f_i(x)$ and the consumption process γ_i defined as

$$\gamma_i(x) = \begin{cases} \left[f_i(x) - B_i \, \mathsf{E}\left\{ \dfrac{u_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right]^+, \ i \in S, \\ 0, \ i \notin S. \end{cases}$$

Thus, all the results obtained in this section for discrete-time American options can be carried over to Bermudan options. For example, if $v_i(x)$ is a lower bound on the true option price $u_i(x)$, the price $V_i(x)$ of the European option with the payoff function $f_i(x)$ and with the consumption process

$$\gamma_{i,v}(x) = \begin{cases} \left[f_i(x) - B_i \, \mathsf{E}\left\{ \dfrac{v_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right]^+, \ i \in S, \\ 0, \ i \notin S. \end{cases}$$

is an upper bound: $V_i(x) \geq u_i(x)$.

18.3 The Main Procedure

The difficulties mentioned in Subsection 2.4 can be avoided by using an increasing sequence of simple lower bounds.

18.3.1 Local Lower Bounds

The trivial lower bound is $f_i(x)$ and the simplest nontrivial one is given by

$$v_i^{i+1}(x) = \max\left\{ f_i(x), B_i \, \mathsf{E}\left\{ \dfrac{f_{i+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right\}.$$

The function $v_i^{i+1}(x)$ is the price of the American option at the position (t_i, x) on the time interval $[t_i, t_{i+1}]$. It takes into account the behavior of assets at one step ahead. Let $v_i^{i+k}(x)$ be the price of the American option at the position (t_i, x) on the time interval $[t_i, t_{i+k}]$. The function $v_i^{i+k}(x)$ corresponds to an analysis of the market over k steps ahead. The calculation of $v_i^{i+k}(x)$ can be done iteratively. Indeed, the price of the American option on the interval $[t_i, t_{i+k+1}]$ with $k+1$ exercise periods can be calculated using the American options on the interval $[t_{i+1}, t_{i+k+1}]$ with k exercise periods

$$v_i^{i+k+1}(x) = \max \left\{ f_i(x), B_i \, \mathsf{E} \left\{ \frac{v_{i+1}^{i+k+1}(X_{i+1})}{B_{i+1}} | X_i = x \right\} \right\}. \qquad (18.12)$$

We see that $v_i^{i+k+1}(x)$ is, as a rule, much more expensive than $v_i^{i+k}(x)$. The direct formula (18.12) can be too laborious even for $k \geq 3$. As an example of a simpler lower bound, let us consider the maximum of the American option on the interval $[t_i, t_{i+k}]$ and the European option on the interval $[t_i, t_{i+k+1}]$:

$$\bar{v}_i^{i+k}(x) = \max \left\{ v_i^{i+k}(x), B_i \, \mathsf{E} \left\{ \frac{f_{i+k+1}(X_{i+k+1})}{B_{i+k+1}} | X_i = x \right\} \right\}.$$

This lower bound is not so expensive as $v_i^{i+k+1}(x)$. Clearly

$$v_i^{i+k}(x) \leq \bar{v}_i^{i+k}(x) \leq v_i^{i+k+1}(x).$$

Different combinations consisting of European, American, and Bermudan options can give other simple lower bounds.

The success of the main procedures (see below) exceedingly depends on a choice of lower bounds. Therefore their efficient construction is of great importance. To this aim one can use the known methods and among them the method from Belomestny and Milstein (2004).

We emphasize again (see Introduction) that if after using some lower bound it is established that the position belongs to $\mathcal{C}$, then this position does not need any further analysis. Therefore, at the beginning the simplest nontrivial lower bound $v_i^{i+1}(x)$ should be applied and then other lower bounds should be used adaptively in the order of increasing complexity.

18.3.2 The Main Procedure for Constructing Upper Bounds for the Initial Position (Global Upper Bounds)

Aiming to estimate the price of the American option at a fixed position (t_0, x_0), we simulate the independent trajectories $X_{i,m}$, $i = 1, ..., L$, $m =$

$1, ..., M$, of the process X_i, starting at the instant $t = t_0$ from x_0 : $X_0 = x_0$. Let $v_i(x)$ be a lower bound and $(t_i, X_{i,m})$ be the position on the m-th trajectory at the time instant t_i. We calculate the lower continuation value

$$c_{i,v}(X_{i,m}) = B_i \, \mathsf{E} \left\{ \frac{v_{i+1}(X_{i+1,m})}{B_{i+1}} | \mathcal{F}_i \right\} \qquad (18.13)$$

at the position $(t_i, X_{i,m})$. If

$$f_i(X_{i,m}) < c_{i,v}(X_{i,m}), \qquad (18.14)$$

then $(t_i, X_{i,m}) \in \mathcal{C}$ (see (18.4)) and we move one step ahead along the trajectory to the next position $(t_{i+1}, {}_m X_{i+1})$. Otherwise if

$$f_i(X_{i,m}) \geq c_{i,v}(X_{i,m}), \qquad (18.15)$$

then we cannot say definitely whether the position $(t_i, X_{i,m})$ belongs to $\mathcal{C}$ or to E. In spite of this fact we do one step ahead in this case as well. Let us recall that the true consumption at (t_i, x) is equal to

$$\gamma_i(x) = [f_i(x) - C_i(x)]^+ \qquad (18.16)$$

(see (18.7) and (18.3)). Thus, it is natural to define the upper consumption $\gamma_{i,v}$ at any position $(t_i, X_{i,m})$ by the formula

$$\gamma_{i,v}(X_{i,m}) = [f_i(X_{i,m}) - c_{i,v}(X_{i,m})]^+. \qquad (18.17)$$

Obviously, $c_{i,v} \leq C_i$ and hence $\gamma_{i,v} \geq \gamma_i$. Therefore, the price $V_i(x)$ of the European option with payoff function $f_i(x)$ and upper consumption process $\gamma_{i,v}$ is an upper bound on the price $u_i(x)$ of the original American option. In the case (18.14) $\gamma_{i,v}(X_{i,m}) = \gamma_i(X_{i,m}) = 0$ and we do not get any error. If (18.15) holds and besides $c_{i,v}(X_{i,m}) < C_i(X_{i,m})$, we get an error. If $\gamma_{i,v}(X_{i,m})$ is large, then it is in general impossible to estimate this error, but if $\gamma_{i,v}(X_{i,m})$ is small, the error is small as well.

Having found $\gamma_{i,v}$, we can construct an estimate $\widehat{V}_0(x_0)$ of the upper bound $V_0(x_0)$ for $u_0(x_0)$ by the formula

$$\widehat{V}_0(x_0) = \frac{1}{M} \sum_{m=1}^{M} \frac{f_L(X_{L,m})}{B_L} + \frac{1}{M} \sum_{i=0}^{L-1} \sum_{m=1}^{M} \frac{\gamma_{i,v}(X_{i,m})}{B_i}. \qquad (18.18)$$

Note that for the construction of an upper bound V_0 one can use different local lower bounds depending on a position. This opens various opportunities for adaptive procedures. For instance, if $\gamma_{i,v}(X_{i,m})$ is large, then it is reasonable to use a more powerful local instrument at the position $(t_i, X_{i,m})$.

18.3.3 The Main Procedure for Constructing Lower Bounds for the Initial Position (Global Lower Bounds)

Let us proceed to the estimation of a lower bound $v_0(x_0)$. We stress that both $V_0(x_0)$ and $v_0(x_0)$ are estimated for the initial position $\{t_0, x_0\}$ only. Since we are interested in obtaining as large as possible lower bound, it is reasonable to calculate different not too expensive lower bounds at the position $\{t_0, x_0\}$ and to take the largest one. Let us fix a local lower bound v. We denote by $t_0 \leq \tau_1^{(m)} \leq L$ the first time when either (18.15) is fulfilled or $\tau_1^{(m)} = L$. The second time $\tau_2^{(m)}$ is defined in the following way. If $\tau_1^{(m)} < L$, then $\tau_2^{(m)}$ is either the first time after $\tau_1^{(m)}$ for which (18.15) is fulfilled or $\tau_2^{(m)} = L$. So, $t_0 \leq \tau_1^{(m)} < \tau_2^{(m)} \leq L$. In the same way we can define θ times

$$0 \leq \tau_1^{(m)} < \tau_2^{(m)} < ... < \tau_\theta^{(m)} = L. \tag{18.19}$$

The number θ depends on the m-th trajectory: $\theta = \theta^{(m)}$ and can vary between 1 and $L+1 : 1 \leq \theta \leq L+1$. We put by definition $\tau_{\theta+1}^{(m)} = \tau_\theta^{(m)} = L$, $\tau_{\theta+2}^{(m)} = ... = \tau_{L+1}^{(m)} = L$. Thus, we get times $\tau_1, ..., \tau_{L+1}$ which are connected with the considered process X_i. For any $1 \leq k \leq L+1$ the time τ_k does not anticipate the future because at each point X_i at time t_i the knowledge of X_j, $j = 0, 1, ..., i$, is sufficient to define it uniquely. So, the times $\tau_1, ..., \tau_{L+1}$ are stopping rules and the following lower bound can be proposed

$$v_0(x_0) = \max_{1 \leq k \leq L+1} \mathsf{E} \frac{f_{\tau_k}(X_{\tau_k})}{B_{\tau_k}}$$

which can be in turn estimated as

$$\widehat{v}_0(x_0) = \max_{1 \leq k \leq L+1} \frac{1}{M} \sum_{m=1}^{M} \frac{f_{\tau_k^{(m)}}(X_{\tau_k^{(m)}, m})}{B_{\tau_k^{(m)}}}.$$

Of course, $v_0(x_0)$ depends on the choice of the local lower bound v. Clearly, increasing the local lower bound implies increasing the global lower bound $v_0(x_0)$.

REMARK 18.1 It is reasonable instead of the stopping criterion (18.15) to use the following criterion

$$\gamma_{i,v}(X_{i,m}) \geq \varepsilon \tag{18.20}$$

for some $\varepsilon > 0$. On the one hand, $\gamma_{i,v} \geq \gamma_i$ and hence the stopping criterion with $\varepsilon = 0$ can lead to earlier stopping and possibly to a large error when

$\gamma_{i,v} > 0$ but $\gamma_i = 0$. On the other hand, if $0 < \gamma_{i,v}(X_{i,m}) < \varepsilon$ we can make an error using criterion (18.20). Indeed, in this case we continue and if $\gamma_i > 0$ then $(t_i, X_{i,m}) \in \mathcal{E}$ and the true decision is to stop. Since the price of the option at $(t_i, X_{i,m})$ upon the continuation is $C_i(X_{i,m})$ and

$$f_i(X_{i,m}) - C_i(X_{i,m}) = \gamma_i \leq \gamma_{i,v} < \varepsilon,$$

the error due to the wrong decision at $(t_i, X_{i,m})$ is small as long as ε is small. It is generally difficult to estimate the influence of many such wrong decisions on the global lower bound. Fortunately, any $\varepsilon > 0$ leads to a sequence of stopping times (18.19) and, consequently, to a global lower bound $v_0(x_0)$. What the global upper bound is concerned, we have $0 \leq \gamma_{i,v} - \gamma_i < \varepsilon$ when $\gamma_{i,v} < \varepsilon$ and hence the error in estimating V_0 is small due to (18.18). The choice of ε can be based on some heuristics and the empirical analysis of overall errors in estimating true γ_i's.

18.3.4 Kernel Interpolation

The computational complexity of the whole procedure can be substantially reduced by using methods from the interpolation theory. As discussed in the previous sections, the set of independent paths

$$\mathcal{P}_M \overset{\text{def}}{=} \{X_{i,m}, \ i = 1, ..., L, \ m = 1, ..., M\}$$

and the sequence of local lower bounds $\{v_i^1, ..., v_i^l\}$ deliver the set of the upper consumption values $\{\gamma_{i,v}(_mX_i), \ i = 0, ..., L, \ m = 1, ..., M\}$, where $v_i \overset{\text{def}}{=} \max\{v_i^1, ..., v_i^l\}$. If M is large one may take a subset $\mathcal{P}_{\widetilde{M}}$ of $\mathcal{P}_M$ containing first $\widetilde{M} \ll M$ trajectories

$$\mathcal{P}_{\widetilde{M}} \overset{\text{def}}{=} \{X_{i,m}, \ i = 1, ..., L, \ m = 1, ..., \widetilde{M}\} \tag{18.21}$$

and compute $\{\gamma_{i,v}(X_{i,m}), \ i = 0, ..., L, \ m = 1, ..., \widetilde{M}\}$. The remaining consumption values $\gamma_{i,v}(_nX_i)$ for $n = \widetilde{M} + 1, ..., M$ can be approximated by

$$\widehat{\gamma}_{i,v}(X_{i,n}) \overset{\text{def}}{=} \sum_{\left\{m:X_{i,m}\in\mathcal{B}^k_{\mathcal{P}_{\widetilde{M}}}(_nX_i)\right\}} w_{n,m}\gamma_{i,v}(_mX_i),$$

where $\mathcal{B}^k_{\mathcal{P}_{\widetilde{M}}}(_nX_i)$ is the set of k nearest neighbors of $_nX_i$ lying in the $\mathcal{P}_{\widetilde{M}}$ for fixed exercise date t_i and

$$w_{n,m} \overset{\text{def}}{=} \frac{K(\|_nX_i - X_{i,m}\|/h)}{\sum_{\left\{m:X_{i,m}\in\mathcal{B}^k_{\mathcal{P}_{\widetilde{M}}}(_nX_i)\right\}} K(\|_nX_i - X_{i,m}\|/h)}$$

with $K(\cdot)$ being a positive kernel. A bandwidth h and the number of nearest neighbors k are chosen experimentally. Having found $\widehat{\gamma}_{i,v}(_nX_i)$, we get the global upper bound at (t_0, x_0) according to (18.18) by plugging estimated values $\widehat{\gamma}_{i,v}(_mX_i)$ with $m = \widetilde{M}+1, ..., M$ in place of the corresponding $\gamma_{i,v}(_mX_i)$
.

The simulations show that an essential reduction of computational time can be sometimes achieved at small loss of precision. The reason for the success of kernel methods is that the closeness of the points in the state space implies the closeness of the corresponding consumption values.

18.4 Simulations

18.4.1 Bermudan Max Calls on d Assets

This is a benchmark example studied in Broadie and Glasserman (1997), Haugh and Kogan (2004) and Rogers (2001) among others. Specifically, the model with d identical assets is considered where each underlying has dividend yield δ. The risk-neutral dynamic of assets is given by

$$\frac{dX_t^k}{X_t^k} = (r - \delta)dt + \sigma dW_t^k, \quad k = 1, ..., d, \tag{18.22}$$

where W_t^k, $k = 1, ..., d$, are independent one dimensional Brownian motions and r, δ, σ are constants. At any time $t \in \{t_0, ..., t_L\}$ the holder of the option may exercise it and receive the payoff

$$f(X_t) = (\max(X_t^1, ..., X_t^d) - K)^+.$$

In applying the method developed in this paper we take $t_i = iT/L$, $i = 0, ..., L$, with $T = 3$, $L = 9$ and simulate $M = 50000$ trajectories

$$\mathcal{P}_M = \{X_{i,m}, i = 0, ..., L\}_{m=1}^M$$

using Euler scheme with a time step $h = 0.1$. Setting $\widetilde{M} = 500$, we define the set $\mathcal{P}_{\widetilde{M}}$ as in (18.21) and compute adaptively the lower continuation values for every point in $\mathcal{P}_{\widetilde{M}}$. To this end we simulate $N = 100$ points

$$_nX_{i+1}^{(t_i, X_{i,m})}, \quad 1 \le n \le N,$$

from each point $(t_i, X_{i,m})$ with $i < L$ and $m \le \widetilde{M}$. For any natural l such that $0 \le l \le L - i - 1$, values

$$v_{i+1}^{(j)}(_nX_{i+1}^{(t_i, X_{i,m})}), \quad 0 \le j \le l,$$

based on local lower bounds of increasing complexity, can be constructed as follows. First, $v_{i+1}^{(0)}\big(_nX_{i+1}^{(t_i, X_{i,m})}\big) = f\big(_nX_{i+1}^{(t_i, X_{i,m})}\big)$ and $v_{i+1}^{(j)}$ for $j = 1, 2$ are values of the American option on the intervals $[t_{i+1}, t_{i+1+j}]$. If $j > 2$ then $v_{i+1}^{(j)}$ is defined as value of the Bermudan option with three exercise instances at time points $\{t_{i+1}, t_{i+j}, t_{i+j+1}\}$. Now, we estimate the corresponding lower continuation value by

$$\widehat{c}_{i,l}(X_{i,m}) = \frac{e^{-r(t_{i+1}-t_i)}}{N} \sum_{n=1}^{N} \max_{0 \le j \le l} \left\{ v_{i+1}^{(j)}\big(_nX_{i+1}^{(t_i, X_{i,m})}\big) \right\}.$$

Clearly, $\widehat{c}_{i,l}$ is the Monte-Carlo estimate of $c_{i,v}$, where $v = \max_{0 \le j \le l} v_{i+1}^{(j)}$. Let us fix a maximal complexity l^*. Sequentially increasing l from 0 to $l_i^* = \min\{l^*, L - i - 1\}$, we compute $\widehat{c}_{i,l}$ until $l \le l_*$, where

$$l_* \overset{\text{def}}{=} \min\{l : f_i(X_{i,m}) < \widehat{c}_{i,l}(X_{i,m})\}$$

or $l_* \overset{\text{def}}{=} l_i^*$ if

$$f_i(X_{i,m}) \ge \widehat{c}_{i,l}(X_{i,m}), \quad l = 1, \ldots, l_i^*.$$

Note, that in the case $l_* < l_i^*$ the numerical costs are reduced as compared to the non-adaptive procedure while the quality of the estimate $\widehat{c}_{i,v_*}$, where $v_* = \max_{0 \le j \le l_*} v_{i+1}^{(j)}$ is preserved. The estimated values $\widehat{c}_{i,v_*}(X_{i,m})$ allow us, in turn, to compute the estimates for the corresponding upper consumptions $\gamma_{i,v_*}(X_{i,m})$ with $m = 1, \ldots, \widetilde{M}$. The upper consumptions values for $m = \widetilde{M} + 1, \ldots M$ are estimated using kernel interpolation with an exponential kernel (see Subsection 3.4). In Table 18.1 the corresponding results are presented in dependence on l^* and x_0 with $X_0 = (X_0^1, \ldots, X_0^d)^T$, $X_0^1 = \ldots = X_0^d = x_0$. The true values are quoted from Glasserman (2004). We see that while the quality of bounds increases significantly from $l^* = 1$ to $l^* = 3$, the crossover to $l^* = 6$ has a little impact on it. It means that either the true value is achieved (as for $x_0 = 90$) or deeper analysis is needed (as for $x_0 = 100$).

18.4.2 Bermudan Basket-Put

In this example we consider again the model with d identical assets driven by independent identical geometrical Brownian motions (see (18.22)) with $\delta = 0$. Defining the basket at any time t as $\bar{X}_t = (X_t^1 + \ldots + X_t^d)/d$, let us consider the Bermudan basket put option granting the holder the right to sell this basket for a fixed price K at time $t \in \{t_0, \ldots, t_L\}$ getting the profit given by $f(\bar{X}_t) = (K - \bar{X}_t)^+$. We apply our method for constructing lower and upper bounds on the true value of this option at the initial point (t_0, X_0). In

l^*	x_0	Lower Bound $v_0(X_0)$	Upper Bound $V_0(X_0)$	True Value
1	90	7.892±0.1082	8.694±0.0023	8.08
	100	12.872±0.1459	15.2568±0.0042	13.90
	110	19.275±0.1703	23.8148±0.0062	21.34
3	90	8.070±0.1034	7.900±0.0018	8.08
	100	13.281±0.1434	14.241±0.0038	13.90
	110	19.526±0.1852	21.807±0.0058	21.34
6	90	8.099±0.1057	7.914±0.0018	8.08
	100	13.196±0.1498	13.844±0.0038	13.90
	110	19.639±0.1729	21.411±0.0056	21.34

Table 18.1. Bounds (with 95% confidence intervals) for the 2-dimensional Bermudan max call with parameters $K = 100$, $r = 0.05$, $\sigma = 0.2$, $L = 9$ and l^* varying as shown in the table.

order to construct local lower bounds we need to compute the prices of the corresponding European style options $v_t^{t+\theta}(x) = e^{-r\theta} \mathsf{E}(f(\bar{X}_{t+\theta})|X_t = x)$ for different θ and t. It can be done in principle by Monte-Carlo method since the closed form expression for $v_t^{t+\theta}(x)$ is not known. However, in this case it is more rational to use the so-called moment-matching procedure from Brigo, Mercurio, Rapisarda and Scotti (2002) and to approximate the distribution of the basket $\bar{X}_{t+\theta}$ by a log-normal one with parameters $\tilde{r} - \tilde{\sigma}^2/2$ and $\tilde{\sigma}\theta^{1/2}$, where $\tilde{r}$ and $\tilde{\sigma}$ are chosen in a such way that the first two moments of the above log-normal distribution coincide with the true ones. In our particular example $\tilde{r} = r$ and

$$\tilde{\sigma}^2 = \frac{1}{\theta} \log \left\{ \frac{\sum_{i,j=1}^{d} X_t^i X_t^j \exp(\mathbf{1}_{\{i=j\}}\sigma^2\theta)}{\left[\sum_{i=1}^{d} X_t^i\right]^2} \right\}. \tag{18.23}$$

In Table 18.2 the results of simulations for different maximal complexity l^* and initial values $x_0 = X_0^1 = ... = X_0^d$ are presented. Here, overall $M = 50000$ paths are simulated and on the subset of $\widetilde{M} = 500$ trajectories the local analysis is conducted. Other trajectories are handled with the kernel interpolation method as described in Subsection 3.4. Similar to the previous example, significant improvements are observed for $l^* = 2$ and $l^* = 3$. The difference between the upper bound and lower bound for $l^* > 3$ is less than 5%.

l	x_0	Lower Bound $v_0(X_0)$	Upper Bound $V_0(X_0)$	True Value
	100	2.391±0.0268	2.985±0.0255	2.480
1	105	1.196±0.0210	1.470±0.0169	1.250
	110	0.594±0.0155	0.700±0.0105	0.595
	100	2.455±0.0286	2.767±0.0238	2.480
2	105	1.210±0.0220	1.337±0.0149	1.250
	110	0.608±0.0163	0.653±0.0094	0.595
	100	2.462±0.0293	2.665±0.0228	2.480
3	105	1.208±0.0224	1.295±0.0144	1.250
	110	0.604±0.0166	0.635±0.0090	0.595
	100	2.473±0.0200	2.639±0.0228	2.480
6	105	1.237±0.0231	1.288±0.0142	1.250
	110	0.611±0.0169	0.632±0.0089	0.595
	100	2.479±0.0300	2.627±0.0226	2.480
9	105	1.236±0.0232	1.293±0.0144	1.250
	110	0.598±0.0167	0.627±0.0087	0.595

Table 18.2. Bounds (with 95% confidence intervals) for the 5-dimensional Bermudan basket put with parameters $K = 100$, $r = 0.05$, $\sigma = 0.2$, $L = 9$ and different l^*.

18.5 Conclusions

In this paper a new Monte-Carlo approach towards pricing discrete American and Bermudan options is presented. This approach relies essentially on the representation of an American option as the European one with the consumption process involved. The combination of the above representation with the analysis of the market over a small number of time steps ahead provides us with a lower as well an upper bound on the true price at a given point. Additional ideas concerning adaptive computation of the continuation values and the use of interpolation techniques help reducing the computational complexity of the procedure. In summary, the approach proposed has following features:

⊡ It is Monte-Carlo based and is applicable to the problems of medium dimensionality.

⊡ The propagation of errors is transparent and the quality of final bounds can be easily assessed.

⊡ It is adaptive that is its numerical complexity can be tuned to the accuracy needed.

⊡ Different type of sensitivities can be efficiently calculated by combining the current approach with the method developed in Milstein and Tretyakov (2005).

Bibliography

L. Andersen and M. Broadie (2004). A primal-dual simulation algorithm for pricing multi-dimensional American options. *Management Science* **50**, no. 9: 1222–1234.

V. Bally, G. Pagès and J. Printems (2005). A quantization tree method for pricing and hedging multidimensional American options. *Mathematical Finance* **15**, No. 1, 119-168.

D. Belomestny, and G.N. Milstein, (2004). Monte Carlo evaluation of American options using consumption processes. *WIAS-Preprint* No. 930, Berlin

P. Boyle, M. Broadie and P. Glasserman (1997). Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control* **21**: 1267-1321.

M. Broadie and P. Glasserman (1997). Pricing American-style securities using simulation. *J. of Economic Dynamics and Control* **21**: 1323-1352.

D.Brigo, F. Mercurio, F. Rapisarda and R. Scotti (2002). Approximated moment-matching dynamics for basket-options simulation. Working paper.

E. Clément, D. Lamberton and P. Protter (2002). An analysis of a least squares regression algorithm for American option pricing. *Finance and Stochastics* **6**: 449-471.

P. Glasserman (2004). *Monte Carlo Methods in Financial Engineering* Springer.

M. Haugh and L. Kogan (2004). Pricing American options: a duality approach. *Opeations Research* **52**, No. 2: 258–270.

F. Jamshidian (2003). Minimax optimality of Bermudan and American claims and their Monte Carlo upper bound approximation. Working paper.

A. Kolodko and J. Schoenmakers (2004). Iterative construction of the optimal Bermudan stopping time. *WIAS-Preprint* No. 926, Berlin.

D. Lamberton and B. Lapeyre (1996). *Intoduction to Stochastic Calculus Applied to Finance* Chapman & Hall.

F.A. Longstaff and E.S. Schwartz (2001). Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies* **14**, 113-147.

G.N. Milstein and M.V. Tretyakov (2005). Numerical Analysis of Monte Carlo evaluation of Greeks by finite differences. *J. of Computational Finance* **8**, No. 3.

L.C.G. Rogers (2001). Monte Carlo valuation of American options. *Mathematical Finance* **12**, 271-286.

A.N. Shiryaev (1999). Essentials of Stochastic Finance: Facts, Models, Theory. *World Scientific*

19 High-Frequency Volatility and Liquidity

Nikolaus Hautsch and Vahidin Jeleskovic

19.1 Introduction

Due to the permanently increasing availability of high-frequency financial data, the empirical analysis of trading behavior and the modelling of trading processes has become a major theme in modern financial econometrics. Key variables in empirical studies of high-frequency data are price volatilities, trading volume, trading intensities, bid-ask spreads and market depth as displayed by an open limit order book. A common characteristic of these variables is that they are positive-valued and persistently clustered over time.

To capture the stochastic properties of positive-valued autoregressive processes, *multiplicative error models* (MEMs) have been proposed. The basic idea of modelling a positive-valued process in terms of the product of positive-valued (typically i.i.d.) innovation terms and an observation-driven and/or parameter driven dynamic function is well-known in financial econometrics and originates from the model structure of the *autoregressive conditional heteroscedasticity* (ARCH) model introduced by Engle (1982) or the *stochastic volatility* (SV) model proposed by Taylor (1982). Engle and Russell (1997, 1998) introduced the *autoregressive conditional duration* (ACD) model to model autoregressive (financial) duration processes in terms of a multiplicative error process and a GARCH-type parameterization of the conditional duration mean. The term 'MEM' is ultimately introduced by Engle (2002) who discusses this approach as a general framework to model any kind of positive-valued dynamic process. Manganelli (2005) proposes a multivariate MEM to jointly model high-frequency volatilities, trading volume and trading intensities. Hautsch (2008) generalizes this approach by introducing a common latent dynamic factor serving as a subordinated process driving the individual trading components. The resulting model combines features of a

GARCH type model and an SV type model and is called *stochastic* MEM. Engle and Gallo (2006) apply MEM specifications to jointly model different volatility indicators including absolute returns, daily range, and realized volatility. Recently, Cipollini et al. (2007) extend the MEM by a copula specification in order to capture contemporaneous relationships between the variables.

Given the growing importance of MEMs for the modelling of high-frequency trading processes, liquidity dynamics and volatility processes, this chapter gives an introduction to the topic and an overview of the current literature. Given that the ACD model is the most popular specification of a univariate MEM, we will strongly rely on this string of the literature. Finally, we will present an application of the MEM to jointly model the multivariate dynamics of volatilities, trade sizes, trading intensities, and trading costs based on limit order book data from the Australian Stock Exchange (ASX).

The chapter is organized as follows: Section 19.2 presents the major principles and properties of univariate MEMS. In Section 19.3, we will introduce multivariate specifications of MEMs. Estimation and statistical inference is illustrated in Section 19.4. Finally, Section 19.5 gives an application of the MEM to model high-frequency trading processes using data from the ASX.

19.2 The Univariate MEM

Let $\{Y_t\}$, $t = 1, \ldots, T$, denote a non-negative (scalar) random variable. Then, the univariate MEM for Y_t is given by

$$Y_t = \mu_t \varepsilon_t,$$
$$\varepsilon_t | \mathcal{F}_{t-1} \sim \text{i.i.d. } \mathbb{D}(1, \sigma^2),$$

where $\mathcal{F}_t$ denotes the information set up to t, μ_t is a non-negative conditionally deterministic process given $\mathcal{F}_{t-1}$, and ε_t is a unit mean, i.i.d. variate process defined on non-negative support with variance σ^2. Then, per construction we have

$$\mathsf{E}\left[Y_t | \mathcal{F}_{t-1}\right] \overset{\text{def}}{=} \mu_t, \qquad (19.1)$$
$$\mathsf{Var}\left[Y_t | \mathcal{F}_{t-1}\right] = \sigma^2 \mu_t^2. \qquad (19.2)$$

The major principle of the MEM is to parameterize the conditional mean μ_t in terms of a function of the information set $\mathcal{F}_{t-1}$ and parameters θ. Then,

the basic linear MEM(p,q) specification is given by

$$\mu_t = \omega + \sum_{j=1}^{p} \alpha_j Y_{t-j} + \sum_{j=1}^{q} \beta_j \mu_{t-j}, \qquad (19.3)$$

where $\omega > 0$, $\alpha_j \geq 0$, $\beta_j \geq 0$. This specification corresponds to a generalized ARCH model as proposed by Bollerslev (1986) as long as Y_t is the squared (de-meaned) log return between t and $t-1$ with μ_t corresponding to the conditional variance. Accordingly, the process (19.3) can be estimated by applying GARCH software based on $\sqrt{Y_t}$ (without specifying a conditional mean function). Alternatively, if Y_t corresponds to a (financial) duration, such as, e.g., the time between consecutive trades (so-called trade durations) or the time until a cumulative absolute price change is observed (so-called price durations), the model is referred to an ACD specification as introduced by Engle and Russell (1997, 1998).

The unconditional mean of Y_t is straightforwardly computed as

$$\mathsf{E}[Y_t] = \omega/(1 - \sum_{j=1}^{p} \alpha_j - \sum_{j=1}^{q} \beta_j). \qquad (19.4)$$

The derivation of the unconditional variance is more cumbersome since it requires the computation of $\mathsf{E}[\mu_t^2]$. In the case of an MEM(1,1) process, the unconditional variance is given by (see, e.g., Hautsch (2004))

$$\mathsf{Var}[Y_t] = \mathsf{E}[Y_t]^2 \sigma^2 (1 - \beta^2 - 2\alpha\beta)/(1 - (\alpha + \beta)^2 - \alpha^2 \sigma^2) \qquad (19.5)$$

corresponding to

$$\mathsf{Var}[Y_t] = \mathsf{E}[Y_t]^2 (1 - \beta^2 - 2\alpha\beta)/(1 - \beta^2 - 2\alpha\beta - 2\alpha^2) \qquad (19.6)$$

in case of $\sigma^2 = 1$ which is, e.g., associated with a standard exponential distribution. Correspondingly, the model implied autocorrelation function is given by

$$\rho_1 \overset{\text{def}}{=} \mathsf{Corr}[Y_t, Y_{t-1}] = \alpha(1 - \beta^2 - \alpha\beta)/(1 - \beta^2 - 2\alpha\beta), \qquad (19.7)$$

$$\rho_j \overset{\text{def}}{=} \mathsf{Corr}[Y_t, Y_{t-j}] = (\alpha + \beta)\rho_{j-1} \quad \text{for } j \geq 2. \qquad (19.8)$$

Similarly to the GARCH model, the MEM can be represented in terms of an ARMA model for Y_t. Let $\eta_t \overset{\text{def}}{=} Y_t - \mu_t$ denote a martingale difference, then the MEM(p,q) process can be written as

$$Y_t = \omega + \sum_{j=1}^{max(p,q)} (\alpha_j + \beta_j) Y_{t-j} - \sum_{j=1}^{q} \beta_j \eta_{t-j} + \eta_t. \qquad (19.9)$$

The weak stationarity condition of a MEM(1,1) process is given by $(\alpha+\beta)^2 - \alpha^2\sigma^2 < 1$ ensuring the existence of $\mathsf{Var}[Y_t]$.

Relying on the GARCH literature, the linear MEM specification can be extended in various forms. A popular form is a logarithmic specification of a MEM ensuring the positivity of μ_t without imposing parameter constraints. This is particularly important whenever the model is augmented by explanatory variables or when the model has to accommodate negative (cross-) autocorrelations in a multivariate setting. Two versions of logarithmic MEM's have been introduced by Bauwens and Giot (2000) in the context of ACD models and are given (for simplicity for $p = q = 1$) by

$$\log \mu_t = \omega + \alpha g(\varepsilon_{t-1}) + \beta \log \mu_{t-1}, \tag{19.10}$$

where $g(\cdot)$ is given either by $g(\varepsilon_{t-1}) = \varepsilon_{t-1}$ (type I) or $g(\varepsilon_{t-1}) = \log \varepsilon_{t-1}$ (type II). The process is covariance stationary if $\beta < 1$, $\mathsf{E}[\varepsilon_t \exp\{\alpha g(\varepsilon_t)\}] < \infty$ and $\mathsf{E}[\exp\{2\alpha g(\varepsilon_t)\}] < \infty$. For more details, see Bauwens and Giot (2000). Notice that due to the logarithmic transformation, the news impact function, i.e., the relation between Y_t and ε_{t-1} is not anymore linear but is convex in the type I case and is concave in the type II parameterization. I.e., in the latter case, the sensitivity of Y_t to shocks in ε_{t-1} is higher if ε_{t-1} is small than in the case where it is large.

A more flexible way to capture nonlinear news responses is to allow for a kinked news response function

$$\log \mu_t = \omega + \alpha\{|\varepsilon_{t-1} - b| + c(\varepsilon_{t-1} - b)\}^\delta + \beta \log \mu_{t-1}, \tag{19.11}$$

where b gives the position of the kink while δ determines the shape of the piecewise function around the kink. For $\delta = 1$, the model implies a linear news response function which is kinked at b resembling the EGARCH model proposed by Nelson (1991). For $\delta > 1$, the shape is convex while it is concave for $\delta < 1$. Such a specification allows to flexibly capture asymmetries in responses of Y_t to small or large lagged innovation shocks, such as, e.g., shocks in liquidity demand, liquidity supply or volatility. A similar specification is considered by Cipollini et al. (2007) to capture leverage effects if Y_t corresponds to a volatility variable. For more details on extended MEM specifications in the context of ACD models, see Hautsch (2004) or Bauwens and Hautsch (2008).

The error term distribution of ε_t is chosen as a distribution defined on positive support and standardized by its mean. If Y_t is the squared (de-meaned) log return, then $\sqrt{\varepsilon_t} \sim N(0,1)$ yields the Gaussian GARCH model. If Y_t denotes a liquidity variable (such as trade size, trading intensity, bid-ask spread or market depth), a natural choice is an exponential distribution.

Though the exponential distribution is typically too restrictive to appropriately capture the distributional properties of trading variables, it allows for a quasi maximum likelihood (QML) estimation yielding consistent estimates irrespective of distributional misspecifications. For more details, see Section 19.4. More flexible distributions are, e.g., the Weibull distribution, the (generalized) gamma distribution, the Burr distribution or the generalized F distribution. The latter is proposed in an ACD context by Hautsch (2003) and is given in standardized form (i.e., with unit mean) by the p.d.f.

$$f_\varepsilon(x) = [a\{x/\zeta(a,m,\eta)\}^{am-1}[\eta + \{x/\zeta(a,m,\eta)\}]^{(-\eta-m)}\eta^\eta]/\mathcal{B}(m,\eta), \quad (19.12)$$

where a, m, and η are distribution parameters, $\mathcal{B}(m,\eta) = \Gamma(m)\Gamma(\eta)/\Gamma(m+\eta)$, and

$$\zeta(a,m,\eta) \stackrel{\text{def}}{=} \{\eta^{1/a}\Gamma(m+1/a)\Gamma(\eta-1/a)\}/\{\Gamma(m)\Gamma(\eta)\}. \quad (19.13)$$

The generalized F-distribution nests the generalized gamma distribution for $\eta \to \infty$, the Weibull distribution for $\eta \to \infty$, $m = 1$, the log-logistic distribution for $m = \eta = 1$, and the exponential distribution for $\eta \to \infty$, $m = a = 1$. For more details, see Hautsch (2004).

19.3 The Vector MEM

Consider in the following a k-dimensional positive-valued time series, denoted by $\{\mathbf{Y}_t\}$, $t = 1\ldots,T$, with $\mathbf{Y}_t \stackrel{\text{def}}{=} (Y_t^{(1)},\ldots,Y_t^{(k)})$. Then, the so-called vector MEM (VMEM) for $\mathbf{Y}_t$ is defined by

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu}_t \odot \boldsymbol{\varepsilon}_t \\ &= \text{diag}(\boldsymbol{\mu}_t)\boldsymbol{\varepsilon}_t, \end{aligned}$$

where $\odot$ denotes the Hadamard product (element-wise multiplication) and $\boldsymbol{\varepsilon}_t$ is a k-dimensional vector of mutually and serially i.i.d. innovation processes, where the j-th element is given by

$$\varepsilon_t^{(j)}|\mathcal{F}_{t-1} \sim \text{i.i.d. } \mathbb{D}(1,\sigma_j^2), \quad j = 1,\ldots,k.$$

A straightforward extension of the univariate linear MEM as proposed by Manganelli (2005) is given by

$$\boldsymbol{\mu}_t = \boldsymbol{\omega} + \mathcal{A}_0\mathbf{Y}_t + \sum_{j=1}^p \mathcal{A}_j\mathbf{Y}_{t-j} + \sum_{j=1}^q \mathcal{B}_j\boldsymbol{\mu}_{t-j}, \quad (19.14)$$

where $\boldsymbol{\omega}$ is a $(k\times 1)$ vector, and $\boldsymbol{\mathcal{A}}_0$, $\boldsymbol{\mathcal{A}}_j$, and $\boldsymbol{\mathcal{B}}_j$ are $(k\times k)$ parameter matrices. The matrix $\boldsymbol{\mathcal{A}}_0$ captures contemporaneous relationships between the elements of $\mathbf{Y_t}$ and is specified as a matrix where only the upper triangular elements are non-zero. This triangular structure implies that $Y_t^{(i)}$ is predetermined for all variables $Y_t^{(j)}$ with $j < i$. Consequently, $Y_t^{(i)}$ is conditionally i.i.d. given $\{Y_t^{(j)}, \mathcal{F}_{t-1}\}$ for $j < i$.

The advantage of this specification is that contemporaneous relationships between the variables are taken into account without requiring multivariate distributions for $\boldsymbol{\varepsilon}_t$. This eases the estimation of the model. Furthermore, the theoretical properties of univariate MEMs as discussed in the previous section can be straightforwardly extended to the multivariate case. However, an obvious drawback is the requirement to impose an explicit ordering of the variables in $\mathbf{Y}_t$ which is typically chosen in accordance with a specific research objective or following economic reasoning. An alternative way to capture contemporaneous relationships between the elements of $\mathbf{Y}_t$ is to allow for mutual correlations between the innovation terms $\varepsilon_t^{(j)}$. Then, the innovation term vector follows a density function which is defined over non-negative k-dimensional support $[0, +\infty)^k$ with unit mean $\boldsymbol{\iota}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e.,

$$\boldsymbol{\varepsilon}_t | \mathcal{F}_{t-1} \sim \text{i.i.d. } \boldsymbol{\mathcal{D}}(\boldsymbol{\iota}, \boldsymbol{\Sigma})$$

implying

$$\mathsf{E}\left[\mathbf{Y}_t | \mathcal{F}_{t-1}\right] = \boldsymbol{\mu}_t,$$
$$\mathsf{Var}\left[\mathbf{Y}_t | \mathcal{F}_{t-1}\right] = \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top \odot \boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\mu}_t) \boldsymbol{\Sigma} \text{ diag}(\boldsymbol{\mu}_t).$$

Finding an appropriate multivariate distribution defined on positive support is a difficult task. As discussed by Cipollini et al. (2007), a possible candidate is a multivariate gamma distribution which however imposes severe restrictions on the contemporaneous correlations between the errors $\varepsilon_t^{(i)}$. Alternatively, copula approaches can be used as, e.g., proposed by Heinen and Rengifo (2006) or Cipollini et al. (2007).

In correspondence to the univariate logarithmic MEM, we obtain a logarithmic VMEM specification by

$$\log \boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\mathcal{A}}_0 \log \mathbf{Y}_t + \sum_{j=1}^{p} \boldsymbol{\mathcal{A}}_j g(\boldsymbol{\varepsilon}_{t-j}) + \sum_{j=1}^{q} \boldsymbol{\mathcal{B}}_j \log \boldsymbol{\mu}_{t-j}, \qquad (19.15)$$

where $g(\boldsymbol{\varepsilon}_{t-j}) = \boldsymbol{\varepsilon}_{t-j}$ or $g(\boldsymbol{\varepsilon}_{t-j}) = \log \boldsymbol{\varepsilon}_{t-j}$, respectively. Generalized VMEMs can be specified accordingly to Section 19.2.

A further generalization of VMEM processes has been introduced by Hautsch (2008) and captures mutual (time-varying) dependencies by a subordinated

common (latent) factor jointly driving the individual processes. The so-called *stochastic* MEM can be compactly represented as

$$\mathbf{Y}_t = \boldsymbol{\mu}_t \odot \boldsymbol{\lambda}_t \odot \boldsymbol{\varepsilon}_t, \tag{19.16}$$

where $\boldsymbol{\lambda}_t$ is a $(k \times 1)$ vector with elements $\{\lambda_t^{\delta_i}\}$, $i = 1, \ldots, k$,

$$\log \lambda_t = a \log \lambda_{t-1} + \nu_t, \quad \nu_t \sim \text{i.i.d. } N(0, 1), \tag{19.17}$$

and ν_t is assumed to be independent of $\boldsymbol{\varepsilon}_t$. Hence, λ_t serves as a common dynamic factor with process-specific impacts δ_i. Then, the elements of $\boldsymbol{\mu}_t$ represent 'genuine' (trade-driven) effects given the latent factor. They are assumed to follow (19.15) with $g(\boldsymbol{\varepsilon}_t) = \mathbf{Y}_t \odot \boldsymbol{\mu}_t^{-1}$. The model corresponds to a mixture model and nests important special cases, such as the SV model by Taylor (1982) or the stochastic conditional duration model by Bauwens and Veredas (2004). Applying this approach to jointly model high-frequency volatilities, trade sizes and trading intensities, Hautsch (2008) shows that the latent component is a major driving force of cross-dependencies between the individual processes.

19.4 Statistical Inference

Define $f(Y_t^{(1)}, Y_t^{(2)}, \ldots, Y_t^{(k)} | \mathcal{F}_{t-1})$ as the joint conditional density given $\mathcal{F}_{t-1}$. Without loss of generality the joint density can be decomposed into

$$\begin{aligned}
f(Y_t^{(1)}, Y_t^{(2)}, \ldots, Y_t^{(k)} | \mathcal{F}_{t-1}) = & f(Y_t^{(1)} | Y_t^{(2)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1}) \\
& \times f(Y_t^{(2)} | Y_t^{(3)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1}) \\
& \times f(Y_t^k | \mathcal{F}_{t-1}).
\end{aligned} \tag{19.18}$$

Then, the log likelihood function is defined by

$$\mathcal{L}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \sum_{t=1}^{T} \sum_{j=1}^{k} \log f(Y_t^{(j)} | Y_t^{(j+1)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1}). \tag{19.19}$$

For instance, if $Y_t^{(j)} | Y_t^{(j+1)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1}$ follows a generalized F distribution with parameters $a^{(j)}$, $m^{(j)}$ and $\eta^{(j)}$, the corresponding log likelihood contribution is given by

$$\begin{aligned}
\log f(Y_t^{(j)} | & Y_t^{(j+1)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1}) \\
& = \log[\Gamma(m^{(j)} + \eta^{(j)}) / \{\Gamma(m^{(j)})\Gamma(\eta^{(j)})\}] + \log a^{(j)} \\
& \quad - a^{(j)} m^{(j)} \log \widetilde{\mu}_t^{(j)} + (a^{(j)} m^{(j)} - 1) \log Y_t^{(j)} \\
& \quad - (\eta^{(j)} + m^{(j)}) \log \left(\eta^{(j)} + Y_t^{(j)} / \widetilde{\mu}_t^{(j)}\right) + \eta^{(j)} \log(\eta^{(j)}), \quad (19.20)
\end{aligned}$$

where $\widetilde{\mu}_t^{(j)} = \mu_t^{(j)}/\zeta(a^{(j)}, m^{(j)}, \eta^{(j)})$ and $\zeta(\cdot)$ defined as above.

Constructing the likelihood based on an exponential distribution leads to the quasi likelihood function with components

$$\log f(Y_t^{(j)}|Y_t^{(j+1)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1}) = -\sum_{t=1}^{T} \left(\log \mu_t^{(j)} + Y_t^{(j)}/\mu_t^{(j)} \right),$$

where the score and Hessian are given by

$$\frac{\partial \log f(Y_t^{(j)}|Y_t^{(j+1)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1})}{\partial \theta^{(j)}} = -\sum_{t=1}^{T} \frac{\partial \mu_t^{(j)}}{\partial \theta^{(j)}} \frac{1}{\mu_t^{(j)}} \left(\frac{Y_t^{(j)}}{\mu_t^{(j)}} - 1 \right),$$

$$\frac{\partial^2 \log f(Y_t^{(j)}|Y_t^{(j+1)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1})}{\partial \theta^{(j)} \partial \theta^{(j)\top}} = \sum_{t=1}^{T} \left\{ \frac{\partial}{\partial \theta^{(j)\top}} \left(\frac{1}{\mu_t^{(j)}} \frac{\partial \mu_t^{(j)}}{\partial \theta^{(j)}} \right) \left(\frac{Y_t^{(j)}}{\mu_t^{(j)}} - 1 \right) \right.$$
$$\left. - \frac{1}{\mu_t^{(j)}} \frac{\partial \mu_t^{(j)}}{\partial \theta^{(j)}} \frac{\partial \mu_t^{(j)}}{\partial \theta^{(j)\top}} \frac{Y_t^{(j)}}{\mu_t^{(j)2}} \right\}.$$

Building on the results by Bollerslev and Wooldridge (1992) and Lee and Hansen (1994), Engle (2000) shows the consistency and asymptotic normality of $\widehat{\boldsymbol{\theta}}$, where the asymptotic covariance corresponds to the Bollerslev and Wooldridge (1992) QML covariance matrix.

Model evaluation can be straightforwardly performed by testing the dynamic and distributional properties of the model residuals

$$e_t^{(j)} \stackrel{\text{def}}{=} \widehat{\varepsilon}_t^{(j)} = Y_t^{(j)}/\widehat{\mu}_t^{(j)}. \tag{19.21}$$

Under correct model specification, the series $e_t^{(j)}$ must be i.i.d. with distribution $\mathbb{D}(\cdot)$. Portmanteau statistics such as the Ljung-Box statistic (Ljung and Box (1978)) based on (de-meaned) MEM residuals can be used to analyze whether the specification is able to capture the dynamic properties of the process. The distributional properties can be checked based on QQ-plots. Engle and Russell (1998) propose a simple test for no excess dispersion implied by an exponential distribution using the statistic

$$\sqrt{n} \left\{ (\widehat{\sigma}_{e^{(j)}}^2 - 1)/\widetilde{\sigma}^{(j)} \right\},$$

where $\widehat{\sigma}_{e^{(j)}}^2$ is the sample variance of $e_t^{(j)}$ and $\widetilde{\sigma}^{(j)}$ is the standard deviation of $(\varepsilon_t^{(j)} - 1)^2$. Under the null hypothesis of an exponential distribution, the test statistic is asymptotically normally distributed with $\widehat{\sigma}_{e^{(j)}}^2 = 1$ and $\left(\widetilde{\sigma}^{(j)} \right)^2 = \sqrt{8}$.

Alternatively, probability integral transforms can be used to evaluate the in-sample goodness-of-fit, see, e.g., Bauwens et al. (2004). Building on the work by Rosenblatt (1952), Diebold et al. (1998) show that

$$q_t^{(j)} \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f_{e^{(j)}}(s)ds$$

must be i.i.d. $\mathbb{U}[0, 1]$. Alternative ways to evaluate MEMs are Lagrange Multiplier tests as proposed by Meitz and Teräsvirta (2006), (integrated) conditional moment tests as discussed by Hautsch (2006) or nonparametric tests as suggested by Fernandes and Grammig (2006).

19.5 High-Frequency Volatility and Liquidity Dynamics

In this section, we will illustrate an application of the VMEM to jointly model return volatilities, average trade sizes, the number of trades as well as average trading costs in intra-day trading. We use a data base extracted from the electronic trading of the Australian Stock Exchange (ASX) which is also used by Hall and Hautsch (2006, 2007). The ASX is a continuous double auction electronic market where the continuous trading period between 10:09 a.m. and 4:00 p.m. is preceded and followed by a call auction. During continuous trading, any buy (sell) order entered that has a price that is greater than (less than) or equal to existing queued buy (sell) orders, will execute immediately and will result in a transaction as long as there is no more buy (sell) order volume that has a price that is equal to or greater (less) than the entered buy (sell) order. In case of partial execution, the remaining volume enters the limit order queues. Limit orders are queued in the buy and sell queues according to a strict price-time priority order and may be entered, deleted and modified without restriction. For more details on ASX trading, see Hall and Hautsch (2007).

Here, we use data from completely reconstructed order books for the stocks of the National Australian Bank (NAB) and BHP Billiton Limited (BHP) during the trading period July and August 2002 covering 45 trading days. In order to reduce the impact of opening and closure effects, we delete all observations before 10:15 a.m. and after 3:45 p.m. To reduce the complexity of the model we restrict our analysis to equi-distant observations based on one-minute aggregates. For applications of MEMs to irregularly spaced data, see Manganelli (2005) or Engle (2000).

Table 19.1 shows summary statistics for log returns, the average trade size,

the number of trades, and the average (time-weighted) trading costs. The log returns correspond to the residuals of an MA(1) model for differences in log transaction prices. This pre-adjustment removes the effects of the well-known bid-ask bounce causing negative first-order serial correlation, see Roll (1984). The trading costs are computed as the hypothetical trading costs of an order of the size of 10,000 shares in excess to the trading costs which would prevail if investors could trade at the mid-quote. They are computed as a time-weighted average based on the average ask and bid volume pending in the queues and yield natural measures of transaction costs induced by a potentially lacking liquidity supply. Conversely, trade sizes and the number of trades per interval indicate the liquidity demand in the market.

We observe that high-frequency log returns reveal similar stochastic properties as daily log returns with significant overkurtosis and slight left-skewness. For the average trade size and the number of trades per interval we find a highly right-skewed distribution with a substantial proportion of observations being zero. These observations stem from tranquil trading periods, where market orders do not necessarily occur every minute. As illustrated below, these periods typically happen around noon causing the well-known 'lunch-time dip'. On the other hand, we also find evidence for very active trading periods resulting in a high speed of trading and large average trade sizes. On average, the number of trades per one-minute interval is around 2.5 and 3.5 for NAB and BHP, respectively, with average trade sizes of approximately 2,300 and 5,800 shares, respectively. The excess trading costs associated with the buy/sell transaction of 10,000 shares are on average around 60 ASD for BHP and 188 ASD for NAB. Hence, on average, excess trading costs for NAB are significantly higher than for BHP which is caused by a higher average bid-ask spread and a lower liquidity supply in the book. The Ljung-Box statistics indicate the presence of a strong serial dependence in volatilities and all liquidity variables, and thus reveal the well-known clustering structures in trading processes. The significant Ljung-Box statistics for log returns are induced by the bid-ask bounce effect causing significantly negative first order autocorrelation. Obviously, the MA(1) filter does not work very well in the case of NAB data. Alternatively, one could use higher order MA-filter. The empirical autocorrelations (ACFs) shown in Figure 19.1 confirm a relatively high persistence in volatilities and liquidity variables indicated by the Ljung-Box statistics. A notable exception is the process of trade sizes for NAB revealing only weak serial dependencies. Figure 19.2 displays the cross-autocorrelation functions (CACFs) between the individual variables. It turns out that squared returns are significantly positively (cross-)autocorrelated with the number of trades and excess trading costs, and – to less extent – with the average trade size. This indicates strong dynamic

	BHP				NAB			
	LR	TS	NT	TC	LR	TS	NT	TC
Obs.	14520	14520	14520	14520	14503	14503	14503	14503
Mean	6.81E-7	5811.52	3.53	60.20	-3.19E-4	2295.24	2.69	188.85
S.D.	7.41E-2	8378.09	3.20	18.47	3.83E-2	7228.38	2.72	97.37
Min	-0.50	0	0	2.99	-0.31	0	0	16.52
Max	0.44	250000	24	231.38	0.38	757500.50	23	1043.35
q10	-0.10	0	0	5.00	-0.04	0	0	84.48
q90	0.10	13475	8	8.80	0.04	5150	6	317.48
Kurtosis	5.23	-	-	-	9.85	-	-	-
LB_{20}	29.61	1585.04	34907.77	22422.32	939.05	95.94	22825.72	23786.09
LB_{20}(SR)	2073.77	-	-	-	2808.75	-	-	-

Table 19.1. Descriptive statistics of log returns (LR), trade sizes (TS), number of trades (NT), and trade costs (TC) for BHP and NAB. Evaluated statistics: mean value, standard deviation (S.D.), minimum and maximum, 10%- and 90%-quantile (q10 and q90, respectively), kurtosis, and the Ljung-Box statistic (associated with 20 lags). LB_{20}(SR) represents the Ljung-Box statistic computed for the squared log returns (SR).
XFGdescriptive

interdependencies between volatility and liquidity demand as well as supply. Similarly, we also observe significantly positive CACFs between trade sizes and the speed of trading. Hence, periods of high liquidity demand are characterized by both high trade sizes and a high trading intensity. Conversely, the CACFS between trading costs and trade sizes as well as between trading costs and the trading intensity are significantly negative. Ceteris paribus this indicates that market participants tend to exploit periods of high liquidity supply, i.e. they trade fast and high volumes if the trading costs are low and thus liquidity supply is high.

A typical feature of high-frequency data is the strong influence of intra-day seasonalities which is well documented by a wide range of empirical studies. For detailed illustrations, see Bauwens and Giot (2001) or Hautsch (2004). One possibility to account for intra-day seasonalities is to augment the specification of $\boldsymbol{\mu}_t$ by appropriate regressors. An alternative way is to adjust for seasonalities in a first step. Though the effect of a pre-adjustment on the final parameter estimates is controversially discussed in the literature (see e.g. Veredas et al. (2001)), most empirical studies prefer the two-stage method since it reduces model complexity and the number of parameters to be estimated in the final step. Here, we follow this proceeding and adjust the individual variables $Y_t^{(i)}$ for deterministic intraday-seasonalities based on cubic spline regressions with 30-minute nodes between 10:30 and 15:30. Figure 19.3 shows the resulting estimated seasonality components. Confirming empirical findings from other markets, we observe that the liquidity demand follows a distinct U-shape pattern with a clear dip around lunch time. How-
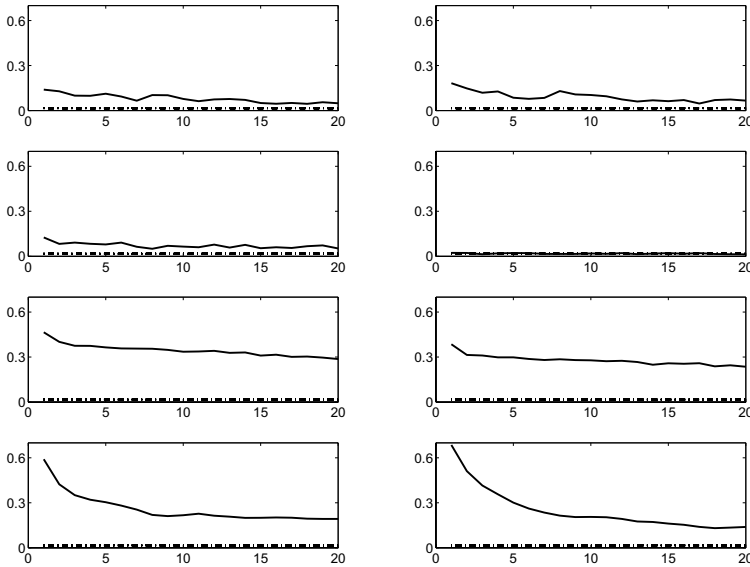
Figure 19.1. Sample ACF of squared log returns (SR), trade size (TS), number of trades (NT), and trade costs (TC)(from top to bottom) for BHP (left) and NAB (right). The x-axis shows the lags. The broken line shows the asymptotic 95% confidence intervals. **Q** XFGacf

ever, a clearly different picture is revealed for the trading costs. Obviously, the liquidity supply is lowest during the morning and around noon inducing high trading costs. Then, (excess) trading costs decline during the after-noon and reach a minimum before market closure. This indicates that not only liquidity demand but also liquidity supply is highest before the end of the trading period. For volatility, we observe a rather typical picture with the highest volatility after the opening of the market and (but to less ex-tent) before closure. The high volatility at morning is clearly associated with information processing during the first minutes of trading.

Conceptual difficulties are caused by the relatively high number of zeros in the liquidity demand variables which cannot be captured by a standard MEM re-quiring positive random variables. In order to account for zeros, we augment
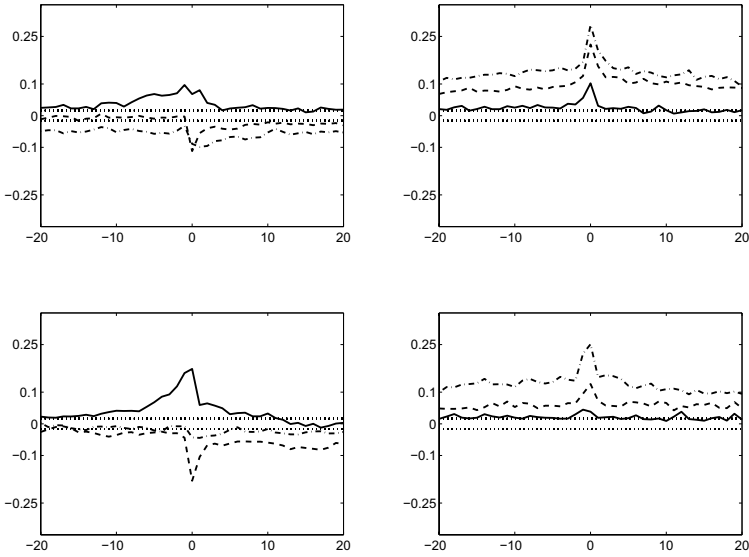
Figure 19.2. Sample CACF for BHP (top) and NAB (bottom). The solid, dash-dotted and dashed lines show the CACF between TC and SR, TC and TS, TC and NT, respectively, on the left side and between SR and TS, SR and NT, TS and NT, respectively, on the right side. The dotted line shows the asymptotic 95% confidence interval. The x-axis shows the lags. ◻ XFGcacf

a Log-VMEM by corresponding dummy variables:

$$\log \boldsymbol{\mu}_t = \boldsymbol{\omega} + \boldsymbol{\mathcal{A}}_0[(\log \mathbf{Y_t}) \odot \mathbf{1}\{Y_t > 0\}] + \boldsymbol{\mathcal{A}}_0^0 \odot \mathbf{1}\{Y_t = 0\} \tag{19.22}$$

$$+ \sum_{j=1}^{p} \boldsymbol{\mathcal{A}}_j[g(\boldsymbol{\varepsilon}_{t-j}) \odot \mathbf{1}\{Y_{t-1} > 0\}] + \sum_{j=1}^{p} \boldsymbol{\mathcal{A}}_j^0 \odot \mathbf{1}\{Y_{t-1} = 0\} \tag{19.23}$$

$$+ \sum_{j=1}^{q} \boldsymbol{\mathcal{B}}_j \log \boldsymbol{\mu}_{t-j}, \tag{19.24}$$

where $\mathbf{1}\{Y_t > 0\}\}$ and $\mathbf{1}Y_t = 0$ denote $k \times 1$ vectors of indicator variables indicating non-zeor and zero realizations ,respectively, and $\boldsymbol{\mathcal{A}}_j^0$, $j = 0, \dots, p$, are corresponding $k \times k$ parameter matrices.
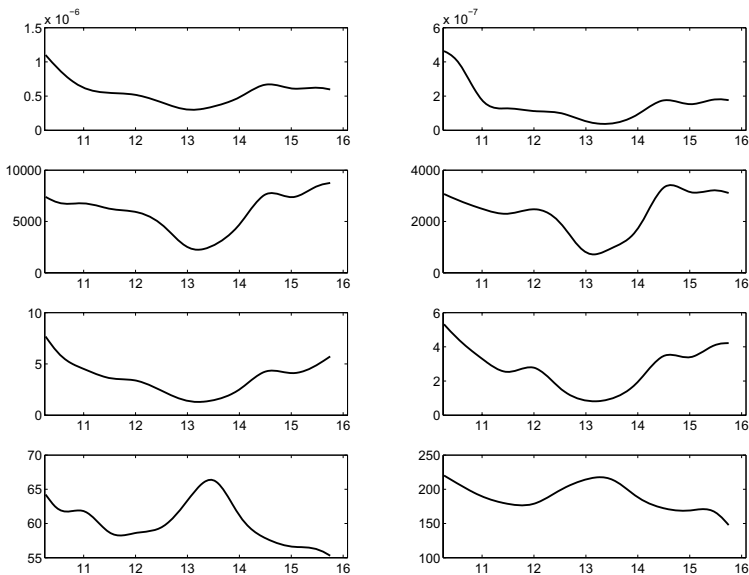
Figure 19.3. Deterministic intra-day seasonality patterns for SR, TS, NT and TC (from top to bottom) for BHP (left) and NAB (right). The seasonality components are estimated using cubic spline functions based on 30-minute nodes. The x-axis gives the time of the day. **Q** `XFGseasonality`

Then, the log likelihood function is split up into two parts yielding

$$\mathcal{L}(\boldsymbol{\theta}) = \sum_{t=1}^{T} \sum_{j=1}^{k} \log f(Y_t^{(j)} | Y_t^{(j+1)}, \ldots, Y_t^{(k)}; Y_t^{(j)} > 0, \mathcal{F}_{t-1}) \qquad (19.25)$$

$$\times \log \mathrm{P}[Y_t^{(j)} > 0 | Y_t^{(j+1)}, \ldots, Y_t^{(k)}; \mathcal{F}_{t-1}]. \qquad (19.26)$$

If both likelihood components have no common parameters, the second part can be maximized separately based on a binary choice model including past (and contemporaneous) variables as regressors. Then, the first log likelihood component is associated only with positive values and corresponds to the log likelihood given by (19.22).

We estimate a four-dimensional Log-VMEM for squared log returns, trade sizes, the number of trades and transaction costs standardized by their corresponding seasonality components. For simplicity and to keep the model tractable, we restrict our analysis to a specification of the order $p = q = 1$. The innovation terms are chosen as $g(\boldsymbol{\varepsilon_t}) = \boldsymbol{\varepsilon_t}$. For the process of squared

returns, $Y_t^{(1)} = r_t^2$, we assume $Y_t^{(1)}|Y_t^{(2)}, \dots, Y_t^{(4)}, \mathcal{F}_{t-1} \sim \mathrm{N}(0, \mu_t^{(1)})$. Accordingly, for $Y_t^{(j)}$, $j \in \{2, 3, 4\}$, we assume $Y_t^{(j)}|Y_t^{(j+1)}, \dots, Y_t^{(4)}, \mathcal{F}_{t-1} \sim Exp(\mu_t^{(j)})$. Though it is well-known that both the normal and the exponential distribution are not flexible enough to capture the distributional properties of high-frequency trading processes, they allow for a QML estimation of the model.

Hence, the adjustments for zero variables have to be done only in the liquidity components but not in the return component. Moreover, note that there are no zeros in the trading cost component. Furthermore, zero variables in the trade size and the number of trades per construction always occur simultaneously. Consequently, we can only identify the $(2, 3)$-element in $\mathcal{A}_0^0$ and one of the two middle columns in $\mathcal{A}_1^0$, where all other parameters in $\mathcal{A}_0^0$ and $\mathcal{A}_1^0$ are set to zero.

For the sake of brevity we do not show the estimation results of the binary choice component but restrict our analysis to the estimation of the MEM. Figure 19.2 shows the estimation results for BHP and NAB based on a specification with fully parameterized matrix $\mathcal{A}_1$ and diagonal matrix $\mathcal{B}_1$.

We can summarize the following major findings: First, we observe significant mutual correlations between nearly all variables. Confirming the descriptive statistics above, volatility is positively correlated with liquidity demand and liquidity supply. Hence, active trading as driven by high volumes and high trading intensities is accompanied by high volatility. Simultaneously, as indicated by significantly negative estimates of $\mathcal{A}_{24}^0$ and $\mathcal{A}_{34}^0$, these are trading periods which are characterized by low transaction costs.

Second, as indicated by the diagonal elements in $\mathcal{A}_1$ and the elements in $\mathcal{B}_1$, all trading components are strongly positively autocorrelated but are not very persistent. As also revealed by the descriptive statistics, the strongest first order serial dependence is observed for the process of trading costs. The persistence is highest for trade sizes and trading intensities.

Third, we find Granger causalities from liquidity variables to future volatility. High trade sizes predict high future return volatilities. However, the impact of trading intensities and trading costs on future volatility is less clear. Here, we find contradictive results for both stocks. Conversely, we do not observe any predictability of return volatility for future liquidity demand and supply. For both stocks all corresponding coefficients are insignificant.

Fourth, trade sizes are significantly negatively driven by past trading intensities and past trading costs. This finding indicates that a high speed of trading tends to reduce trade sizes over time. Similarly, increasing trading

	BHP		NAB	
Coeff.	Coeff.	Std. err.	Coeff.	Std. err.
ω_1	-0.0673	0.0663	0.0023	0.0302
ω_2	0.1921	0.0449	0.1371	0.0254
ω_3	-0.4722	0.1009	-0.1226	0.0432
ω_4	-0.4914	0.1066	-0.5773	0.0485
$A_{0,12}$	0.0549	0.0092	0.1249	0.0056
$A_{0,13}$	0.3142	0.0173	0.6070	0.0122
$A_{0,14}$	0.4685	0.0489	0.7876	0.0094
$A_{0,23}$	0.0673	0.0074	0.0531	0.0070
$A_{0,24}$	-0.1002	0.0289	0.0176	0.0093
$A_{0,34}$	-0.2181	0.0618	-0.0235	0.0123
$A_{0,12}^0$	-3.8196	0.0402	-1.5086	0.0176
$A_{1,11}$	0.1446	0.0080	0.0804	0.0038
$A_{1,12}$	0.0043	0.0090	0.0804	0.0041
$A_{1,13}$	-0.0939	0.0173	0.2036	0.0125
$A_{1,14}$	0.1487	0.0602	-0.0833	0.0214
$A_{1,21}$	0.0004	0.0034	-0.0002	0.0015
$A_{1,22}$	0.0488	0.0049	0.0259	0.0025
$A_{1,23}$	-0.0377	0.0115	-0.0116	0.0093
$A_{1,24}$	-0.1911	0.0398	-0.1329	0.0226
$A_{1,31}$	0.0100	0.0053	-0.0022	0.0020
$A_{1,32}$	0.0095	0.0071	0.0045	0.0031
$A_{1,33}$	0.1088	0.0152	0.0894	0.0109
$A_{1,34}$	0.3420	0.0932	0.0341	0.0377
$A_{1,41}$	0.0064	0.0113	0.0044	0.0067
$A_{1,42}$	0.0091	0.0163	0.0081	0.0081
$A_{1,43}$	0.0524	0.0321	0.0537	0.0249
$A_{1,44}$	0.4256	0.0898	0.5105	0.0431
$A_{1,21}^0$	1.1467	0.0911	-0.5181	0.0204
$A_{1,22}^0$	0.1497	0.0212	0.0341	0.0134
$A_{1,23}^0$	0.0946	0.0318	0.0985	0.0132
$A_{1,24}^0$	-0.0006	0.0755	0.0115	0.0579
$B_{1,11}$	0.4027	0.0252	0.2616	0.0078
$B_{1,22}$	0.7736	0.0179	0.9109	0.0081
$B_{1,33}$	0.9731	0.0074	0.9673	0.0070
$B_{1,44}$	0.5369	0.1024	0.7832	0.0374
Log Likelihood	-60211		-58622	
BIC	-60378		-58790	

Table 19.2. Quasi-maximum likelihood estimation results of a MEM for seasonally adjusted (i) squared (bid-ask bounce adjusted) log returns, (ii) average trade sizes, (iii) number of trades, and (iv) average trading costs per one-minute interval. Standard errors are computed based on the OPG covariance matrix. **Q** XFGestimates

Descriptive statistics of seasonally adjusted data							
	BHP				NAB		
	SR	TS	NT	TC	SR	TS	NT	TC
Mean	1.000	1.001	1.000	1.000	1.002	1.001	1.000	0.999
S.D.	1.963	1.528	0.834	0.300	3.152	2.644	0.991	0.507
LB_{20}	1159.456	202.001	8782.762	19210.412	800.808	124.806	3775.762	19707.831
Descriptive statistics of MEM-residuals							
	BHP				NAB		
	SR	TS	NT	TC	SR	TS	NT	TC
Mean	1.000	1.000	1.000	1.000	1.000	0.999	1.001	1.000
S.D.	1.568	1.348	0.629	0.228	2.595	2.280	0.675	0.386
LB_{20}	63.559	61.388	519.348	1568.428	63.455	14.201	751.317	163.426

Table 19.3.   Summary statistics of the seasonality adjusted time series and the corresponding MEM residuals for BHP and NAB. 🔍XFGdiagnostic

costs deplete the incentive for high order sizes but on the other hand increase the speed of trading. Hence, market participants observing a low liquidity supply reduce trade sizes but trade more often. A possible explanation for this finding is that investors tend to break up large orders into sequences of small orders.

Fifth, (excess) transaction costs depend only on their own history but not on the lagged volatility or liquidity demand. This indicates that liquidity supply is difficult to predict based on the history of the trading process.

Sixth, as shown by the summary statistics of the MEM residuals, the model captures a substantial part of the serial dependence in the data. This is indicated by a significant reduction of the corresponding Ljung-Box statistics. Nevertheless, for some processes, there is still significant remaining serial dependence in the residuals. This is particularly true for the trading cost component for which obviously higher order dynamics have to be taken into account. For the sake of brevity we refrain from showing results of higher parameterized models. Allowing for more dynamic and distributional flexibility further improves the goodness-of-fit, however, makes the model less tractable and less stable for out-of-sample forecasts.

In summary, we find strong dynamic interdependencies and causalities between high-frequency volatility, liquidity supply, and liquidity demand. Such results might serve as valuable input for trading strategies and (automated) trading algorithms.

Bibliography

Bauwens, L. and Giot, P. (2000). The Logarithmic ACD Model: An Application to the Bid/Ask Quote Process of two NYSE Stocks. *Annales d'Economie et de Statistique*: **60**, 117-149.

Bauwens, L. and Giot, P. (2001). Econometric Modelling of Stock Market Intraday Activity, Springer Verlag.

Bauwens, L., Giot, P., Grammig, J. and Veredas, D. (2004). A Comparison of Financial Duration Models Via Density Forecasts. *International Journal of Forecasting*: **20**, 589-609.

Bauwens, L. and Hautsch, N.(2008). Modelling Financial High Freuquency Data Using Point Processes, *in* T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch (eds), *Handbook of Financial Time Series*, Springer Verlag.

Bauwens, L. and Veredas, D. (2004). The Stochastic Conditional Duration Model: A Latent Factor Model for the Analysis of Financial Durations. *Journal of Econometrics*: **119**, 381-412.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*: **31**, 307-327.

Bollerslev, T. and Wooldridge, J. M. (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*: **11**, 143-172.

Cipollini, F., Engle, R. F. and Gallo, G. M. (2007). Vector Multiplicative Error Models: Representation and Inference. *Technical Report, NBER*.

Diebold,F. X., Gunther, T. A. and Tay, A. S. (1998). Evaluating Density Forecasts, with Applications to Financial Risk Management. *International Economic Review*: **39**, 863-883.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*: **50**, 987-1008.

Engle, R. F. (2000). The Econometrics of Ultra-High Frequency Data. *Econometrica*: **68**, 1-22.

Engle, R. F. (2002). New frontiers for ARCH models. *Journal of Applied Econometrics*: **17**, 425-446.

Engle, R. F. and Gallo, G. M. (2006). A Multiple Indicators Model For Volatility Using Intra-Daily Data. *Journal of Econometrics*: **131**, 3-27.

Engle, R. F. and Russell, J. R. (1997). Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with the Autoregressive Conditional Duration Model. *Journal of Empirical Finance*: **4**, 187-212.

Engle, R. F. and Russell, J. R. (1998). Autoregressive Conditional Duration: A New Model for Irregularly Spaced Transaction Data. *Econometrica*: **66**, 1127-1162.

Fernandes, M. and Grammig, J. (2006). A Family of Autoregressive Conditional Duration Models. *Journal of Econometrics*: **130**, 1-23.

Hall, A.D. and Hautsch, N. (2006). Order Aggressiveness and Order Book Dynamics. *Empirical Economics*: **30**, 973-1005.

Hall, A.D. and Hautsch, N. (2007). Modelling the Buy and Sell Intensity in a Limit Order Book Market. *Journal of Financial Markets*: **10(3)**, 249-286.

Hautsch, N. (2003). Assessing the Risk of Liquidity Suppliers on the Basis of Excess Demand Intensities. *Journal of Financial Econometrics*: **1(2)**, 189-215.

Hautsch, N. (2004). *Modelling Irregularly Spaced Financial Data - Theory and Practice of Dynamic Duration Models, vol. 539 of Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin.

Hautsch, N. (2006). Testing the Conditional Mean Function of Autoregressive Conditional Duration Models. *Technical Report, Finance Research Unit, Department of Economics, University of Copenhagen,*.

Hautsch, N. (2008). Capturing Common Components in High-Frequency Financial Time Series: A Multivariate Stochastic Multiplicative Error Model. *Journal of Economic Dynamics & Control, forthcoming*.

Heinen, A. and Rengifo, E. (2006). Multivariate Autoregressive Modelling of Time Series Count Data Using Copulas. *Journal of Empirical Finance*: **14**, 564-583.

Lee, S. and Hansen, B. (1994). Asymptotic Theory for the GARCH(1,1) Quasi-Maximum Likelihood Estimator. *Econometric Theory*: **10**, 29-52.

Ljung, G. M. and Box, G. E. P. (1978). On a Measure of Lack of Fit in Time Series Models. *Biometrika*: **65**, 297-303.

Manganelli, S. (2005). Duration, Volume and Volatility Impact of Trades. *Journal of Financial Markets*: **131**, 377-399.

Meitz, M. and Teräsvirta, T. (2006). Evaluating models of autoregressive conditional duration. *Journal of Business and Economic Statistics*: **24**, 104-124.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Journal of Econometrics*: **43**, 227-251.

Roll, R. (1984). A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market. *Journal of Finance*: **39**, 1127-1139.

Rosenblatt, M. (1952). Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*: **23**, 470-472.

S. J. Taylor(1982). Financial Returns Modelled by the Product of Two Stochastic Processes - a Study of Daily Sugar Prices, *in* O. D. Anderson (eds), *Time Series Analysis: Theory and Practice*, North-Holland, Amsterdam.

Veredas,D., Rodriguez-Poo, J. and Espasa, A. (2001). On the (Intradaily) Seasonality, Dynamics and Durations Zero of a Financial Point Process. *Universidad Carlos III, Departamento de Estad ́istica y Econometr ́ia*.

20 Statistical Process Control in Asset Management

Vasyl Golosnoy and Wolfgang Schmid

20.1 Introduction

Statistical process control (SPC) suggests tools for the on-line detection of changes in the parameters of the process of interest. For the purpose of the data analysis the observations are divided into two parts, historical and on-line observations. The historical observations are used to make statements about the distributional properties of the process. Such results are necessary for the calculation of the design of control charts, which are the most important monitoring instruments in SPC. Every newly incoming information is immediately exploited. The new observations are analyzed on-line. It is examined at each time point whether the process parameters identified from the historical observations remain unchanged. The control chart gives a signal if the process parameters have changed in a statistically significant way. A decision maker should carefully analyze possible causes and consequences of any signal.

Although the methods of SPC have been used for a while in engineering and medical applications Montgomery (2005), they have only recently been applied to economic and financial problem settings. Theodossiou (1993) proposed tools for predicting business failures, Schmid and Tzotchev (2004) introduced control procedures for the parameters of the Cox-Ingersoll-Ross term structure model, Andersson et al. (2005) considered the surveillance of business cycle turning points, while Andreou and Ghysels (2006) monitor the variance of risky assets. A comprehensive review of SPC methods in finance is provided in Frisén (2007).

This chapter provides an overview about possible applications of SPC in asset management. Financial decisions are based on the knowledge of the process parameters, usually estimated from historical data. However, the parameters

may change over time. SPC suggests methods for a quick detection of important changes in the parameters. This presentation intends to show how both active and passive portfolio investors can successfully apply the tools of SPC for their activities. A passive portfolio investor buys a well-diversified fund hoping that the fund manager can beat the market on the basis of his high professional skills. Such investor needs a tool for the on-line monitoring of the manager's performance. As long as the manager performs well, the investor should hold this fund. However, as soon as his performance can be considered unsatisfactory, the investor should reconsider his wealth allocation decisions, and probably choose another investment fund. In order to differentiate between a satisfactory and a non-satisfactory manager's performance we discuss a method proposed by Yashchin et al. (1997) and Philips et al. (2003). They make use of a cumulative sum (CUSUM) chart to monitor the success of the portfolio manager.

An active portfolio manager does not hire a fund manager but composes his portfolio on his own. Hence he needs to monitor the optimal portfolio proportions maximizing his objective function. Here the surveillance of the global minimum variance portfolio (GMVP) weights is investigated. The GMVP purely consists of risky assets and exhibits the lowest attainable portfolio variance in the mean-variance space. Since the vector of the GMVP weights is high-dimensional, multivariate control charts are needed for this task. A signal indicates on possible changes in the GMVP proportions. We consider two multivariate exponentially weighted moving average (EWMA) charts. The presentation of the charts which are suitable for an active portfolio investor goes along the lines of Golosnoy and Schmid (2007).

The rest of the chapter is organized as follows. Section 20.2 provides a brief review of the terminology, instruments, and procedures of SPC. Section 20.3 discusses the application of SPC in asset management for passive as well as for active portfolio investors. In particular, Section 20.3.1 deals with monitoring a fund manager's performance, while in Section 20.3.2 the monitoring the GMVP weights is discussed. Section 20.4 concludes.

20.2 Review of Statistical Process Control Concepts

The methods of SPC are of great importance in many fields of science. SPC deals with the question whether the process under investigation shows a supposed behavior or not. Because the data are sequentially analyzed a change in the parameters of the process of interest (target values) can be detected

quicker than by using conventional statistical fixed-sample tests.

The first sequential monitoring procedure was suggested by Walter Shewhart about 80 years ago. Juran (1997) described in his memories the beginning of SPC by the remarkable sentence "Shewhart invented the control chart on May 16, 1924". Shewhart control charts have gained widespread applications in industry. Because they are able to find an error in a production process at an early point in time their application allows to reduce production costs dramatically. Nowadays control charts refer to the most important and widely used devices in applied statistics Stoumbos et al. (2000).

The work of Shewhart has been the starting point of many new fruitful developments. Because Shewhart charts make exclusively use of the present sample, they are not effective for small and moderate changes. In such cases, control schemes with memory like, e.g., the chart of Page (1954) and the EWMA scheme of Roberts (1959) provide better results. Many further control schemes have been introduced in the literature Montgomery (2005), Schmid (2007) for considering among others autocorrelated processes, multivariate extensions etc.

A control chart is a rule dealing with a decision problem. It should provide a differentiation between the two hypotheses H_0 and H_1. H_0 says that there is no change in the pre-identified process while H_1 means that there is a change at a certain unknown point in time $q \geq 1$. Thus the hypothesis H_0 is equivalent to $q = \infty$ and H_1 to $q < \infty$. Frequently the parameter of interest is a location parameter. Assuming a change point model the hypothesis H_0 states that $\mathsf{E}(X_t) = \mu_0$ for all $t \in \mathbb{N}$. Then under H_1 it holds that $\mathsf{E}(X_t) = \mu_0$ for $1 \leq t < q$ and $\mathsf{E}(X_t) = \mu_0 + a$ for $t \geq q$. Here the parameter $a \neq 0$ describes the size of the shift. The parameter μ_0 is called target value, which is assumed to be known in the majority of applications. If the null hypothesis remains valid the observed process is said to be in control, otherwise it is denoted as being out of control.

In classical statistics a sample of fixed size is taken to differentiate between the hypotheses H_0 and H_1. In SPC, however, a sequential procedure is chosen. At each time point $t \geq 1$ it is analyzed whether based on the information contained in the first t observations the null hypothesis can be rejected or not. If at time point t the null hypothesis is not rejected then the analysis continues and the decision problem at time $t + 1$ using the first $t + 1$ observations is analyzed. Otherwise, if the alternative hypothesis is accepted at time t the procedure stops. This shows that a control chart is a sequential tool with a random sample size.

A decision rule is based on a control statistic $T_t, t \geq 1$ and a non-rejection

area $\mathcal{A}$. In practice, $\mathcal{A}$ is frequently an interval of the type $\mathcal{A} = [-c, c]$ or $\mathcal{A} = (-\infty, c]$, and c is called control limit. If $T_t \in \mathcal{A}$ then it is concluded that the observed process is in control. The analysis continues and the next observation is taken. Else, if $T_t \notin \mathcal{A}$, the process is considered to be out of control. The monitoring procedure stops. Note that there exists a possibility of a false alarm, i.e. a signal is given although the observed process is actually in control. The following actions depend on the specific situation. In engineering applications the machine would be maintained while in finance the portfolio would be adjusted.

The behavior of a chart heavily depends on the choice of the control limit c. If c is small then the chart will give a signal very fast, however, the rate of false alarms would be high. If c is large, the rate of false signals is smaller but the chart will react on an actual change at a later point in time. The choice of c is closely related with the choice of a performance criterion of a control chart. Note that a control chart is based on a sequential approach. Consequently it is a priori unknown how many observations will be analyzed. This is a great difference with respect to classical statistics where the underlying sample size is fixed. For that reason it is also not possible to assess the performance of a control chart with the criteria commonly used in testing theory, e.g., the power function. In SPC all relevant information about the control chart performance is contained in the distribution of the run length, defined as $N = \inf\{t \in \mathbb{N} : T_t \notin \mathcal{A}\}$. The performance criteria are based on the moments of the run length. The most popular measure is the average run length (ARL). The in-control ARL is defined as $\mathsf{E}_\infty(N)$, while the out-of-control ARL is given by $\mathsf{E}_{a,q=1}(N)$. Here the notation $\mathsf{E}_{a,q}(.)$ means that the expectation is calculated with respect to the change point model presented above and $\mathsf{E}_\infty(N)$ is used for $q = \infty$. Note that the out-of-control ARL is calculated under the assumption that the change already arises at the beginning, i.e. for $q = 1$.

The control limit is usually chosen such that the in-control ARL, i.e. $\mathsf{E}_\infty(N)$, is equal to a pre-specified value, say ξ. The choice of ξ depends on the amount of data and the nature of the monitoring task. In engineering ξ is frequently chosen equal to 500. In financial applications this value is too large if daily data are considered. Then a smaller value fits better like $\xi = 60$ or $\xi = 120$ which roughly corresponds to three and six months of the daily observations at the stock exchange, respectively. Then the control limit c is the solution of $\mathsf{E}_\infty[N(c)] = \xi$. For each chart its own control limit has to be determined. This step is similar to the determination of the critical values for a significance test. Based on these control limits different control charts can be compared with each other in the out-of-control case.

There is a huge discussion about performance criteria of control charts and statements about optimal charts in literature. A recent review can be found in Frisén (2003). The most frequently applied performance measure is the out-of-control average run length. An extensive comparison of the Shewhart, EWMA, and CUSUM mean charts for an independent sample is given in Lucas and Saccucci (1990). The authors give recommendations about the memory parameters of the EWMA and CUSUM charts as well. These quantities are denoted as smoothing parameters for the EWMA scheme and reference values for the CUSUM chart. The main disadvantage of the out-of-control ARL lies in the assumption that the process is out of control already from the beginning at $q = 1$. For this reason many authors prefer to work with measures based on the average delay. The average delay is equal to the average number of observations after the change, i.e. $\mathsf{E}_{a,q}(N - q + 1 | N \geq q)$. Since this quantity depends on q the supremum over q or the limit for q tending to infinity is considered in practice.

Contrary to testing theory, it has turned out that there is no globally optimal control scheme. Statements about local optimality could be obtained only for the most simple control schemes, see Lorden (1971) and Moustakides (1986). There are several papers as well dealing with the asymptotic optimality of certain procedures, see Srivastava and Wu (1993), Lai (1995) and Lai (2004). This evidence shows that each monitoring problem requires a separate analysis.

The practical calculation of the performance measures turns out to be complicated. Explicit formulas are only known for the no memory Shewhart chart. The Markov chain approach of Brook and Evans (1972) has turned out to be quite successful for the EWMA and CUSUM schemes. However in general, e.g. for multivariate processes or for dependent data, this method cannot be used. Then the critical values as well as the performance criteria can only be estimated by a simulation study.

20.3 Applications of SPC in Asset Management

The application of control charts to financial problems has been recently suggested in literature Schmid and Tzotchev (2004), Andreou and Ghysels (2006), Frisén (2007). Our presentation focuses on possible applications of SPC for financial management problems. Both actively trading and passively holding investors need sequential monitoring instruments for making on-line decisions concerning their wealth allocation.

A passive investor, holding a well-diversified fund, is interested in sequential

procedures for the evaluation of the fund manager's abilities. As soon as the signal is given, the performance of the manager is considered to be unsatisfactory. This issue is investigated in the studies of Yashchin et al. (1997) and Philips et al. (2003).

Different from a passive investor, an active investor is willing to select his portfolio proportions on his own. He estimates the optimal portfolio weights using historical information. Since suboptimal portfolio holdings may cause significant economic losses, he requires instruments to check whether his portfolio proportions are still optimal at a later time point. For this task suitable schemes have been developed by Golosnoy (2004) and Golosnoy and Schmid (2007).

This chapter provides a review of monitoring procedures for the problems described above. For each approach we introduce hypotheses to be checked in a sequential manner. Then the appropriate monitoring schemes (control charts) are discussed. Each case is illustrated with studies based on Monte Carlo simulations. The instruments used for the control of a fund manager's performance are univariate control charts. On the contrary, the tools for the surveillance of the optimal portfolio proportions are multivariate schemes. These examples provide a brief illustration of SPC methods which can be useful in asset management.

20.3.1 Monitoring Active Portfolio Managers

An investor following a passive trading strategy buys a fund hoping that a fund-manager can achieve high returns on the portfolio by taking a reasonable level of risk. Thus it is important to evaluate the level of proficiency of the fund manager.

Monitoring Problem

The passive portfolio investor evaluates the performance of a fund manager relative to some pre-specified benchmark portfolio. Usually, an appropriate stock market index can serve as a benchmark. Thus the results of the manager's activity must be evaluated sequentially compared to the benchmark performance. If the manager succeeds in outperforming the benchmark in the long run, i.e. the return on his portfolio $1 + r_M$ is significantly larger than on the benchmark $1 + r_B$, his performance is believed to be satisfactory. In case that he underperforms compared to the benchmark, the performance

of the manager is said to be unsatisfactory. If no clear decision can be taken, the manager performance is considered to be neither satisfactory nor unsatisfactory. In this case the evaluation of the manager should be proceeded.

Yashchin et al. (1997) suggest to evaluate the performance of fund managers at the end of the period t using the information ratio R_t. The information ratio denotes the excess return over the benchmark normalized by its standard deviation. It is given by

$$R_t = \mathsf{E}(\Delta_t)/\sqrt{\mathsf{Var}(\Delta_t)} \quad \text{with} \quad \Delta_t = \log \frac{1 + r_{M,t}}{1 + r_{B,t}}.$$

This measure has a clear interpretation due to its link to the Sharpe ratio, moreover, it is adjusted for heteroscedasticity. An estimator of this quantity is given by

$$\widehat{R}_t = \frac{\Delta_t}{\widehat{\sigma}_{t-1}}. \tag{20.1}$$

The investor willing to monitor the information ratio of his fund manager should differentiate at each point of time t between the two hypotheses:

$$H_{0,t}: \ \mathsf{E}(\widehat{R}_t) = R_0 \qquad \text{against} \qquad H_{1,t}: \ \mathsf{E}(\widehat{R}_t) = R_a. \tag{20.2}$$

The information ratio R_0 corresponds to the desired level of the manager's performance, while the alternative R_a represents a non-satisfactory performance. Accordingly, the process of interest $\{\widehat{R}_t\}$ is considered to be in control for $\mathsf{E}(\widehat{R}_t) = R_0$, and to be out of control for $\mathsf{E}(\widehat{R}_t) = R_a$.

Yashchin et al. (1997) suggest to use a conditional standard deviation in (20.1) given by the recursion

$$\widehat{\sigma}_t^2 = (1 - \lambda)\widehat{\sigma}_{t-1}^2 + \lambda(\Delta_t - \Delta_{t-1})^2/2, \qquad t \geq 2 \tag{20.3}$$

with the starting values $\widehat{\sigma}_0^2 = \widehat{\sigma}_1^2 = \sigma_0^2$ and the memory parameter $\lambda \in (0, 1]$. The one period lagged standard deviation $\widehat{\sigma}_{t-1}$ is chosen in (20.1) in order to reduce the autocorrelation in the observed $\widehat{R}_t$. The resulting process is weakly autocorrelated Philips et al. (2003). The parameter λ reflects the impact of innovations on the volatility estimate. A small value of λ implies a long memory and only a minor impact of incoming news on the volatility measure, while the value $\lambda = 1$ determines the case of no memory. Of course, other approaches (e.g. GARCH) can be applied for estimating the conditional covariances as well.

Univariate CUSUM Scheme

Since the aim is to quickly detect an unsatisfactory performance of the fund manager, a univariate one-sided CUSUM procedure is mostly appropriate for this task. The CUSUM approach originates from the classical likelihood ratio test Ghosh and Sen (1991). The first CUSUM control chart has been introduced by Page (1954).

For an independent sequence $\{X_t\}$ the CUSUM control statistic is computed via the recursion for $t \geq 1$

$$T_t = \max\left\{0, T_{t-1} + \log\frac{f(x_t|\text{process is out-of-control})}{f(x_t|\text{process is in-control})}\right\} \quad T_0 = 0, \ (20.4)$$

where $f(\cdot)$ denotes the probability density function of X_t.

In our case the process of interest is $\{\widehat{R}_t\}$. Assuming normality the formula (20.4) can be written as

$$T_t = \max\{0, T_{t-1} - \widehat{R}_t + h\} \quad \text{for} \quad t \geq 1, \ T_0 = 0, \tag{20.5}$$

where $h = (R_0 + R_a)/2$ and R_0 and R_a are the means of the process in the acceptable and unacceptable states, respectively. The recursion (20.5) is suitable to differentiate between a satisfactory and a non-satisfactory performance. The scheme provides a signal if $T_t > c > 0$, where the control limit c is chosen to achieve a desired trade-off between the rate of false alarms and the sensitivity of the control chart.

Application for Manager Performance Evaluation

The CUSUM procedure in (20.5) is in a certain sense optimal if R_0 and R_a are both known. Following the analysis of Yashchin et al. (1997), p. 199, assume that a satisfactory annualized information ratio is $R_0 = 0.5$. This choice implies that the probability to outperform the benchmark is about 0.807 at the end of a three year period.

In order to calculate h we have to fix R_0 as well as R_a. A natural choice of R_a is $R_a = 0$ which can be seen as a boundary between a satisfactory and an unsatisfactory performance. Consequently the reference value is equal to $h = R_0/2$. In this chapter the neutral (or boundary) performance information ratio $R_a = 0$ as well as the unsatisfactory information ratio $R_a^* = -0.5$ are taken as the parameters in the out-of-control state.

For the parameter $\lambda = 1$, i.e. the case of no memory, the in-control informa-
tion ratio $\{\widehat{R}_t\}$ is an independent and normally distributed random variable
with variance equal to one. Since the parameter value λ should be taken
smaller than 1 for the purpose of volatility smoothing, the resulting process
$\{\widehat{R}_t\}$ is positively autocorrelated. In this case numerical techniques are re-
quired for calculating both the in-control and the out-of-control ARLs. Since
the sequence $\{\widehat{R}_t\}$ is assumed to be Gaussian, Monte Carlo simulations are
appropriate for this task.

A monthly time scale is a natural choice to illustrate the usefulness of the pro-
cedure. If there is no signal then the investor considers the fund manager's
performance as satisfactory, while a signal indicates that the performance
may not be satisfactory any more. According to `RiskMetrics` recommenda-
tion the smoothing parameter for the conditional variance in (20.3) is selected
to be $\lambda = 0.1$.

The last step of the procedure is to determine the control limit c. This is
done in such a way that the in-control ARL exhibits a pre-specified value.
Philips et al. (2003) provide control limits for various in- and out-of-control
ARLs (see Table 20.1).

c	$R_0 = 0.5$	$R_a = 0$	$R_a^* = -0.5$
11.81	24	16	11
15.00	36	22	15
17.60	48	27	18
19.81	60	32	21
21.79	72	37	23
23.59	84	41	25

Table 20.1. Control limit c and ARLs (months) for various
values of the information ratio for the reference value $h = 0.25$.

The control limits c reported in the first column of Table 20.1 are calculated
in order to achieve the desired in-control ARL which is given in the second
column. The out-of-control ARLs are calculated for the neutral performance
$R_a = 0$ and the unsatisfactory performance with $R_a^* = -0.5$. The results are
given in the third and fourth column, respectively. Table 20.1 reports that
the signal for the unsatisfactory performance is given much quicker as for the
neutral case. For example, in the case of an in-control ARL of 60 months, the
chart would give a signal on average after 21 monthly observations for the
information ratio $R_a^* = -0.5$. Thus the introduced control chart allows to get
a signal more than two times earlier in case of an unsatisfactory performance

of the fund manager.

This example illustrates the importance of the introduced monitoring tool for an investor willing to evaluate the performance of the fund manager. Further discussion on this issue can be found in Yashchin et al. (1997) and Philips et al. (2003).

20.3.2 Surveillance of the Optimal Portfolio Proportions

An active investor wants to manage his portfolio by his own. In order to make decisions about his wealth allocation he has to estimate the optimal portfolio proportions, which may change over time. For this reason he requires a tool for the sequential detection of changes in the optimal portfolio weights. The following presentation relies on ideas and results given in Golosnoy (2004) and Golosnoy and Schmid (2007).

Monitoring Problem

Let us consider an investor who wants to invest his money. The k-dimensional vector $\mathbf{X}_t$ denotes the vector of the asset returns at time point t. The portfolio theory of Markowitz (1952) requires the knowledge of the expectation and the variance of the asset returns, denoted by $\mathsf{E}(\mathbf{X}) = \boldsymbol{\mu}$ and $\mathrm{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$, for any investment decisions. The covariance matrix $\boldsymbol{\Sigma}$ is assumed to be positive definite. The unknown true moments of the normal distribution have to be estimated using historical information. However, the portfolio performance is often severely hampered by errors arising due to the estimation of the expected returns Best and Grauer (1991).

A pure volatility timing investor selects the global minimum variance portfolio (GMVP). In that case the portfolio proportions do not depend on the mean vector $\boldsymbol{\mu}$ at all. The GMVP allows to get the smallest attainable level of risk. Choosing the GMVP implies that the investor has no reliable information about the expected returns, but hopes to profit from the allocation into risky assets. The GMVP weights $\mathbf{w}$ are obtained by minimizing the portfolio variance subject to the constraint $\mathbf{w}^\top \mathbf{1} = 1$, where $\mathbf{1}$ is a vector of ones of corresponding dimension. The vector of the GMVP weights $\mathbf{w}$ is given by

$$\mathbf{w} = \frac{\boldsymbol{\Sigma}^{-1}\mathbf{1}}{\mathbf{1}^\top \boldsymbol{\Sigma}^{-1}\mathbf{1}} \; . \tag{20.6}$$

The unknown covariance matrix $\boldsymbol{\Sigma}$ must be estimated as well. Hereafter it is replaced by the sample estimator, based on n period returns $\mathbf{X}_{t-n+1}, ..., \mathbf{X}_t$,

i.e.

$$\widehat{\boldsymbol{\Sigma}}_{t,n} = \frac{1}{n-1} \sum_{j=t-n+1}^{t} (\mathbf{X}_j - \overline{\mathbf{X}}_{t,n})(\mathbf{X}_j - \overline{\mathbf{X}}_{t,n})^{\top}, \quad \overline{\mathbf{X}}_{t,n} = n^{-1} \sum_{v=t-n+1}^{t} \mathbf{X}_v \quad (20.7)$$

Here usually n is assumed to be fixed. The matrix $\boldsymbol{\Sigma}$ is estimated using an equally-weighted rolling window. Alternative weighting schemes have been proposed by, e.g., Foster and Nelson (1996) and Andreou and Ghysels (2002) but stay beyond the scope of the presentation. The weights of the GMVP are estimated at time t as a function of $\widehat{\boldsymbol{\Sigma}}_{t,n}$ by

$$\widehat{\mathbf{w}}_{t,n} = \frac{\widehat{\boldsymbol{\Sigma}}_{t,n}^{-1} \mathbf{1}}{\mathbf{1}^{\top} \widehat{\boldsymbol{\Sigma}}_{t,n}^{-1} \mathbf{1}}, \quad (20.8)$$

where $\widehat{\boldsymbol{\Sigma}}_{t,n}$ is given in equation (20.7). Assuming that the returns $\{\mathbf{X}_t\}$ follow a stationary Gaussian process with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ Okhrin and Schmid (2006) prove that the estimated optimal weights $\widehat{\mathbf{w}}_{t,n}$ are asymptotically normally distributed. Moreover, they derive the exact distribution of $\widehat{\mathbf{w}}_{t,n}$ under the assumption that the underlying returns on risky assets are independent and normally distributed random variables. The joint distribution of $\widehat{\mathbf{w}}_{t,n}$ is degenerated. It holds that

$$\mathsf{E}(\widehat{\mathbf{w}}_{t,n}) = \mathbf{w}, \quad \mathrm{Cov}(\widehat{\mathbf{w}}_{t,n}) = \boldsymbol{\Omega} = \frac{1}{n-k-1} \frac{\mathbf{Q}}{\mathbf{1}^{\top} \boldsymbol{\Sigma}^{-1} \mathbf{1}}, \quad (20.9)$$

$$\mathbf{Q} = \boldsymbol{\Sigma}^{-1} - \frac{\boldsymbol{\Sigma}^{-1} \mathbf{1} \mathbf{1}^{\top} \boldsymbol{\Sigma}^{-1}}{\mathbf{1}^{\top} \boldsymbol{\Sigma}^{-1} \mathbf{1}}. \quad (20.10)$$

Because from one point in time to the next one there are $n-1$ overlapping values, the process of the estimated weights $\{\widehat{\mathbf{w}}_{t,n}\}$ is strongly autocorrelated, although the process of returns $\{\mathbf{X}_t\}$ is assumed to be i.i.d. Golosnoy and Schmid (2007) propose to approximate the autocovariance function of the estimated GMVP weights. It is shown that for fixed $s \geq 1$ and large n it holds that

$$\mathrm{Cov}_0(\widehat{\mathbf{w}}_{t,n}, \widehat{\mathbf{w}}_{t-s,n}) \approx \frac{n-s-1}{n-1} \frac{1}{n-k-1} \frac{\mathbf{Q}}{\mathbf{1}^{\top} \boldsymbol{\Sigma}^{-1} \mathbf{1}}. \quad (20.11)$$

Further investigations show that this approximation is reasonable for $n \geq 30$ and $s \ll n$. For small values of n this approximation seems to be strongly biased and, consequently, it is not suitable.

An active investor is interested in monitoring the optimal portfolio proportions because he makes his investment decisions in terms of the weights. He

should detect changes in the optimal portfolio weights as soon as possible in order to adjust his asset allocation and to minimize utility losses due to suboptimal portfolio holdings. Next we formalize the task of the surveillance of the optimal portfolio weights.

The estimator of the optimal portfolio weights $\widehat{\mathbf{w}}_{t,n}$ is unbiased if the underlying process is equal to the target process, i.e. $\mathsf{E}_\infty(\widehat{\mathbf{w}}_{t,n}) = \mathbf{w}$. The actual (observed) process $\{\widehat{\mathbf{w}}_t\}$ is considered to be "in control" if the actual process is equal to the target process, else it is denoted to be "out of control".

In the out-of-control situation the returns are assumed to be independent and normally distributed, however, the parameters of the distribution have changed. The changes have occurred in the covariance matrix of the asset returns. Note that not every change in the covariance matrix $\boldsymbol{\Sigma}$ leads to alterations in the GMVP weights. The time points of the changes are unknown and should be determined. Further it is assumed that there are no changes in the process up to time point $t = 0$. The investor requires an instrument for deciding between the two hypotheses at each time point $t \geq 1$

$$H_{0,t} : \mathsf{E}(\widehat{\mathbf{w}}_{t,n}) = \mathbf{w} \qquad \text{against} \qquad H_{1,t} : \mathsf{E}(\widehat{\mathbf{w}}_{t,n}) \neq \mathbf{w} \ . \tag{20.12}$$

The validity of the hypothesis $H_{0,t}$ is analyzed sequentially. Due to the restriction $\widehat{\mathbf{w}}_{t,n}^\top \mathbf{1} = 1$, it is sufficient to monitor the vector of the first $k-1$ components of $\widehat{\mathbf{w}}_{t,n}$, denoted as $\widehat{\mathbf{w}}_{t,n}^*$. The $(k-1) \times (k-1)$-dimensional matrix $\boldsymbol{\Omega}^*$ is obtained from the matrix $\boldsymbol{\Omega}$ by deleting the last row and column. Next we discuss sequential procedures for differentiating between the hypotheses (20.12). Both considered multivariate schemes belong to the family of the exponentially weighted moving average (EWMA) control charts.

Multivariate Control Charts

A Chart Based on the Mahalanobis Distance

The Mahalanobis distance measures the distance between the estimated GMVP weights $\widehat{\mathbf{w}}_{t,n}^*$ and the target weights $\mathbf{w}^* = \mathsf{E}_\infty(\widehat{\mathbf{w}}_{t,n}^*)$. It is given by

$$T_{t,n} = (\widehat{\mathbf{w}}_{t,n}^* - \mathbf{w}^*)^\top \boldsymbol{\Omega}^{*-1} (\widehat{\mathbf{w}}_{t,n}^* - \mathbf{w}^*) \quad , \quad t \geq 1 \ .$$

The quantities $T_{1,n}, ..., T_{t,n}$ are exponentially smoothed. The EWMA recursion is given by

$$Z_{t,n} = (1-\lambda)Z_{t-1,n} + \lambda T_{t,n} \quad \text{with} \quad Z_{0,n} = k-1, \tag{20.13}$$

since $Z_{0,n} = \mathsf{E}_\infty(T_{t,n}) = \mathrm{tr}[\boldsymbol{\Omega}^{*-1}\ \mathsf{E}_\infty\{(\widehat{\mathbf{w}}_{t,n}^* - \mathbf{w}^*)(\widehat{\mathbf{w}}_{t,n}^* - \mathbf{w}^*)^\top\}] = k - 1$. Again, the parameter $\lambda \in (0,1]$ is a smoothing parameter. A small value of λ leads to a larger influence of past observations. The special case $\lambda = 1.0$ corresponds to the classical no-memory Shewhart control chart. The control chart gives a signal at time t if $Z_{t,n} > c_1$. The constant $c_1 > 0$ is a given value determined via Monte Carlo simulations as a solution of (20.16).

Note that previous values $\mathbf{X}_{1-n}, .., \mathbf{X}_1$ are required for calculating $T_{1,n}$. Hence, a starting problem is present. Since our surveillance procedure is started at time $t = 1$ we need these past observations. However, in financial applications this is usually no problem. If all observations at times $t \le 0$ are realizations from the target process, only the first observation used for the calculation of the GMVP weights may be contaminated at time $t = 1$. Consequently, its influence on the estimated weights may be small. This evidence may cause delays in the reaction on changes. Thus this chart suffers under an inertia property.

A Chart Based on the Multivariate EWMA Statistic

A multivariate EWMA (MEWMA) chart has been initially suggested for i.i.d. observations by Lowry et al. (1992). In our case each component of $\widehat{\mathbf{w}}_{t,n}^*$ is exponentially smoothed by an own smoothing factor. The $k - 1$-dimensional EWMA recursion $\mathbf{Z}_{t,n}$ can be written using the matrix notation as follows

$$\mathbf{Z}_{t,n} = (\mathbf{I} - \mathbf{R})\mathbf{Z}_{t-1,n} + \mathbf{R}\widehat{\mathbf{w}}_{t,n}^*, \qquad t \ge 1. \tag{20.14}$$

$\mathbf{I}$ denotes the $(k-1) \times (k-1)$ identity matrix and $\mathbf{R} = \mathrm{diag}(r_1, ..., r_{k-1})$ is a $(k-1) \times (k-1)$ diagonal matrix with elements $0 < r_i \le 1$ for $i \in \{1, ..., k-1\}$. The initial value $\mathbf{Z}_{0,n}$ is taken as $\mathbf{Z}_{0,n} = \mathsf{E}_\infty(\widehat{\mathbf{w}}_{t,n}^*) = \mathbf{w}^*$. Then the vector $\mathbf{Z}_{t,n}$ can be presented as

$$\mathbf{Z}_{t,n} = (\mathbf{I} - \mathbf{R})^t\mathbf{Z}_{0,n} + \mathbf{R}\sum_{v=0}^{t-1}(\mathbf{I} - \mathbf{R})^v\widehat{\mathbf{w}}_{t-v,n}^*.$$

It holds that $\mathsf{E}_\infty(\mathbf{Z}_{t,n}) = \mathbf{w}^*$. The covariance matrix of the multivariate EWMA statistic $\mathbf{Z}_{t,n}$ in the in-control state is given by

$$\mathrm{Cov}_\infty(\mathbf{Z}_{t,n}) = \mathbf{R}\ \left(\sum_{i,j=0}^{t-1}(\mathbf{I} - \mathbf{R})^i\ \mathrm{Cov}_0(\widehat{\mathbf{w}}_{t-i,n}^*, \widehat{\mathbf{w}}_{t-j,n}^*)\ (\mathbf{I} - \mathbf{R})^j\right)\ \mathbf{R}.$$

Since the evaluation of this formula is not straightforward, Golosnoy and Schmid (2007) derive an approximation to $\mathrm{Cov}_\infty(\mathbf{Z}_{t,n})$ as n tends to ∞. However, from the practical point of view, it is often more suitable to estimate the required finite sample matrix within a Monte Carlo study.

A corresponding control chart is constructed by considering the distance between $\mathbf{Z}_{t,n}$ and $\mathsf{E}_\infty(\mathbf{Z}_{t,n}) = \mathbf{w}^*$ which is measured by the Mahalanobis statistic. The chart provides a signal at time t if

$$\{\mathbf{Z}_{t,n} - \underset{\infty}{\mathsf{E}}(\mathbf{Z}_{t,n})\}^\top \; \underset{\infty}{\mathrm{Cov}}(\mathbf{Z}_{t,n})^{-1} \; \{\mathbf{Z}_{t,n} - \underset{\infty}{\mathsf{E}}(\mathbf{Z}_{t,n})\} \; > c_2. \qquad (20.15)$$

Again, the in-control ARL of the control chart is taken equal to a pre-specified value and c_2 is determined as the solution of this equation.

Application for the Optimal Portfolio Weights

Next we illustrate an application of the methodology described above. We restrict ourselves to the case of equally smoothed coefficients, i.e. $r_1 = \cdots = r_{k-1} = \lambda$. The control limits c_1, c_2 for both control procedures are required for monitoring the optimal portfolio weights. They are obtained by setting the desired in-control ARL equal to a predetermined value ξ conditional on the given values of λ, $\boldsymbol{\Sigma}$, n and k, i.e.

$$ARL(c|\boldsymbol{\Sigma}, n, \lambda, k) \; = \; \xi. \qquad (20.16)$$

Due to the complexity of the underlying process of the estimated GMVP weights, the equation (20.16) can only be solved with respect to the control limits c_1 and c_2 by combining a numerical algorithm for solving a nonlinear equation with Monte Carlo simulations. In our study we make use of the Regula falsi algorithm and a Monte Carlo study based on 10^5 replications.

In order to illustrate the application of this methodology, we consider an example with $k = 10$ assets in the portfolio, see Golosnoy (2004). It should be noted that the number of assets k has a significant influence on the control limits resulting from (20.16).

The simulation study is organized as follows. The investment decisions are made on the basis of daily data. The annualized diagonal elements of the $k = 10$ dimensional in-control covariance matrix $\boldsymbol{\Sigma}$ are chosen to be $\sigma_i^2 = 0.1 + (i - 1)/100$, the correlation coefficient is $\rho = 0.3$ for all pairs of assets. These parameter values are typical for asset allocation problems in practice Ang and Bekaert (2002). The optimal portfolio weights $\widehat{\mathbf{w}}_{t,n}$ are calculated for $n = 60$. The control limits for the considered control charts are given in Table 20.2 for the in-control ARLs $\xi = 100$ and $\xi = 200$ for values $\lambda \in \{0.1, 0.25, 0.4, 0.55, 0.7, 0.85, 1.0\}$. The control limits are calculated to achieve an ARL precision of $\pm 0.5\%$.

There exists a large number of different possibilities to model the out-of-control situation. For illustration purposes we provide an example where

λ	Mahalanobis ARL=100	MEWMA ARL=100	Mahalanobis ARL=200	MEWMA ARL=200
0.10	11.72	9.700	14.07	12.16
0.25	13.37	12.61	16.07	15.33
0.40	14.15	13.72	16.97	16.57
0.55	14.64	14.37	17.52	17.26
0.70	15.01	14.83	17.92	17.76
0.85	15.32	15.21	18.27	18.20
1.00	15.61	15.61	18.62	18.62

Table 20.2. Control limits c_1 and c_2 for different in-control ARLs and different values of λ.

the calm time on the market is changed to the turmoil period. Such perturbations are usually characterized by a sudden simultaneous increase both of the volatilities and the correlations of the risky assets Ang and Bekaert (2002). The detection of changes from calm to turmoil periods is of immense importance for portfolio investors.

Although we monitor changes in the GMVP weights, the alterations are modeled in the elements of the covariance matrix of the asset returns. In particular, changes in the variance of the ith asset are assessed as

$$\sigma_{i,1}^2 = \delta_i \sigma_i^2, \qquad \delta_i = 1 + \delta + \delta \log(i)$$

where the factor δ captures the dynamics of the change for all assets. According to this model, the largest increase in the variance occurs for the most volatile asset. Changes in the correlations are given by

$$\rho_{ij,1} = \theta \rho_{ij} \quad \text{for} \quad i \neq j.$$

Thus, we capture the transition of the market from calm to turmoil situations with two parameters, δ and θ. The values $\delta = 0$ and $\theta = 1$ describe the in-control situation. The out-of-control cases are modeled by taking $\theta \in \{1.0, 1.3, 1.7, 2.0\}$ and $\delta \in \{0, 0.2, 0.6, 1.0\}$. The corresponding out-of-control ARLs for the in-control ARL=200 days are reported in Table 20.3. The best smoothing values λ and r are provided in parentheses.

The results in Table 20.3 can be characterized as follows. Large changes can be easily detected, while it is relatively hard to detect small changes in the volatility only, see case $\delta = 0.2, \theta = 1$. Both charts are equivalent for $\lambda = 1$, which is the case of the no-memory Hotelling procedure Montgomery (2005). As expected, the choice $\lambda = 1$ is the best one for large changes, while $\lambda < 1$ is more appropriate for small ones. The important large perturbations could

δ/θ	1.0	1.3	1.7	2.0
0	Mahalanobis	84.4 (0.85)	40.2 (1.0)	30.2 (1.0)
	MEWMA	84.0 (1.0)	40.1 (1.0)	30.2 (1.0)
0.2	201.7 (0.1)	68.3 (0.1)	31.7 (0.7)	22.2 (1.0)
	188.8 (0.1)	70.5 (0.1)	31.7 (1.0)	22.2 (1.0)
0.6	77.5 (0.1)	36.1 (0.7)	17.8 (1.0)	12.9 (1.0)
	78.8 (0.1)	36.1 (1.0)	17.8 (1.0)	12.9 (1.0)
1.0	16.0 (1.0)	19.2 (1.0)	11.2 (1.0)	8.56 (1.0)
	15.9 (1.0)	19.1 (1.0)	11.2 (1.0)	8.55 (1.0)

Table 20.3. Out-of-control ARLs, for increasing volatilities δ and correlations θ, $n = 60$, in-control ARL=200 calculated based on 10^5 replications. The best smoothing parameters are given in parentheses.

be detected within 8-12 days on average, which corresponds to about 2 weeks of observations. Both schemes seem to be nearly equivalent with a slightly better performance of the MEWMA chart. Compared to an in-control ARL of 200, this speed of detection advocates the usage of the introduced monitoring instruments for the GMVP problem.

In general, SPC provides no recommendation concerning the actions after the signal has occurred. The portfolio investor should carefully analyze all available market information and then make his wealth allocation decisions. Control charts primarily serve as instruments for the detection of statistically significant changes in the parameters of interest.

20.4 Summary

The procedures of statistical process control allow to quickly detect changes in the parameters of interest which are required in asset management. Although this field of research is relatively new, the main instruments of SPC, called control charts, are already available for various investment problems of practical relevance. This chapter presents two possibilities of applying SPC tools to wealth allocation decisions.

First, a scheme for monitoring the performance of a fund manager is considered. This issue is of great interest for passive investors concerned with the proficiency of their fund manager. A signal indicates that the fund manager's performance is no more satisfactory. Another example presents the tools for

monitoring the optimal portfolio weights. A signal suggests that the optimal portfolio proportions may have changed. An active investor requires such analysis for adjusting his portfolio and for avoiding non-optimal positions. Methodological issues for both cases are illustrated with examples based on Monte Carlo simulations.

The described applications stress the necessity of sequential monitoring for practical financial management. Neglecting on-line surveillance approaches may lead to suboptimal investment decisions and cause significant financial losses. Much further research should be done in order to establish this field of analysis in financial practice.

Bibliography

Andersson, E., Bock, D., and M. Frisén(2002). Statistical surveillance of cyclical processes with application to turns in business cycles, *Journal of Forecasting, 24, 465-490.*

Ang, A. and Bekaert, G. (2002). International asset allocation with regime shifts, *Review of Financial Studies, 15, 1137-1187.*

Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics, *Journal of Applied Econometrics, 17, 579-600.*

Andreou, E. and Ghysels, E. (2006). Monitoring disruptions in financial markets, *Journal of Econometrics, 135, 77-124.*

Best, M. and Grauer, R. (1991). On the sensitivity of meanvariance-efficient portfolios to changes in asset means: some analytical and computational results, *Review of Financial Studies, 4, 315-342.*

Brook, D. and Evans, D. (1972). An approach to the probability distribution of CUSUM run length, *Biometrika, 59, 539-549.*

Foster, D. P. and Nelson, D. (1996). Continuous record asymptotics for rolling sample variance estimators, *Econometrica, 64, 139-174.*

Frisén, M. (2003). Statistical surveillance: optimality and methods, *International Statistical Review 71, 403-434.*

Frisén, M., ed. (2007). *Financial Surveillance*, John Wiley & Sons, New York.

Ghosh, B.K. and Sen, P.K., eds. (1991). *Handbook of Sequential Analysis*, Marcel Dekker, New York.

Golosnoy, V. (2004). *Sequential Control of Optimal Portfolio Weights*, Ph.D. Thesis, Frankfurt (Oder), Germany.

Golosnoy, V. and Schmid, W. (2007). EWMA control charts for optimal portfolio weights, *Sequential Analysis, 26, 195-224.*

Juran, J. M. (1997). Early SQC: a historical supplement, *Quality Progress 30, 73-81.*

Lai, T. L. (1995). Sequential changepoint detection in quality-control and dynamical systems, *Journal of the Royal Statistical Society B, 57, 613-658.*

Lai, T. L. (2004). Likelihood ratio identities and their applications to sequential analysis, *Sequential Analysis 23, 467-497.*

Lorden, G. (1971). Procedures for reacting to a change in distribution, *Annals of Mathematical Statistics, 41, 1897-1908.*

Lowry, C.A., Woodall, W.H., Champ, C.W., and Rigdon, S.E. (1992). A multivariate exponentially weighted moving average control chart, *Technometrics 34, 46-53.*

Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: Properties and enhancements, *Technometrics, 32, 1-12.*

Markowitz, H. (1952). Portfolio selection, *Journal of Finance, 7, 77-91.*

Montgomery, D. C. (2005). *Introduction to Statistical Quality Control,* 5th edn. John Wiley & Sons, New York.

Moustakides, G. V. (1986). Optimal stopping times for detecting changes in distributions, *Annals of Statistics, 14, 1379-1387.*

Page, E. S. (1954). Continuous inspection schemes, *Biometrika, 41, 100-114.*

Philips, T. K., Yashchin, E. and Stein, D. M. (2003). Using statistical process control to monitor active managers, *Journal of Portfolio Management, 30, 86-95.*

Roberts, S. W. (1959). Control chart tests based on geometric moving averages, *Technometrics, 1, 239-250.*

Okhrin, Y. and Schmid, W. (2006). Distributional properties of optimal portfolio weights, *Journal of Econometrics, 134, 235-256.*

Schmid, W. (2007). Eighty years of control charts, *Sequential Analysis, 26, 117-122.*

Schmid, W. and Tzotchev, D. (2004). Statistical surveillance of the parameters of a onefactor Cox-Ingersoll-Ross model, *Sequential Analysis, 23, 379-412.*

Srivastava, M. S. and Wu, Y. (1993). Comparison of EWMA, CUSUM and Shiryaev-Roberts procedures for detecting a shift in the mean, *The Annals of Statistics, 21, 645-670.*

Stoumbos, Z. G., Reynolds Jr, M. R., Ryan, T. P., and Woodall, W. H. (2000). The state of statistical process control as we proceed into the 21st century, *Journal of the American Statistical Association, 95, 992-998.*

Theodossiou, P. (1993). Predicting shifts in the mean of a multivariate time series process: an application in predicting business failures, *Journal of the American Statistical Association, 88, 441-449.*

Yashchin, E., Philips, T. K., and Stein, D. M. (1997). Monitoring active portfolios using statistical process control, in *Computational Approaches to Economic Problems*, selected papers from the 1st Conference of the Society for Computational Economics, eds. H. Amman et al., Kluwer Academic, Dordrecht, pp. 193-205.

21 Canonical Dynamics Mechanism of Monetary Policy and Interest Rate

Jenher Jeng, Wei-Fang Niu, Nan-Jye Wang, and Shih-Shan Lin

21.1 Introduction

Interest rates are the fundamental elements of financial and economic activities, and their movements are the major risk factors driving the global capital flows. In the United States, the central bank (Federal Reserve Bank) uses the Fed Funds Rate (FFR, the overnight borrowing rate between banks) as the key tool to anchor its monetary policy for maintaining both sustainable growth and price stability. It has been a sophisticated art and science for the Federal Open Market Committee (FOMC, the primary unit of the FRB for setting the FFR) to balance growth and inflation by tuning the FFR. Among a few models trying to quantitatively assess the FOMC's efforts on FFR determination is the popular Taylor Rule (Taylor (1993)) for best outlining the thoughts of arguments from the beginning.

The Taylor Rule formulates the FFR with the weighted estimated real GDP output gap and inflation bias (as measured by the deviation of the GDP deflator from the 2% target level) by a linear model. However, a detailed examination on the difference between the actual FFR series (01/1958 - 12/2007) and the expected FFR series estimated based on the Taylor Rule reveals the fact that the linearity assumption leads Taylor's response model to seriously overestimate the FOMC's moves at critical economic junctures in a systematically biased manner(see Figure 21.3). Despite the simplicity and rigidity of the Taylor Rule, it indeed sheds some lights on several crucial questions about monetary-policy making, such as "how does the Fed adapt to the business cycles for balancing growth and inflation?", "which set of economic measures is the key to determine the FFR?", "how does the Fed perceive major macroeconomic risks – inflation crisis and depression?", "how does the FFR decision-feedback affect the on-going business cycles?", and so on.

To carefully approach these questions in a rapidly changing economic and financial system nowadays since the new global financial infrastructure emerges based on the internet after the late 1990s, we need to establish a much more sophisticated statistical modeling framework which is capable of pattern recognition and structural interpretation on the nonlinear dynamic nature of the macroeconomic game (between the FOMC and the economy, and even the capital markets). Toward this goal, we will start with a basic nonparametric statistical methodology called Projection Pursuit Regression (PPR).

Besides the above academic questions on monetary-policy decision-making from perspectives of economists, there are some practical issues surrounding the movement of the FFR from perspectives of capital markets. In the capital markets, along with the growth of the hedge fund industry, interest-rate speculation (especially by the practice of the "macro" hedge funds) is getting increasingly furious, and thus making the Fed's job more and more difficult. Therefore, hedging the interest-rate risk has become a major task of asset management. Currently, the popular financial instruments for hedging inflation and interest-rate risk have all focused on the Consumer Price Index (CPI), such as TIPS (Treasury Inflation Protection Security) and CPI futures. Therefore, from the practical viewpoint, it would be urgent to develop an FFR-response model with a dynamic indicator based on the CPI rather than the GDP deflator.

Based on a dynamic indicator as our primary gauge of inflation implicitly deduced as a linear combination of the two CPI components – core and non-core – from the principal component analysis of the PPR framework , a nonlinear 3-phase structure of the interest-rate response curve is discovered in this study. The term "3-phase" means that the FOMC implicitly adopts various strategies during three different inflation scenarios – *careful*, *tense* and *panic*. These different regimes of inflation are quantitatively sectored by the new inflation measure, and the FOMC's behavioral patterns in FFR decision-making are then shaped in the geometrical structure of the response curve. We found that such a PPR modeling interpretation seems matching quite well with most FOMC meeting statements and thus provide a way of getting deeper insights into the Fed's views and actions toward inflation bias and the Fed's strategies on keeping business cycles on track.

21.2 Statistical Technology

In this section we review the popular monetary-policy formulation by Taylor (1993), the nonparametric multiple regression technique Projection Pursuit Regression and some classical original mean-reversion interest-rate models.

The Taylor Rule   Taylor (1993) proposed a simple mathematical formula, for estimating the GDP-based feedback FFR in the following way:

$$r_t = \pi_t + r^* + p_1(\pi_t - \pi^*) + p_2 y_t, \tag{21.1}$$

where r_t is the estimated FFR, r^* is the (subjective) equilibrium real FFR (2%), π_t is the inflation rate (to be gauged by GDP deflator), π^* is the (subjective) target inflation rate, y_t is the output gap, and the feedback weights p_1 and p_2 are (subjectively) equally set to be 0.5. First of all, we note that, obviously, this formula is actually an empirically subjective rule rather than a rigorously statistical result for the parameters here can be more accurately assessed through a statistical linear regression instead of being subjectively set.

Taylor's simple "rule" quite successfully sketches the evolution of the fundamental interest rate in the 1990s; however, its call for the monetary policy makers to set target values of interest rate can be questionable. This draws criticism, for example, from McCallum's comment (1993) that Taylor's formula was not "operational". Another viewpoint was presented by Chairman Greenspan (1997) precisely: "As Taylor himself has pointed out, these types of formulations are at best guideposts to help central banks, not inflexible rules that eliminate discretion."

Following the Taylor Rule, there are many model varieties addressing the problem of formulating monetary policy based on the different aspects of dynamic macroeconomic analysis. A typical example is the New Keynesian model, which has been analyzed by Hansen & Sargent (2002), Giannoni (2001, 2002), and Giannoni & Woodford (2002). This model has purely forward-looking specification for price setting and aggregate demand, and no intrinsic persistence is assumed. On the other hand, the model of Rudebusch and Svensson (1999) has purely backward-looking structure and significant intrinsic persistence. Further (2000) suggested a model that utilizes rational expectations but exhibits substantial intrinsic persistence of aggregate spending and inflation. Moreover, Levin and Williams (2003) considered the question - "is there any simple rule capable of providing robust performance across very divergent representations of the economy?" Their research reveals that a robust outcome is attainable only if the object function places substantial weight on stabilizing both output and inflation.

As a matter of fact, we believe that many questions surrounding the Taylor Rule originate from the philosophical issue – what is the real meaning of the hypothetical response interest-rate estimated from statistical regression based on a set of monetary-policy related macroeconomic indicators? Apparently, as mentioned by the former FOMC Chairman Alan Greenspan, it is very unlikely for the FOMC to simply adopt a rigid mathematical rule for being adaptive to the complex financial world. But then, excepting quantitatively analyzing the FOMC's behavior, what could the hypothetical response function try to tell in a forward-looking sense? According to our analysis in this paper, the answer should be embedded in a canonical framework of monitoring interest-rate dynamics centered on mathematically formulating the core idea of "neutral" interest-rate level, as proposed in sections **??** and 21.3.3.

Projection Pursuit Regression based on Cubic Splines   Since functional nonparametric statistical modeling techniques are not broadly used by most macroeconomic and econometric researchers, we would like to recall some details on the basic methodology Projection Pursuit Regression (PPR), which is first proposed by Friedman and Stuetzle (1981). In general, given the covariate random variable Y and the random vector of input variates $X = (X_1, ..., X_d)^\top$, we can formally model their association in the following way:

$$Y = f(X) + \varepsilon \tag{21.2}$$

for some response curve f with the background (e.g. macroeconomic) noise ε.

For linear regression, f is simply in the shape of a line. For a well-structured nonlinear form, say a polynomial, the modeling task well fall into the parametric framework of Generalized Linear Model (GLM), which is currently popular for building most financial scenario simulation and risk management systems. Unfortunately, the real world is far more complicated from what most researchers assume – in mathematics, the response curve f can be highly nonlinear enough to make the GLM techniques generate serious modeling risks (either systematically underestimate or overestimate). Therefore, we need to keep in mind that a reasonable response function might need infinitely many parameters to be figured out (note that, for the linear model (21.1) of Taylor, only the four parameters (r^*, π^*, p_1, p_2) are needed. On the other hand, we cannot allow the structure of the response curve to be as free (wild) as possible for no advanced statistical estimation techniques can come close without any suitable restrictions based on certain feasible realistic assumptions. The PPR methodology proposes the following model-structural

form

$$f = \sum_{j=1}^{J} f_j(a_j^\top X), f_j \in S \tag{21.3}$$

where $a_j, j = 1, ..., J$ are deterministic vectors in $\mathbb{R}^d$ and S is the functional space of all functions which has a continuous second-order derivative. Note that all we need assume is that all the f_j's have continuous second-order derivatives (this smoothness assumption only asserts that the "response" cannot absurdly fluctuate). Apparently, one cannot find a finite set of parameters to identifiably mark all the members in S. This requirement of infinity in function-parameter dimension is exactly the essence of "nonparametric" modeling.

Such a structural assumption in (21.3) is mainly motivated by the following heuristics. Technically speaking, once we move beyond the territory of GLM, we need face the formidable challenge of "curse of dimensionality" when the regression surface is not a hyper-plane. Such a problem usually results in the situation of data insufficiency for a complicated combination of the non-parametric nature of the response surface and the multiplicity of the input variable. Therefore our first statistical task is essentially "dimension reduction" for which the well-known method is principal component analysis. The PPR modeling technique can be regarded as taking the sum of the separate "nonlinear" effects on the principal components. That is, mathematically, we can express f as $T \circ F \circ A$, where A is the $M \times n$ PCA transformation matrix constituted by the columns a^j's, F is the nonlinear transform $(f_1, ..., f_J)$, and T is simply the sum operator. Now, we are left to deal with the core problem about how to estimate the nonparametric nonlinear component - functions of F. There have been many statistical methods developed in nonparametric estimation, including the most advanced wavelet-based methods (see Donoho, Johnstone, Kerkyacharian and Picard (1995); Jeng, 2002). However, in this paper, since we try to focus on the economical issue of discovering the genuinely nonlinear structure of the response curve rather than the deep statistical issue of estimation efficiency and adaptivity, it is enough for us to adopt the fundamental nonparametric regression technique Penalized Least Square Regression (PLSR) based on cubic splines. For introducing the basic materials about the PLSR, we follow the convention of Green and Silverman (1994).

The *Penalized Least Square Estimator* (PLSE) is defined as following. Let $f \in S_2[a, b]$ where S_2 is the space of the functions that are differentiable on

(a, b) and have absolutely continuous derivative. If $\widehat{f}$ satisfies

$$\widehat{f} = \arg \min_{f \in S_2[a,b]} \sum_i \{y_i - f(x_i)\}^2 + \alpha \int_a^b \{f^*(t)\}^2 dt \qquad (21.4)$$

where $\{(x_i, y_i)\}$ is the sample of data and $f^*(x)$ is an integrable function such that $f'(b) - f'(a) = \int_a^b f^*(t)dt$, then $\widehat{f}$ is called the *penalized least square estimator*. To solve the above functional minimization problem, it is required to gradually (along with the sample size) "parameterize" the functional classes taken as our submodels based on a certain kind of function building blocks. One classical choice of building block is the cubic spline (local cubic polynomial), as defined in the following. $f : [a, b] \rightarrow \mathbb{R}$ is called a *cubic spline* if

(i) $f \in P_3[(t_i, t_{i+1})] i = 0, 1, \ldots, n$, where $a \le t_1 \le \ldots \le t_n \le b$ are called *knots*.

(ii) $f \in C^2[a, b]$

Basically speaking, a cubic spline is a second continuously differentiable function constituted by local 3-degree polynomials from knot to knot.

According to certain theorems of functional and statistical analysis, the penalized LSE $\widehat{f}$ must be a natural cubic spline (NCS) with the additional condition $f''(a) = f''(b) = f'''(a) = f'''(b) = 0$ Thus, a natural cubic spline f has the property that it can be completely determined by the values $f(t_i)$'s and $f''(t_i)$'s. Therefore, with the sample data, we can solve the above functional minimization problem through a quadratic-form equation for any smoothness penalty weight α. Moreover, the smoothing parameter α can be further empirically decided by the sample of data via minimizing the cross validation scores in the way

$$\widehat{\alpha} = \arg \min_{\alpha \in (0,\infty)} N^{-1} \sum_{i=1}^n \sum_{j=1}^{m_i} \{y_{ij} - \widehat{f}(t_i; \alpha)\}^2 \qquad (21.5)$$

where N is the total numbers of the data at knots $\{t_i, i = 1, \cdots, n\}$ and $\widehat{f}^{-(i,j)}(x; \alpha)$ is the minimum of the penalty LSE from all the data omitting y_{ij} under $\alpha^{(2)}$. Finally, it comes down to the aggregation work of simultaneously estimating the PCA part and the PLSR part (as the linear and the nonlinear parts). For simplicity of mathematical discussion, we only describe how this is done for the case $M{=}1$ (the way by which we deal with the data in this paper) via the following iterative algorithm:

Step (1): Initialize current residuals as r_i and the iteration-loop counter K, $r_i \leftarrow y_i$ $K = 0$ where $\sum y_i = 0$

Step (2): Find the cross-validation $\widehat{\alpha}_a$ of $\{r_i\}$ and $\{s_i^a\}$

and the cubic spline $\widehat{f}_a$ between $\{r_i\}$ and $\{s_i^a\}$ under $\widehat{\alpha}_a$.

Step (3): Let $I(a) = 1 - \dfrac{\sum\limits_{t=1}^{N}(r_t - \widehat{f}_{\widehat{a}}(s_t^{\widehat{a}}))^2}{\sum\limits_{t=1}^{N} r_t^2}$ where $\widehat{a} = \arg\min I(a)$ and the corre-

sponding smooth function is $\widehat{f}_{\widehat{a}}$

Step (4): Termination Condition:

If the figure of merit is smaller than a user-specified threshold, stop.

Otherwise, update the initialized condition $r_i \leftarrow y_i - \widehat{f}_{\widehat{a}}(s_i^a)$ $K = K + 1$ and
go to step (2)

Interest-Rate Dynamics   Vasicek (1977) proposes an interest rate model
for treasury debt pricing through a mean-reversion type stochastic differential
equation:

$$dr_t = -k(r_t - \mu)dt + \sigma dW_t \tag{21.6}$$

where k is the constant strength of mean-reversion, μ is the equilibrium level,
σ is the volatility and W_t is a standard Brownian motion. Note that Equation
(21.6) can be solved explicitly and represented as

$$r_t \backsim N\{r_{t-1}e^{-k} + \mu(1 - e^{-k}), \frac{\sigma^2(1 - e^{-2k})}{2k}\} \tag{21.7}$$

In check with the big swing in the trend of the actual FFR time series from
1960 to 2005, the above model of mean-reversion with a constant equilib-
rium is obviously too na¨ive to approach the dynamics of the FFR decision-
making. To be more realistic, Hull and White (1990) proposed the local
mean-reversion model with a dynamic equilibrium level

$$dr_t = -k(dr_t - \theta(t))dt + \sigma dW_t \tag{21.8}$$

Now, the real problem is that, for the FFR, it is extremely difficult to analyze
how the dynamic equilibrium level – or the neutral interest rate level - or
the neutral interest rate level $\theta(t)$ is determined by monetary-policy related
economic conditions. In this paper (see 21.3.3), we will propose a fundamental
canonical framework where the dynamic equilibrium level $\theta(t)$ is determined
by the response curve f on a certain set of exogenous economic factors which
are carefully monitored by the FOMC.

21.3 Principles of the Fed Funds Rate Decision-Making

21.3.1 Fairness of Inflation Gauge

This section presents our original intuition and modeling incentives as the basis for further interpretations and improvements on the modeling works.

The FFR is the key instrument for the FOMC to keep price stable while maintaining sustainable growth. However, balancing between inflation and growth is a very subtle task, especially when it is compounded with the question – "what is a fair gauge of inflation?". Among what have been proposed by many central bank officials and economists, Consumer Price Index (CPI), Personal Consumption Expenditures (PCE) and GDP deflator are the most commonly referred as measurements of inflation. However, even these three "standard" inflation indices may still have quite divergent movements from time to time, especially at critical junctures (See Figure 21.1). Theoretically, at the very beginning of modeling, it is very difficult to judge which index should serve as the best primary dynamic indicator for modeling the FOMC's FFR decision-making behavior according to the comparison result by Rich and Steindel (2005). Perhaps, within a general canonical nonparametric statistical modeling framework, we can try to compare the best models under certain statistical criteria based on the different indices for selecting the optimal inflation indicators. But this could involve the very challenging issue of sharp adaptation in nonparametric modeling (see Jeng, 2002). Therefore, at this point, we turn to focus on the practical end. Currently, the popular financial instruments under certain statistical criteria of hedging inflation and interest-rate risk are still mainly focused on the CPI, such as TIPS (Treasury Inflation Protection Security) and the CPI futures issued by CME. Therefore, from the practical viewpoint, it would be urgent to develop an FFR-response model with a dynamic indicator based on the CPI rather than the GDP deflator, as suggested in the Taylor Rule.

As one of the premier indices closely monitored by the FOMC for watching inflation. the CPI weights on a broad range of goods prices which are mainly divided into two component categories – the core CPI and the non-core CPI (see Figure 21.2). The non-core component consists of only food and energy prices. One important reason why the economists like to focus on the core part is that the prices of food and energy are very volatile from time to time (mainly due to short-term non-economical factors, such as weather and commodity futures market speculation). Hence, it is usually concerned that the unstable non-core part of the CPI could mislead the prospect on inflation.
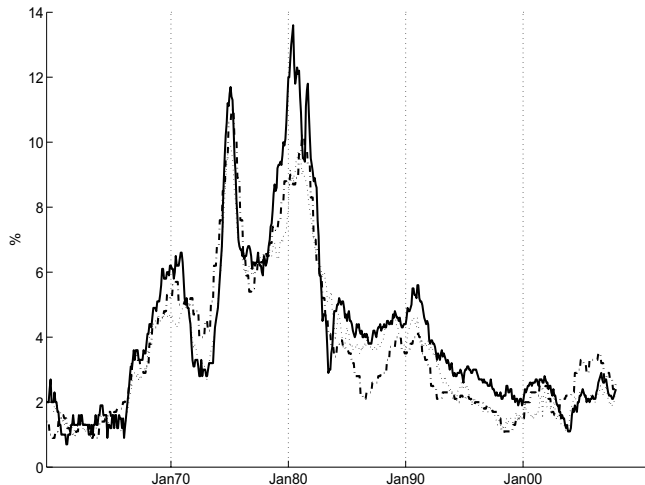
Figure 21.1. Inflation Indices (Jan. 1958 - Dec. 2007):
PCE (dotted), core CPI (solid) and GDP deflator (dashed).
◨ XFGinflation

However, we believe that the high volatility of the non-core CPI should not make the FOMC completely ignore the potential contribution of the non-core CPI into the inflation trend (as usually mentioned as the *spill-over* effect). Then, how does the FOMC weight the core and the non-core components of the CPI (either explicitly by a mathematical rule or implicitly in their sophisticated minds)? In this study, we simply formulate a "fair" gauge of inflation q_t^* as a linear combination of the two CPI components:

$$q_t^* = \beta_{p,1} q_{1,t} + \beta_{p,2} q_{2,t} \tag{21.9}$$

where $q_{1,t}$ and $q_{2,t}$ stand for the core and the non-core CPI respectively. We call q_t^* the *principal dynamic indicator* for the PPR. Its meanings together with the arguments on the term "fair" will be further clarified in 21.3.3 and supported by our data analysis in 21.4.

21.3.2 Neutral Interest Rate Based on Fair Gauge of Inflation

In this study, for the sake of emphasizing model-structural meanings rather than detailed macroeconomic factor analysis, we assume (following the classical viewpoint) that containing inflation is the FOMC's primary focus for
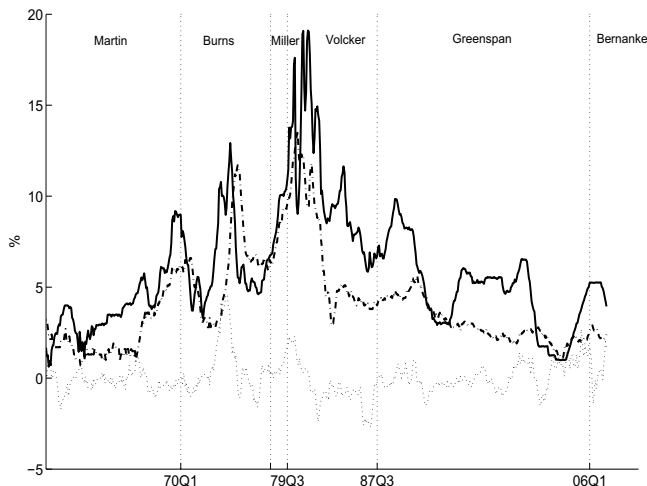
Figure 21.2. The Fed Funds Rate (solid), Core (dot-dash-dot)
and non-Core CPI (dotted) Q XFGrate

setting the FFR.Therefore, it is statistically feasible to formulate the FFR
response of the FOMC on our inflation gauge in the following compact way

$$r_t = f(q_t^*) + \varepsilon \qquad (21.10)$$

where f is called the inflation response curve of the FOMC and the extra
stochastic term ε simply represents the aggregated effect on the FFR ad-
justment decision from all the other secondary economic conditions, such as
unemployment rates, exchange rates, asset prices, geopolitical and financial
crises, and so on. Apparently, the geometrical structure of the response curve
f is supposed to quantitatively depict the FOMC's behavioral patterns on
monetary-policy decision-making. Of course, it would be naïve to presume
any geometrical structure with a couple of parameters for the FOMC could
act in a very flexible and subtle way at some critical junctures of inflation
trends. Moreover, note that the term "response" could be misleading in some
sense. It is clear that the FOMC does not exactly respond in the way which
f describes because there are certainly other factors to distract them from
what the response curve indicates based on the implicit inflation gauge q_t^*,
but the response curve indeed formulates the sense that, without the other
factors' appearances, the FOMC should have set the FFR at the level as
indicated by the response curve based on the implicit inflation gauge q_t^*, and
no matter what happens, the FOMC will eventually corrects back onto the

"fair track" once the effects of the non-primary factors subside in the long run. Note that this is the pivot idea supporting our modeling framework as suggested in the title of the paper. In 21.3.3, we will explain why such an idea is fundamental to analyze the FFR dynamics.

Now, combining (21.9) and (21.10), we build a simple (basic functional non-parametric) PPR framework for modeling the FFR decision-making based on the CPI indices:

$$r = f(\beta_{p,1}q_{1,t} + \beta_{p,2}q_{2,t}) + \varepsilon \qquad (21.11)$$

To interpret this model in plain terminology, the inflation gauge $\beta_{p,1}q_{1,t} + \beta_{p,2}q_{2,t}$ measures the FOMC's consensus over inflation and the response function f depicts the FOMC's mind-set of configuring strategies for maintaining prices stability through accommodative-tightening cycles.

Furthermore, for the interest of *comparative pattern recognition*, we also consider a comparison model via bivariate linear regression (BLR) to heuristically distinguish the interesting nonlinear geometrical features of the PPR model. The comparison model is designed as

$$r = \beta_{b,1}q_{1,t} + \beta_{b,2}q_{2,t} + \varepsilon \qquad (21.12)$$

where $\beta_{b,1}$ and $\beta_{b,2}$ are the weights on core and non-core CPI, and α is the interception. Inside the general framework of 21.10, the BLR model can be regarded as a special case with the imposed modeling bias of linearity. That is,

$$r = f_b(\beta_{b,1}q_{1,t} + \beta_{b,2}q_{2,t}) + \varepsilon \qquad (21.13)$$

where, similarly, we have $\beta_{b,1}q_{1,t} + \beta_{b,2}q_{1,t}$ as another "inflation gauge" with the identity response function, which simply means that the FFR should be set to be the inflation gauge plus a inflation-risk control premium α (or "real neutral" interest rate).

21.3.3 Monetary Policy-Making as Tight-Accommodative Cycles Along Neutral Level as Dynamic Principal

In 21.4, by plotting together the actual FFR and the estimated response FFR series

$$r_t^* = \widehat{f}(\widehat{\beta}_{b,1}q_{1,t} + \widehat{\beta}_{b,2}q_{1,t}), \qquad (21.14)$$

where $(\widehat{f}, \widehat{\beta}_{b,1}, \widehat{\beta}_{b,2})$ are the estimated PPR model based on the real data, we found that the time series r_t actually fluctuates along the track of the PPR time series r_t^* in a quite perfect manner of mean-reversion. Thus, at the

very beginning point, the simplest quantitative way of describing the mean-reversion phenomenon is quite intuitive to assume that the deviation spread or residue $z_t = r_t - r_t^*$ follows an *Ornstein-Uhlenbeck* process

$$dz_t = -kz_t dt + \sigma dW_t \qquad (21.15)$$

Under the proposition that the model-projection level r_t^* represents the "neutral" interest rate, the mean-reversion phenomenon of the actual-projection spread z_t then indeed quantitatively formulates the FOMC's language on monetary policy – the policy cycle is *accommodative* when the spread is negative while an ongoing positive spread means a *tightening* cycle. Note that, for exploring the dynamic of the FFR, this observation reveals a canonical framework which directly generalizes the model of Hull and White (1990) in the following way. The FFR actually follows a process with the dynamic local mean-reversion equilibrium level r_t^* which is completely determined by the *principal dynamic indicator* q_t^* through the dynamic mechanism f. Note that if we, in addition, assume that r_t^* also follows some stochastic diffusion process, simply say

$$dr_t^* = \mu_t dt + \delta dY_t \qquad (21.16)$$

then it is formally straightforward to reach the stochastic version of Hull and White model

$$dr_t = (\theta_t - kr_t)dt + \sigma' dZ_t \qquad (21.17)$$

where the stochastic process $\theta_t = kr_t^* + \mu_t$ is stochastic (rather than a deterministic $\theta(t)$).

21.4 Response Curve Structure and FOMC Behavioral Analysis

The section presents all the empirical results by data analysis with our interpretations.

21.4.1 Data Analysis and Regressive Results

Based on the 571 monthly data of the effective Fed Funds rate and the CPI (from March 1958 to September 2005), the PPR results with the BLR comparison are listed as following: Projection Pursuit Regression (PPR):

$$r = f(0.9593q_1 + 0.2824q_2) \qquad (21.18)$$

where f is the non-parametric function plotted in Figure 21.3.

Bivariate Linear Regression (BLR):

$$r = 2.0670 + 0.9427q_1 + 0.3097q_2 \qquad (21.19)$$

Interestingly, in the two different models, first note that the two inflation gauges are very close to each other with the nonlinear one weighting a bit more on the core. Figure 21.3 shows the estimated PPR response curve and BLR regression line - note that, for convenience of comparison, we use the same x-axis to indicate the two slightly different inflation gauges. Or precisely, the response surfaces with respect to (q_1, q_2) can be compared in Figure 21.4 and 21.5. Moreover, the three FFR series (the actual and the three model-projected) are plotted together in Figure 21.6. .
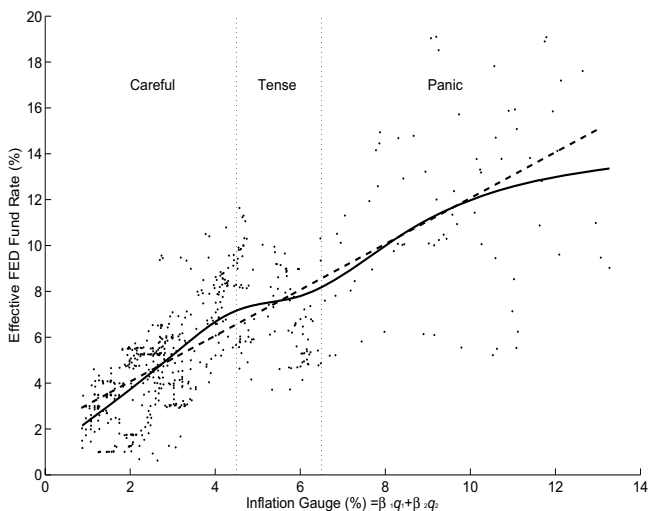
Figure 21.3. The Response Functions: by Bivariate Linear Regression (dashed) and PPR method (solid) ◘ XFGresponse

21.4.2 The Structure of the FOMC's Response Curve – Model Characteristics, Interpretations

To comprehend the complicated reasoning behind the PPR modeling, we, first of all, note the long-lasting debate, both in the academic and Wall Street, on the issue V what are the proper weights should be put on the
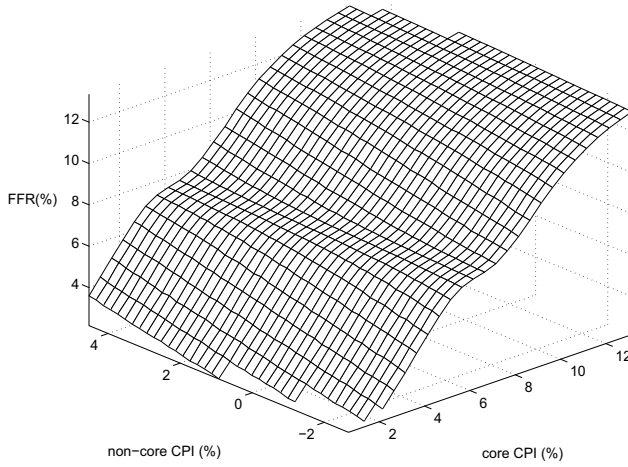
Figure 21.4. The Response Surface ⊙ XFGsurface
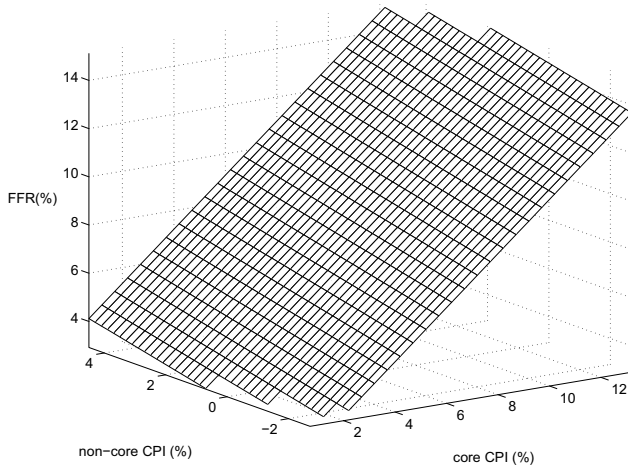
Figure 21.5. The Response Surface ⊙ XFGsurface2

different price components for watching inflation? For giving credit to the principal component analysis embedded in the PPR modeling, note that the
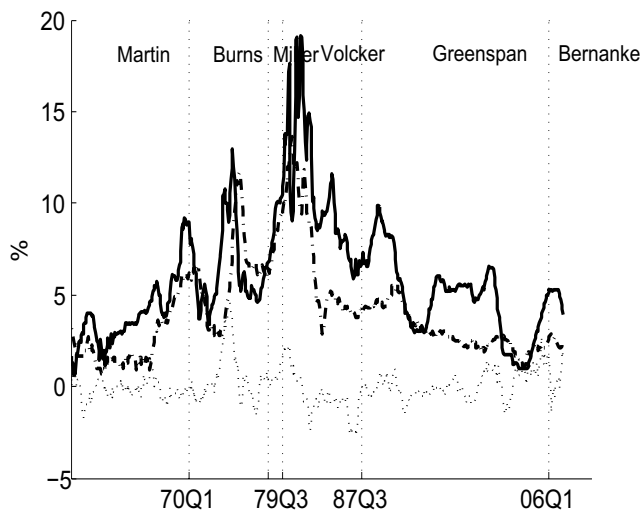
Figure 21.6. Effective and Model-projected Neutral Fed Fund Rates: Effective Fed Fund Rate (dashed line), by Taylor rule (dot-dash-dot line), by PPR method projected (solid line) and by Bilinear Regression (dotted line) 🔍 XFGrate2

PPR inflation-gauge weighting $(0.9699q_1 + 0.2437q_2)$ almost fully takes the highly focused core CPI and puts a much less significant, but not ignorable, weighting factor on the non-core CPI. This suggests that, in the FOMC's mindset, although the volatile non-core CPI cannot be taken too seriously from time to time, but still not negligible (especially deserving a close watch on the "spill-over" effect from the non-core to the core at critical junctures). More importantly, the BLR inflation gauge $(0.9605q_1 + 0.2513q_2)$ comes very close to the PPR measurement. This interesting but not very surprising fact actually implies that this inflation gauge is robust under different models from linear to nonlinear.

Secondly, despite the two models' coincidence on gauging inflation, the PPR response curve has a very subtle geometrical structure distinct from a straight line – we call it a *3-phase* structure. The twisted non-parametric curve can be divided into 3 phases, including *careful*, *tense* and *panic*, meaning FOMC's different response sentiments and actions inside three distinct ranges of inflation gauge, as shown in Figure 21.3.

The *careful* phase, with the inflation gauge less than 4.3% or so and FFR approximately below 7%, reflects the FOMC's mindset for rate adjustment against the inflation movement in the "normal" economic situation. In this

phase, the steepest (most aggressive) FFR ascending along with the inflation increment indicates the FOMC's sensitive and determinate will on "inflation targeting" (quick action at a spark). Moreover, in comparison with the BLR model, note that the BLR model tends to overestimate the rate-hike speed when the inflation is low (or when the Fed likes to give the capital markets a break) while underestimating the Fed's aggressiveness when the inflation is about to get out of the gate.

The *tense* phase, with the PPR inflation gauge ranging within $4.5\% \sim 6.5\%$ and the FFR hanging around $7\% \sim 8\%$, has a flattened response curve segment. In this phase, it seems quite clear that the FOMC's mind-set is getting nervous and reluctant to further hike the FFR when the rate reaches the level 7%. It is reasonable to infer from this pattern that the FOMC should have learned from the past lessons that even a bid further aggressive action beyond this level could trigger a recession anytime soon. These "lessons" can be checked in Figure 21.6 where we notice that a stay near or over the 8% level is always followed by a sharp FFR drop which must have been reacting to a economic recession. Also, notice that, in Figure 21.3, the divergence (variance) of the scatter diagram from the regression curve indeed increase significantly – note that the dispersion of the scatter diagram is nicely homogeneous in the inflation range of the *careful phase*. As a matter of fact, the economy is usually going through a high-growth period in the second phase. The Fed should be glad to wait and see and try not to ruin the economic party once the inflation gauge can be controlled within 6.5%. Of course, the BLR model completely misses this interesting decision-making sentiment feature.

The *panic* phase, with the inflation gauge over 6.5% and the FFR above 8%, shows the Fed's determination to fight the possibly out-of-control inflation once the inflation gauge is heating up over the 6.5% level. The scatter diagram in this phase exhibits a much larger variance than those in the other phases and the inflation is usually far beyond controlled. In this phase, the reason for the FOMC to rekindle the rate-hike from the halting pace is that the cost to pay for falling into a recession is far smaller than taking the risk of stagflation or even depression due to an inflation crisis.

21.4.3 The Dynamics of the FFR – Model Implications

The 3-phase structure subtly reveals the intriguing fact that it is a much tougher game than usually thought in the "linear" way for the FOMC to balance growth and inflation, especially at critical junctures which usually takes place in *the tense phase*. At these junctures, the FOMC often overre-

acted and then cut back the rate in a hustle. Such consequences are clearly shown as the spikes of the actual FFR time series. It leaves us to ponder what if the FOMC took the suggestion of the PPR model - could these absurd spikes be just removed so that the economy thus becomes healthier and the capital market turns more stable? Through these research works from static regression to dynamic analysis, based on the pivot idea of dynamic equilibrium principal, the key observation is that, as shown in Figure 21.7, the actual FFR time series fluctuates along with the PPR-projection series (the FFR series projected by the PPR model) with a quite robust *tight-accommodative* (up-down mean-reversion) cyclic pattern. Thus, it is simply intuitive to take the PPR-projection series as the *dynamic equilibrium level* (see Section 21.3.3), from which the cyclic pattern of deviation is obviously created by the self-correcting efforts of the FOMC.
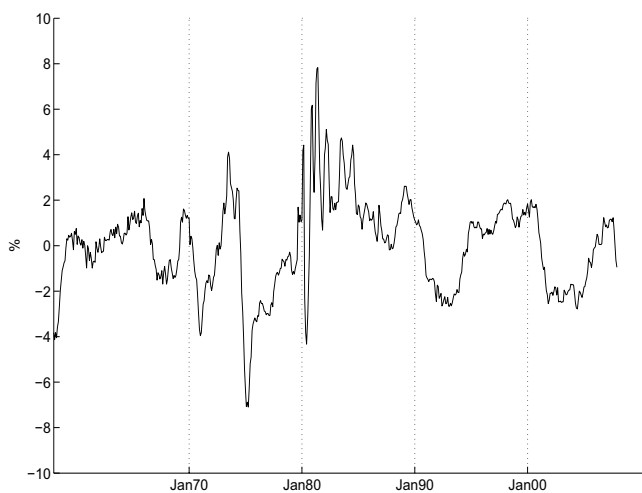
Figure 21.7. The FFR series projected by the PPR model
XFGffr

Now, the really intriguing question is: if the dynamic equilibrium level has done a good statistical job to precisely shape the so-called "neutral interest rate level" in the FOMC's mindset under changing economic conditions, why would the FOMC have not just followed this "neutral track" more closely? The following two possible reasons might be able to provide some crucial heuristics for modeling the dynamics of the FFR:

(I) The dynamic equilibrium level precisely pinpoints the consensus of the FOMC members on the "neutral interest rate level", but there is still the

variance of opinions among the members and even the uncertainty around each member's judgment. For the sake of risk control, such a variance and those personal uncertainties could push the policy makers to do a bid more or less than the "neutral" estimate. In a nonlinear dynamic system, such a bid action could eventually push the whole system away from the equilibrium and thus need the follow-up self-correcting efforts from the FOMC, and, therefore, the cycles are formed.

(II) The second reason provides a deeper insight into the dynamics of the FFR, and thus a more active incentive for not simply keeping the actual FFR on the "neutral track". Since the FOMC's mission is not only controlling inflation, but also minimizing unemployment rate, it becomes an art and science to balance growth and inflation. Regarding the monetary-policy making task as a game between the FOMC and the economy, the Octave Diagram shown in Figure 21.8 illustrates the tight-accommodative cyclic mechanism in the FFR dynamics:
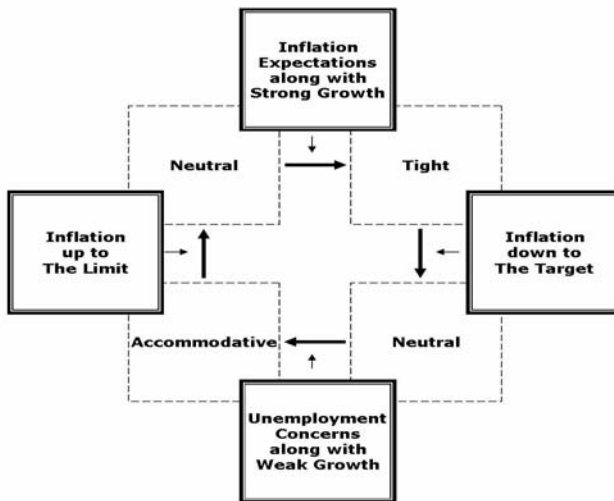
Figure 21.8. Tight-accommodative cyclic mechanism in the FFR dynamics

Note that how the trend of the CPI evolves must depend on the strategy of the FOMC. The dynamic equilibrium level only indicates the "neutral rate" level in the FOMC's mindset to provide a principal guidance for tightening or loosing monetary policy. In the model (21.13), it is more adequate to call f the FOMC's "neutral" response curve, and ε represents the FOMC's strategic spread to form the tight-accommodative cycle for conditional economical

purposes. Just like a box attached to a spring which carries non-zero momentum whenever reaching the equilibrium position in a harmonic motion, the FFR must naturally exhibit the cyclic pattern according to the dynamic mechanism in the Octave Diagram. If the FOMC naively holds its targeted FFR at the neutral level long enough, the inflation could quickly get out of control as the growth momentum is positive; on the other hand, the economy could soon dip into recession or even stall to dive into a depression as the growth momentum is negative. By carefully checking the mean-reversion pattern of the spread (the difference between the actual FFR time series and the dynamic equilibrium level series, i.e. the PPR-projected FFR time series) in Figure 21.7, we may reach the conclusion that a healthy interest rate cycle is essential for price stability.

To reason with the reality, the key relies on the justification of the concept of "neutral real interest rate", as focused for researching monetary policy in the classic booklet of Blinder (1999). In Blinder's viewpoint, the neutral real interest rate is difficult to estimate and impossible to know with precision, but is most usefully thought of as a concept rather than as a number, as a way of thinking about monetary policy rather than as the basis for a mechanical rule. However, as a significant consequence of our modeling works in this paper, the dynamics of the inflation indeed reveals a sharp pattern reflecting on monetary policy through a mathematical transform introduced by the perception of neutral real interest rate. Here we define the neutral real interest rate as the difference

$$p_t^* = r_t^* - q_t^* = f(q_t^*) - q_t^*, \tag{21.20}$$

Figure 21.9 illustrates the process of real neutral rate of interest from 1958 to 2007. The patterns shown in the process of p_t^* could provide lots of clues to many debates over monetary policies in the past 50-plus years. To comprehend some important features in the process, first we propose the following basic hypotheses implied by the simple original modeling framework set in the Taylor Rule:

(I) 2-percent annual inflation rate is a commonly favorable control target for central bankers in normal economic conditions.

(II) If the inflation measure stays at the level of 2 percent, the real rate sufficient for keeping the inflation process at equilibrium is 2 percent.

(III) When the inflation measure deviates from the 2-percent equilibrium target level, the "neutral" real rate will incorporate an extra right effort for bringing inflation back on track while reflecting the current inflation scenario with growth prospect.
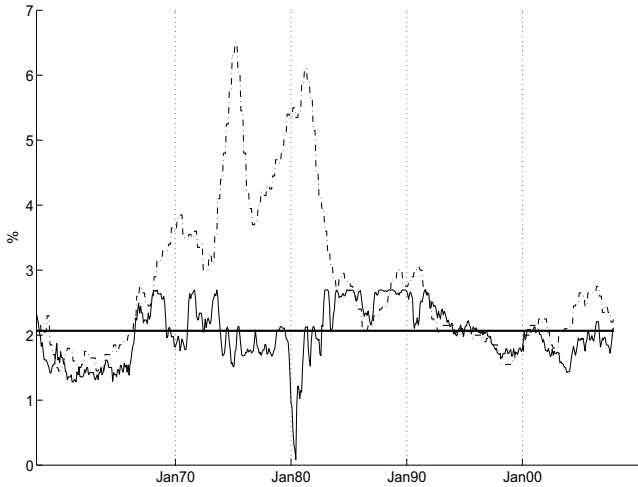
Figure 21.9. Neutral Real Interest Rate: by Bilinear Regression (bold solid), by PPR method (solid) and by Taylor Rule (dot-dash-dot). **Q** XFGneutral

Based on the above hypotheses and our interpretation about the response curve, we compare the three neutral real rate processes in Figure 21.9. In comparison with the Taylor model, the BLR takes the simplest assumption that the neutral real rate is an all-time constant of estimation value 2.067 with respect to the BLR inflation measure in 21.19. Clearly, contrary to what being suggested by the first and third hypotheses, the statistical bias of compact formulation by BLR modeling has completely ignored the inflation bias of the central banker due to inflation targeting. Thus, the general neutral real rate level is nothing more than just a number which neither guarantees stabilization of inflation (as suggested by the second hypothesis) nor provides clues for figuring out thoughts of the central bankers in face of various inflation scenarios. When we turn to the more sophisticated PPR model, Figure 21.9 shows that it provides rich clues more than just a number for understanding the central bankers mindsets on monetary policy-making. Due to the intriguing structural pattern of formulating the neutral real rate, the PPR model indeed sheds some light on Hypothesis (III) which suggests a canonical framework of setting inflation-control principal strategies toward various inflation scenarios. To fully comprehend how the PPR modeling deeply digs into Hypothesis (III), we establish a theory of co-integrating the dynamic mechanisms of interest rate and inflation in the following section.

21.4.4 General Dynamic Mechanism for Long-Run Dependence of Interest Rate and Inflation

The movements of the FFR and the CPI numbers can be regarded as the sequential outcomes of the constantly interactive game between the FOMC and the economy. The basic goal of the FOMCs game strategy can be mathematically described as the optimization problem - maximizing employment under price stability

$$\max_{I \leqq K} E, \tag{21.21}$$

where I and E stand for inflation and employment rate respectively, and K is the inflation-targeting bound of the so-called comfort zone. Then, according to the findings of the PPR modeling works, we can decompose the general strategy of the FOMC into two parts V the principal part for dealing with the primary inflation condition and the adaptive part for dealing with the secondary economic and financial conditions

$$\begin{cases} r_t^* = f(q_t^*) \\ q_t^* = h(\{r_s - r_t^*\}|_{s=t-l}^{s=t-1}) + \varepsilon^* \end{cases} \tag{21.22}$$

where f outlines the FOMC principal strategy for setting the neutral interest rate and the $r - r^*$ sequence details the FOMCs adaptive strategy Γ in tight-accommodative cycles in response to the reaction (the game-strategy) of the economy as described by h (mainly due to inflation expectation) plus a random factor, and the efforts of combining these two components are set to achieve

$$\arg\max_{I_h(f,\Gamma) \leqq K} E_h(f, \Gamma) \tag{21.23}$$

For figuring out the interactive dynamic mechanism of inflation and interest rate and constructing quantitative monetary policy-making strategy based on the interaction mechanism, the key relies on the decomposition

$$r = [r - f(q^*)] + [f(q^*) - q^*] + q^* \tag{21.24}$$

that is, the interaction of r and q^* is intermediately dominated by the principal strategy on the neutral real rate and the adaptive strategy along tight-accommodative cycles. Therefore, the canonical formalism of 21.24 induces the following five layers of problems about monetary policy-making by the above works:

(I) *Nature of Economy* – nonparametric modeling of the strategic function h of the economy via co-integrative autoregression of the three processes (see Figure 21.10) in the above decomposition.
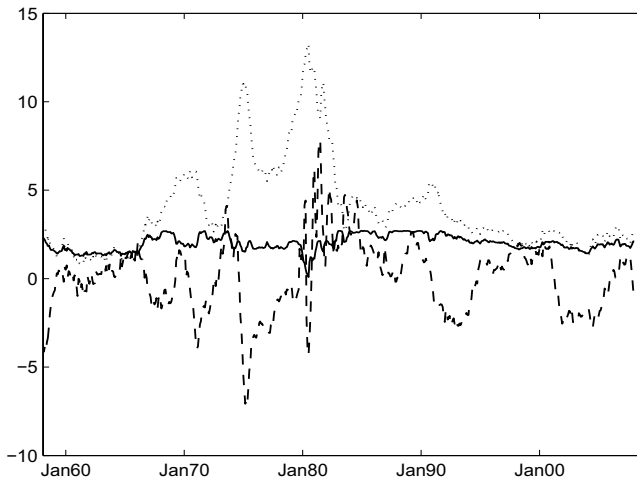
Figure 21.10. Tight-accommodative Cycles: Real Interest
Rate (dashed), Neutral Real Interest Rate (solid) and Infla-
tion Gage (dotted) ⬛XFGcycles

(II) *Goal of Monetary Policy* V set a suitable q_0^* as the upper limit of the
comfort-zone of inflation in 21.4.2, then under the assumption that inflation
normally rises as economy grows, the central banker would like to have the
neutral real rate staying at the equilibrium level $c = f(q_0^*) - q_0^*$.

(III) *Mechanism of Strategic Design* V suppose h does not depend on f,
under some regular conditions on f, there exists a stationary solution q_s, with
$q^* = q_0^*$ as the mean-reversion level, corresponding to a well-set adaptive
strategic sequence r_s for the equation $f(q_t^*) - h(\{r_s - f(q_s)\}_s) = c$.

(IV) *Interaction of Interest Rate and Inflation* V combine a model like 21.15
for setting an autoregressive mechanical rule of adaptive strategy, we can
iteratively simulate the interactive evolutions of interest rate and inflation.

(V) *Match of Principal and Adaptive* V formulate sophisticated inflation-
control functional I and economic performance measure functional E so that
a sound optimization framework 21.24 emerges to provide a good match of
the principal component and the adaptive component of a monetary policy-
making strategy for ensuring sustainable growth and price stability.

Obviously, the above problems have gone far beyond the three former hy-
potheses about thinking monetary policy. However, the complex framework
for comprehending the art and science of monetary policy-making indeed re-

veals a simple motto V better principal strategy implies more stable inflation and smaller tight-accommodative cycle.

21.5 Discussions and Conclusions

Although the PPR-oriented nonlinear modeling methodology remarkably establishes a benchmark of neutral interest rate which is beyond the scope of the Taylor Rule for providing deep principal insights into sentiments and strategies of the FOMC, it relies on many deeper technical explorations in data analysis and further understandings of the nature of nonlinear dynamic systems. The foundations are in the following list of confounded economical and statistical issues surrounding the intriguing ε in 21.10, which represents the tight-accommodative cycles along the benchmark:

- ⊡ Extension to a multi-pursuit ($M > 1$) regression over unemployment rate, exchange rates and asset price changes.

- ⊡ Autocorrelation and memory process modeling on the FOMC's self-correcting efforts.

- ⊡ Hetereoscedasticity in the variance of ε, as shown in the three different phases of the PPR modeling.

- ⊡ The co-integrated evolution of the trio ($r_t, q_{1,t}, q_{2,t}$) with a dynamically more adaptive inflation gauge.

This paper is only the beginning of a new series of research works toward building a more advanced and complete framework than the fundamental one proposed in Section 21.3.3. Unfortunately, the statistical modeling challenge to solve these problems is ar beyond the reach of the PPR modeling based on cubic splines.

Note that, due to the fact of long-range dependency in ε within the regression framework in 21.10, the PPR estimation accuracy of f can be seriously questionable (see Opsomer, Wang and Yang (2001)). However, while the request of statistical rigor needs more efforts, the practical interpretations seems giving good credits to the PPR modeling results as the mean-reversion nature might well offset the estimation bias. Limitation of smoothing technique and correlation structure of uncertainty are the two major factors for affecting estimation accuracy on pin-pointing the singular structure of the response curve which corresponds to the reaction of the FOMC to critical times of the inflation process.

In the near future, we can employ the advanced wavelet-based PPR modeling methodology to crack these core problems of interest-rate dynamics. Although, in this paper, the fundamental PPR modeling framework is surely not perfect due to the problems presented above, it indeed heuristically demonstrates the need and power of functional nonparametric modeling. Without the PPR modeling, we would have not found out the systematical risk of overestimating the FFR by the Taylor Rule and thus miss the phenomenon of cyclic mean-reversion along the dynamic equilibrium level of FFR driven by a fair inflation gauge via the response mechanism. Moreover, without entering the world of functional nonparametric modeling, the 3-phase response curve structure, which is the key to comprehend the sentiments and strategies of central bankers toward balancing growth and inflation, could have never been systematically discovered.

Bibliography

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995). Wavelet Shrinkage: Asyptotia?, *Journal of Royal Statistical Society Series B* **31**(2): 301-369.

Federal Reserve Bank (2005). *FOMC Statement 11/01/2005*, http://www.federalreserve.gov /boarddocs/press/monetary/2005/20051101/default.htm.

Friedman, J. H. and Stuetzle, W. (1981). Projection Pursuit Regression, *Journal of the American Statistical Association*, **76**(376): 817-823.

Further, J.C. (2000). Habit formation in consumption and its implication for monetary-policy models, *American Economic Review*, **90** (3): 367-390.

Giannoni, M.P. (2001). Robust optimal monetary policy in a forward-looking model with parameter and shock uncertainty,Manuscript, Princeton University.

Giannoni, M.P. (2002). Does model uncertainty justify caution? Robust optimal monetary policy in a forward-looking model, *Macroeconomic Dynamics*, **6**(1): 111-144.

Giannoni,M.P. and Woodford, M. (2002) Optimal interest-rate rules: I. General theory, Manuscript, Princeton University.

Green, P.J. and Silverman B.W. (1994) *Nonparametric regression and generalized linear models- a roughness penalty approach*, Chapman and Hall.

Greenspan, A. (1997)Rules vs. discretionary monetary policy, Remarks at the 15th Anniversary Conference of the Center for Economic Policy Research at Standard University, Sep. 5.

Hansen, L.P. and Sargent, T.J. (2002)Robust control and model uncertainty in economics, New York University.

Hull, J. and White, A. (1990)Pricing Interest-rate-derivatives securities, *Review of Financial Studies*, **3** (4) : 573-592.

Jeng, Jenher (2002) Wavelet Methodology for Advanced Nonparametric Curve Estimation: from Confidence Band to Sharp Adaptation, PhD Thesis, Statistics, University of California at Berkeley

Judd, John P. and Rudebusch, Gleen D. (1998) Taylor's Rule and the Fed: 1970-1997, *FRBSC Economic Review*, **3**: 3-16.

Levin, A. T. and Williams, J. C. (2003) Robust monetary policy with competing reference models, *Journal of Monetary Economics*, **50** (5): 945-975.

Orphanides, A. (2003) Historical monetary policy analysis and the Taylor Rule, *Journal of Monetary Economics*, **50** (5): 983-1022.

Rudebusch, G. D. and Svensson, L.E.O. (1999) Policy rules for inflation targeting, *Monetary Policy Rules* (ed. By Taylor, J.B.), 203-253, University of Chicago.

Taylor, John B. (1993) Discretion versus policy rules in practice, *Carnegie-Rochester Conference Series on Public Policy*, **39**: 195-214.

Opsomer, J., Wang, Y. and Yang, Y.(2001) Nonparametric regression with correlated errors, *Statistical Science*, **16** (2): 134-153.

Vasicek, O. E. (1977) An equilibrium characterization of the term structure, *Journal of Financial Economics*, **5**: 177-188.

Index