ISSN 0005-1179, Automation and Remote Control, 2018, Vol. 79, No. 9, pp. 1687–1702. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © A.V. Dobrovidov, V.E. Tevosian, 2017, published in Problemy Upravleniya, 2017, No. 4, pp. 26–36.

Nonparametric Estimation of Volatility and Its Parametric Analogs

A. V. Dobrovidov[*,a] **and V. E. Tevosian**[*,b]

[*] *Trapeznikov Institute of Control Sciences, Russian Academy of Sciences, Moscow, Russia*
e-mail: [a] *dobrovidov@gmail.com,* [b] *vachik.tevosian@gmail.com*

Received June 15, 2017

Abstract—This paper suggests a nonparametric method for stochastic volatility estimation and its comparison with other widespread econometric algorithms. A major advantage of this approach is that the volatility can be estimated even in the case of its completely unknown probability distribution. As demonstrated below, the new method has better characteristics against the popular parametric algorithms based on the GARCH model and Kalman filter.

Keywords: stochastic volatility, nonparametric estimation of signals, Kalman filter, GARCH, Taylor model

DOI: 10.1134/S0005117918090126

1. INTRODUCTION

Prior to describing the stochastic volatility models, consider this notion in more detail. Different financial phenomena can be understood under the term *volatility*. Depending on context and viewpoint, volatility means the variability of prices in form of the standard deviation of return, financial risk parameter or price model, or a random process of certain type. The notion of volatility was modified with the appearance of the Black–Scholes–Merton option pricing model [1] in 1973. Black and Scholes applied the continuous-time geometric Brownian motion to construct a mathematical model of pricing. The success of their paper promoted a wide use of stochastic models in econometrics. In our paper, the term volatility will be associated with a random process that describes the variability of prices.

As indicated by the return analysis of American investment funds, the main distinctive feature of successful funds is the capability to control portfolio volatility for reducing the share of most volatile assets during the periods of high variability of asset prices [2]. Also note that the active investment strategies based on dynamic volatility measurements guarantee considerable yield [3, 4]. All these conclusions testify that the knowledge of volatility is crucial for efficient portfolio management, investment strategy design and company risk management, along with other issues.

Some publications on nonparametric volatility estimation have appeared recently. For example, consider volatility prediction using local exponential estimation for the Yao–Tong model [5] suggested by Ziegelmann [6], or parameter estimation for the stochastic Duffie–Pan–Singleton model [7] and also for the Jacquier–Polson–Rossi model [8] using infinitesimal moments in the paper [9] by Bandi and Reno. In contrast to these publications, our paper relies on Wald's theory of decision making, more specifically, on risk function minimization as the most widespread criterion in statistics. We study the volatility model in the form of an unobservable stochastic process with unknown distribution and unknown state equation. Under certain conditions, this process can be estimated using the empirical Bayesian approach developed within the theory of nonparametric filtering, interpolation, and prediction of partially observable Markov processes [10]. In this theory, the stochastic state models of useful signals are assumed unknown while the observation models that describe investigator's instruments are assumed completely known. If a useful signal is unobservable in pure form, then generally its distribution cannot be restored. Hence, the optimal Bayesian procedure is not directly applicable to useful signal restoration based on observations. In comparison with the above-mentioned parametric estimation methods, the main advantage of this approach consists in the feasibility of considering volatility processes of arbitrary frequency without any hypotheses about the model of the unobservable component.

2. DESCRIPTIVE MODELS FOR STOCHASTIC VOLATILITY

In a financial market, the asset price S_t at a time t is positive by definition and often considered in the representation $S_t = S_0 e^{R_t}$, where

$$R_t = \sum_{i=1}^{t} r_i, \quad t = 0, 1, 2, \ldots.$$

The value r_t defined by

$$r_t = \log \frac{S_t}{S_{t-1}}$$

is called the logarithmic rate of return at the time t.

The simplest assumption about the probabilistic nature of r_t is that its model is described by the stochastic equation $r_t = \sigma\varepsilon_t$, where $\varepsilon = (\varepsilon_t)$ denotes a sequence of independent random variables with the same distribution $\mathcal{N}(0,1)$ and σ is a parameter termed volatility in financial literature.

Further research demonstrated that volatility is "volatile" itself and its behavior well fits the framework of random time-varying functions. Volatility as a characteristic of variability is not observed but can be treated as a random process (σ_t) in the observable rate-of-return model

$$r_t = \mu_t + \sigma_t \varepsilon_t, \tag{1}$$

where μ_t and σ_t specify given random processes. In this paper, we will consider two parametric conditional Gaussian models of volatility in discrete time within the efficient market concept, namely, Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) suggested by Bollerslev [11] and Stochastic Volatility (SV) developed by Taylor [12]. The former model is widely used in practice because it describes current volatility and also predicts its value for one step forward. However, in this model the volatility σ_t is a given function of the observable rate of return r_t, which slightly restricts its possible modifications. The latter model has a feature that the volatility σ_t is written as a separate stochastic process independent of the rate r_t [13]. This assumption extends the behavioral variety of volatility, and hence a rich arsenal of methods from the theories of random processes and martingales [14] become applicable here. The two models will be considered in detail in Sections 2.1 and 2.2 below.

2.1. GARCH Model

Suppose the sequence $\varepsilon = (\varepsilon_t)_{t\geq 1}$ in model (1) is a unique source of randomness in the market while the conditional mean and variance have the form

$$\mu_t = E(r_t | r_{t-1}, \ldots, r_1) = 0, \quad \sigma_t^2 = E\left(r_t^2 | r_{t-1}, \ldots, r_1\right) = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2, \tag{2}$$

where $\alpha_0 > 0,\ \alpha_i \geq 0,\ i = 1, \ldots, p$, and $r_{1-p}, \ldots, r_0$ are given initial constants. The conditional mean μ_t is zero because the daily average rate of return and the average intraday rates of return for stocks and currencies are zero [15].

This model was introduced by Engle in 1982 and called the AutoRegressive Conditional Heteroskedastic Model, ARCH(p). It well explains several nontrivial properties of financial time series such as the clustering effect of the values r_t.

Definition 1. ARCH(p) is a sequence $r = (r_t)$ of the form $r_t = \sigma_t \varepsilon_t$, where $\varepsilon = (\varepsilon_t)$ represents a sequence of independent random variables with the standard Gaussian distribution, $\varepsilon_t \sim \mathcal{N}(0, 1)$, while σ_t^2 satisfies the recursive Eq. (2).

The obvious success of ARCH(p) soon resulted in the appearance of its different extensions and modifications. One such extension is the Generalized AutoRegressive Conditional Heteroskedastic (GARCH) model suggested by Bollerslev in 1986. GARCH (p, q) has the following advantage over ARCH(p): for adjustment to real data, the latter needs large values p whereas the former works well with small values p and q.

As before, assume $\mu_t = 0$ but, unlike formula (2), let σ_t satisfy the equation

$$\sigma_t^2 = E\left(r_t^2 | r_{t-1}, \ldots, r_1\right) = \alpha_0 + \sum_{i=1}^{p} \alpha_i r_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2, \tag{3}$$

with $\alpha_0 > 0$, $\alpha_i, \beta_i \geq 0$, and initial conditions $(r_{1-p}, \ldots, r_0)$ and $(\sigma_{1-q}^2, \ldots, \sigma_0^2)$, which can be constants or generated in advance.

Definition 2. GARCH(p, q) is a sequence $r = (r_t)$ of the form $r_t = \sigma_t \varepsilon_t$, where $\varepsilon = (\varepsilon_t)$ represents a sequence of independent random variables with the same distribution $\mathcal{N}(0, 1)$ while σ_t obeys (3).

2.2. Taylor Model

The ARCH-type models contain a single source of noise defined by a sequence of independent Gaussian random variables $\varepsilon = (\varepsilon_t)_{t \geq 1}$. By assumption, the variance of the rate-of-return process depends in some way on its past realizations. An alternative approach is to describe price dynamics by a simple model (e.g., a differential equation, like in the Black–Scholes–Merton option pricing model) with volatility considered as a separate stochastic process. This leads to two independent sources of randomness. Stochastic volatility models include another source of noise $\delta = (\delta_t)_{t \geq 1}$, which also consists of the independent random variables with the same distribution $\mathcal{N}(0, 1)$ in the simplest case.

As in the previous models, assume the first condition in (2) holds. Then

$$r_t = \sigma_t \varepsilon_t, \tag{4}$$

where ε_t are random variables with the distribution $\mathcal{N}(0, 1)$.

Let

$$\sigma_t = e^{\frac{1}{2}\Delta_t}, \tag{5}$$

which implies that σ_t is a positive random variable. The models of dependent random variables are most widespread here, in which the sequence $(\Delta_t)_{t \geq 1}$ satisfies the autoregression model

$$\Delta_t = a_0 + a_1 \Delta_{t-1} + \ldots + a_p \Delta_{t-p} + c\delta_t, \tag{6}$$

where c denotes the variance of noise component. It follows from (4)–(6) that the conditional distribution of r_t with fixed previous values $r_{t-1}, \ldots, r_1$ is Gaussian with the parameters 0 and σ_t^2, i.e.,

$$f(r_t | r_{t-1}, \ldots, , r_1) = \mathcal{N}\left(0, \sigma_t^2\right).$$

Definition 3. The stochastic volatility model is a sequence $r = (r_t)$ of the form $r_t = \sigma_t \varepsilon_t$, where $\varepsilon = (\varepsilon_t)$ represents a sequence of independent random variables with the same distribution $\mathcal{N}(0, 1)$ while σ_t is given by the autoregression equation of Δ_t induced by the white noise $\mathcal{N}(0, 1)$.

3. KALMAN FILTER

The Kalman filter is one of the most popular volatility estimation methods in modern econometrics [16, 17]. Consider this filter subject to stochastic volatility estimation. It is applicable to the linear Gaussian models only and hence Eq. (4) must be linearized. Using formula (5), we obtain the expression

$$r_t = \sigma_t \varepsilon_t = e^{\frac{1}{2}\Delta_t}\varepsilon_t.$$

Taking the logarithm of its square yields

$$\log\left(r_t^2\right) = \log\left(e^{\Delta_t}\varepsilon_t^2\right) = \log\left(e^{\Delta_t}\right) + \log\left(\varepsilon_t^2\right) = \Delta_t + \log\left(\varepsilon_t^2\right).$$

Introduce the notations $b_t = \log\left(r_t^2\right)$ and $\xi_t = \log\left(\varepsilon_t^2\right)$. Then Eq. (4) is transformed into

$$b_t = \Delta_t + \xi_t. \tag{7}$$

For the sake of simplicity, Eq. (6) will be written as the first-order autoregression with the zero mean:

$$\Delta_t = a_1\Delta_{t-1} + c\delta_t. \tag{8}$$

A prerequisite for using the Kalman filter is the Gaussian distribution of the noisy components. However, the random variable $\xi_t = \log(\varepsilon_t^2)$ in the state model has another distribution; so it is replaced by the Gaussian counterpart ξ_t^g with the mean $E\xi_t = E\xi_t^g = -1.27$ and the variance $D\xi_t = D\xi_t^g = 4.93 = \pi^2/2$ [18] (see the Appendix for details). Then system (7), (8) takes the form

$$\begin{cases} b_t = \Delta_t + \xi_t^g \\ \Delta_t = a_1\Delta_{t-1} + c\delta_t. \end{cases}$$

As a result, the Kalman filter for the stochastic volatility model is described by the system of equations

$$\begin{aligned} \hat{\Delta}_t &= a_1\hat{\Delta}_{t-1} + \frac{2c^2 + 2a_1^2\gamma_{t-1}}{\pi^2 + 2c^2 + 2a_1^2\gamma_{t-1}}\left(b_n + 1.27 - a_1\hat{\Delta}_{t-1}\right), \\ \gamma_t &= \frac{\pi^2\left(c^2 + a_1^2\gamma_{t-1}\right)}{\pi^2 + 2c^2 + 2a_1^2\gamma_{t-1}} \end{aligned} \tag{9}$$

with the initial conditions

$$\hat{\Delta}_1 = \frac{2c^2}{\pi^2(1-a_1^2) + 2c^2}b_1,$$

$$\gamma_1 = \frac{\pi^2 c^2}{\pi^2(1-a_1^2) + 2c^2}.$$

In this system, the estimate $\hat{\Delta}_t$ is recursive, which facilitates calculations. At the same time, the substitution for the Gaussian noise may cause additional errors.

Finally, the solution of system (9) is used in formula (5) for obtaining the volatility estimate

$$\hat{\sigma}_t^2 = \exp\left(a_1\hat{\Delta}_{t-1} + \frac{2c^2 + 2a_1^2\gamma_{t-1}}{\pi^2 + 2c^2 + 2a_1^2\gamma_{t-1}}\left(b_n + 1.27 - a_1\hat{\Delta}_{t-1}\right)\right).$$

4. OPTIMAL FILTERING EQUATION WITH UNKNOWN DISTRIBUTION OF USEFUL SIGNAL

A lot of research efforts were made in the 1960–1970s to develop rigorous learning and self-learning methods for automatic systems. A line of investigations was focused on self-learning problems under statistical processing of signals with unknown state models. This line yielded the theory of nonparametric estimation of signals, which stems from a pioneering paper [19] published in 1983. In this theory, the stochastic state models of useful signals are assumed unknown while the observation models that describe investigator's devices are assumed completely known. If a useful signal is unobservable in pure form, then generally its distribution cannot be restored. Hence, the optimal Bayesian procedure is not directly applicable to useful signal restoration based on observations. Parametric models involve different mathematical models of unobservable signals. For example, GARCH and Stochastic Volatility include the state Eqs. (3) and (6), respectively. Nonparametric estimation allows obtaining an estimate of an unobservable signal even with a completely unknown distribution, which forms its major advantage. In other words, an unobservable signal is filtered without introducing any descriptive models of the process and assumptions about its character.

Let $(X_t, S_t)_{t\geq 1}$ be a partially observable two-component process, where $(X_t \in R)_{t\geq 1}$ and $(S_t \in R)_{t\geq 1}$ denote the observable and unobservable components that are statistically connected with each other by a conditional density function $f(x_t|s_t) = f(x_t|S_t = s_t)$. Conditional density is defined by specifying two objects, namely, 1) the distribution of noises η_t and 2) an observation model, i.e., an equation relating an observation X_t to the useful signal S_t and noise η_t, $X_t = \phi(S_t, \eta_t)$. Different classes of filtering problems correspond to different forms of the observation model and distribution of noises. Consider a class of problems in which the statistical connection equation and distribution of noises are chosen so that the conditional density function $f(x_t|s_t)$ belongs to the exponential family, i.e.,

$$f(x_t|s_t) = \hat{C}(s_t)g(x_t)\exp\{T(x_t)Q(s_t)\} \tag{10}$$

for a fixed useful signal $S_t = s_t$. Here T, Q, and g are given Borel functions while $\hat{C}(s_t)$ denotes a normalization factor. First of all, the parametric family (10) contains the Gaussian density function, a single density with known optimal mean-square estimates in explicit form. Also it includes the χ^2-distribution, the beta-distribution, some of the Pearson distributions, etc. If the random variable S_t has probabilistic sense, then this parametrization is termed *natural.* The family of distributions (10) allows a more simpler parametrization called *canonical* in which the role of parameter is played by the term $\theta_t = Q(S_t)$ that enters linearly in exponent index. In this case,

$$f(x_t|\theta_t) = C(\theta_t)g(x_t)\exp\{T(x_t)\theta_t\}, \tag{11}$$

and the normalization factor takes the form

$$C^{-1}(\theta_t) = \int\limits_R g(x_t)\exp\{T(x_t)\theta_t\}dx_t. \tag{12}$$

The set of all values θ_t for which there exists the exponential density function $f(x_t|\theta_t)$ is restricted by the set of those θ_t for which integral (12) is finite.

Here the main idea of the estimation algorithm consists in the following. Instead of the process (S_t), at first estimate the process $(\theta_t = Q(S_t))_{t\geq 1}$ that induces the simpler canonical parametrization in the family of distributions (11). Secondly obtain the estimate $\widehat{S}$ in the form $Q^{-1}(\hat{\theta}_t)$ using the inverse function Q^{-1}, which is assumed to exist. The resulting estimate $\widehat{S}$ is also optimal but with the loss function $\left(Q(\theta_t) - Q\left(\hat{\theta}_t\right)\right)^2$.

The process θ_t will be estimated under the hypothesis that the unobservable sequence $(S_t)_{t\geq 1}$ is a Markov sequence. Hence the two-component process $(Z_t)_{t\geq 1} = (S_t, X_t)_{t\geq 1}$ is also a Markov process.

Under completely known statistical information, the process θ_t is estimated using the optimal Bayesian estimate in form of the conditional mean

$$\hat{\theta}_t = \int\limits_{\Theta_t} \theta_t w_t\left(\theta_t|x_1^t\right) d\theta_t, \tag{13}$$

where Θ_t denotes the value set of the process θ_t while w_t is the posterior probability satisfying the recursive equation

$$w_t(\theta_t|x_1^t) = \frac{f(x_t|\theta_t)}{f\left(x_t|x_1^{t-1}\right)} w_t\left(\theta_t|x_1^{t-1}\right), \quad t \geq 2, \tag{14}$$

where

$$w_t\left(\theta_t|x_1^{t-1}\right) = \int\limits_{\Theta_{t-1}} \tilde{p}(\theta_t|\theta_{t-1}) w_{t-1}\left(\theta_{t-1}|x_1^{t-1}\right) d\theta_{t-1}$$

gives the predicted posterior density function and $\tilde{p}(\theta_t|\theta_{t-1})$ is the transition density function of the Markov sequence $(\theta_t)_{t\geq 1}$.

Using the Bayes' formula, calculate the initial condition for Eq. (14):

$$w_1(\theta_1|x_1) = \frac{f(x_1|\theta_1)\tilde{p}(\theta_1)}{\int\limits_{\Theta_t} f(x_1|\theta_1)\tilde{p}(\theta_1)d\theta_1}. \tag{15}$$

Under incomplete statistical information, for optimal posterior estimation the integrals in Eqs. (14) and (15) are replaced by sums. This yields the Kushner–Stratonovich filter [20, 21].

Now, integrate both sides of Eq. (14) over θ_t and rearrange the term that depends on observations only to the left side:

$$f\left(x_t|x_1^{t-1}\right) = \int\limits_{\Theta_t} f(x_t|\theta_t) w_t\left(\theta_t|x_1^{t-1}\right) d\theta_t. \tag{16}$$

Differentiate this expression with respect to x_t:

$$f'_{x_t}\left(x_t|x_1^{t-1}\right) = \int\limits_{\Theta_t} f'_{x_t}(x_t|\theta_t) w_t\left(\theta_t|x_1^{t-1}\right) d\theta_t. \tag{17}$$

As mentioned earlier, the conditional density function $f(x_t|\theta_t)$ is assumed to belong to the exponential family. Then

$$\begin{aligned}\left(\frac{f(x_t|x_1^{t-1})}{g(x_t)}\right)'_{x_t} &= \frac{\partial}{\partial x_t}\int\limits_{\Theta_t} C(\theta_t)\exp(T(x_t)\theta_t)w_t\left(\theta_t|x_1^{t-1}\right)d\theta_t \\ &= \int\limits_{\Theta_t} T'_{x_t}(x_t)\theta_t C(\theta_t)\exp(T(x_t)\theta_t)w_t\left(\theta_t|x_1^{t-1}\right)d\theta_t,\end{aligned} \tag{18}$$

where prime denotes differentiation with respect to x_t.

Multiply the left and right sides of Eq. (18) by $g(x_t)/f\left(x_t|x_1^{t-1}\right)$:

$$\frac{\partial}{\partial x_t}\left(\ln\frac{f\left(x_t|x_1^{t-1}\right)}{g(x_t)}\right) = \frac{T'_{x_t}(x_t)}{f\left(x_t|x_1^{t-1}\right)}\int\limits_{\Theta_t}\theta_t C(\theta_t)g(x_t)\exp(T(x_t)\theta_t)w_t\left(\theta_t|x_1^{t-1}\right)d\theta_t.$$

Using the definition of Bayesian estimation, write the optimal filtering equation for the estimate $\hat{\theta}_t$:

$$T'_{x_t}(x_t)\hat{\theta}_t = \frac{\partial}{\partial x_t}\left(\ln\frac{f\left(x_t|x_1^{t-1}\right)}{g(x_t)}\right). \tag{19}$$

This equation does not explicitly depend on the prior and transition densities of the unobservable sequence S_t. So it can be used for obtaining the optimal estimate $\hat{\theta}_t$ based on the values of the observable component X_t without explicit specification of the transition density function.

For the class of models (11), the only unknown characteristic in the optimal filtering equation is the conditional density function $f(x_t|x_1^{t-1})$ of the observation x_n under fixed previous observations $x_1^{t-1} = (x_1, x_2, \ldots, x_{t-1})$. Since the process (X_t) is observable, its realization can be used for restoring this characteristic by nonparametric kernel methods. Note that the functional form of the conditional density $f\left(x_t|x_1^{t-1}\right)$ is not required. For kernel estimation, we have to specify just two parameters—the smoothing factor (bandwidth) and the parameter of regularization. In this paper, for these purposes we will employ smoothed cross-validation [22] and the Tikhonov regularization [23], respectively. With the estimation methods for the smoothing factors and the parameters of regularization together with nonparametric filtering algorithms at our disposal, we get an efficient tool for extracting unknown useful signals against the background of noises, which depends on the observable sample only. Therefore, the nonparametric signal extraction methods can be called automatic or self-learned.

Under the unknown state equation, the conditional density function $f\left(x_t|x_1^{t-1}\right)$ of observations cannot be calculated precisely. However, we may construct its approximation using nonparametric kernel estimation based on dependent observations x_1^t with given accuracy. In accordance with this procedure, the unknown density function $f\left(x_t|x_1^{t-1}\right)$ is replaced by the "truncated" density function $f\left(x_t|x_{t-\tau}^{t-1}\right)$, where τ means the degree of dependence of the observable process (X_t). As a matter of fact, τ represents the order of connectivity for the Markov process that approximates the non-Markov process (X_t), see [10]. By definition, $f\left(x_t|x_{t-\tau}^{t-1}\right) = f\left(x_{t-\tau}^t\right)/f\left(x_{t-\tau}^{t-1}\right)$. Then

$$\ln\left(f\left(x_t|x_{t-\tau}^{t-1}\right)\right)'_{x_n} = \frac{f'_{x_t}\left(x_{t-\tau}^t\right)}{f\left(x_{t-\tau}^t\right)}. \tag{20}$$

The denominator of this formula is a $(\tau+1)$-dimensional density function.

For such a function, the nonparametric kernel estimate takes the form

$$\hat{f}\left(x_{t-\tau}^{t}\right)=\frac{1}{(t-\tau-1)h^{\tau+1}}\sum_{i=1}^{t-\tau-1}\prod_{j=1}^{\tau+1}K\left(\frac{x_{t-j+1}-x_{t-j-i+1}}{h_t}\right). \tag{21}$$

For the numerator of (20), the nonparametric estimate is written as

$$\hat{f'}_{x_t}\left(x_{t-\tau}^{t}\right)=\frac{1}{(t-\tau-1)h_{1t}^{\tau+3}}\sum_{i=1}^{t-\tau-1}K'\left(\frac{x_t-x_{t-i}}{h_{1t}}\right)\prod_{j=2}^{\tau+1}K\left(\frac{x_{t-j+1}-x_{t-j-i+1}}{h_{1t}}\right), \tag{22}$$

where $\hat{f'}_{x_t}$ and K' denotes the partial derivative with respect to the variable x_t.

Consequently, the nonparametric estimate for the logarithmic derivative of the density function has the formula

$$\ln\left(\hat{f}\left(x_t|x_{t-\tau}^{t-1}\right)\right)'_{x_n}=\frac{\hat{f}'\left(x_{t-\tau}^{t}\right)}{\hat{f}\left(x_{t-\tau}^{t}\right)}.$$

This is the so-called plug-in estimate. For calculating ratio (20), it suffices to choose the smoothing factors h_t (for the density function (21)) and h_{1t} (for its partial derivative (22)), see [24].

5. NONPARAMETRIC ESTIMATION OF VOLATILITY

For the linear case of the Taylor model (7), (8), under fixed Δ_t the conditional density function b_t takes the following form (see the details in the Appendix):

$$p(b_t|\Delta_t)=\frac{\exp(b_t/2)}{\sqrt{2\pi}\exp(\Delta_t/2)}\exp\left(-\frac{\exp(b_t)}{2\exp(\Delta_t)}\right). \tag{23}$$

Formula (23) plays crucial role for the nonparametric estimation of volatility because one of the prerequisites is that the conditional distribution belongs to the exponential family.

As easily verified, the density function (23) is a member of the exponential family (11), where

$$g(b_t)=\frac{\exp(b_t/2)}{\sqrt{2\pi}};\quad g'_{b_t}(b_t)=\frac{1}{2}\frac{\exp(b_t/2)}{\sqrt{2\pi}}=\frac{1}{2}g(b_t);$$

$$C(\Delta_t)=\frac{1}{\exp(\Delta_t/2)};$$

$$T(b_t)=-\frac{\exp(b_t)}{2};\quad T'_{b_t}(b_t)=-\frac{\exp(b_t)}{2};$$

$$\theta(\Delta_t)=\frac{1}{\exp(\Delta_t)}.$$

Then the optimal filtering Eq. (19) can be written as

$$\theta(\Delta_t)\times T'_{b_t}(b_t)=\frac{\partial}{\partial b_t}\left(\ln\left(\frac{f\left(b_{t-\tau}^{t}\right)}{g(b_t)}\right)\right)=\frac{f'(b_{t-\tau}^{t})}{f(b_{t-\tau}^{t})}-\frac{g'(b_t)}{g(b_t)}. \tag{24}$$

Substituting the above expressions into Eq. (24) yields

$$\frac{\exp(b_t)}{2\exp(\Delta_t)}=\frac{1}{2}-\frac{f'(b_{t-\tau}^{t})}{f(b_{t-\tau}^{t})}.$$

As a result, the nonparametric estimate takes the form

$$\Delta_t = \log\left(\exp(b_t)\Big/\left(1 - \frac{2f'(b_{t-\tau}^t)}{f(b_{t-\tau}^t)}\right)\right),$$

and the volatility estimate is given by

$$\sigma_t = \exp\left(\frac{\Delta_t}{2}\right) = \sqrt{\exp(b_t)\Big/\left(1 - \frac{2f'(b_{t-\tau}^t)}{f(b_{t-\tau}^t)}\right)}. \tag{25}$$

Now, demonstrate that the denominator in the rooted expression in the (25) has positive values only, i.e.,

$$1 - \frac{2f'(b_{t-\tau}^t)}{f(b_{t-\tau}^t)} > 0.$$

To this end, multiply both sides of this inequality by $f(b_{t-\tau}^t)$:

$$f(b_{t-\tau}^t) - 2f'(b_{t-\tau}^t) > 0.$$

Next, divide the resulting expression by $f(b_{t-\tau}^{t-1})$:

$$\frac{f(b_{t-\tau}^t)}{f(b_{t-\tau}^{t-1})} - \frac{2f'(b_{t-\tau}^t)}{f(b_{t-\tau}^{t-1})} = f(b_t|b_{t-\tau}^{t-1}) - 2f'(b_t|b_{t-\tau}^{t-1}) > 0.$$

Using formulas (16) and (17), the last inequality can be reduced to

$$\int\limits_{\Delta_t} p(b_t|\Delta_t)w_t\left(\Delta_t|b_1^{t-1}\right)d\Delta_t - 2\int\limits_{\Delta_t} p'(b_t|\Delta_t)w_t\left(\Delta_t|b_1^{t-1}\right)d\Delta_t$$
$$= \int\limits_{\Delta_t} (p(b_t|\Delta_t) - 2p'(b_t|\Delta_t))w_t\left(\Delta_t|b_1^{t-1}\right)d\Delta_t > 0,$$

where $p'(b_t|\Delta_t)$ denotes the derivative with respect to b_t. Since $w_t(\cdot) \geq 0$, it suffices to prove that

$$p(b_t|\Delta_t) - 2p'(b_t|\Delta_t) > 0.$$

Taking advantage of formula (23), we obtain

$$\frac{\exp(b_t/2)}{\sqrt{2\pi}\exp(\Delta_t/2)}\exp\left(-\frac{\exp(b_t)}{2\exp(\Delta_t)}\right) - \frac{\exp(b_t/2)}{\sqrt{2\pi}\exp(\Delta_t/2)}\exp\left(-\frac{\exp(b_t)}{2\exp(\Delta_t)}\right)$$
$$+\frac{\exp(3b_t/2)}{\sqrt{2\pi}\exp(3\Delta_t/2)}\exp\left(-\frac{\exp(b_t)}{2\exp(\Delta_t)}\right) = \frac{\exp(3b_t/2)}{\sqrt{2\pi}\exp(3\Delta_t/2)}\exp\left(-\frac{\exp(b_t)}{2\exp(\Delta_t)}\right) > 0.$$

All the functions entering this expression are positive; hence the inequality holds for any values b_t and Δ_t.

For the nonparametric case of the exact Eq. (25), the functions $f(b_{t-\tau}^t)$ and $f'(b_{t-\tau}^t)$ should be replaced by their nonparametric estimates $\hat{f}(b_{t-\tau}^t)$ and $\hat{f}'(b_{t-\tau}^t)$, see formulas (21) and (22). As the result of this replacement, the denominator in the rooted expression (25) may become nonpositive. For avoiding this, apply the following additional condition of separability from nonpositive values: $(\upsilon)^+ = ((\upsilon \text{ for } \upsilon > 0) \vee (\varepsilon_0^+ \text{ for } \upsilon \leq 0))$, where ε_0^+ is a small positive value. Then the nonparametric estimate of volatility takes the form

$$\hat{\sigma}_t = \sqrt{\exp(b_t)\Big/\left(1 - \frac{2\hat{f}'(b_{t-\tau}^t)}{\hat{f}(b_{t-\tau}^t)}\right)^+}. \tag{26}$$

6. NUMERICAL IMPLEMENTATION

For a numerical implementation of the developed estimation algorithm, first we generated $t = 1500$ values of true volatility using the Taylor model (8) and then obtained the rate-of-return data (4) based on these values. The first 1000 values were employed for model learning while the other 500 for testing. For each method, the quality of the volatility estimate $\hat{\sigma}_i$ was assessed during 100 replicated trials by calculating the root-mean-square deviation (RMSD) from the true volatility σ_i:

$$\mathrm{RMSD}(\hat{\sigma}) = \sqrt{\frac{\sum\limits_{i=1000}^{1500} (\sigma_i - \hat{\sigma}_i)^2}{t}}.$$

This assessment criterion is used below for comparing different estimation methods, see the table.

6.1. Nonparametric Estimate

The nonparametric volatility estimate (26) (light plot) against the true volatility (dark plot) is shown in Fig. 1.

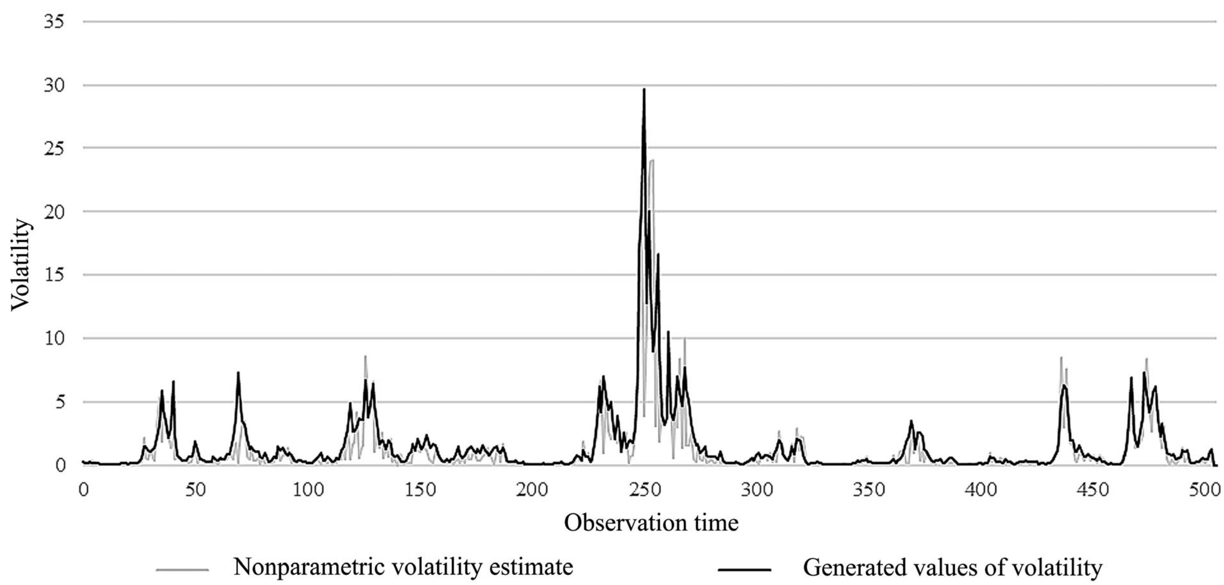

Fig. 1. Nonparametric volatility estimate.

Note that the nonparametric estimate well captured all clustering (inertial) segments of volatility, which are typical for economic crises and indicate of strong long-term fluctuations of rates of return during these periods. For 100 different trials, the average RMSD value was 2.8471. The algorithm was numerically implemented with $\varepsilon_0^+ = 0.005$.

6.2. Kushner–Stratonovich Filter

The volatility estimate using the Kushner–Stratonovich filter (13), (14) (light plot) against the true volatility (dark plot) is presented in Fig. 2. In this case, $\mathrm{RMSD} = 2.4674$, which is better than for the nonparametric estimate. This result was obvious from the very beginning, since the Kushner–Stratonovich filter yields the optimal estimate under complete statistical information.

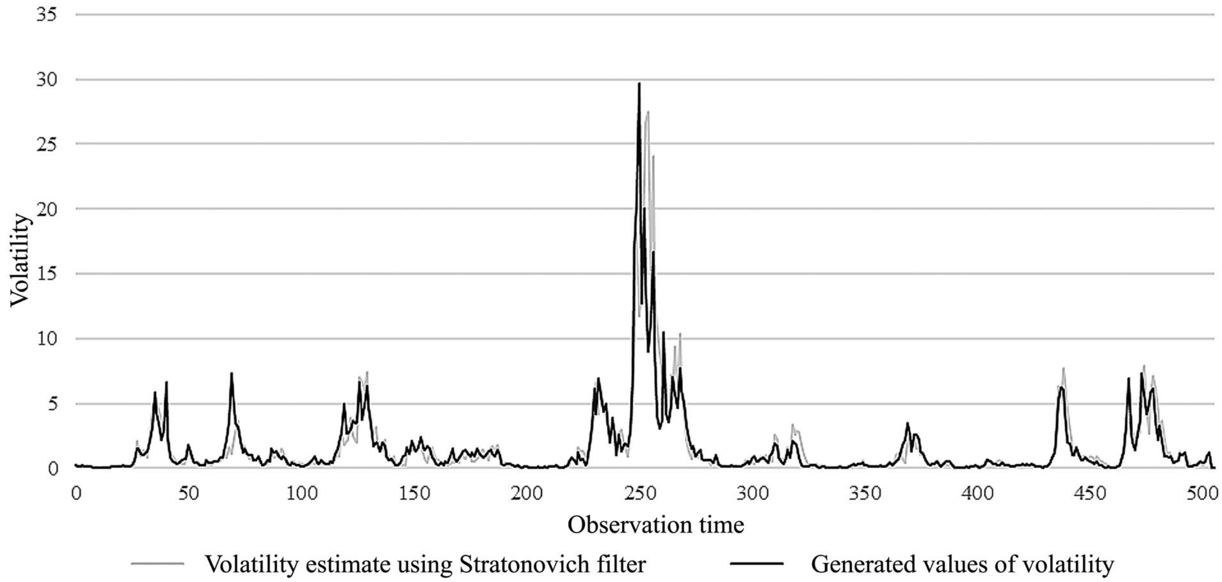

Fig. 2. Volatility estimate using Kushner–Stratonovich filter.

6.3. Kalman Filter

The volatility estimate yielded by the Kalman filter is demonstrated in Fig. 3 (light plot). Like in the two previous cases, all clustering segments were successfully captured but RMSD = 3.1614, which is almost two and a half times higher than for the nonparametric analog. Probably this result has two causes as follows.

—For application of the Kalman filter, we replaced the true distribution with the Gaussian distribution with the same first and second moments.

—We estimated not volatility but its logarithmic values.

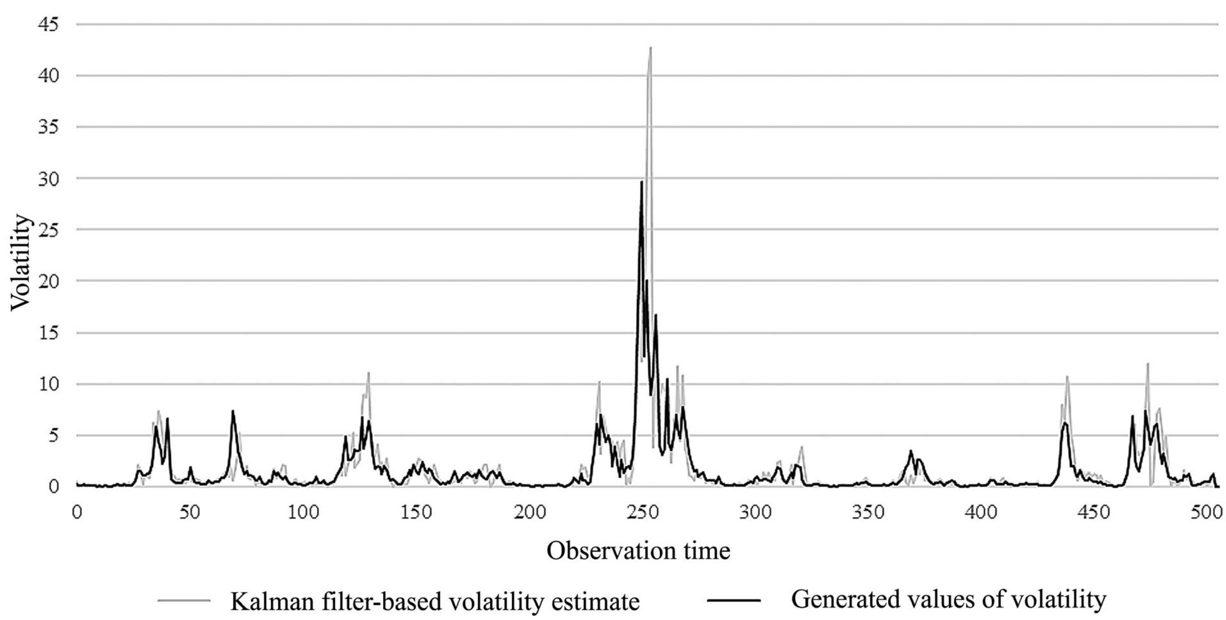

Fig. 3. Volatility estimate using Kalman filter.

6.4. GARCH

Finally, consider volatility estimate using the GARCH model from subsection 2.1. Numerical implementation was performed using the standard Matlab functions *garchset* and *garchfit*; the former setts the model in accordance with the input values p and q while the latter estimates volatility for this model and the input vector of rate-of-return values. The volatility estimate yielded by the GARCH model (light plot) against the true volatility (dark plot) is shown in Fig. 4. Here the RMSD value reached 4.0727, so GARCH (like the Kalman filter) is worse than nonparametric estimation. Note that the numerical algorithm selected the values of p and q with the minimal RMSD.

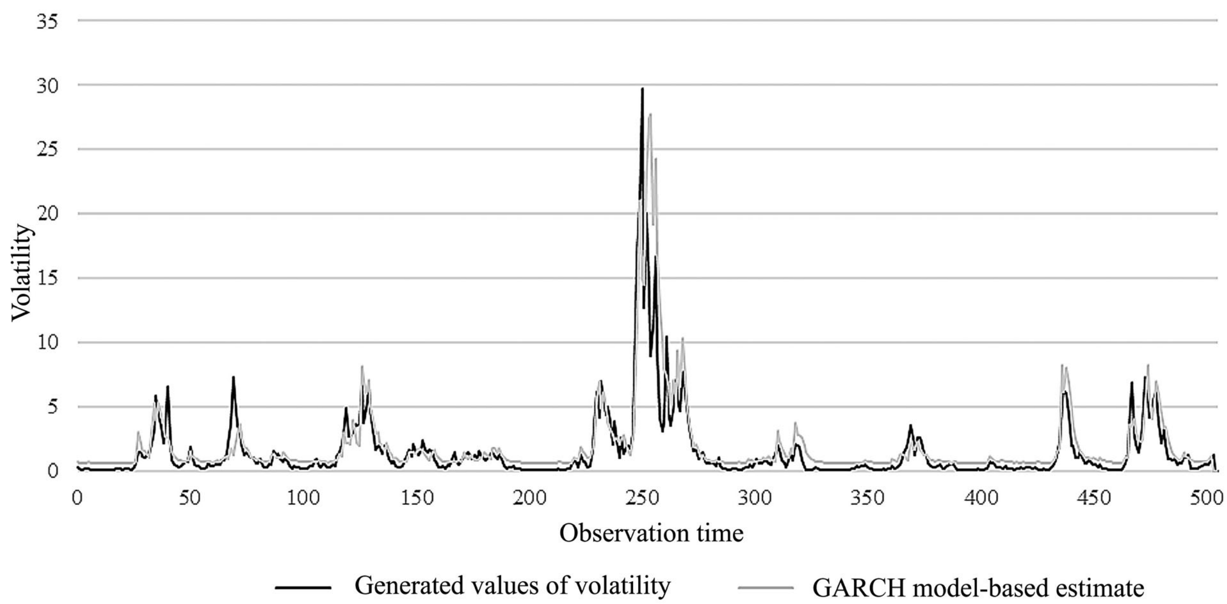

Fig. 4. Volatility estimate using GARCH model.

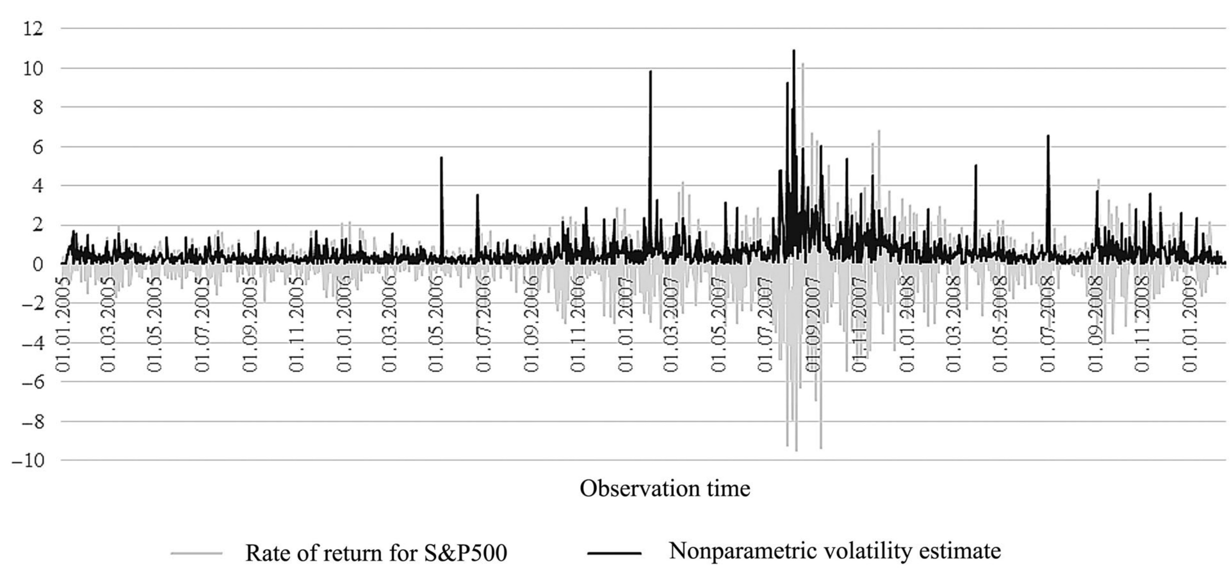

Fig. 5. Nonparametric estimate of volatility for S&P 500 index.

6.5. Nonparametric Estimation on Real Data

We also tested the nonparametric estimation methods on real data to be sure of their efficiency. Consider the daily values of S&P 500 for the period from January 1, 2005 to December 31, 2010. Recall that strong fluctuations occurred in the stock market during the 2008 World Economic Crisis. Between days 900 and 1100, the period corresponding to the global recession of 2009, the nonparametric estimate has a clustering segment, see Fig. 5. Therefore, the developed nonparametric estimation methods yield adequate results for real data too.

7. CONCLUSIONS

The table below compares three different methods with the Kushner–Stratonovich filter (denoted by K–S) in terms of RMSD, i.e.,

$$Res_i = \frac{\mathrm{RMSD}_i - \mathrm{RMSD}_{\mathrm{K\text{–}S}}}{\mathrm{RMSD}_{\mathrm{K\text{–}S}}},$$

where $i = 1, 2, 3$ corresponds to the nonparametric estimate, Kalman filter, and GARCH model, respectively. As easily seen, the best estimate of volatility was obtained using the nonparametric method. This result seems natural because nonparametric estimation can be applied even under unknown distribution of volatility (there is no need to construct rather inaccurate models of volatility). For estimation the parametric models involve state equations of the useful signal. In the case of the Kalman filter, also the non-Gaussian noises have to be replaced by the Gaussian ones with corresponding parameters. At the same time, the nonparametric method demonstrates good results under unknown distribution of the signal. Moreover, nonparametric estimation can be nonlinear, which also explains this result.

Comparison of different estimation methods of volatility with Kushner–Stratonovich filter

Method	Nonparametric estimate	Kalman filter	GARCH model
Res_i	0.15	0.28	0.65

Despite the satisfactory result obtained in this paper, our analysis has covered the diffusion component of the rate of return only; the discontinuous component as well as the memory effects of past periods have been neglected. These aspects will be studied in future research using the theory of nonparametric estimation of useful signals developed in this paper.

APPENDIX

A.1. DISTRIBUTION $\xi_t = \log(\varepsilon_t^2)$

In Eq. (7), the random variable is $\xi_t = \log(\varepsilon_t^2)$, where $\varepsilon_t \sim \mathcal{N}(0, 1)$. Introduce the random variable $\zeta_t = \varepsilon_t^2$, which has the χ^2-distribution with one degree of freedom. For calculating the distribution of $\xi_t = \log(\zeta_t)$, use the density transformation

$$p(\xi_t)d\xi_t = \phi(\zeta_t)d\zeta_t,$$

where $\phi(\zeta_t)$ is the density function of the χ^2-distribution. This density function has the form

$$\phi(\zeta_t(\xi_t)) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\xi_t}{2}\right)\exp\left(-\frac{\exp(\xi_t)}{2}\right).$$

Since the differential is

$$\frac{d\zeta_t}{d\xi_t} = \frac{d(\exp(\xi_t))}{d\xi_t} = \exp(\xi_t),$$

then

$$p(\xi_t) = \frac{\exp(\xi_t/2)}{\sqrt{2\pi}}\exp\left(-\frac{\exp(\xi_t)}{2}\right).$$

From Eq. (7) it follows that $b_t - \Delta_t = \xi_t$. In this case,

$$\frac{d\xi_t}{db_t} = 1,$$

and hence under fixed Δ_t the conditional density of all observations b_t can be written as

$$p(b_t|\Delta_t) = \frac{\exp((b_t-\Delta_t)/2)}{\sqrt{2\pi}}\exp\left(-\frac{\exp(b_t-\Delta_t)}{2}\right)$$
$$= \frac{\exp(b_t/2)}{\sqrt{2\pi}\exp(\Delta_t/2)}\exp\left(-\frac{\exp(b_t)}{2\exp(\Delta_t)}\right).$$

A.2. PARAMETERS OF RESULTING DISTRIBUTION

The mean and variance of ξ_t are easily calculated through the generating moment function $M_\xi(t) = E[\exp(t\xi)]$, i.e.,

$$E[\xi_t^k] = \frac{d^k M_\xi(t)}{dt^k}|_{t=0}.$$

The generating function for $\xi_t = \log(\zeta_t)$ has the form

$$M_\xi(t) = E[\exp(t\log(\zeta))] = E[\zeta^t].$$

In turn, the generating function for χ^2 with m degrees of freedom is given by

$$M_\zeta(t) = (1-2t)^{-m/2},$$

while the tth moment by

$$E(\zeta^t) = 2^t\frac{\Gamma(m/2\ +\ t)}{\Gamma(m/2)}.$$

Then

$$E(\xi_t) = \frac{dM_\xi(t)}{dt}|_{t=0} = \frac{dE(\zeta^t)}{dt}|_{t=0} = \frac{d\left(2^t\frac{\Gamma(1/2+t)}{\Gamma(1/2)}\right)}{dt}|_{t=0}$$
$$= \left(2^t\log 2\frac{\Gamma(1/2+t)}{\Gamma(1/2)} + 2^t\frac{\Gamma(1/2+t)\psi(1/2+t)}{\Gamma(1/2)}\right)|_{t=0} = \log 2 + \psi(1/2) \approx -1.2704,$$

where $\psi(1/2)$ denotes the digamma function.

For calculating the variance, first find $E(\xi_n^2)$:

$$E(\xi_t^2) = \frac{d^2 M_\xi(t)}{dt^2}|_{t=0} = \left(\log 2\left(2^t \log 2 \frac{\Gamma(1/2+t)}{\Gamma(1/2)} + 2^t \frac{\Gamma(1/2+t)\psi(1/2+t)}{\Gamma(1/2)}\right)\right.$$

$$\left. + \psi(1/2+t)\left(2^t \log 2 \frac{\Gamma(1/2+t)}{\Gamma(1/2)} + 2^t \frac{\Gamma(1/2+t)\psi(1/2+t)}{\Gamma(1/2)}\right) + 2^t \frac{\Gamma(1/2+t)}{\Gamma(1/2)}\psi'(1/2+t)\right)|_{t=0}$$

$$= \log 2(\log 2 + \psi(1/2)) + \psi(1/2)(\log 2 + \psi(1/2)) + \psi'(1/2) = (\log^2 + \psi'(1/2))^2 + \psi'(1/2).$$

In final analysis,

$$D(\xi_t) = E(\xi_t^2) - (E(\xi_t))^2 = \psi'(1/2) = \pi^2/2 \approx 4.9348.$$

REFERENCES

1. Merton, E., Theory of Rational Optiopricing, *Bell J. Econom. Manage. Sci.*, 1973, vol. 4, pp. 141–183.

2. Busse, J., Volatility Timing in Mutual Funds: Evidence from Daily Returns, *Rev. Financ. Stud.*, 1999, vol. 12, no. 5, pp. 1009–1041.

3. Fleming, J., Kirby, C., and Ostdiek, B., The Economic Value of Volatility Timing, *J. Finance*, 2001, vol. 56, no. 1, pp. 329–352.

4. Fleming, J., Kirby, C., and Ostdiek, B., The Economic Value of Volatility Timing Using "Realized" Volatility, *J. Finance*, 2003, vol. 67, no. 3, pp. 473–509.

5. Yao, Q. and Tong, H., Quantifying the Influence of Initial Values on Nonlinear Prediction, *J. Royal Statist. Soc.*, 1994, ser. B56, pp. 701–725.

6. Ziegelmann, F., Nonparametric Estimation of Volatility Functions: The Local Exponential Estimator, *Econometr. Theory*, 2002, vol. 18, pp. 985–91.

7. Duffie, D., Pan, J., and Singleton, K., Transform Analysis and Asset Pricing for Affine Jump-Diffusions, *Econometrica*, 2000, vol. 68, no. 6, pp. 1343–1376.

8. Jacquier, E., Polson, N., and Rossi, P., Bayesian Analysis of Stochastic Volatility Models, *J. Business Econom. Statist.*, 1994, vol. 12, pp. 371–389.

9. Bandi, F. and Reno, R., Nonparametric Stochastic Volatility, Working Paper, 2016. https://papers. ssrn. com/sol3/papers.cfm?abstract_id=1158438 (Accessed July 4, 2017).

10. Dobrovidov, A.V., Koshkin, G.M., and Vasiliev, V.A., *Non-parametric Models and Statistical Inference from Dependent Observations*, USA: Kendrick Press, 2012.

11. Bollerslev, T., Generalized Autoregressive Conditional Heteroskedasticity, *J. Econometr.*, 1986, vol. 31, pp. 307–327.

12. Taylor, S., *Modelling Financial Times Series*, New York: Wiley, 1986.

13. Shiryaev, A.N., *Osnovy stokhasticheskoi finansovoi matematiki* (Fundamentals of Stochastic Financial Mathematics), Moscow: Fazis, 1998.

14. Shiryaev, A.N., *Veroyatnost'* (Probability), Moscow: Nauka, 1989, 2nd. ed.

15. Mills, T.C., *The Econometric Modelling of Financial Time Series*, Cambridge: Cambridge Univ. Press, 1999, 2nd ed.

16. Harvey, A.C., *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge Univ. Press, 1989.

17. Kalman, R., A New Approach to Linear Filtering and Prediction Problems, *J. Basic Eng.*, 1960, vol. 82, pp. 34–45.

18. Theoret, E. and Eacicot, F., *Forecasting Stochastic Volatility Using the Kalman Filter: An Application to Canadian Interest Rates and Price-Earnings Ratio*, Munich Personal EePEc Archive, 2010. https://mpra.ub.uni-muenchen.de/35911/1/MPRA_paper_35911.pdf (Accessed July 4, 2017).

19. Dobrovidov, A.V., Nonparametric Methods of Nonlinear Filtering of Stationary Random Sequences, *Autom. Remote Control*, 1983, no. 6, pp. 757–768.

20. Kushner, H.J., On the Dynamical Equations of Conditional Probability Density Functions with Applications to Optimal Stochastic Control Theory, *J. Math. Anal. Appl.*, 1964, vol. 8, no. 2, pp. 332–344.

21. Stratonovich, R.L., *Uslovnye markovskie protsessy i ikh primenenie k teorii optimal'nogo upravleniya* (Conditional Markov Processes and Their Application in Optimal Control Theory), Moscow: Mosk. Gos. Univ., 1966.

22. Tikhonov, A.N., Ill-posed Problems of Linear Algebra and a Stable Method for Their Solution, *Dokl. Akad. Nauk SSSR*, 1965, vol. 163, no. 3, pp. 591–594.

23. Hall, P., Marron, J., and Park, B., Smoothed Cross-Validation, *Probab. Theory Relat. Fields*, 1992, vol. 90, pp. 1–20.

24. Dobrovidov, A.V. and Rud'ko, I.M., Bandwidth Selection in Nonparametric Estimator of Density Derivative by Smoothed Cross-Validation Method, *Autom. Remote Control*, 2010, vol. 71, no. 2, pp. 209–224.