12

Nonlinear Time Series Analysis Based on Markov Switching Models

12.1 Introduction

In practical time series analysis, an important aspect is properties of the marginal distribution of Y_t as well as properties of the one-step ahead predictive density $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$, implied by the chosen time series model. Typical stylized facts of the marginal distribution of practical time series are asymmetry and nonnormality with rather fat tails, and autocorrelation not only in the level Y_t, but also in the squared process Y_t^2. Properties of the predictive distribution are nonlinear effects of past observation on the mean and conditional heteroscedasticity.

It is well known that standard ARMA models (Box and Jenkins, 1970) often are not able to capture stylized facts of practical time series. Some unrealistic features of ARMA models based on normal errors are normality of the predictive as well as the marginal density, linearity of the expectation $\mathrm{E}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ in the past observation $y_1, \ldots, y_{t-1}$, and homoscedasticity of $\mathrm{Var}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ (Brockwell and Davis, 1991; Hamilton, 1994b). Numerous nonlinear time series models such as GARCH models, threshold autoregressive models, and many others have been designed to reproduce empirical features of practical time series (Tong, 1990; Granger and Teräsvirta, 1993; Franses and van Dijk, 2000).

This chapter discusses Markov switching models that constitute another very flexible class of nonlinear time series models and are able to capture many features of practical time series by appropriate modifications of the basic Markov switching model introduced in Subsection 10.3.1. Section 12.2 deals with the Markov switching autoregressive model and Section 12.3 considers the related Markov switching dynamic regression model. Section 12.4 shows that Markov switching models give rise to very flexible predictive distributions. Section 12.5 deals with Markov switching conditional heteroscedasticity and switching ARCH models are introduced. Section 12.6 studies further extensions, namely hidden Markov chains with time-varying transition probabilities and hidden Markov models for longitudinal data and multivariate time series.

12.2 The Markov Switching Autoregressive Model

It has been discussed in Subsection 10.2.4 that a Markov mixture model introduces autocorrelation in the process Y_t even for the basic Markov switching model, where conditionally on knowing the states, the process Y_t is uncorrelated. In this section the Markov switching autoregressive model is introduced that deals with autocorrelation in a more flexible way than the basic Markov switching model.

12.2.1 Motivating Example

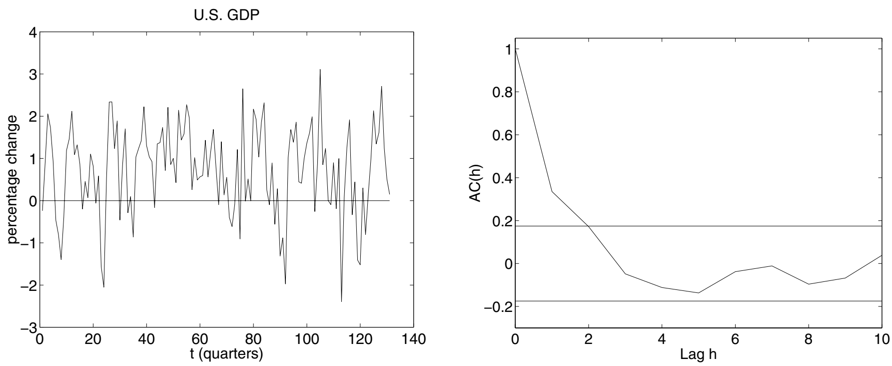

Fig. 12.1. GDP DATA, quarterly data 1951.II to 1984.IV, left: time series plot of y_t in comparison to level 0, right: empirical autocorrelogram

A standard time series that has been analyzed in numerous papers is the percentage growth rate of the U.S. quarterly real GDP series:

$$Y_t = 100(\log(\mathrm{GDP}_t) - \log(\mathrm{GDP}_{t-1})), \tag{12.1}$$

$t = 1, \dots, T$. Figure 12.1 shows a time series plot of the data for the period 1951.II to 1984.IV, together with empirical autocorrelation. First, we fit various AR(p) models to these data to capture autocorrelation in this time series. Figure 12.2, comparing the unconditional distribution of Y_t, implied by each of the fitted AR(p) models with the empirical histogram of y_t, reveals a striking difference between the empirical histogram which evidently shows bimodality, and any of the implied marginal distributions which are unimodal and, by the way, show surprisingly little difference for the different model orders.

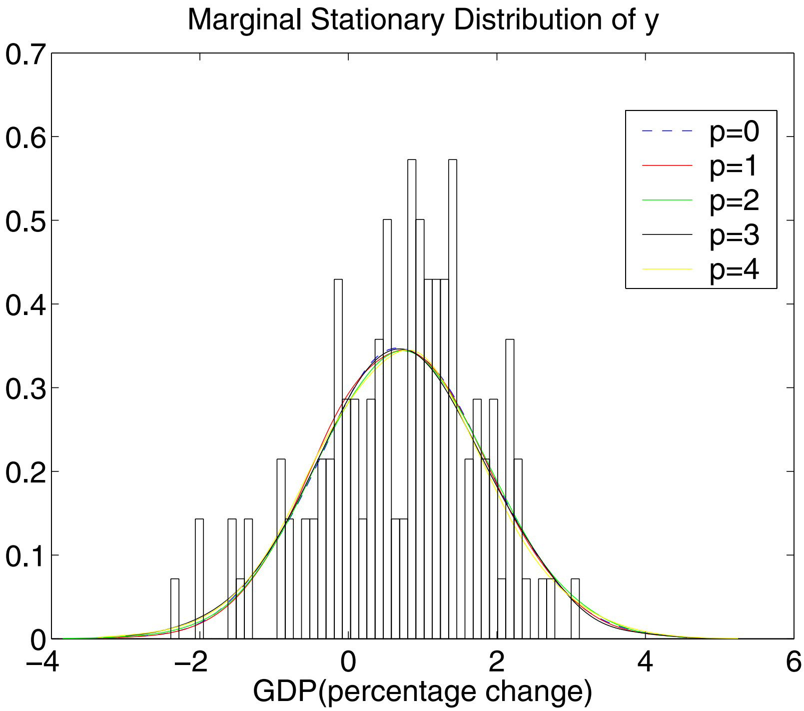

Fig. 12.2. GDP DATA, modeled by an AR(p) model with $p = 0, 1, \ldots, 4$; implied unconditional distribution of Y_t (full line) in comparison to the empirical marginal distribution of Y_t (histogram)

From where does this bimodality in the empirical time series come? Figure 12.1 displays the growth rate of the U.S. GDP series in comparison to the zero line. Evidently periods of positive growth rate, where $y_t > 0$, are followed by periods of negative growth rate, where $y_t < 0$. What we find here is known by economists as the business cycle. Macro-economic variables such as the GDP are influenced by the state of the economy and follow different processes, depending on whether the economy is in a boom or in a recession. Figure 12.1 suggests that the marginal distribution of Y_t is a mixture distribution with different means and possibly different variances. If we fitted a standard mixture of two normal distributions, the implied marginal distribution of Y_t would be in fact multimodal, but marginally Y_t would be a process that is uncorrelated over time. To capture both multimodality and autocorrelation for such time series, Hamilton (1989) introduced the Markov switching autoregressive model.

12.2.2 Model Definition

The standard model to capture autocorrelation is the AR(p) model,

$$Y_t - \mu = \delta_1(Y_{t-1} - \mu) + \cdots + \delta_p(Y_{t-p} - \mu) + \varepsilon_t, \tag{12.2}$$

where $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, which is equivalent to model

$$Y_t = \delta_1 Y_{t-1} + \cdots + \delta_p Y_{t-p} + \zeta + \varepsilon_t, \tag{12.3}$$

with $\zeta = \mu(1 - \delta_1 - \cdots - \delta_p)$.

An important extension of the basic Markov switching model is the Markov switching autoregressive (MSAR) model, where a hidden Markov chain is introduced into model (12.2). This model was used independently in the work of Neftçi (1984) and Sclove (1983), and became popular in econometrics for analyzing economic time series such as the GDP data introduced in Subsection 12.2.1 through the work of Hamilton (1989) who allowed for a random shift in the mean level of process (12.2) through a two-state hidden Markov chain:

$$Y_t - \mu_{S_t} = \delta_1(Y_{t-1} - \mu_{S_{t-1}}) + \cdots + \delta_p(Y_{t-p} - \mu_{S_{t-p}}) + \varepsilon_t. \tag{12.4}$$

An important alternative to model (12.4) was suggested by McCulloch and Tsay (1994b), who introduced the hidden Markov chain into (12.3) rather than into (12.2), by assuming that the intercept is driven by the hidden Markov chain rather than the mean level:

$$Y_t = \delta_1 Y_{t-1} + \cdots + \delta_p Y_{t-p} + \zeta_{S_t} + \varepsilon_t. \tag{12.5}$$

Although the parameterization (12.2) and (12.3) are equivalent for the standard AR model, a model with a Markov switching intercept turns out to be different from a model with a Markov switching mean level. In (12.4), after a one-time change from S_{t-1} to $S_t \neq S_{t-1}$, an immediate mean level shift from $\mu_{S_{t-1}}$ to μ_{S_t} occurs. In (12.5), however, the mean level approaches the new value smoothly over several time periods.

Both models violate assumption **Y4** as the one-step ahead predictive density $p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \boldsymbol{\vartheta})$ depends on past values $\mathbf{y}^{t-1}$. For a model with switching mean level it is evident from (12.4) that the predictive density $p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \boldsymbol{\vartheta})$ depends not only on S_t, but also on the past values $S_{t-1}, \ldots, S_{t-p}$ of the hidden Markov chain fulfilling only assumption **Y2** stated in Subsection 10.3.4. On the other hand for a model with switching intercept the predictive density $p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \boldsymbol{\vartheta})$ depends only on S_t and such a process fulfills the stronger condition **Y3**. As discussed in Subsection 11.2.5, condition **Y3** essentially influences the complexity of econometric inference about the hidden Markov chain S_t. As a result, econometric inference for an MSAR model with switching intercept is not more complicated than for the basic Markov switching model, whereas for an MSAR model with switching mean inference on the hidden Markov chain S_t is far more involved.

In its most general form the MSAR model allows that the autoregressive coefficients are also affected by S_t (Sclove, 1983; Holst et al., 1994; McCulloch and Tsay, 1994b):

$$Y_t = \delta_{S_t,1} Y_{t-1} + \cdots + \delta_{S_t,p} Y_{t-p} + \zeta_{S_t} + \varepsilon_t. \tag{12.6}$$

The assumption that the autoregressive parameters switch between the two states implies different dynamic patterns in the various states, and introduces asymmetry over time. Asymmetry over time between the states is introduced also through the hidden Markov chain as different persistence probabilities imply different state duration; see (10.13). This combined asymmetry leads to a rather flexible model that is able to capture asymmetric patterns observed in economics time series, such as the fast rise and the slow decay in the U.S. quarterly unemployment rate.

In any of these models the variance may be assumed to be constant, irrespective of the state of S_t, or it is possible to assume a shift in the variance, $\varepsilon_t \sim \mathcal{N}\left(0, \sigma^2_{\varepsilon,S_t}\right)$.

Subsequently the notation MS(K)-AR(p) is used occasionally to denote a Markov switching autoregressive model with K states and autoregressive order p. A more subtle notation that also differentiates between homo- and heteroscedastic variances, switching in the mean level or in the intercept as well as between invariant and switching autoregressive parameters is introduced in Krolzig (1997).

Related Models

The mixture autoregressive model (Juang and Rabiner, 1985; Wong and Li, 2000) defines the one-step ahead predictive $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ directly as a mixture of normal distributions with an AR structure in the mean:

$$p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}) = \sum_{k=1}^{K} \eta_k f_N(y_t; \mu_{k,t}, \sigma^2_k), \tag{12.7}$$

where $\mu_{k,t} = \mathrm{E}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}_k) = \delta_{k,1} y_{t-1} + \cdots + \delta_{k,p} y_{t-p} + \zeta_k$. This model results as that special of an MSAR model, where S_t is an i.i.d. process, with each row of the transition matrix $\boldsymbol{\xi}$ being equal to the weight distribution in (12.7). Because autocorrelation in Y_t is introduced only through the observation equation this model is not able to capture spurious autocorrelation that disappears once we condition on the state of S_t.

MSAR models are related to the self-exciting threshold autoregressive (SETAR) models (Jalali and Pemberton, 1995; Clements and Krolzig, 1998) which are themselves that special case of a threshold autoregressive (TAR) model (Tong, 1990), where the mean and the autoregressive parameters switch according to the level of the threshold variable $z_t = Y_{t-d}$:

$$Y_t = \begin{cases} \delta_{1,1} Y_{t-1} + \cdots + \delta_{1,p} Y_{t-p} + \zeta_1 + \varepsilon_t, & Y_{t-d} \leq r, \\ \delta_{2,1} Y_{t-1} + \cdots + \delta_{2,p} Y_{t-p} + \zeta_2 + \varepsilon_t, & Y_{t-d} > r, \end{cases}$$

with $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$. Consider, for instance, a first-order SETAR model, where $p = 1$ and $d = 1$ and define an indicator S_t such that

$$S_t = \begin{cases} 1, & Y_{t-1} \leq r, \\ 2, & Y_{t-1} > r. \end{cases}$$

Then S_t follows a first-order Markov process with transition matrix $\boldsymbol{\xi}$ given by

$$\boldsymbol{\xi} = \begin{pmatrix} \Phi(r_1) & 1 - \Phi(r_1) \\ \Phi(r_2) & 1 - \Phi(r_2) \end{pmatrix},$$

with $\Phi(\cdot)$ being the standard normal distribution, and $r_k = (r - \mu_k)/\sigma_\varepsilon$. Therefore the first-order SETAR model with $d = 1$ corresponds to a two-state Markov switching autoregressive model with a restricted transition matrix, which has a single free parameter, once μ_1, μ_2, and σ_ε^2 are known.

12.2.3 Features of the MSAR Model

The Markov switching autoregressive model is a special case of a dynamic stochastic system with stochastic autoregressive parameters for which it is not straightforward to find conditions under which the process Y_t is strictly stationary and certain moments exist (Tjøstheim, 1986; Karlsen, 1990; Bougerol and Picard, 1992b). Results on the stationarity of Markov switching autoregressive models can be found in Holst et al. (1994), Krolzig (1997), Yao and Attali (2000), and Francq and Zakoian (2001). Timmermann (2000) illustrates how the variance and higher-order moments of a process generated by an MSAR model may be computed explicitly provided the process is stationary.

The Markov switching autoregressive model introduces autocorrelation both through the hidden Markov chain as well as through the observation equation, leading to rather flexible autocorrelation structures. The autocorrelation function may be computed explicitly provided that the process is second-order stationary (Timmermann, 2000, Proposition 4). For an MS(2)-AR(1) model with switching mean, fixed variance, and fixed AR coefficients, for instance, the autocorrelation function of Y_t reads:

$$\rho_{Y_t}(h|\boldsymbol{\vartheta}) = \frac{1}{\mathrm{Var}(Y_t|\boldsymbol{\vartheta})} \left(\lambda^h (\mu_1 - \mu_2)^2 \eta_1 \eta_2 + \delta_1^h \frac{\sigma_\varepsilon^2}{1 - \delta_1^2} \right), \tag{12.8}$$

with $\lambda = \xi_{11} - \xi_{21}$ being the second eigenvalue of the transition matrix $\boldsymbol{\xi}$ and the unconditional variance $\mathrm{Var}(Y_t|\boldsymbol{\vartheta})$ being equal to

$$\mathrm{Var}(Y_t|\boldsymbol{\vartheta}) = (\mu_1 - \mu_2)^2 \eta_1 \eta_2 + \frac{\sigma_\varepsilon^2}{1 - \delta_1^2}.$$

The autocorrelation function fulfills, for $h > 2$, the following recursion,

$$\rho_{Y_t}(h|\boldsymbol{\vartheta}) = (\delta_1 + \lambda)\rho_{Y_t}(h-1|\boldsymbol{\vartheta}) - \delta_1 \lambda \rho_{Y_t}(h-2|\boldsymbol{\vartheta}), \tag{12.9}$$

and corresponds to the autocorrelation function of an ARMA$(2,1)$ model, but has a nonnormal unconditional distribution.

Krolzig (1997) derived general results on the relation between Markov switching autoregressive models and nonnormal ARMA models. A process generated by an MS(K)-AR(p) model with switching intercept, but fixed variances and AR coefficients, for instance, possesses an ARMA$(K+p-1, K-1)$ representation (Krolzig, 1997, Proposition 3), whereas an ARMA$(K+p-1, K+p-2)$ representation results, if a switching mean is considered, rather than a switching intercept (Krolzig, 1997, Proposition 4).

12.2.4 Markov Switching Models for Nonstationary Time Series

The work of Nelson and Plosser (1982) started a discussion in econometrics, as to whether macro-economic time series contain a deterministic or a stochastic trend, the latter typically being a unit root in the autoregressive representation of the time series. This is tested by applying a unit root test to Y_t which often leads to nonrejection of the unit root null hypothesis. Perron (1989, 1990) found evidence for spurious unit roots in real interest rates under structural breaks in the trend level and the growth rate.

Markov switching models are to a certain degree able to deal with spurious unit roots caused by structural breaks. To illustrate this point consider a process Y_t, generated by a two-state Markov mixture of normal distributions with $\mu_2 \neq \mu_1$ and a highly persistent transition matrix where ξ_{11} and ξ_{22} are close to one, pushing the second eigenvalue $\lambda = \xi_{11} - \xi_{21}$ toward 1. It is evident from the autocorrelation function of Y_t, derived in (10.20), that high autocorrelation in the marginal process Y_t is present, although there exists no autocorrelation within the two regimes. Furthermore the autocorrelation increases as the size $|\mu_2 - \mu_1|$ of the shift in the mean increases. This may lead to detecting a spurious unit root because a unit root test applied to Y_t is biased toward nonrejection of the unit root hypothesis under a sudden change in the mean with increasing rate of nonrejection as the size $|\mu_2 - \mu_1|$ of the break increases. Garcia and Perron (1996), by modeling interest rates by a three-state MSAR model with state-invariant autocorrelation and heteroscedastic variances, show that the autocorrelation actually nearly disappears in the various regimes.

This raises the question as to whether a Markov switching model should be applied to the level or to the growth rate of a nonstationary time series. Hamilton (1989), following the standard ARIMA modeling approach, which is based on autoregressive modeling of the growth rate, applied the MSAR model to the growth rate of a nonstationary time series such as the GDP. In terms of the (log) level Y_t the model reads:

$$\begin{aligned}
Y_t &= \mu_t + Z_t, \\
\mu_t &= \mu_{t-1} + \zeta_{S_t}, \\
\boldsymbol{\delta}(L)\triangle Z_t &= \boldsymbol{\delta}(L)(1-L)Z_t = \varepsilon_t,
\end{aligned} \tag{12.10}$$

where L is the lag operator, $\boldsymbol{\delta}(L) = 1-\delta_1 L - \cdots - \delta_p L^p$ and all roots of $\boldsymbol{\delta}(L)$ lie outside the unit circle. This model is also called the Markov switching trend model, because the untransformed time series Y_t has a stochastic trend with a drift that is switching according to a hidden Markov chain.

Specification (12.10) assumes that Y_t has a unit root, however, as noted by Lam (1990), the results of Perron (1989, 1990) suggest that the unit root in Y_t disappears once occasional shifts in the deterministic trend are allowed for. Lam (1990) assumes that Y_t is trend stationary around a Markov switching trend:

$$\begin{aligned}
Y_t &= \mu_t + Z_t, \\
\mu_t &= \mu_{t-1} + \zeta_{S_t}, \\
\boldsymbol{\delta}(L) Z_t &= \varepsilon_t,
\end{aligned} \tag{12.11}$$

where all roots of $\boldsymbol{\delta}(L)$ lie outside the unit circle. In this model the predictive density $p(y_t|\mathbf{S}, \mathbf{y}^{t-1})$ depends on the whole history of S_t (assumption **Y1**) and estimation has to be carried within the framework of switching state space models; see Chapter 13.

As a compromise between these two models, Hall et al. (1999) consider a model based on the Dickey–Fuller regression (Dickey and Fuller, 1981) and allow for regression parameter switching according to a two-state hidden Markov chain:

$$\triangle Y_t = \zeta_{S_t} + \psi_{S_t} Y_{t-1} + \sum_{j=1}^{p} \delta_{S_t,j} \triangle Y_{t-j} + \varepsilon_t. \tag{12.12}$$

In (12.12), Y_t is the (log) level of the observed process, whereas $\triangle Y_t = Y_t - Y_{t-1}$ is the growth rate. If $\psi_1 = \psi_2 = 0$ in both regimes then a unit root is present in Y_t, and the Markov switching trend model of Hamilton (1989) results. On the other hand, if $\psi_1 \neq 0$ and $\psi_2 \neq 0$, then Y_t is stationary around a trend with Markov switching slope, leading to the model of Lam (1990).

Model (12.12) allows that Y_t has a unit root in one state ($\psi_1 = 0$), whereas Y_t is stationary in the other state ($\psi_2 \neq 0$). This model has been found useful in applied time series analysis, for instance, in economics for modeling the GDP (McCulloch and Tsay, 1994a), in finance for modeling interest rates (Ang and Bekaert, 2002), as well as in geophysics (Karlsen and Tjøstheim, 1990).

Several authors investigate the power of unit root tests when the data arise from particular Markov switching alternatives (Nelson et al., 2001; Psaradakis, 2001, 2002).

12.2.5 Parameter Estimation and Model Selection

ML estimation is usually carried out through the EM algorithm (Hamilton, 1990; Holst et al., 1994). Asymptotic properties of the ML estimator for MSAR models are established in Francq and Roussignol (1998), Krishnamurthy and Rydén (1998), and Douc et al. (2004).

Bayesian estimation of the MSAR model relies on data augmentation and MCMC (Albert and Chib, 1993; McCulloch and Tsay, 1994b; Chib, 1996; Frühwirth-Schnatter, 2001b). For an MSAR model where all coefficients, including the intercept and the variance, are switching, MCMC estimation is carried out along the lines indicated in *Algorithm 11.3*, with step (a2) being the only model-specific part. Sampling the model parameters $\boldsymbol{\vartheta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K, \sigma^2_{\varepsilon,1}, \dots, \sigma^2_{\varepsilon,K})$, with $\boldsymbol{\beta}_k = (\delta_{k,1}, \dots, \delta_{k,p}, \zeta_k)$, in combination with the conjugate priors

$$\boldsymbol{\beta}_k \sim \mathcal{N}_{p+1}(\mathbf{b}_0, \mathbf{B}_0), \qquad \sigma^2_{\varepsilon,k} \sim \mathcal{G}(c_0, C_0),$$

is closely related to sampling these parameters for a finite mixture regression model as in steps (a2) and (a3) of *Algorithm 8.1*. An MSAR model, where only some parameters are switching, may be considered as a special case of a Markov switching dynamic regression model, which is introduced in Section 12.3, where Bayesian estimation is discussed in Subsection 12.3.2.

The presence of the lagged values $y_{t-1}, \dots, y_{t-p}$, however, causes certain technical problems that are avoided if inference is carried out conditional on the first p values. For an unconditional analysis as in Albert and Chib (1993), the first p values are considered to be random draws from the stationary distribution $p(y_1, \dots, y_p|\boldsymbol{\vartheta})$. An undesirable effect of an unconditional analysis is that the posterior of $\boldsymbol{\vartheta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K, \sigma^2_{\varepsilon,1}, \dots, \sigma^2_{\varepsilon,K})$ no longer has a standard form, as the stationary distribution depends on these parameters in a nonconjugate manner. Albert and Chib (1993) suggest using rejection sampling to sample from this posterior.

The most commonly occurring model selection problems for MSAR models is selecting the number of states of the hidden Markov chains well as order selection. Frühwirth-Schnatter (2004) shows that it is important to consider these model selection problems jointly in order to avoid underfitting the number of states while overfitting the AR order; see also Subsection 12.2.6.

12.2.6 Application to Business Cycle Analysis of the U.S. GDP Data

The motivating example studied in Subsection 12.2.1 demonstrated one of the key features of the business cycle, namely that periods of expansion and contraction are quite different. Whereas in expansion periods the output growth rate is high and the economy is booming, growth rates are typically negative in contraction periods, where the economy is in a recession. An important feature of macro-economic time series such as the GDP or industrial production is persistence of the respective states. Once the economy is in a certain state it tends to remain there for more than one period. Furthermore there is some asymmetry in this persistency, as longer periods of positive growth rates are followed by shorter periods of negative growth rates. This asymmetry over the business cycle has been captured using a basic Markov switching model for unemployment rates (Neftçi, 1984) and GDP, investment, and productivity (Falk, 1986).

Markov switching autoregressive models, often also called regime switching models by economists, became extremely popular in business cycle analysis since Hamilton's (1989) paper, and further applications include Goodwin (1993), Sichel (1994), Clements and Krolzig (1998), and Kaufmann (2000), among many others. For a theoretical justification of why Markov switching might be sensible models for the economy we refer to Hamilton and Raj (2002) and Raj (2002) and the references therein.

Model Selection for the GDP Data

We return to modeling the U.S. quarterly GDP series introduced in Subsection 12.2.1 within the framework of MSAR models, by comparing different Markov switching models. The first model is the K-state MSAR model with switching intercept, but state-independent AR parameters and state-independent variances, defined in (12.5), which has been applied by Chib (1996). The second is the K-state MSAR model with switching intercept, switching AR parameters, and switching error variance ("totally switching"), defined in (12.6) which was applied by McCulloch and Tsay (1994b). The priors are selected to be rather vague and state-independent. We assume no prior correlation among the regression parameters. The prior on the switching intercept is $\mathcal{N}(0,1)$; the prior both on switching and state-independent AR parameters is $\mathcal{N}(0,0.25)$. The prior both on switching and state-independent variances is $\mathcal{G}^{-1}(2,0.5)$. The prior on the rows $\boldsymbol{\xi}_{k\cdot}$ of the transition matrix is, for all k, $\mathcal{D}(e_{k1},\dots,e_{kK})$ with $e_{kk}=2$ and $e_{kk'}=1/(K-1)$, if $k \neq k'$.

We compare the Markov switching models (12.5) and (12.6), where K is equal to 2 or 3, with the classical AR(p) model, which corresponds to $K=1$, using marginal likelihoods. We assume that p varies between 0 and 4, leading to a total of 25 different models. The marginal likelihoods are estimated from the MCMC output of a random permutation sampler ($M=6000$ after a burn-in phase of 1000 simulations) using the "optimal" bridge sampling estimator described in Subsection 5.4.6, where the construction of the importance density $q(\boldsymbol{\vartheta})$ according to (11.29) is based on $S=100\cdot K!$ components.

From Table 12.1, reporting the log of the estimated marginal likelihoods, we find that the two-state totally switching MSAR model of order $p=2$ has the highest marginal likelihood. This result is interesting for various reasons: first, we were able to produce evidence in favor of Markov switching heterogeneity from univariate time series observations of the GDP alone, without the need to include other time series as in Kim and Nelson (2001). Second, the

Table 12.1. GDP DATA, modeled by an AR(p) model ($\mathcal{M}_1$) and different Markov switching models ($\mathcal{M}_2$ ... switching intercept, $\mathcal{M}_3$ ... totally switching) with different order p and different number of states K; log of marginal likelihoods $\log p(\mathbf{y}|\mathcal{M}_j, K, p)$ (from Frühwirth-Schnatter (2004) with permission granted by the Royal Economic Society)

	$\mathcal{M}_1$	$\mathcal{M}_2$		$\mathcal{M}_3$	
p	$K = 1$	$K = 2$	$K = 3$	$K = 2$	$K = 3$
0	–199.71	–193.54	–192.25	–194.25	–193.10
1	–194.22	–192.54	–192.75	–193.58	–194.71
2	–196.30	–194.15	–194.38	**–191.62**	–194.33
3	–197.26	–194.59	–194.74	–193.67	–196.78
4	–199.18	–195.70	–195.72	–195.34	–199.88

evidence in favor of the hypothesis that the dynamic pattern of the economy is different between contraction and expansion periods confirms the empirical results of McCulloch and Tsay (1994b).

Testing for Markov switching heterogeneity is highly influenced by selecting the appropriate model order. If we compare in Table 12.1 the AR(1) model, which has highest marginal likelihood among all AR(p) models considered, with a two-state totally switching model of order four, which is the model considered by McCulloch and Tsay (1994b), we end up with evidence in favor of *no* Markov switching heterogeneity. For a two-state totally switching MSAR model, however, the optimal model order is $p = 2$ rather than $p = 4$. Only if we compare the AR(1) model with a two-state switching model with p close to the optimal order, will we end up with evidence *in favor of* Markov switching heterogeneity. These results indicate the importance of simultaneously testing for Markov switching heterogeneity and selecting the appropriate model order and might explain why other studies, reviewed in Kim and Nelson (2001), have produced somewhat conflicting evidence concerning the presence or absence of Markov switching heterogeneity in this time series.

Exploratory Bayesian Analysis

A number of exploratory cues with regard to model selection are available from the point process representations of the MCMC output of the various models. We start with the point process representations of various bivariate marginal distributions for the three-state totally switching MSAR model of order four. Although we allowed for three states, the scatter plots in Figure 12.3 indicate that a model with three states is overfitting. If we compare this figure with the simulations of a two-state totally switching MSAR model of order four in Figure 12.4, we obtain a similar picture, with fuzziness being reduced due to the smaller number of parameters; nevertheless the two states are not very clearly separated. The bivariate marginal density of the autoregressive parameters $\delta_{k,3}$ and $\delta_{k,4}$ clusters around 0 for all states, suggesting reducing

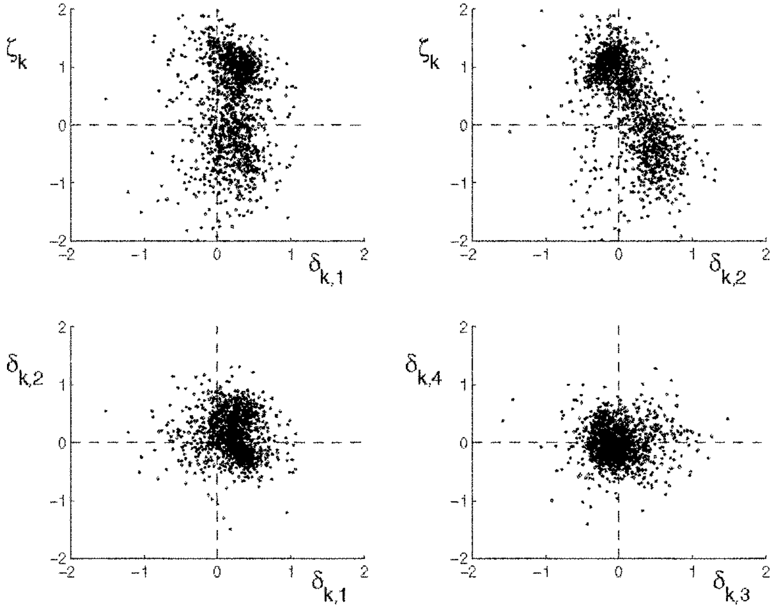

Fig. 12.3. GDP DATA, totally Markov switching model with $K = 3$ and $p = 4$; MCMC simulations from various bivariate marginal densities obtained from random permutation sampling

the model order p to 2. The point process representations MCMC simulations for the two-state totally switching MSAR model of order 2 in Figure 12.5 show a much clearer picture. As $\delta_{k,2}$ has two simulation clusters, one of which is shifted away from 0, there is no exploratory evidence that we should reduce the model order further. Furthermore the two simulation clusters provide evidence in favor of a totally switching rather than a switching intercept MSAR model.

On the whole, exploratory Bayesian analysis using projections of the point process representations of the MCMC draws supports the findings from formal model selection using marginal likelihood.

Parameter Estimation for the "Best" Model

To identify the two-state totally switching MSAR model of order two, we use the identifiability constraint $\zeta_1 < \zeta_2$, as the growth rate in the two states is expected to be different. This choice is supported by point process representation in Figure 12.5, showing that the simulations of ζ_k cluster around two points, one having an intercept bigger, the other having an intercept smaller than zero.

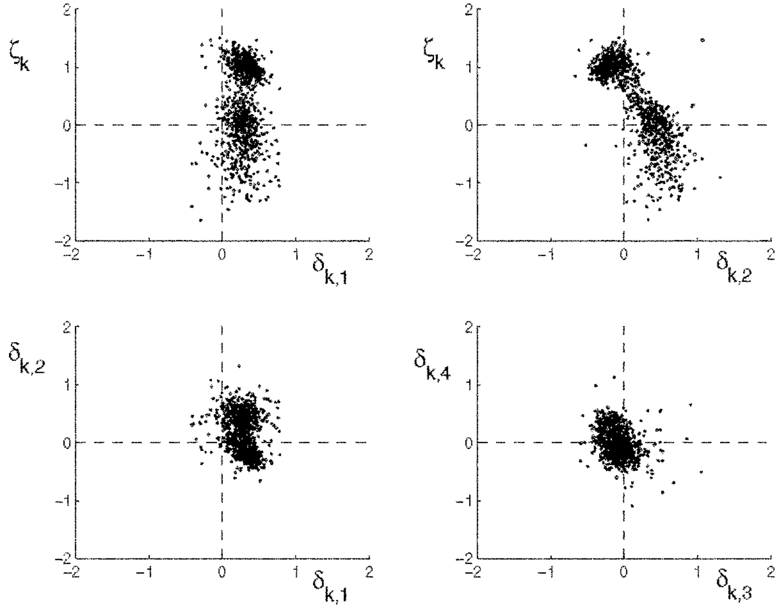

Fig. 12.4. GDP Data, totally Markov switching model with $K = 2$ and $p = 4$; MCMC simulations from various bivariate marginal densities obtained from random permutation sampling

To produce simulations under the identifiability constraint we apply the permutation sampler by reordering the MCMC output according to the constraint $\zeta_1 < \zeta_2$. If the constraint is violated for any MCMC draw, $\zeta_1^{(m)} > \zeta_2^{(m)}$, we permute the labels of all state-dependent parameters with $\rho(1) = 2$ and $\rho(2) = 1$. This is the basic idea behind permutation sampling under an identifiability constraint. It has been proven in Frühwirth-Schnatter (2001b) that due to the invariance of the posterior distribution to relabeling the states, this is a valid strategy to produce a sample from the constrained Markov mixture posterior distribution.

The resulting parameter estimates are summarized in Table 12.2. Positive growth in expansion is followed by negative growth in contraction. The dynamic behavior of the U.S. GDP growth rate is different between contraction and expansion with reaction to a percentage change of the GDP growth being faster in expansion than in contraction. The expected duration of expansion is longer than that of contraction.

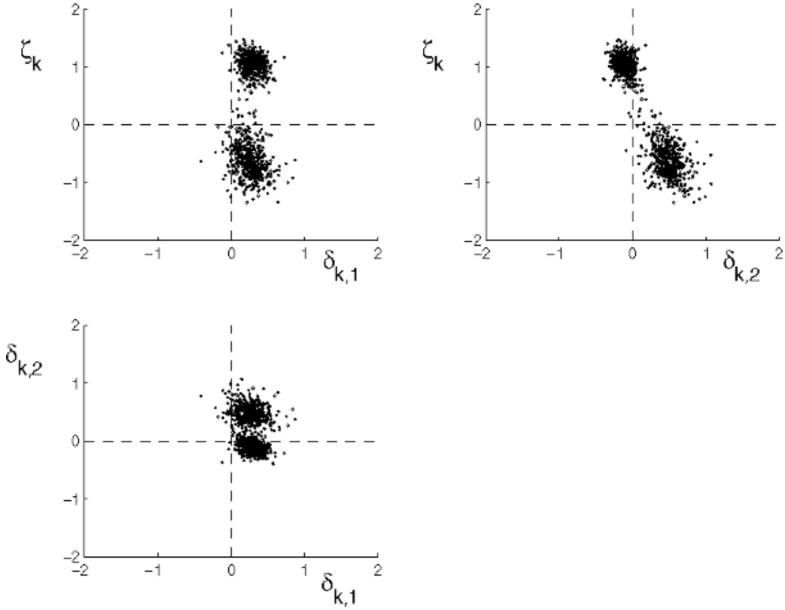

Fig. 12.5. GDP DATA, totally Markov switching model with $K = 2$ and $p = 2$; MCMC simulations from various bivariate marginal densities obtained from random permutation sampling (from Frühwirth-Schnatter (2001b) with permission granted by the American Statistical Association)

Table 12.2. GDP DATA, totally Markov switching model with $K = 2$ and $p = 2$, identified through $\zeta_1 < \zeta_2$; parameters estimated by posterior means; standard errors given by posterior standard deviations in parentheses

Parameter	Contraction ($k = 1$)	Expansion ($k = 2$)
$\delta_{k,1}$	0.249 (0.164)	0.295 (0.116)
$\delta_{k,2}$	0.462 (0.164)	–0.114 (0.098)
ζ_k	–0.557 (0.322)	1.060 (0.175)
$\sigma_{\varepsilon,k}$	0.768 (0.161)	0.692 (0.115)
$\xi_{kk'}$	0.489 (0.165)	0.337 (0.145)

12.3 Markov Switching Dynamic Regression Models

An important extension both of Markov switching autoregressive models and the Markov switching regression model, discussed in Subsection 10.3.2, is the Markov switching dynamic regression model.

12.3.1 Model Definition

The MSAR model (12.4) has been extended in the following way to deal with the presence of exogenous variables $\boldsymbol{z}_t = \begin{pmatrix} z_{t1} & \cdots & z_{td} \end{pmatrix}$ (Cosslett and Lee, 1985; Albert and Chib, 1993),

$$
\begin{aligned}
Y_t - \mu_{S_t} - \boldsymbol{z}_t\boldsymbol{\beta} = \delta_1(Y_{t-1} - \mu_{S_{t-1}} - \boldsymbol{z}_{t-1}\boldsymbol{\beta}) + \cdots \\
+ \ \delta_p(Y_{t-p} - \mu_{S_{t-p}} - \boldsymbol{z}_{t-p}\boldsymbol{\beta}) + \varepsilon_t,
\end{aligned}
$$

where the regression coefficient $\boldsymbol{\beta}$ is considered to be unaffected by S_t. In the following dynamic regression model all parameters, including the regression coefficient $\boldsymbol{\beta}$, are affected by endogenous regime shifts following a hidden Markov chain (McCulloch and Tsay, 1994b),

$$
Y_t = \delta_{S_t,1}Y_{t-1} + \cdots + \delta_{S_t,p}Y_{t-p} + \boldsymbol{z}_t\boldsymbol{\beta}_{S_t} + \zeta_{S_t} + \varepsilon_t.
$$

For estimation it is useful to view this model as a Markov switching regression model as in Subsection 10.3.2, without distinguishing between endogenous variables, exogenous variables, and the intercept:

$$
Y_t = \mathbf{x}_t\boldsymbol{\beta}_{S_t} + \varepsilon_t, \tag{12.13}
$$

where $\mathbf{x}_t = \begin{pmatrix} y_{t-1} & \cdots & y_{t-p} & z_{t1} & \cdots & z_{td} & 1 \end{pmatrix}$. In the mixed-effects Markov switching dynamic regression model only certain elements of the parameter $\boldsymbol{\beta}_{S_t}$ in (12.13) actually depend on the state of the hidden Markov chain, and others are state independent (McCulloch and Tsay, 1994b):

$$
Y_t = \mathbf{x}_t^f\boldsymbol{\alpha} + \mathbf{x}_t^r\boldsymbol{\beta}_{S_t} + \varepsilon_t, \tag{12.14}
$$

where $\mathbf{x}_t^f$ are those columns of $\mathbf{x}_t$ that correspond to the state-independent parameters $\boldsymbol{\alpha}$ whereas the columns of $\mathbf{x}_t^r$ correspond to the state-dependent parameters. Any of these models may be combined with homoscedastic variances, $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$, or heteroscedastic variances, where the error variances are different in the various states, $\varepsilon_t \sim \mathcal{N}\left(0, \sigma_{\varepsilon,S_t}^2\right)$.

12.3.2 Bayesian Estimation

Bayesian estimation of the Markov switching dynamic regression model along the lines indicated in *Algorithm 11.3* is closely related to Bayesian estimation of finite mixtures of regression models. Sampling the parameters

$(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,K})$ conditional on a known trajectory $\mathbf{S}$ of the hidden Markov chain in step (a2) of *Algorithm 11.3* is exactly the same as for a mixed-effects finite mixture regression model and may be implemented as in *Algorithm 8.2*. As the Markov switching dynamic regression model includes lagged values of Y_t, inference is usually carried out conditional on the first p observations $y_1, \ldots, y_p$ and t runs from $t_0 = p+1$ to T. To adapt the formulae of Subsection 8.4.4, in particular (8.36) and (8.37), to the slightly different notation used here, note that i corresponds to $t-p$, whereas N corresponds to $T-p$.

Usually an independence prior is applied where location and scale parameters are assumed to be independent a priori (Albert and Chib, 1993; McCulloch and Tsay, 1994b):

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,K}) = p(\boldsymbol{\alpha}) \prod_{k=1}^{K} p(\boldsymbol{\beta}_k) p(\sigma^2_{\varepsilon,k}), \tag{12.15}$$
$$\boldsymbol{\alpha} \sim \mathcal{N}_r(\mathbf{a}_0, \mathbf{A}_0), \qquad \boldsymbol{\beta}_k \sim \mathcal{N}_d(\mathbf{b}_0, \mathbf{B}_0), \qquad \sigma^2_{\varepsilon,k} \sim \mathcal{G}(c_0, C_0).$$

Conditionally conjugate priors exist only for two special cases of model (12.14); first, for a model with homoscedastic variances, namely

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma^2_{\varepsilon}) = p(\sigma^2_{\varepsilon}) p(\boldsymbol{\alpha}|\sigma^2_{\varepsilon}) \prod_{k=1}^{K} p(\boldsymbol{\beta}_k|\sigma^2_{\varepsilon}), \tag{12.16}$$
$$\boldsymbol{\alpha}|\sigma^2_{\varepsilon} \sim \mathcal{N}_r\left(\mathbf{a}_0, \sigma^2_{\varepsilon}\mathbf{A}_0\right), \qquad \boldsymbol{\beta}_k|\sigma^2_{\varepsilon} \sim \mathcal{N}_d\left(\mathbf{b}_0, \sigma^2_{\varepsilon}\mathbf{B}_0\right), \qquad \sigma^2_{\varepsilon} \sim \mathcal{G}(c_0, C_0),$$

and, second, for a model with heteroscedastic variances and no common parameters, where $\mathbf{x}^f_t \boldsymbol{\alpha}$ vanishes in (12.14), namely

$$p(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K, \sigma^2_{\varepsilon,1}, \ldots, \sigma^2_{\varepsilon,K}) = \prod_{k=1}^{K} p(\boldsymbol{\beta}_k|\sigma^2_{\varepsilon,k}) p(\sigma^2_{\varepsilon,k}), \tag{12.17}$$
$$\boldsymbol{\beta}_k|\sigma^2_{\varepsilon,k} \sim \mathcal{N}_d\left(\mathbf{b}_0, \sigma^2_{\varepsilon,k}\mathbf{B}_0\right), \qquad \sigma^2_{\varepsilon,k} \sim \mathcal{G}(c_0, C_0).$$

With increasing number K of states joint sampling of all regression parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K)$ may be rather time consuming, especially for regression models with high-dimensional parameter vectors, and further blocking may be applied (Albert and Chib, 1993; McCulloch and Tsay, 1994b; Kim and Nelson, 1999).

12.4 Prediction of Time Series Based on Markov Switching Models

12.4.1 Flexible Predictive Distributions

The predictive distribution of a Markov switching model is much more flexible than the predictive distribution of more traditional time series models. Consider the one-step ahead predictive density $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ of a Markov switching model that reads

$$p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}) = \sum_{k=1}^{K} p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}_k)\Pr(S_t = k|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}), \tag{12.18}$$

if at least assumption **Y2** holds. Various features of (12.18) are worth mentioning.

First, the one-step ahead predictive density $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ is a finite mixture distribution and potentially nonnormal, even if the component densities $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}_k)$ are normal. The weights of this mixture density are given by the one-step ahead predictive probabilities $\Pr(S_t = k|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}), k = 1, \ldots, K$, which are determined recursively by the filter derived in Subsection 11.2.2, and are dynamic, depending on the past values of $\mathbf{y}^{t-1}$, as long as S_t does not reduce to an i.i.d. process. Additional important features of the predictive density are nonlinearity of $\mathrm{E}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ in the past values $\mathbf{y}^{t-1}$, and conditional heteroscedasticity, meaning that $\mathrm{Var}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ depends on the past. Also higher-order moments of $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ are dynamic and depend on the past.

These features are made more explicit for a two-state Markov switching model with normal component densities, $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}_k) = f_N(y_t; \mu_{k,t}, \sigma^2_{k,t})$, with $\mu_{k,t} = \mathrm{E}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}_k)$ and $\sigma^2_{k,t} = \mathrm{Var}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\theta}_k)$ being the conditional mean and the conditional variance. Obviously from (12.18), the predictive density $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ is a mixture of two normal distributions,

$$\begin{aligned} p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}) &= w_{t-1}(\mathbf{y}^{t-1}) f_N(y_t; \mu_{1,t}, \sigma^2_{1,t}) \\ &\quad + (1 - w_{t-1}(\mathbf{y}^{t-1})) f_N(y_t; \mu_{2,t}, \sigma^2_{2,t}), \end{aligned} \tag{12.19}$$

where $w_{t-1}(\mathbf{y}^{t-1}) = \Pr(S_t = 1|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ is the predictive probability of $S_t = 1$ given time series observations up to $t - 1$. Using the filter equations given in Subsection 11.2.3, it is possible to show that $w_{t-1}(\mathbf{y}^{t-1})$ is a nonlinear function of the past values $\mathbf{y}^{t-1}$. From (11.5) follows

$$w_{t-1}(\mathbf{y}^{t-1}) = (1 - \lambda)\eta_1 + \lambda \Pr(S_{t-1} = 1|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}), \tag{12.20}$$

where the filter equation (11.2) implies that the odds ratio for the filter probability $\Pr(S_{t-1} = 1|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ is given by

$$\begin{aligned} &\mathrm{logit}\ \Pr(S_{t-1} = 1|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}) = -.5 \\ &\times \left(\frac{(y_{t-1} - \mu_{1,t-1})^2}{\sigma^2_{1,t-1}} - \frac{(y_{t-1} - \mu_{2,t-1})^2}{\sigma^2_{2,t-1}} + \log \frac{\sigma^2_{1,t-1}}{\sigma^2_{2,t-1}} \right) + \mathrm{logit}\ w_{t-2}(\mathbf{y}^{t-1}). \end{aligned}$$

Consequently, the right-hand side of (12.20) is a nonlinear function of y_{t-1}, and by recursion, of all other previous values $\mathbf{y}^{t-2}$. Hence, the mean of the one-step ahead predictive distribution $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ of a two-state hidden Markov model, which is given by

$$\mathrm{E}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}) = \mu_{1,t} w_{t-1}(\mathbf{y}^{t-1}) + \mu_{2,t}(1 - w_{t-1}(\mathbf{y}^{t-1})),$$

is nonlinear in the past $\mathbf{y}^{t-1}$, even if the conditional means $\mu_{1,t}$ and $\mu_{2,t}$ are linear as for the MSAR model.

Furthermore, the dependence of the weights $w_{t-1}(\mathbf{y}^{t-1})$ on past observations through the nonlinear function (12.20) introduces conditional heteroscedasticity, even if the predictive densities are homoscedastic. For a two-state hidden Markov model the variance of the one-step ahead predictive distribution $p(y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ is given by

$$\begin{aligned}\mathrm{Var}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}) = \sigma^2_{1,t} w_{t-1}(\mathbf{y}^{t-1}) + \sigma^2_{2,t}(1 - w_{t-1}(\mathbf{y}^{t-1}))\\ + 2\mu_{1,t}\mu_{2,t} w_{t-1}(\mathbf{y}^{t-1})(1 - w_{t-1}(\mathbf{y}^{t-1})),\end{aligned}$$

$w_{t-1}(\mathbf{y}^{t-1})$ depends on past observations through the nonlinear function (12.20). Thus the conditional variance $\mathrm{Var}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ of a Markov switching model is in general a nonlinear function of past squared errors and able to capture conditional heteroscedasticy observed in financial time series; see Section 12.5.

12.4.2 Forecasting of Markov Switching Models via Sampling-Based Methods

Predictors $\hat{y}_{T+1}, \ldots, \hat{y}_{T+h}$ of a time series $\mathbf{y} = (y_1, \ldots, y_T)$ which are optimal with respect to the mean-squared prediction error criterion may be computed recursively for most Markov switching models (Krolzig, 1997; Clements and Krolzig, 1998).

Bayesian forecasting of future observations $y_{T+1}, \ldots, y_{T+h}$ of a time series $\mathbf{y} = (y_1, \ldots, y_T)$ is based on the predictive density $p(y_{T+1}, \ldots, y_{T+h}|\mathbf{y})$ which is not available in closed form for most time series models, even for simple AR(p) models (Schnatter, 1988a). Sampling-based forecasting procedures that have been applied to AR models (Thompson and Miller, 1986) and to ARCH models (Geweke, 1992) were extended to deal with Markov switching autoregressive models (Albert and Chib, 1993).

The following algorithm shows how forecasting by a sampling-based approach is implemented for arbitrary Markov switching models fulfilling at least assumption **Y3** whereas S_t only needs to fulfill **S1**.

Algorithm 12.1: Forecasting of a Markov Switching Time Series Model For each MCMC draw $(\boldsymbol{\vartheta}^{(m)}, S_1^{(m)}, \ldots, S_T^{(m)})$ from the joint posterior $p(\mathbf{S}, \boldsymbol{\vartheta}|\mathbf{y})$ carry out the following steps to sample from the posterior predictive density $p(y_{T+1}, \ldots, y_{T+h}|\mathbf{y})$.

(a) Starting with $S_T^{(m)}$, sample a future path of the hidden Markov chain by sampling $S_{T+s}^{(m)}$ recursively for $s = 1, \ldots, h$ from the discrete distribution $p(S_{T+s}|S_{T+s-1}^{(m)}, \boldsymbol{\vartheta}^{(m)})$. For a homogeneous Markov chain, this distribution is equal to the kth row of $\boldsymbol{\xi}^{(m)}$, if $S_{T+s-1}^{(m)}$ takes the value k.

(b) Given $\boldsymbol{\vartheta}^{(m)}$ and $S_{T+1}^{(m)}, \dots, S_{T+h}^{(m)}$, sample $y_{T+1}^{(m)}$ from the predictive density $p(y_{T+1}|\mathbf{y}, \boldsymbol{\vartheta}, S_{T+1}^{(m)})$, and for $s = 2, \dots, h$, sample $y_{T+s}^{(m)}$ recursively from the predictive density $p(y_{T+s}|y_{T+s-1}^{(m)}, \dots, y_{T+1}^{(m)}, \mathbf{y}, \boldsymbol{\vartheta}, S_{T+s}^{(m)})$.

To implement step (b) for the MSAR model, for instance, one samples future paths $y_{T+1}^{(m)}, \dots, y_{T+h}^{(m)}$ recursively from:

$$
\begin{aligned}
& y_{T+1}|S_{T+1}^{(m)} = k_1, \mathbf{y}, \boldsymbol{\vartheta}^{(m)} \sim \\
& \qquad \mathcal{N}\left(\zeta_{k_1}^{(m)} + \delta_{k_1,1}^{(m)} y_T + \dots + \delta_{k_1,p}^{(m)} y_{T-p}, \sigma_{\varepsilon,k_1}^{(2,m)}\right) \\
& y_{T+2}|S_{T+2}^{(m)} = k_2, y_{T+1}^{(m)}, \mathbf{y}, \boldsymbol{\vartheta}^{(m)} \sim \\
& \qquad \mathcal{N}\left(\zeta_{k_2}^{(m)} + \delta_{k_2,1}^{(m)} y_{T+1}^{(m)} + \delta_{k_2,2}^{(m)} y_T + \dots, \sigma_{\varepsilon,k_2}^{(2,m)}\right) \\
& y_{T+3}|S_{T+3}^{(m)} = k_3, y_{T+2}^{(m)}, y_{T+1}^{(m)}, \mathbf{y}, \boldsymbol{\vartheta}^{(m)} \sim \\
& \qquad \mathcal{N}\left(\zeta_{k_3}^{(m)} + \delta_{k_3,1}^{(m)} y_{T+2}^{(m)} + \delta_{k_3,2}^{(m)} y_{T+1}^{(m)} + \delta_{k_3,3}^{(m)} y_T + \dots, \sigma_{\varepsilon,k_3}^{(2,m)}\right), \\
& \dots
\end{aligned}
$$

where $\delta_{k,l}^{(m)} = 0$ for $l > p$.

12.5 Markov Switching Conditional Heteroscedasticity

12.5.1 Motivating Example

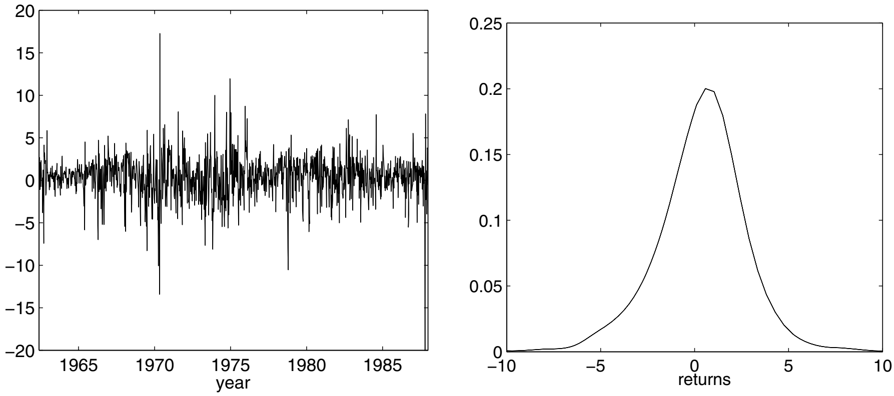

Fig. 12.6. New York Stock Exchange Data, left: time series plot; right: smoothed histogram of the marginal distribution

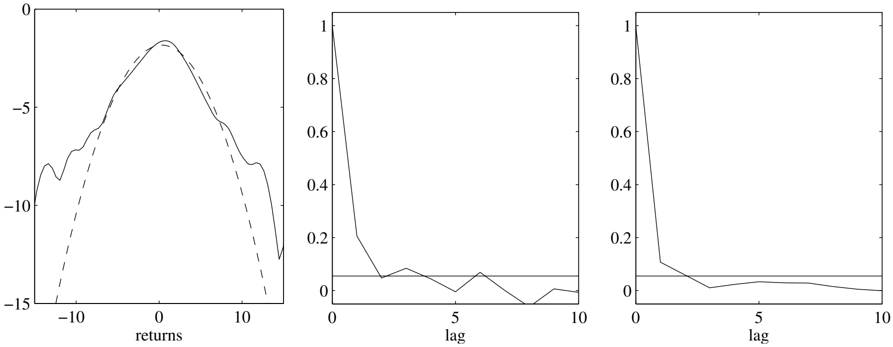

Fig. 12.7. New York Stock Exchange Data, left: log of the smoothed histogram (solid line) in comparison to the log of a normal distribution with same mean and variance (dashed line); middle: empirical autocorrelogram of the returns; right: empirical autocorrelogram of the squared returns

Figure 12.6 shows the weekly New York Stock Exchange Data investigated in Hamilton and Susmel (1994). The series originates from the CRISP data tapes and consists of a value-weighted portfolio of stocks traded on the New York Stock Exchange and starts with the week ending Tuesday, July 3, 1962 and ends with the week ending Tuesday, December 29, 1987, making in total 1330 observations. The smoothed histogram of the marginal distribution indicates asymmetry and fat tails. The empirical skewness coefficient and excess kurtosis are given by −1.2923 and 17.6394, respectively.

A central topic of econometrics of financial markets is the question of how to model the distribution of such returns, and how to estimate the variability, usually termed volatility, of financial time series. The returns are in general defined as $y_t = \log p_t - \log p_{t-1}$, where p_t is the price of a financial asset or a stock index. Two important stylized facts of financial time series, known as fat tails and volatility clustering, were discovered in the 1960s. Fama (1965, p.48), when studying 30 stocks from the Dow Jones industrial average index, summarized:

> *In any case the empirical distributions are more peaked than the normal in the centre and have longer tails than the normal distribution.*

Departure from normality also occurs for the returns of the New York Stock Exchange Data. Nonnormal tail behavior is evident in particular from the left plot in Figure 12.7, comparing the log of the smoothed histogram with the log of a normal distribution with the same mean and the same variance.

Concerning the second stylized fact, Mandelbrot (1963, p.418) states,

> *Large changes tends to be followed by large changes — of either sign — and small changes tend to be followed by small changes.*

This kind of volatility clustering is also evident for the returns of the NEW YORK STOCK EXCHANGE DATA from the time series plot in Figure 12.6. The presence of volatility clusters if often tested by analyzing the autocorrelation in the squared process. Many studies find significant serial correlation in the squared values of financial time series; see also the right-hand side plot in Figure 12.7 for the returns of the NEW YORK STOCK EXCHANGE DATA.

Later on researchers realized various asymmetric effects also called leverage effects. As Engle and Ng (1995, p.173) summarize,

> *Overall, these results show a greater impact on volatility of negative, rather than positive return shocks.*

12.5.2 Capturing Features of Financial Time Series Through Markov Switching Models

Markov switching models are often used by researchers to account for specific features of financial time series such as asymmetries, fat tails, and volatility clusters.

To deal with skewness and excess kurtosis in the unconditional distribution of daily stock returns standard finite mixtures of normal distributions have been applied quite frequently (Fama, 1965; Granger and Orr, 1972; Kon, 1984; Tucker, 1992). Such a modeling approach, however, is appropriate for time series data only if the processes Y_t and Y_t^2 do not exhibit autocorrelation, as by the results of Subsections 10.2.4 and 10.2.5 a standard finite mixture model implies zero autocorrelation in Y_t and Y_t^2.

Volatility clustering implies persistence of states of high volatility and leads to the rejection of standard time series models in favor of time series models that allow the conditional variance $\mathrm{Var}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta})$ to depend on the history $y_{t-1}, y_{t-2}, \ldots$ of the observed process such as the autoregressive conditionally heteroscedastic (ARCH) model (Engle, 1982), where

$$\mathrm{Var}(Y_t|\mathbf{y}^{t-1}, \boldsymbol{\vartheta}) = \gamma_t + \alpha_1 y_{t-1}^2 + \cdots + \alpha_m y_{t-m}^2,$$

and the generalized autoregressive conditionally heteroscedastic (GARCH) model (Bollerslev, 1986). The popularity of ARCH models, in particular if they are based on the t_ν-error distributions (Bollerslev et al., 1992), can certainly be explained by their ability to generate processes with serial correlation in Y_t^2, whereas the introduction of t_ν-error helps to capture the tail behavior appropriately. Tsay (1987) considered random coefficient autoregressive models which are another example of a conditional heteroscedastic time series model and showed that the ARCH process is a special case of this model class. More recently, stochastic volatility models have been increasingly applied to financial time series (Shephard, 1996; Kim et al., 1998; Chib et al., 2002).

As an alternative to these models, Markov mixture models where the variance of a location-scale family is driven by a hidden Markov chain have been applied to financial time series (Engel and Hamilton, 1990; McQueen and

Thorely, 1991; Rydén et al., 1998). Although these models introduce autocorrelation in Y_t^2, while preserving nonnormality of the marginal distribution (see again Subsection 10.2.5), a closer inspection of the autocorrelation function of Y_t^2 reveals that the basic Markov switching model generates only limited persistence in the squared process. For a two-state model, where $\mu_1 = \mu_2$, for instance, there exists a strong relationship between the fatness of the tails, measured by the excess kurtosis, and autocorrelation of the squared process.

Far more general autocorrelation functions of Y_t^2 are possible if Y_t is generated by an MSAR model with or without switching AR coefficients; see Timmermann (2000, Proposition 5). Hence the MSAR model has been applied to a number of financial time series (Hamilton, 1988; Turner et al., 1989; Cecchetti et al., 1990; Engel, 1994; Gray, 1996; Ang and Bekaert, 2002).

To obtain even more flexibility in the autocorrelation of Y_t^2, for a given marginal distribution of Y_t, Hamilton and Susmel (1994), Cai (1994), and Gray (1996) proposed to combine ARCH and Markov switching effects to formulate the switching ARCH model, which is defined in Subsection 12.5.3 as a highly flexible, nonlinear time series model. Bekaert and Harvey (1995) introduced a model that combines Markov switching models with multivariate ARCH models to allow for time-dependence in the integration of emerging markets. Francq et al. (2001) considered the switching GARCH model; see Subsection 12.5.5.

Smith (2002) extends Markov switching models further, by incorporating a regime-dependent variance parameter, when modeling stochastic volatility in interest rates.

12.5.3 Switching ARCH Models

A simple model to capture volatility clusters in financial time series is the ARCH model (Engle, 1982) which may be written as

$$\begin{aligned} Y_t &= \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1), \\ \sigma_t^2 &= \gamma_t + \alpha_1 Y_{t-1}^2 + \cdots + \alpha_m Y_{t-m}^2 \end{aligned} \tag{12.21}$$

with $\gamma_t \equiv \gamma$. An alternative parameterization of this model reads:

$$\begin{aligned} Y_t &= \sqrt{\gamma_t} h_t \varepsilon_t, \\ h_t^2 &= 1 + \frac{\alpha_1}{\gamma_{t-1}} Y_{t-1}^2 + \cdots + \frac{\alpha_m}{\gamma_{t-m}} Y_{t-m}^2. \end{aligned} \tag{12.22}$$

The two parameterizations are equivalent if $\gamma_t \equiv \gamma$, however, they generate different processes if γ_t is time dependent. The switching ARCH model results by allowing time dependence of γ_t through a hidden K-state Markov chain S_t: $\gamma_t = \gamma_{S_t}$.

Such a switching parameter was introduced by Hamilton and Susmel (1994) into parameterization (12.22):

$$Y_t = \sqrt{\gamma_{S_t}} h_t \varepsilon_t,$$
$$h_t^2 = 1 + \frac{\alpha_1}{\gamma_{S_{t-1}}} Y_{t-1}^2 + \cdots + \frac{\alpha_m}{\gamma_{S_{t-m}}} Y_{t-m}^2,$$

whereas Cai (1994) introduced a two-state and Kaufmann and Frühwirth-Schnatter (2002) a K-state switching parameter into parameterization (12.21):

$$Y_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \gamma_{S_t} + \alpha_1 Y_{t-1}^2 + \cdots + \alpha_m Y_{t-m}^2.$$

Gray (1996) introduced switching into all coefficients of the ARCH process, represented by (12.21):

$$Y_t = \sigma_t \varepsilon_t,$$
$$\sigma_t^2 = \gamma_{S_t} + \alpha_{S_t,1} Y_{t-1}^2 + \cdots + \alpha_{S_t,m} Y_{t-m}^2. \tag{12.23}$$

A special case of model (12.23) is the mixture autoregressive conditional heteroscedastic model (Wong and Li, 2001), where S_t is an i.i.d. process rather than a Markov process. Francq et al. (2001) provide conditions under which model (12.23) is second-order stationary; see also the discussion in Subsection 12.5.5.

The switching ARCH model may be combined with a Markov switching autoregressive model for the mean equation that includes the same hidden Markov chain (Gray, 1996):

$$Y_t = \zeta_{S_t} + \delta_{S_t,1} Y_{t-1} + u_t,$$
$$u_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1),$$
$$\sigma_t^2 = \gamma_{S_t} + \alpha_{S_t,1} u_{t-1}^2 + \cdots + \alpha_{S_t,m} u_{t-m}^2. \tag{12.24}$$

The switching ARCH model has been extended by including a leverage effect into the ARCH specification (Hamilton and Susmel, 1994; Kaufmann and Frühwirth-Schnatter, 2002) to deal with asymmetries in the marginal distribution:

$$Y_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1),$$
$$\sigma_t^2 = \gamma_{S_t} + \alpha_1 y_{t-1}^2 + \cdots + \alpha_m y_{t-m}^2 + \varrho d_{t-1} y_{t-1}^2, \tag{12.25}$$

where $d_t = 1$ if $y_t \leq 0$, $d_t = 0$ if $y_t > 0$ and $\varrho > 0$.

Further applications of switching ARCH models in financial econometrics include modeling of stock market returns (Hamilton and Lin, 1996; Fong, 1997), interest rates (Cai, 1994; Gray, 1996; Ang and Bekaert, 2002), and exchange rate data (Klaasen, 2002).

Spurious Persistency in Squared Returns

A common finding when fitting GARCH models to high-frequency financial data is the somewhat unexpected persistence of shocks to the variance implied by the estimated coefficients which led to the development of the class of integrated generalized autoregressive conditional heteroscedasticity (IGARCH) models (Engle and Bollerslev, 1986). Lamoureux and Lastrapes (1990) investigated the possibility that the appearance of a unit root in the GARCH model may be due to time-varying GARCH parameters. They show that a deterministic structural shift in the unconditional variance, caused by exogenous shocks such as changes in the monetary policy, will increase persistency of squared residuals, however, when the structural break is accounted for, persistency often decreases dramatically.

Introducing a hidden Markov chain into a variance model helps to explain spurious persistence in squared returns. Consider, for illustration, a simple Markov mixture of two normal distributions with $\mu_1 = \mu_2$ and $\sigma_1^2 \neq \sigma_2^2$ driven by a highly persistent transition matrix $\boldsymbol{\xi}$ with $\lambda = \xi_{11} - \xi_{21}$ being close to 1. Together with $\sigma_2^2 - \sigma_1^2$ being large this leads to slowly decaying persistence in Y_t^2:

$$\rho_{Y_t^2}(h|\boldsymbol{\vartheta}) = \frac{\eta_1\eta_2(\sigma_1^2 - \sigma_2^2)^2}{\mathrm{E}(Y_t^4|\boldsymbol{\vartheta}) - \mathrm{E}(Y_t^2|\boldsymbol{\vartheta})^2}\lambda^h$$

(see again (10.24)), although the squared returns are uncorrelated within each regime. Also for the more general switching ARCH model, Hamilton and Susmel (1994) attribute part of the high marginal persistence in Y_t^2, which is typically much larger than autocorrelation of Y_t^2 in the various regimes, to this effect.

12.5.4 Statistical Inference for Switching ARCH Models

Parameter estimation for switching ARCH models may be carried out by ML estimation (Hamilton and Susmel, 1994; Francq et al., 2001). Hamilton and Susmel (1994) report extreme difficulties with maximizing the likelihood function for the New York Stock Exchange Data, and only by restricting seven transition probabilities to 0 were they able to run the optimization procedure and to report standard errors for their final model.

Bayesian estimation of the switching ARCH model as exemplified in Kaufmann and Frühwirth-Schnatter (2002) has the advantage of coping with the near boundary space problem by imposing a proper prior on the transition matrix $\boldsymbol{\xi}$ as discussed in Subsection 11.5.1, in which case the posterior density is proper also for unobserved transitions, and standard errors and confidence regions are directly available. MCMC sampling may be carried out along the lines indicated in *Algorithm 11.3*, however, the Metropolis–Hastings algorithm is needed to implement step (a2) due to the nonlinear structure of the underlying model which does not lead to simple conditional densities.

Consider, as an example, the following special case of the switching AR-ARCH model,

$$y_t = \zeta + \delta_1 y_{t-1} + u_t, \tag{12.26}$$
$$u_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1), \tag{12.27}$$
$$\sigma_t^2 = \gamma_{S_t} + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2,$$

where step (a2) in *Algorithm 11.3* requires sampling the AR parameters $\boldsymbol{\phi}_1 = (\zeta, \delta_1)$ and the ARCH parameters $\boldsymbol{\phi}_2 = (\gamma_1, \ldots, \gamma_K, \alpha_1, \ldots, \alpha_m)$ from the appropriate conditional densities. To this aim, Kaufmann and Frühwirth-Schnatter (2002) developed the following two-block Metropolis–Hastings step, building on Nakatsuma (2000).

(a2-1) Sample the AR parameters $\boldsymbol{\phi}_1 = (\zeta, \delta_1)$ from the conditional posterior $p(\boldsymbol{\phi}_1|\mathbf{S}, \boldsymbol{\phi}_2, \mathbf{y})$ using a Metropolis–Hastings algorithm with proposal density $q(\boldsymbol{\phi}_1^{new}|\boldsymbol{\phi}_1^{old})$.

(a2-2) Sample the ARCH parameters $\boldsymbol{\phi}_2 = (\gamma_1, \ldots, \gamma_K, \alpha_1, \ldots, \alpha_m)$ from the conditional posterior $p(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1, \mathbf{S}, \mathbf{y})$ using a Metropolis–Hastings algorithm with proposal density $q(\boldsymbol{\phi}_2^{new}|\boldsymbol{\phi}_2^{old})$.

Due to the presence of ARCH errors in regression model (12.26) no direct method of sampling the AR parameters $\boldsymbol{\phi}_1$ is available even if the ARCH parameters $\boldsymbol{\phi}_2$ are known and a normal prior $\boldsymbol{\phi}_1 \sim \mathcal{N}(\mathbf{b}_0, \mathbf{B}_0)$ is assumed (Bauwens and Lubrano, 1998; Kim et al., 1998; Nakatsuma, 2000). The crucial point is that the error variance σ_t^2 depends on $\boldsymbol{\phi}_1 = (\zeta, \delta_1)$ through the lagged residuals $u_{t-1}, \ldots, u_{t-m}$:

$$\begin{aligned}\sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) = \gamma_{S_t} &+ \alpha_1 (y_{t-1} - \zeta - \delta_1 y_{t-2})^2 + \cdots \\ &+ \alpha_m (y_{t-m} - \zeta - \delta_1 y_{t-m-1})^2.\end{aligned}$$

Because model (12.26) is a standard regression model with heteroscedastic errors, if $\sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ is independent of $\boldsymbol{\phi}_1$, the following normal proposal density results when substituting $\sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)$ by $\sigma_t^2(\boldsymbol{\phi}_1^{old}, \boldsymbol{\phi}_2)$ (Kaufmann and Frühwirth-Schnatter, 2002),

$$\begin{aligned}
q(\boldsymbol{\phi}_1^{new}|\boldsymbol{\phi}_1^{old}) &= f_N(\boldsymbol{\phi}_1^{new}; \mathbf{b}_N(\boldsymbol{\phi}_1^{old}), \mathbf{B}_N(\boldsymbol{\phi}_1^{old})), \\
\mathbf{b}_N(\boldsymbol{\phi}_1) &= \mathbf{B}_N(\boldsymbol{\phi}_1) \left(\sum_{t=m+2}^{T} \frac{1}{\sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)} \mathbf{x}_t^{'} y_t + \mathbf{B}_0^{-1} \mathbf{b}_0 \right), \\
\mathbf{B}_N(\boldsymbol{\phi}_1) &= \left(\sum_{t=m+2}^{T} \frac{1}{\sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)} \mathbf{x}_t^{'} \mathbf{x}_t + \mathbf{B}_0^{-1} \right)^{-1},
\end{aligned}$$

where $\mathbf{b}_0$ and $\mathbf{B}_0$ are the prior parameters and $\mathbf{x}_t = \begin{pmatrix} 1 & y_{t-1} \end{pmatrix}$.

Also the conditional posterior of the ARCH parameters $\boldsymbol{\phi}_2$ is not of any closed form. To derive a proposal density $q(\boldsymbol{\phi}_2^{new}|\boldsymbol{\phi}_2^{old})$, the switching ARCH model is reformulated in Kaufmann and Frühwirth-Schnatter (2002) as a generalized linear model. From (12.27), $u_t^2 = \sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)\varepsilon_t^2$, where ε_t^2 is a χ_1^2 random variable that may be expressed as $\varepsilon_t^2 = 1 + \tilde{\varepsilon}_t$ with $\mathrm{E}(\tilde{\varepsilon}_t) = 0$ and $\mathrm{Var}(\tilde{\varepsilon}_t) = 2$. Therefore:

$$
\begin{aligned}
u_t^2 = \gamma_1 D_t^1 + \cdots + \gamma_K D_t^K + u_{t-1}^2 \alpha_1 + \cdots \\
+ \; u_{t-m}^2 \alpha_m + \sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2)\tilde{\varepsilon}_t,
\end{aligned}
\tag{12.28}
$$

where $D_t^k = 1$ iff $S_t = k$. A normal proposal for $\boldsymbol{\phi}_2$ has been derived in Kaufmann and Frühwirth-Schnatter (2002) from model (12.28) by substituting the nonnormal errors by normal ones with variance $2(\sigma_t^2(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2^{old}))^2$.

Note that both the basic Markov switching model with heterogeneous variances as well as the ARCH model are nested within the switching ARCH model. Therefore model selection may be used to test for the usefulness of the combined model as well as the correct model order. Francq et al. (2001) show that the AIC and Schwarz criteria do not underestimate the correct order of the switching ARCH model. Kaufmann and Frühwirth-Schnatter (2002) use marginal likelihoods to select both the number of states as well as the model order of a switching ARCH model.

Application to the New York Stock Exchange Data

For illustration, we return to the New York Stock Exchange Data. To account for the autocorrelation found in y_t and y_t^2, as well as for the fat tails and the asymmetry observed in the marginal distribution, Kaufmann and Frühwirth-Schnatter (2002) fitted the following switching AR-ARCH model to these data, which includes a leverage term,

$$
y_t = \zeta + \delta_1 y_{t-1} + u_t, \tag{12.29}
$$

$$
u_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1),
$$

$$
\sigma_t^2 = \gamma_{S_t} + \alpha_1 u_{t-1}^2 + \cdots + \alpha_m u_{t-m}^2 + \varrho d_{t-1} y_{t-1}^2. \tag{12.30}
$$

Table 12.3. New York Stock Exchange Data, modeled by a switching AR-ARCH model with leverage with different numbers of states K and different model orders m; log of the marginal likelihoods under different priors on the switching ARCH intercept (from Kaufmann and Frühwirth-Schnatter (2002) with permission granted by Blackwell Publisher Ltd.)

		$\log p(\mathbf{y}\|K,m)$	
K	m	(prior 1)	(prior 2)
3	2	–2858.5	–2858.0
3	3	–2858.2	–2857.7
3	4	**–2857.1**	–2856.4
4	2	–2861.0	–2859.7
4	3	–2860.7	–2859.4
4	4	–2859.1	**–2855.9**

Table 12.3 summarizes the marginal likelihoods $p(\mathbf{y}|K,m)$ for different numbers of states K and different model orders m. The marginal likelihoods

are estimated using bridge sampling as described in Section 5.4.6. Kaufmann and Frühwirth-Schnatter (2002) noted sensitivity of the model selection procedure with respect to the prior on γ_k. Selecting the model order m is unaffected by this prior and yields $m = 4$. Depending on the prior, the marginal likelihood would favor either a model with three or four states; see Table 12.3. This sensitivity may be explained by the fact that for a four-state model one of the states corresponds to a single outlier. Thus little information on the parameters of the fourth state is available from the likelihood and the prior dominates the posterior distribution.

12.5.5 Switching GARCH Models

Francq et al. (2001) consider the following switching GARCH(m, n) model, where all coefficients are switching,

$$\begin{aligned} &Y_t = \sigma_t \varepsilon_t, \qquad \varepsilon_t \sim \mathcal{N}(0,1), \\ &\sigma_t^2 = \gamma_{S_t} + \alpha_{S_t,1} y_{t-1}^2 + \cdots + \alpha_{S_t,m} y_{t-m}^2 + \delta_{S_t,1} \sigma_{t-1}^2 + \cdots + \delta_{S_t,n} \sigma_{t-n}^2 . \end{aligned} \tag{12.31}$$

By recursive substitution it becomes evident that the predictive density $p(y_t|\mathbf{y}^{t-1}, \mathbf{S}^t, \boldsymbol{\vartheta})$ depends on the whole history of S_t. For the switching GARCH$(1, 1)$ model, for instance, the variance of the predictive density reads:

$$\begin{aligned} \sigma_t^2 = {} & \gamma_{S_t} + \alpha_{S_t,1} y_{t-1}^2 + \delta_{S_t,1} (\gamma_{S_{t-1}} + \alpha_{S_{t-1},1} y_{t-2}^2) \\ & + \delta_{S_t,1} \gamma_{S_{t-1}} (\gamma_{S_{t-2}} + \alpha_{S_{t-2},1} y_{t-3}^2) + \cdots . \end{aligned}$$

Thus the model obeys only the weakest assumption **Y1** defined in Subsection 10.3.4. Due to the work of Francq et al. (2001), the theoretical properties of the switching GARCH models are well understood.

First, Francq et al. (2001) establish necessary and sufficient conditions ensuring the existence of a strictly stationary solution by rewriting (12.31) as a stochastic dynamic system and considering the Lyapunov exponent of this system as in Bougerol and Picard (1992a). For the switching GARCH$(1, 1)$ model, for instance, this condition reads:

$$\sum_{k=1}^{K} \eta_k \mathrm{E}(\log\left(\alpha_{k,1} \varepsilon_t^2 + \delta_{k,1}\right)) < 0,$$

which reduces for $K = 1$ to the result given by Nelson (1990) for the standard GARCH$(1, 1)$ model. This condition, however, does not guarantee the existence of the unconditional variance of Y_t.

Francq et al. (2001) establish necessary and sufficient conditions for the existence of second-order stationary solutions, which reduce to the requirement that the spectral radius of a matrix derived from the stochastic dynamic system mentioned above is strictly less than one. For a GARCH(m, n) model where only the intercept is switching, this condition reduces to:

$$\sum_{j=1}^{m} \alpha_j + \sum_{j=1}^{n} \delta_j < 1,$$

which is equal to the condition given by Bollerslev (1986) for a standard GARCH(m, n) model.

Finally, Francq and Zakoian (1999) establish necessary and sufficient conditions for the existence of higher-order moments of Y_t^2. They show that Y_t^2 admits a linear ARMA representation where the orders depend on m, n, and the model coefficients, extending the well-known result that a GARCH(m, n) process has the same autocorrelation as an ARMA$(\max(m, n), m)$ process. Similar ARMA representations are also derived for powers of Y_t^2.

Practical application of switching GARCH models include stock market returns (Dueker, 1997) and exchange rate data (Klaasen, 2002).

12.6 Some Extensions

12.6.1 Time-Varying Transition Matrices

Whereas the transition matrix $\boldsymbol{\xi}$ of the hidden process S_t is time invariant under assumption **S3** or **S4**, the transition probability from S_{t-1} to S_t may depend on exogenous variables under assumption **S1** or **S2**, as suggested by Goldfeld and Quandt (1973).

For a two-state Markov switching model, the transition probabilities ξ_{S_{t-1},S_t} may be reparameterized through a logit model in the following way,

$$\xi_{S_{t-1},S_t} = \frac{\exp(\kappa_{S_{t-1},1})}{1 + \exp(\kappa_{S_{t-1},1})}, \qquad S_t \neq S_{t-1}.$$

A univariate exogenous variable z_t may then be included as in Subsection 8.6.2:

$$\xi_{S_{t-1},S_t} = \frac{\exp(\kappa_{S_{t-1},1} + z_t \kappa_{S_{t-1},2})}{1 + \exp(\kappa_{S_{t-1},1} + z_t \kappa_{S_{t-1},2})}, \qquad S_t \neq S_{t-1}, \tag{12.32}$$

with $\kappa_{j,1}$ and $\kappa_{j,2}$, $j = 1, 2$ being unknown parameters. Note that the transition probability ξ_{S_{t-1},S_t} not only depends on z_t, but also on the state of S_{t-1}. The logit transform could be substituted by another increasing function $F(\cdot)$,

$$\xi_{S_{t-1},S_t} = F(\kappa_{S_{t-1},1} + z_t \kappa_{S_{t-1},2}), \qquad S_t \neq S_{t-1}, \tag{12.33}$$

for instance, the standard normal distribution. If z_t is equal to a lagged value of Y_t, $z_t = y_{t-d}$ for some $d > 0$, then the so-called endogenous selection MSAR model (Krolzig, 1997, Subsection 10.3.2) results. If, in addition, the parameters of model (12.32) or (12.33) are independent of the state of S_{t-1}, $\kappa_{1,1} = \kappa_{2,1}$, $\kappa_{1,2} = \kappa_{2,2}$, then the resulting model is closely related to the smooth transition autoregressive model (Teräsvirta and Anderson, 1992; Granger and Teräsvirta, 1993). Extensions to multiple exogenous variables $\boldsymbol{z}_t = (z_{t1}, \dots, z_{tr})$ and to more than two states are possible.

Models with time-varying transition matrices found applications in hydrology (Zucchini and Guttorp, 1991; Pfeiffer and Jeffries, 1999), in financial econometrics (Diebold et al., 1994; Peria, 2002; Schaller and van Norden, 2002; Ang and Bekaert, 2002), and in business cycle analysis to capture duration dependence, meaning that the transition probability between recession and boom depends on how long the economy remained within the same regime (Durland and McCurdy, 1994; Filardo, 1994; Filardo and Gordon, 1998).

A model with time-varying transition matrices may be estimated through the EM algorithm (Diebold et al., 1994) or through MCMC methods (Filardo and Gordon, 1998).

12.6.2 Markov Switching Models for Longitudinal and Panel Data

Some recent papers combine clustering methods and longitudinal analysis using hidden Markov models. In a health state model comparing the effectiveness of two different medications for schizophrenia, Scott et al. (2005) assume that the observed response $\mathbf{y}_{it}$ for patient i at time t follows a multivariate Student-t distribution,

$$\mathbf{y}_{it}|S_{it} = k \sim t_{\nu_k}\left(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k\right), \tag{12.34}$$

depending on a latent health state S_{it}. The health state is assumed to be a hidden Markov chain with treatment-dependent transition matrix. Estimation of this model is carried out using MCMC and BIC was used to select the number of health states.

Frühwirth-Schnatter and Kaufmann (2006b) combine clustering and Markov switching models in economic panel data analysis by assuming that K hidden groups are present in a panel and that within each group the parameters may switch according to a hidden Markov chain. Consider, for example, the mixed-effects model defined in Subsection 8.5.2,

$$y_{it} = \mathbf{x}_{it}^f \boldsymbol{\alpha} + \mathbf{x}_{it}^r \boldsymbol{\beta}_{it}^s + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right). \tag{12.35}$$

$\boldsymbol{\beta}_{it}^s$ depends on two latent discrete indicators, first on a group indicator S_i. Second, within each group the regression coefficient corresponding to $\mathbf{x}_{it}^r$ may switch between two states, commonly thought of as the state of the economy, depending on a group-specific hidden Markov chain $I_{t,k}$ with group-specific transition matrix $\boldsymbol{\xi}_k$:

$$\boldsymbol{\beta}_{it}^s = \boldsymbol{\beta}_k + (I_{t,k} - 1)\boldsymbol{\gamma}_k, \quad S_i = k.$$

This model allows pooling all time series within each group and is robust against structural changes through including the hidden Markov chain. A simplified version where the hidden Markov chain is group independent has been considered by Frühwirth-Schnatter and Kaufmann (2006a). Estimation is carried out using MCMC methods.

For $K = 1$, this model reduces to the panel data Markov switching model that has been applied to analyze the lending behavior of banks over the business cycle (Asea and Blomberg, 1998; Kaufmann, 2002):

$$y_{it} = \mathbf{x}_{it}^f \boldsymbol{\alpha} + \mathbf{x}_{it}^r \boldsymbol{\beta}_{S_t} + \varepsilon_{it}, \qquad \varepsilon_{it} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right), \tag{12.36}$$

which allows a shift in the regression coefficient corresponding to $\mathbf{x}_{it}^r$ between the two states of S_t, commonly thought of as the state of the economy. Estimation of model (12.36) may be carried out using the EM algorithm (Asea and Blomberg, 1998) or MCMC methods (Kaufmann, 2002; Frühwirth-Schnatter and Kaufmann, 2006a).

12.6.3 Markov Switching Models for Multivariate Time Series

Hidden Markov models have been extended in several ways to deal with multivariate time series $\{\mathbf{Y}_t, t = 1, \ldots, T\}$, where $\mathbf{Y}_t$ is random vector of r different variables, for instance, the GDP from different countries. Common multivariate time series models are the vector autoregressive (VAR) model (Sims, 1980) and cointegration models (Engle and Granger, 1987); see also Shumway and Stoffer (2000, Chapter 4) for a review of multivariate time series analysis.

To analyze the growth rate of GDP in a two-country set-up, Phillips (1991) generalized the univariate MSAR model (Hamilton, 1989) by introducing a hidden Markov chain into a bivariate VAR(1) model:

$$\mathbf{Y}_t - \boldsymbol{\mu}_{\mathbf{S}_t} = \boldsymbol{\Phi}(\mathbf{Y}_{t-1} - \boldsymbol{\mu}_{\mathbf{S}_{t-1}}) + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_r\left(\mathbf{0}, \boldsymbol{\Sigma}\right),$$

with $\boldsymbol{\Phi}$ and $\boldsymbol{\Sigma}$ being (2×2) matrices, and $\boldsymbol{\mu}_{\mathbf{S}_t}$ being a vector of length 2. $\mathbf{S}_t = (S_{t,1}, S_{t,2})$ is a bivariate two-state hidden Markov chain, with $S_{t,j}$ describing the state of the economy in country j, which could be coded as a single Markov chain with four states. The (4×4)-transition matrix $\boldsymbol{\xi}$ of $\mathbf{S}_t$ is unrestricted if the states of the two economies are correlated, a restricted transition matrix results if $S_{t,1}$ and $S_{t,2}$ are assumed to be independent. Hamilton and Lin (1996) apply a related model to analyze jointly growth in industrial production and volatility in stock returns, and discuss restricted transition matrices where one indicator is leading the other.

Krolzig (1997) considered multivariate MS-VAR models, where a single hidden Markov chain S_t may affect the intercept (or the mean level), the matrix containing the AR coefficients as well as the error covariance matrix:

$$\mathbf{Y}_t = \boldsymbol{\Phi}_{S_t} \mathbf{Y}_{t-1} + \boldsymbol{\zeta}_{S_t} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_r\left(\mathbf{0}, \boldsymbol{\Sigma}_{S_t}\right),$$

with $\boldsymbol{\Phi}_{S_t}$ and $\boldsymbol{\Sigma}_{S_t}$ being $(r \times r)$ matrices and $\boldsymbol{\zeta}_{S_t}$ being a vector of length r. Krolzig (1997) discusses ML estimation of this model using the EM algorithm as well as Bayesian estimation using Gibbs sampling.

A related model is applied in Ang and Bekaert (2002) to model interest rates from three different countries, however, $\mathbf{S}_t = (S_{t,1}, S_{t,2}, S_{t,3})$ is a trivariate two-state hidden Markov chain, with $S_{t,j}$ describing the hidden state in country j, coded as a single Markov chain with eight states. Because the states in different countries are assumed to be independent, a restricted transition matrix $\boldsymbol{\xi}$ results for $\mathbf{S}_t$.

Economic theory implies a long-run relationship between certain integrated time series such as consumption and disposable income, implying that the time series are cointegrated. As with unit root tests, discussed in Subsection 12.2.4, common cointegration tests are affected by shifts in the growth rate of the underlying time series (Hall et al., 1997). For this reason several authors considered the introduction of a hidden Markov chain into cointegration models to account for unexpected shifts.

Paap and van Dyck (2003) introduce a multivariate Markov switching trend model that accounts for different growth rates in a bivariate time series $\mathbf{Y}_t$, containing the log of per capita consumption and disposable income:

$$\begin{aligned} \mathbf{Y}_t &= \boldsymbol{\mu}_t + (S_t - 1)\begin{pmatrix} \delta \\ 0 \end{pmatrix} + \mathbf{Z}_t, \\ \boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\beta}_{S_t}, \end{aligned}$$

where S_t is a two-state hidden Markov chain. $\boldsymbol{\beta}_1$ is a vector containing the slopes of the trend function of both time series, if $S_t = 1$ (expansion) and $\boldsymbol{\beta}_2$ contains the slopes if $S_t = 2$ (recession). δ accounts for possible level shifts in the first time series during recession. $\mathbf{Z}_t$ is assumed to follow a standard VAR(p) process. Cointegration analysis based on the vector error correction model is then carried out for $\mathbf{Z}_t$:

$$\triangle \mathbf{Z}_t = \boldsymbol{\Pi} \mathbf{Z}_t + \sum_{j=1}^{p-1} \tilde{\boldsymbol{\Phi}}_j \triangle \mathbf{Z}_{t-j} + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim \mathcal{N}_2(\mathbf{0}, \boldsymbol{\Sigma}). \tag{12.37}$$

Depending on the rank of $\boldsymbol{\Pi}$ three cases arise. If $\boldsymbol{\Pi}$ has rank zero, then the bivariate MS-VAR model for the growth rates results; if $\boldsymbol{\Pi}$ has rank two, then $\mathbf{Z}_t$ is stationary and a generalization of the model of Lam (1990) results; and finally, if $\boldsymbol{\Pi}$ has rank one, then the two time series are cointegrated. Bayesian estimation of this model is carried out in Paap and van Dyck (2003) using MCMC and the Bayes factor is used to test for the cointegration rank (Kleibergen and Paap, 2002). The empirical results of Paap and van Dyck (2003) suggest the existence of a cointegration relationship between U.S. per capita disposable income and consumption.

Related approaches are a single equation cointegration analysis where the parameters are allowed to undergo changes driven by a hidden Markov chain (Hall et al., 1997) and an alternative Markov switching vector error correction (MS-VEC) model (Krolzig and Sensier, 2000; Krolzig, 2001) where a Markov switching intercept is introduced directly into a vector error correction model for $\mathbf{Y}_t$; see also Krolzig (1997, Chapter 13).

Further Markov switching models for multivariate time series, in particular the Markov switching model dynamic factor model (Diebold and Rudebusch, 1996; Kim and Nelson, 1998), are special cases of switching Gaussian state space models which are studied in Chapter 13.