Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 76 (2006) 2007–2016
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Nonparametric estimation of volatility models with serially dependent innovations

Christian M. Dahl[a,*], Michael Levine[b]

[a]Department of Economics, Purdue University, USA
[b]Department of Statistics, Purdue University, USA

Received 22 April 2005; received in revised form 17 April 2006; accepted 23 May 2006
Available online 10 July 2006

Abstract

We are interested in modelling the time series process $y_t = \sigma(x_t)\varepsilon_t$, where $\varepsilon_t = \phi_0\varepsilon_{t-1} + v_t$. This model is of interest as it provides a plausible linkage between risk and expected return of financial assets. Further, the model can serve as a vehicle for testing the martingale difference sequence hypothesis, which is typically uncritically adopted in financial time series models. When x_t has a fixed design, we provide a novel nonparametric estimator of the variance function based on the difference approach and establish its limiting properties. When x_t is strictly stationary on a strongly mixing base (hereby allowing for ARCH effects) the nonparametric variance function estimator by Fan and Yao [1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660] can be applied and seems very promising. We propose a semiparametric estimator of ϕ_0 that is $\sqrt{T}$-consistent, adaptive, and asymptotic normally distributed under very general conditions on x_t.

Keywords: Weak form volatility models; Nonparametric/Semiparametric estimation; Asymptotics

1. Introduction

In this paper we consider estimation of a time series process with an unknown and possibly time varying conditional variance function and serially dependent innovations. By allowing for dependence in the innovation process, the model provides a plausible linkage between risk and expected return of financial assets not previously analyzed. Furthermore, the model provides a vehicle for testing the martingale difference sequence hypothesis, which is typically uncritically assumed in financial time series models, such as ARCH and GARCH.

We characterize the estimated parameters of the serially correlated innovation process as a solution to a weighted least squares (WLS) problem, where the weights are given by a nonparametric estimator of the conditional variance function. This semiparametric estimator belongs to the class of so-called MINPIN estimators. By using the framework of Andrews (1994) the asymptotic properties of the

*Corresponding author. Tel.: +1 765 494 4503; fax: +1 765 496 1778.
E-mail address: dahlc@mgmt.purdue.edu (C.M. Dahl).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2006.05.018

estimated parameters in the innovation process can be established under very general conditions. If the regressors entering the variance function are strictly stationary on an α-mixing base, the nonparametric estimator of the variance function suggested by Fan and Yao (1998) can be used. However, if the design is fixed, a new and in some cases more efficient nonparametric estimator is proposed and its asymptotic properties are established. Based on simulation experiments we show that under a fixed design this novel estimator has better small sample properties than the one proposed by Fan and Yao (1998).

2. The model

Consider the following process for the time series of interest denoted $y_t \in \mathbb{R},\ t = 1, 2, \ldots, T$ where

$$y_t = \sigma_t \varepsilon_t, \tag{1}$$

$$\varepsilon_t = \phi_0 \varepsilon_{t-1} + v_t. \tag{2}$$

Furthermore, assume (i) $v_t \sim$i.i.d.$(0, 1)$, $E(|v_t|^{l+\gamma}) < \infty$ for $l = 1, \ldots, 4$ and for some $\gamma > 0$, (ii) $\phi_0 \in \Theta = (-1, 1)$, (iii) $\sigma_t^2 \equiv \sigma(x_t)^2 \in \mathscr{F} = C^2[0, 1]$, $P(\sigma_t^2 > 0) = 1$ for all $t = 1, 2, \ldots, T$, and finally (iv) ε_t is a strongly mixing sequence with mixing coefficient equal to $-(1 + 2/\delta)$ for $\delta > 0$. σ_t^2 (denoted also as σ^2) will be referred to as the variance function although strictly speaking it does not fully describe the variance structure of the model whenever $\phi_0 \neq 0$. It should be noticed that the model given by (1)–(2) belongs to the general class of function coefficient autoregressive (FAR) models, as can be seen from the following simple re-parameterization:

$$y_t = \sigma_t \sigma_{t-1}^{-1} \phi_0 y_{t-1} + \sigma_t v_t.$$

Here, the functional autoregressive coefficient is given by the term $g(x_t; \phi_0) = \sigma_t \sigma_{t-1}^{-1} \phi_0$. This coefficient is allowed to be numerically larger than unity for certain values of t, and during these periods y_t will exhibit explosive behavior. A second important feature of the model is that an increase in the variance will have a positive (negative) effect on the conditional expectation of y_t provided that ϕ_0 is positive (negative). If y_t are observations on a return series associated with a risky asset, this feature can be interpreted as a tradeoff between risk and expected return. The size and direction of such a tradeoff is of great importance in asset pricing theory and can easily be quantified using our approach. It is important to note that $\phi_0 \neq 0$ implies that the estimator of $\mathrm{var}(y_t | x_t, x_{t-1}, y_{t-1})$ generally will be inconsistent, if based on residuals from a least squares regression of y_t on y_{t-1}, due to the time varying properties of the functional autoregressive coefficient. This potential source of inconsistency has often been ignored (e.g., when estimating (G)ARCH models), due to uncritical adoption of the assumption that the innovation process is a martingale difference sequence. As a by-product of our analysis, a simple parametric test of the martingale difference hypothesis, i.e., $\phi_0 = 0$, is proposed that enables the researcher to avoid this potential pitfall. Our main interest, however, is in the estimation of σ_t^2 and ϕ_0. We will proceed under the following two alternative assumptions regarding the regressor x_t:

Case 1: x_t has a fixed design on the unit interval.

Case 2: x_t is a strictly stationary process with an α-mixing base.

Note that Case 2 encompasses the situation where $x_t = y_{t-1}$, hence allowing for the presence of ARCH effects. The estimation procedure is simple and consists of two stages: In the first stage the estimator of σ_t^2—denoted $\widehat{\sigma}_t{}^2$—is obtained. Secondly, a semiparametric estimator of ϕ_0 is computed using WLS, where the weights are constructed using $\widehat{\sigma}_t{}^2$. This semiparametric estimator belongs to the class of MINPIN estimators introduced by Andrews (1994). In Case 1 we propose a novel nonparametric estimator based on the difference approach, which turns out to have nice asymptotic properties and is very easy to handle computationally. In Case 2 the estimator proposed by Fan and Yao (1998) seems promising. We begin, however, by characterizing the asymptotic properties of the MINPIN estimator.

3. Characterization of $\widehat{\phi}$ and its asymptotics

Consider the objective function $d(\sigma_t, \sigma_{t-1}; \phi) = m(\sigma_t, \sigma_{t-1}; \phi)^2$, where

$$m(\sigma_t, \sigma_{t-1}; \phi) = \sigma_t^{-1}\sigma_{t-1}^{-1} y_t y_{t-1} - \sigma_{t-1}^{-2}\phi y_{t-1}^2 = v_t \varepsilon_{t-1}. \tag{3}$$

Since $d(\sigma_t, \sigma_{t-1}; \phi)$ is unobservable (as σ_t is unknown), a GMM estimator of ϕ_0 can be defined as the minimizer of the sample analog of $E(d(\sigma_t, \sigma_{t-1}; \phi))$, i.e.,

$$\widehat{\phi} = \arg\min_{\phi\in\Theta}(2T)^{-1}\sum_{t=2}^{T} d(\widehat{\sigma}_t, \widehat{\sigma}_{t-1}; \phi) = \left((1/T)\sum_{t=2}^{T}\widehat{\sigma}_{t-1}^{-2} y_{t-1}^2\right)^{-1}\left((1/T)\sum_{t=2}^{T}\widehat{\sigma}_t^{-1}\widehat{\sigma}_{t-1}^{-1} y_t y_{t-1}\right). \tag{4}$$

Before characterizing the asymptotic properties of $\widehat{\phi}$, the following regularity conditions on $m(\sigma_t, \sigma_{t-1}; \phi)$ and its derivative need to be established.

Lemma 1. *$m_t(\sigma; \phi) = m(\sigma_t, \sigma_{t-1}; \phi)$ is twice continuously differentiable in ϕ on Θ, $\forall \sigma \in \mathscr{F}$ and $\forall t \geqslant 1$. $m_t(\sigma; \phi)$ and $(\partial/\partial\phi) m_t(\sigma; \phi)$ satisfy a uniform WLLN on $\Theta \times \mathscr{F}$. Moreover, $m(\sigma; \phi) = \lim_{T\to\infty}(1/T)\sum_{t=2}^{T} E(m(\sigma_t, \sigma_{t-1}; \phi))$ and $M = \lim_{T\to\infty}(1/T)\sum_{t=1}^{T} E(\partial m_t(\sigma; \phi)/\partial\phi)$ exist uniformly over $\Theta \times \mathscr{F}$ and are continuous at (σ, ϕ_0) with respect to some pseudo-metric on $\Theta \times \mathscr{F}$ for which $(\widehat{\sigma}, \widehat{\phi}) \xrightarrow{p} (\sigma, \phi_0)$.*

Proof of Lemma 1. We will begin by verifying that $m_t = m_t(\sigma; \phi)$ satisfies a uniform WLLN on $\Theta \times \mathscr{F}$ following Andrews (1987): Assumption A1 in Andrews (1987) is trivially satisfied. Assumption A2 is satisfied since $m_t = v_t \sum_{i=0}^{\infty} \phi^i v_{t-1-i}$ and consequently $m_t \xrightarrow{p} 0$ uniformly on the interior of $\Theta \times \mathscr{F}$ (not only locally in a closed ball around ϕ). Next, define $m_t^* = m_t(\sigma_t^*, \sigma_{t-1}^*; \phi^*)$ and consider

$$|m_t^* - m_t| = \left| v_t \sum_{i=0}^{\infty} \phi^{*i} v_{t-1-i} - v_t \sum_{i=0}^{\infty} \phi^i v_{t-1-i} \right| \leqslant \sqrt{\sum_{i=0}^{\infty} \phi^{2i} v_{t-1-i}^2 v_t^2}\sqrt{\sum_{i=0}^{\infty}\left(\frac{\phi^{*i} - \phi^i}{\phi^i}\right)^2}.$$

By letting $b_t(v_t, v_{t-1}, \phi) = \sqrt{\sum_{i=0}^{\infty} \phi^{2i} v_{t-1-i}^2 v_t^2}$ and $\rho(\phi^*, \phi) = \sqrt{\sum_{i=0}^{\infty}((\phi^{*i} - \phi^i)/\phi^i)^2}$ this implies that

$$\sup_T (1/T)\sum_{t=2}^{T} E b_t(v_t, v_{t-1}, \phi) \leqslant \sup_T (1/T)\sum_{t=2}^{T}\sqrt{E\left(\sum_{i=0}^{\infty}\phi^{2i} v_{t-1-i}^2 v_t^2\right)} = \sqrt{\frac{1}{1-\phi^2}},$$

and $\rho(\phi^*, \phi) \downarrow 0$ as $\phi^* \to \phi$. Consequently, Assumption 4 in Andrews (1987) holds and accordingly (using Corollary 2 in Andrews, 1987) we can conclude that m_t satisfies the uniform WLLN over $\Theta \times \mathscr{F}$. Next, note that $\partial m_t/\partial\phi = v_t \sum_{i=0}^{\infty} \phi^i v_{t-2-i}$. Using similar steps as above it follows straightforwardly that also for $\partial m_t/\partial\phi$ Assumptions A1, A2, and A4 in Andrews (1987) apply, hence it satisfies the UWLLN uniformly on $\Theta \times \mathscr{F}$. As m_t and $\partial m_t/\partial\phi$ do not depend on σ_t, Corollary 2 in Andrews (1987) also establishes uniform continuity of $m = \lim_{T\to\infty}(1/T)\sum_{t=1}^{T} E(m_t(\phi, \sigma_t))$ and of $M = \lim_{T\to\infty}(1/T)\sum_{t=1}^{T} E(\partial m_t/\partial\phi) = 1/(1-\phi^2)$. Finally, notice that m_t is twice differentiable in ϕ uniformly on Θ which completes the proof. □

Immediately the following asymptotic results can be established.

Theorem 1. *Let data be generated according to model (1)–(2) under Assumptions (i)–(iv) with x_t defined as in Case 1 or Case 2. Let $\widehat{\sigma}_t^2$ be a nonparametric estimator of σ_t^2 and suppose (1) $\sup_{\phi\in\Theta}\|\widehat{\sigma}_t^2 - \sigma_t^2\| \xrightarrow{p} 0$ for some $\sigma_t^2 \in \mathscr{F}$ and (2) $P(\widehat{\sigma}_t^2 \in \mathscr{F}) \xrightarrow{p} 1$. Then $\widehat{\phi} \xrightarrow{p} \phi_0$.*

Proof of Theorem 1. In addition to the requirements (1) and (2), the consistency result requires uniform continuity of $d_t = d(\sigma_t, \sigma_{t-1}; \phi)$ and the existence of a unique minimizer of $d(\sigma_t, \sigma_{t-1}; \phi)$, see, e.g., Theorem A-1 in Andrews (1994). Uniform continuity of d_t follows directly from Lemma 1 as $\sup_{(\phi,\sigma)\in\Theta\times\mathscr{F}}\|m(\sigma; \phi)\| = 0$. The existence of a unique minimizer of d_t on Θ follows from the compactness of Θ, continuity of d_t, and since $\partial^2 d_t/(\partial\phi)^2 = (\sigma_{t-1}^{-2} y_{t-1}^2)^2 > 0$. □

To establish asymptotic normality of $\widehat{\phi}$ define $\overline{m}_T(\sigma; \phi) = (1/T)\sum_{t=1}^{T} m(\sigma_t, \sigma_{t-1}; \phi)$. The regularity conditions on $m(\sigma_t, \sigma_{t-1}; \phi)$, established by Lemma 1, allow a mean value expansion of $\sqrt{T}\overline{m}_T(\widehat{\sigma}; \widehat{\phi})$ about ϕ_0

given as

$$\sqrt{T}\overline{m}_T(\widehat{\sigma};\widehat{\phi}) = \sqrt{T}\overline{m}_T(\widehat{\sigma};\phi_0) + \frac{\partial}{\partial\phi}\overline{m}_T(\widehat{\sigma};\phi^*)\sqrt{T}(\widehat{\phi}-\phi_0), \tag{5}$$

where ϕ^* lies between $\widehat{\phi}$ and ϕ_0. There are basically three steps involved to establish the asymptotic normality of $\sqrt{T}(\widehat{\phi}-\phi_0)$. If (a) $\lim_{T\to\infty}(\partial/\partial\phi)\overline{m}_T(\widehat{\sigma};\phi^*) \xrightarrow{p} M$, where M is given by Lemma 1 then

$$\sqrt{T}(\widehat{\phi}-\phi_0) = -M^{-1}\left(\mathrm{o_p}(1) + \sqrt{T}\overline{m}_T(\widehat{\sigma};\phi_0)\right), \tag{6}$$

since $\sqrt{T}\overline{m}_T(\widehat{\sigma};\widehat{\phi}) = \mathrm{o_p}(1)$, where $\widehat{\phi}$ solves the first order condition $\overline{m}_T(\widehat{\sigma};\widehat{\phi}) = 0$ and $\widehat{\phi}$ belongs to the interior of Θ wp$\to$1. Consequently, the asymptotic normality of $\sqrt{T}(\widehat{\phi}-\phi_0)$ follows if (b) $\sqrt{T}\overline{m}_T(\sigma;\phi_0)$ is asymptotically normally distributed and (c)$\sqrt{T}(\overline{m}_T(\widehat{\sigma};\phi_0) - \overline{m}_T(\sigma;\phi_0)) \xrightarrow{p} 0$. Let $\overline{m}^*_T(\sigma;\phi) = (1/T)\sum_{t=1}^T E(m(\sigma_t,\sigma_{t-1};\phi))$ and $\upsilon_T(\sigma) = \sqrt{T}((1/T)\sum_{t=1}^T m(\sigma_t,\sigma_{t-1};\phi) - \overline{m}^*_T(\sigma;\phi))$ such that

$$\sqrt{T}(\overline{m}_T(\widehat{\sigma};\phi_0) - \overline{m}_T(\sigma;\phi_0)) = \upsilon_T(\widehat{\sigma}) - \upsilon_T(\sigma) - \sqrt{T}\overline{m}^*_T(\widehat{\sigma};\phi).$$

Then, condition (c) is true if $\upsilon_T(\widehat{\sigma})$ is stochastic equicontinuous at σ and $\sqrt{T}\overline{m}^*_T(\sigma;\phi) \xrightarrow{p} 0$. In what follows we first show that $\sqrt{T}\overline{m}^*_T(\sigma;\phi) \xrightarrow{p} 0$ (Lemma 3). Based on this result, it is relatively easy to show that condition (b) holds (Lemma 5).

Lemma 2. *Given the assumptions of model* (1)–(2), $\sqrt{T}\overline{m}_T(\widehat{\sigma};\phi_0) \xrightarrow{p} 0$ *in Case* 1 *and* 2.

Proof of Lemma 2. Note that for any $\varepsilon > 0$

$$\sqrt{T}\overline{m}^*_T(\widehat{\sigma};\phi_0) = \sqrt{T}\overline{m}^*_T(\sigma;\phi_0)1(\rho_{\mathscr{F}}(\widehat{\sigma},\sigma)\leqslant\varepsilon)|_{\sigma=\widehat{\sigma}} + \sqrt{T}\overline{m}^*_T(\widehat{\sigma};\phi_0)1(\rho_{\mathscr{F}}(\widehat{\sigma},\sigma)>\varepsilon),$$

where $\rho_{\mathscr{F}}(\widehat{\sigma},\sigma)$ is a pseudometric defined on $\mathscr{F}$. As $E(m(\sigma_t,\sigma_{t-1};\phi_0)) = 0$ uniformly on $\mathscr{F}$, it follows that $\sqrt{T}\overline{m}^*_T(\widehat{\sigma};\phi_0) = 0 + \mathrm{o_p}(1)$ where the last term is a result of consistency of $\widehat{\sigma}$ with respect to σ. □

Lemma 3. *Define* $\upsilon_T(\sigma) = \sqrt{T}((1/T)\sum_{t=1}^T m(\sigma_t,\sigma_{t-1};\phi) - \overline{m}^*_T(\sigma;\phi))$. *Then*, $\upsilon_T(\sigma) \xrightarrow{d} \mathrm{N}(0, 1/(1-\phi^2))$ *in Case* 1 *and* 2.

Proof of Lemma 3. It follows from Lemma 2 that $\upsilon_T(\sigma) = (1/\sqrt{T})\sum_{t=1}^T v_t\varepsilon_{t-1} + \mathrm{o_p}(1)$ where $v_t\varepsilon_{t-1}$ is a martingale difference sequence with an α-mixing base (given by current and lagged values of v_t) defined uniformly on $\Theta\times\mathscr{F}$ with finite variance $\mathrm{var}(v_t\varepsilon_{t-1}) = 1/(1-\phi^2)$. The result of Lemma 3 then follows in a straightforward manner from, e.g., Theorem 7.11 in Bierens (2005). □

Before showing that $\upsilon_T(\widehat{\sigma})$ is stochastic equicontinuous we need the following result.

Lemma 4. *Given the assumptions of model* (1)–(2), $T^{-1/2}\sum_{t=2}^T \varepsilon^2_{t-1} = \mathrm{O_p}(1)$ *in Case* 1 *and* 2.

Proof of Lemma 4. Define an increasing sequence of σ-fields as $F_t = \sigma((y_t,x_t),(y_{t-1},x_{t-1}),\ldots,(y_1,x_1))$ such that $\{\varepsilon^2_t, F_t\}_t$ is an adaptive stochastic sequence. Since $E(\varepsilon^2_t) = 1/(1-\phi^2) < \infty$, then $\{Z_t, F_t\}_t$ for $Z_t = \varepsilon^2_t - E(\varepsilon^2_t|F_{t-1})$ is a martingale difference sequence on an α-mixing base. Furthermore, note that

$$E(Z^2_t) = E((v^2_t - 1 + 2\phi v_t\varepsilon_{t-1})^2) = E(v^4_t) + 4\phi^2/(1-\phi^2) - 1 < \infty,$$

since $E(v^4_t) = \mu_4 < \infty$ by Assumption (ii) and due to independence of v_t and ε_{t-1}. Consequently, it follows from, e.g., Theorem 7.11 in Bierens (2005) that $\sqrt{T}\sum_{t=2}^T Z_t \xrightarrow{d} \mathrm{N}(0, E(Z^2_t))$. The desired result follows from strict stationarity of ε^2_t and since $E(Z^2_t) < \infty$. □

Lemma 5. *Let* $\upsilon_T(\cdot)$ *be defined as in Lemma* 3. *Then*, $\upsilon_T(\widehat{\sigma})$ *is stochastic equicontinuous at* σ.

Proof of Lemma 5. Write $m(\sigma_t,\sigma_{t-1};\phi) = \boldsymbol{S}'_t\tau$ where

$$\boldsymbol{S}_t = \begin{bmatrix} y_t y_{t-1} \\ \phi y^2_{t-1} \end{bmatrix} = \begin{bmatrix} \sigma_t\sigma_{t-1}\phi\varepsilon^2_{t-1} + \sigma_t\sigma_{t-1}v_t\varepsilon_{t-1} \\ \phi\sigma^2_{t-1}\varepsilon^2_{t-1} \end{bmatrix},$$

and $\tau = (\sigma_t^{-1}\sigma_{t-1}^{-1}, \sigma_{t-1}^{-2})'$. Let $\rho_{\mathscr{F}}(\cdot,\cdot)$ be the Euclidian metric and note that

$$\begin{aligned}\lim_{T\to\infty} P\left(\sup_{\rho_{\mathscr{F}}(\widehat{\sigma},\sigma)<\delta'} |\upsilon_T(\widehat{\sigma}) - \upsilon_T(\sigma)| > \eta\right) &= \lim_{T\to\infty} P\left(\sup_{\rho_{\mathscr{F}}(\widehat{\sigma},\sigma)<\delta'} \left|T^{-1/2}\sum_{t=2}^{T}(\boldsymbol{S}_t - E(\boldsymbol{S}_t)'(\widehat{\tau}-\tau))\right| > \eta\right) \\ &\leqslant \lim_{T\to\infty} P\left(\sup_{\rho_{\mathscr{F}}(\widehat{\sigma},\sigma)<\delta'} \left\|T^{-1/2}\sum_{t=2}^{T}\boldsymbol{S}_t - E(\boldsymbol{S}_t)\right\| > \eta/\delta'\right) \\ &\to 0,\end{aligned}$$

provided that (a)$T^{-1/2}\sum_{t=2}^{T}(\boldsymbol{S}_t - E(\boldsymbol{S}_t)) = \mathrm{O_p}(1)$ and (b) δ' is sufficiently small. In Case 1, condition (a) is satisfied from the results of Lemmas 3 and 4 (implying that $T^{-1/2}\sum_{t=2}^{T}v_t\varepsilon_{t-1} = \mathrm{O_p}(1)$ and $T^{-1/2}\sum_{t=2}^{T}\varepsilon_{t-1}^2 = \mathrm{O_p}(1)$) and because $0<\sigma_t^2<\infty$ for all t. In Case 2, use $T^{-1/2}\sum_{t=2}^{T}\varepsilon_{t-1}^2 = T^{-1/2}\sum_{t=2}^{T}\sigma_t^{-2}y_{t-1}^2 = \mathrm{O_p}(1)$. It then follows that

$$\max_{1<j\leqslant T}(\sigma_j^2)T^{-1/2}\sum_{t=2}^{T}y_{t-1}^2 \leqslant T^{-1/2}\sum_{t=2}^{T}\sigma_t^{-2}y_{t-1}^2 = \mathrm{O_p}(1),$$

hence $T^{-1/2}\sum_{t=2}^{T}y_{t-1}^2 = \mathrm{O_p}(1)$. Furthermore,

$$T^{-1/2}\sum_{t=2}^{T}y_ty_{t-1} = \phi T^{-1/2}\sum_{t=2}^{T}\sigma_t\sigma_{t-1}^{-1}y_{t-1}^2 + T^{-1/2}\sum_{t=2}^{T}\sigma_t\sigma_{t-1}v_t\varepsilon_{t-1} = \mathrm{O_p}(1),$$

due to the previous established result and since $\sigma_t\sigma_{t-1}v_t\varepsilon_{t-1}$ is a martingale difference sequence on an α-mixing base with bounded variance, i.e.,

$$E(\sigma_t^2\sigma_{t-1}^2v_t^2\varepsilon_{t-1}^2) \leqslant \max_t(\sigma_t^2\sigma_{t-1}^2)(1-\phi^2)^{-1} < \infty,$$

as $\sigma_t^2\sigma_{t-1}^2$ is bounded. Consequently, it can be concluded that $\upsilon_T(\widehat{\sigma})$ is stochastic equicontinuous at σ. Condition (b) is trivially satisfied due to consistency of $\widehat{\sigma}_t{}^2$ with respect to σ_t^2. □

Theorem 2. *Let the assumptions of Theorem* 1 *hold. Then, under Case* 1 *and* 2, $\sqrt{T}(\widehat{\phi}-\phi_0) \xrightarrow{d} \mathrm{N}(0, 1-\phi_0^2)$.

Proof of Theorem 2. Consider the mean value expansion given by (5). From Lemma 1 we have that $(\partial/\partial\phi)\overline{m}_T(\widehat{\sigma};\phi^*) \xrightarrow{p} (1-\phi_0^2)^{-1}$. The results of Lemmas 2 and 5 imply that $\sqrt{T}\overline{m}_T(\widehat{\sigma};\phi_0)$ and $\upsilon_T(\sigma) = \sqrt{T}\overline{m}_T(\sigma;\phi_0)$ are asymptotically equivalent and that their asymptotic distribution (from Lemma 3) is given by $\mathrm{N}(0,(1-\phi_0^2)^{-1})$. Consequently,

$$\begin{aligned}\sqrt{T}(\widehat{\phi}-\phi_0) &= -(1-\phi_0^2)(\mathrm{o_p}(1) + \mathrm{N}(0,(1-\phi_0^2)^{-1})) \\ &= \mathrm{N}(0,(1-\phi_0^2)) + \mathrm{o_p}(1),\end{aligned}$$

which completes the proof. □

Theorem 2 gives conditions under which $\widehat{\phi}$ is not dependent on the estimator of σ_t^2 asymptotically and is asymptotically equivalent to the maximum likelihood estimator of ϕ_0 given that ε_t is observable. Consequently, the $\widehat{\phi}$ is asymptotically efficient. In addition, since $\widehat{\phi}$ is an efficient estimator of ϕ_0 and ϕ_0 is the only unknown parameter in the model, the estimator is adaptive, see, e.g., Andrews (1994, p. 59). Finally, it is noteworthy that $\widehat{\phi}$, as many semiparametric estimators, converge to ϕ_0 at the parametric $\sqrt{T}$-rate.

4. Asymptotics of the variance function estimators

To characterize the asymptotic properties of the variance function estimator we will use the asymptotic mean squared error (AIMSE), which consists of the first two terms of a Taylor expansion of integrated mean squared error given as $\mathrm{IMSE} = \int_0^1 E(\widehat{\sigma}_t{}^2 - \sigma_t^2)^2\,\mathrm{d}t$. The minimal value of AIMSE achieved at the optimal (minimizing) bandwidth is referred to as $\mathrm{AIMSE_o}$.[1]

4.1. Case 1: the difference based estimator

Following the so-called difference sequence based approach by Hall et al. (1990) and Levins (2003), we define $\eta_t = (1/2)^{-1/2}(y_t - y_{t-2})$ and consider the local linear estimator $\widehat{\sigma}_t{}^2 = \widehat{\sigma}^2(x_t)$ given as $\widehat{a}$ that solves the problem

$$(\hat{a},\hat{b}) = \arg\min_{a,b} \sum_{t=3}^{T} (\eta_t^2 - a - b(x_t - x))^2 K_h(x_t - x), \tag{7}$$

where $K_h(\cdot)$ is a kernel function. The choice of η_t is motivated by the observation that for any stationary AR(1) time series process the difference between the variance, $\gamma_0 = \mathrm{var}(\varepsilon_t)$, and the covariance, $\gamma_2 = \mathrm{cov}(\varepsilon_t, \varepsilon_{t-2})$, equals unity. We use this property to establish the following consistency result.

Theorem 3. *Let data be generated according to model* (1)–(2) *under Case* 1. *Suppose that $K(u)$ is a second order non-negative kernel function satisfying*: $K(u) \geqslant 0$ *for any* $u \in [-1,1]$, $\mu_1 = \int K(u)\,\mathrm{d}u = 0$, $\sigma_K^2 \equiv \mu_2 = \int u^2 K(u)\,\mathrm{d}u \neq 0$, *and* $R_K = \int K(u)^2\,\mathrm{d}u$. *Then, the estimator given by* (7) *is consistent in mean square with convergence rate* $\mathrm{O}(T^{-4/5})$. *Furthermore, the optimal* (*in the sense of* Parzen, 1962 *and* Rosenblatt, 1956) *bandwidth is* $h = \mathrm{O}(T^{-1/5})$.

Proof of Theorem 3. From (7), $E(\widehat{\sigma}_t{}^2) = \boldsymbol{e}'(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}/T)^{-1}\boldsymbol{X}'\boldsymbol{W}E(\eta_t^2)/T$, where $\boldsymbol{e} = (1,0)'$, $\boldsymbol{W} = \mathrm{diag}(K_h(x_1 - x), \ldots, K_h(x_T - x))$ for $K_h(\cdot) = h^{-1}K(\cdot/h)$, and a typical row of $\boldsymbol{X}$ is $(1, (x - x_t))$. Existence of two continuous derivatives of σ_t^2 guarantees that $\sigma_t^2 = \sigma^2 - \mathrm{D}\sigma^2(x - x_t) + \mathrm{D}^2\sigma^2(x - x_t)^2/2 + \mathrm{o}(h^2)$. Writing $\eta_t^2 = \frac{1}{2}(\sigma_t^2\varepsilon_t^2 + \sigma_{t-2}^2\varepsilon_{t-2}^2 - 2\sqrt{\sigma_t^2\sigma_{t-2}^2}\varepsilon_t\varepsilon_{t-2})$, using the expansions for σ_t^2 and σ_{t-2}^2, and the fact that $\sqrt{1+x} = 1 + x/2 + \mathrm{o}(x)$ for small x gives

$$\begin{aligned} E(\eta_t^2) = {} & (\gamma_0 - \gamma_2)\sigma^2 - \gamma_0 \mathrm{D}\sigma^2[(x - x_t) + (x - x_{t-2})]/2 \\ & + \gamma_0 \mathrm{D}^2\sigma^2[(x - x_t)^2 + (x - x_{t-2})^2]/2 + \gamma_2 \mathrm{D}\sigma^2[(x - x_t) + (x - x_{t-2})]/2 \\ & - \gamma_2 \mathrm{D}^2\sigma^2[(x - x_t)^2 + (x - x_{t-2})^2]/4 - \gamma_2(\mathrm{D}\sigma^2)^2(x - x_t)(x - x_{t-2})/\sigma^2 + \mathrm{o}(h^2), \end{aligned} \tag{8}$$

where $\mathrm{cov}(\varepsilon_t, \varepsilon_{t-l}) \equiv \gamma_l$. Defining $s_r(x;h) = T^{-1}\sum_t (x - x_t)^r K_h(x - x_t)$ and $s_{r,m}(x,h) = T^{-1}\sum_t (x - x_t)^r (x - x_{t-2})^m K_h(x - x_t)$ and noticing that $s_r(x;h) = s_{r-1,1}(x,h) = h^r \int_{-1}^{1} u^r K(u)\,\mathrm{d}u + \mathrm{O}(T^{-1})$ implies

$$(\boldsymbol{X}'\boldsymbol{W}\boldsymbol{X}/T)^{-1} = \begin{bmatrix} 1 + \mathrm{O}(T^{-1}) & \mathrm{O}(T^{-1}) \\ \mathrm{O}(T^{-1}) & h^2\sigma_k^2 + \mathrm{O}(T^{-1}) \end{bmatrix}^{-1}.$$

In addition, as the first entry in $\boldsymbol{X}'\boldsymbol{W}E(\eta^2)/T$ equals

$$(\boldsymbol{X}'\boldsymbol{W}E(\eta_t^2)/T)_1 = \sigma^2 + [\mathrm{D}^2\sigma^2/4 - \gamma_2(\mathrm{D}\sigma^2)^2/\sigma^2]h^2\sigma_K^2/2 + \mathrm{o}(h^2) + \mathrm{O}(T^{-1}),$$

it follows that $\mathrm{Bias}(\widehat{\sigma}_t{}^2) = [\mathrm{D}^2\sigma^2/4 - \gamma_2(\mathrm{D}\sigma^2)^2/\sigma^2]h^2\sigma_K^2/2 + \mathrm{o}(h^2) + \mathrm{O}(T^{-1})$. Using similar techniques, the variance of $\widehat{\sigma}_t{}^2$ can be found as

$$\mathrm{var}(\widehat{\sigma}_t{}^2) = R_K C(\phi_0)\sigma^4(Th)^{-1} + \mathrm{o}((Th)^{-1}), \tag{9}$$

where $C(\phi)$ is a constant that depends on ϕ_0 only. Finally, the optimal bandwidth can be found to be

$$h = T^{-1/5}\left(C(\phi_0)R_K \int \sigma_t^4\,\mathrm{d}t\right)^{1/5} \Bigg/ \left(\sigma_K^4 \int [D^2\sigma_t^2/4 - \gamma_2(\mathrm{D}\sigma_t^2)^2/\sigma_t^2]^2\,\mathrm{d}t\right)^{1/5}, \tag{10}$$

[1]The following notation will be used: $\mathrm{D}^k f(t) = d^k f(t)/(\mathrm{d}t)^k$.

and consequently

$$\mathrm{AIMSE_o} = (5/4)T^{-4/5}\left(R_K C(\phi)\int \sigma_t^4\,\mathrm{d}t\right)^{4/5}\left(\sigma_K^4\int [D^2\sigma_t^2/4 - \gamma_2(\mathrm{D}\sigma_t^2)^2/\sigma_t^2]^2\,\mathrm{d}t\right)^{1/5}.$$

Hence, the optimal AIMSE is of the order $\mathrm{O}(T^{-4/5})$ and the variance estimator $\widehat{\sigma}_t{}^2$ converges in the mean square (pointwise), i.e., $E[(\widehat{\sigma}_t{}^2 - \sigma^2)^2] = \mathrm{O}(T^{-4/5})$ hereby completing the proof. $\square$

A few remarks are in order here. First, note that the quadratic functional

$$\int_0^1 [D^2\sigma_t^2/4 - \gamma_2(\mathrm{D}\sigma_t^2)^2/\sigma_t^2]^2\,\mathrm{d}t, \tag{11}$$

characterizes the degree of curvature of the function σ_t^2 corrected for the correlation present in the data. The larger the expression in (11), the smaller the bandwidth we have to choose. Secondly, the rates of convergence are identical to those obtained for the kernel regression estimator of the mean function under identical smoothness requirements, see, e.g., Simonoff (1996). Thirdly, as an immediate consequence of Theorem 3, $\mathrm{D}^k\widehat{\sigma}_t{}^2 \xrightarrow{p} \mathrm{D}^k\sigma^2$ for any positive integer k. In addition to the result of Theorem 3, these results are very useful in obtaining consistency of $\widehat{\phi}$ as they imply that conditions (1) and (2) in Theorem 1 hold. Finally, notice that since $\widehat{\sigma}_t{}^2$ converges in L_2-sense, it also converges in probability at the rate $\mathrm{O_p}(1/\sqrt{Th})$.

Theorem 4. *Let the assumptions of Theorem* 1 *hold. Then,*

$$\widehat{\sigma}_t{}^2 \xrightarrow{d} \mathrm{N}(E(\widehat{\sigma}_t^2), \mathrm{var}(\widehat{\sigma}_t{}^2)), \tag{12}$$

as $T\to\infty$, $h\to 0$ *and* $Th\to\infty$, *where* $E(\widehat{\sigma}_t{}^2) = \sigma_t^2 + \mathrm{Bias}(\widehat{\sigma}_t{}^2)$ *and where the expression of bias and variance of* $\widehat{\sigma}_t^2$ *are as given in the proof of Theorem* 3.

Proof of Theorem 4. Since $\widehat{\sigma}_t{}^2$ can be written as a partial sum process, i.e., $\widehat{\sigma}_t{}^2 = \sum_{i=3}^T a_T(x_i, x_t)\eta_i^2$ where,

$$a_T(x_t; x) = T^{-1}\frac{(s_2(x;h) - s_1(x;h)(x_t - x))K_h(x_t - x)}{s_2(x;h)s_0(x;h) - s_1(x;h)^2},$$

asymptotic normality of $\widehat{\sigma}_t{}^2$ can be shown using Theorem 2.2(c) in Peligrad and Utev (1997). First, the ‘‘kernel function’’ $a_T(x_t, x)$ must satisfy (2.1) in Peligrad and Utev (1997), which consists of two conditions: (1) it requires that $\max_{1\leqslant t\leqslant T}|a_T(x_t;x)| \to 0$ as $T\to\infty$. This follows immediately from the fact that (a) the kernel function $K(\cdot)$ has bounded support and (b) its first moment is equal to zero. Indeed, asymptotically as $h\to 0$, $T\to\infty$, and $Th\to\infty$, $s_2(x;h)s_0(x;h) - s_1(x;h)^2 \to h^2\sigma_K^2$ and the same is true of $s_2(x;h) - s_1(x;h)(x_t - x)$. Therefore, the entire coefficient asymptotically behaves as $K((x_t - x)/h)/Th$ and therefore $\max_{1\leqslant t\leqslant T}K((x_t - x)/h)/Th \to 0$ as $T\to\infty$ and $Th\to\infty$, as desired. (2) it requires that $\sup_T \sum a_T^2(x_t;x) < \infty$. To establish that this condition holds just note that asymptotically $a_T^2(x_t;x) = (K((x_t - x)/h))^2/(Th)^2$. As the support of the kernel function $K(\cdot)$ is bounded, the infinite sum of $a_T^2(x_t;x)$ will converge and we can conclude that the entire (2.1) in Peligrad and Utev (1997) holds.Next, condition (2.2) in Peligrad and Utev (1997) needs to be verified: Uniform integrability of $\eta_t^{4+\gamma}$ for some $\gamma>0$ follows directly from Shiryaev (1996), with $G(t) = t^{4+\gamma}$ for $\gamma>0$ such that $\lim_{t\to\infty} G(t)/t = \infty$ and $\sup_t E(|\eta_t|^{4+\gamma}) < \infty$. The last condition is due to boundedness of σ^2 and since $E|v_t|^{4+\gamma} < \infty$. The remaining conditions of Theorem 2.2(c) are easily established: η_t^2 is strongly mixing as it is a measurable function of the strongly mixing process ε_t and $\inf_t \mathrm{var}(\eta_t^2) > 0$ follows from the assumption that $\sigma_t^2 > 0$ for $\forall t$. Finally, $\sum_t \alpha(t)t^{2/\delta} < \infty$ follows as ε_t, and therefore—by assumption—η_t^2, has mixing coefficient equal to $-(1+2/\delta)$. $\square$

4.2. Case 2: the Fan and Yao estimator

Here the variance function estimator suggested by Fan and Yao (1998) seems natural to employ with $E(y_t|\boldsymbol{x}_t) = \phi_0 y_{t-1}\sigma(y_{t-1})/\sigma(y_{t-2})$ and $\mathrm{var}(y_t|\boldsymbol{x}_t) = \sigma(y_{t-1})^2$. Under conditions similar to the ones assumed in relation to the model given by (1)–(2), Fan and Yao prove consistency and asymptotic normality of the estimators of $E(y_t|\boldsymbol{x}_t)$ and $\mathrm{var}(y_t|\boldsymbol{x}_t)$, denoted $\widehat{a}_t$ and $\widehat{\alpha}_t$, respectively, and given as

$$(\widehat{a}_t, \widehat{\boldsymbol{b}}_t) = \arg\min_{a_t, \boldsymbol{b}_t} \sum_{s=3}^{T} (y_s - a_t - (\boldsymbol{x}_s - \boldsymbol{x}_t)\boldsymbol{b}_t)^2 K^1_{h_1}(\boldsymbol{x}_s - \boldsymbol{x}_t), \tag{13}$$

$$(\widehat{\alpha}_t, \widehat{\beta}_t) = \arg\min_{\alpha_t, \beta_t} \sum_{s=3}^{T} (\widehat{r}_s^2 - \alpha_t - (y_{s-1} - y_{t-1})\beta_t)^2 K^2_{h_2}(y_{s-1} - y_{t-1}), \tag{14}$$

where $\widehat{r}_t = y_t - \widehat{a}_t$ and $K^1_{h_1}(\cdot)$, $K^2_{h_2}(\cdot)$ are kernel functions. It should be noticed that the Fan and Yao variance function estimator also works in Case 1. However, in this case $\boldsymbol{x}_t = (x_t, x_{t-1}, y_{t-1})$ and convergence can be expected to be relatively slow. The finite sample efficiency relative to the difference based estimator in Case 1 might also be affected adversely by bandwidth selection which has to be performed twice.

5. Simulations

In this section properties of the estimators $\widehat{\sigma}_t{}^2$ and $\widehat{\phi}$ are studied using simulations. We consider the data being generated by model (1)–(2) for alternative choices of variance functions and for various values of ϕ_0. The specification of σ_t^2 in Model 1 is a leading example in econometrics/statistics and can generate ARCH-effects when $x_t = y_{t-1}$. Model 2 is adapted from Fan and Yao (1998). In particular, the choice of σ_t^2 is identical to the variance function in their Example 2. The variance function in Model 3 is from Hardle and Tsybakov (1997). When x_t is i.i.d. U(0, 1) (Case 1) we refer to the models in Table 1 as Models 1–3, respectively. When $x_t = y_{t-1}$ (Case 2) we refer to the models in Table 1 as Models 1e–3e. We consider first the precision of the nonparametric estimator using

$$\mathrm{MSE}(\widehat{\sigma}_t{}^2) = \frac{1}{M}\sum_{s=1}^{M}\left((1/T)\sum_{t=1}^{T}(\widehat{\sigma}^2_{t,s} - \sigma^2_{t,s})^2\right), \tag{15}$$

where M denotes the number of Monte Carlo replications, T equals the sample size, and $\widehat{\sigma}^2_{t,s}$ is the nonparametric estimator of $\sigma^2_{t,s}$ at time t based on the sth Monte Carlo replication. In Fig. 1 the MSE based on the difference estimator and the Fan and Yao (1998) estimator (hereafter, Fan–Yao estimator) are compared for the specifications of σ_t^2 given in Table 1 (Cases 1 and 2) and alternative values of ϕ_0.

Bandwidths for the simulation experiments are chosen such that $\mathrm{MSE}(\widehat{\sigma}_t{}^2)$ is minimized based on appropriate training data set which is feasible because the true data generating process is known. This procedure is adapted to minimize uncertainty related to bandwidth selection.[2] Fig. 1 illustrates that for numerically small values of ϕ_0 the two estimators perform approximately equally well in Case 1. However, for values of $\phi_0 > 0.7$ the difference estimator of $\widehat{\sigma}_t^2$ is clearly more efficient that the Fan–Yao estimator. In Case 2, we have included the difference estimator for easy comparison. We would expect the Fan–Yao estimator to be relatively efficient, which is also the case under Model 2e. However, when applied to Models 1e and 3e the differences are negligible for small values of ϕ_0. In Model 3e the difference estimator actually outperforms the Fan–Yao estimator for larger values of ϕ_0. This may be due to the simplicity of the difference estimator, however, caution is needed when interpreting these results as the asymptotic properties of the difference estimator are unknown under Case 2.

Finally, we consider the sample density of $\widehat{n}_T = \sqrt{T}(\widehat{\phi} - \phi_0)/\sqrt{1 - \phi_0^2}$, which according to Theorem 2 should converge to a standard normal density. In Fig. 2 the density of $\widehat{n}_T$ for each of the Models 1–3 and 1e–3e based on $T = 1000$ and $\phi_0 = 0.5$ is depicted together with the standard normal density. From the illustration, it is clear that the simulation results confirm the prediction of Theorem 2. No severe small sample biases seem to be present in any of the pictures and the small sample approximation to the standard normal distribution in general seems to be very good.

[2] For empirical applications, bandwidths could be chosen using either the approach suggested by Fan and Gijbels (1995) or plug-in methods as the asymptotic variance function of both nonparametric estimators is known.

Table 1
Alternative data generating processes

	Specifications
Model 1	$y_t = \sqrt{0.1 + 0.5x_t^2}\varepsilon_t$
Model 2	$y_t = \sqrt{0.4\ \exp(-2x_t^2) + 0.2}\varepsilon_t$
Model 3	$y_t = \sqrt{\varphi(x_t + 1.2) + 1.5\varphi(x_t - 1.2)}\varepsilon_t$

$\varphi(\cdot)$ is the standard normal c.d.f.

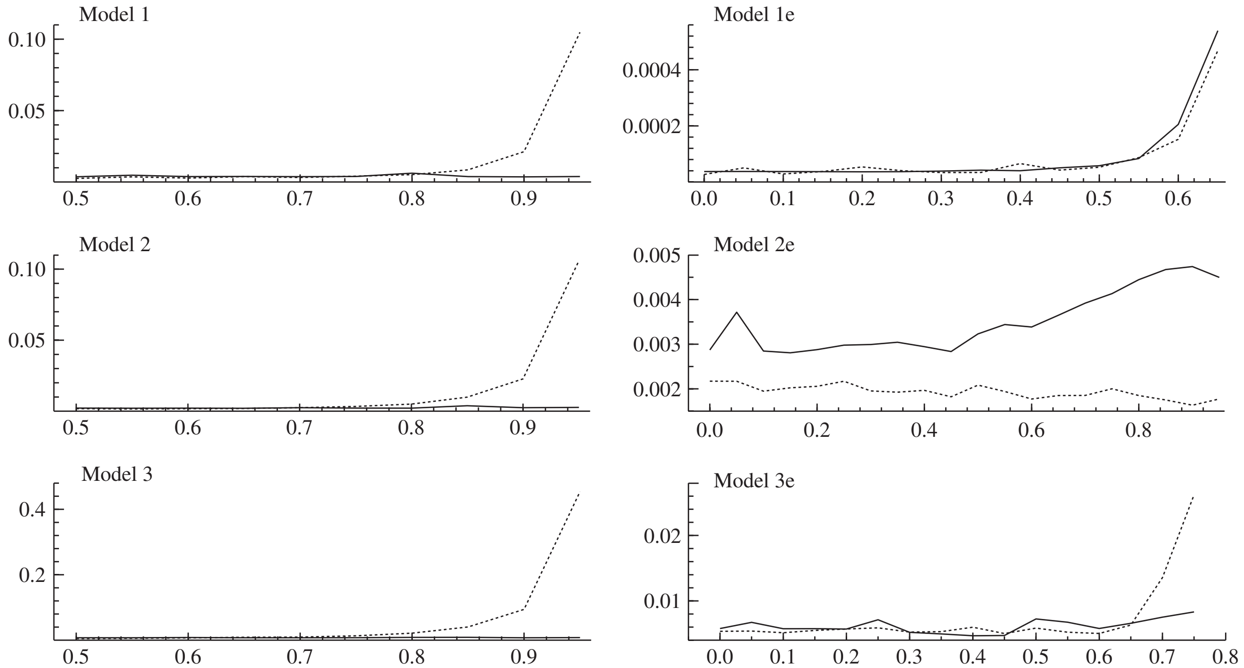

Fig. 1. MSE from the difference based variance function estimator (solid line) and the Fan–Yao estimator (dotted line) under alternative variance function specifications and alternative values of ϕ_0. $T = 1000$ and the number of Monte Carlo replications equals 1000.

6. Conclusion

We introduce and analyze a model that has at least two important implications for research in empirical finance. First, it provides a plausible linkage between risk and expected return of financial assets. Secondly, it can serve as a vehicle for testing the martingale difference sequence hypothesis which typically is uncritically adopted in financial time series models. Under general conditions, we discuss how to estimate the model and establish the asymptotic properties of the proposed estimators. It is important to stress that the present analysis has been limited to a very simple dependence structure in the innovation process. Allowing for a general ARMA structure would be a natural extension and we conjecture that similar results can be obtained. In particular, it is very likely that the difference approach will produce a nonparametric estimator that is inconsistent up to a multiplicative constant and consequently the MINPIN estimators based on WLS will be unaffected asymptotically. Furthermore, if the multiplicative constant depends on the MINPIN parameters, bias-correcting the initial nonparametric estimator will be straightforward. Research addressing these possible extensions is ongoing.

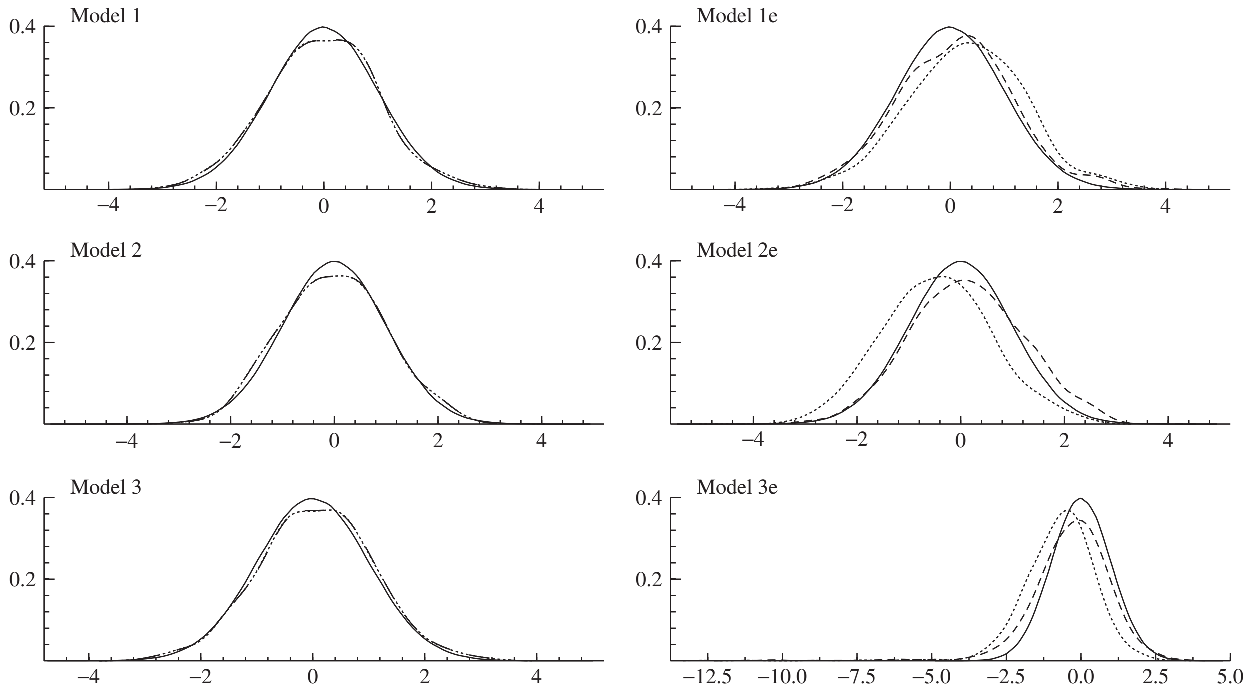

Fig. 2. Finite sample (simulated) densities and the asymptotic density of $\sqrt{T}(\widehat{\phi}-\phi_0)/\sqrt{1-\phi_0^2}$ under alternative variance function specifications for $T=1000$ and $\phi_0=0.5$. Solid line: N(0, 1). Dashed line: Fan–Yao. Dotted line: Difference based. The number of Monte Carlo replications equals 1000.

Acknowledgments

Comments and numerous helpful suggestions from Larry Brown, Gloria Gonzalez-Rivera and an anonymous referee are gratefully acknowledged. The scientific notation follows Abadir and Magnus (2002).

References

Abadir, K.M., Magnus, J.R., 2002. Notation in econometrics: a proposal for a standard. Econom. J. 5, 76–90.

Andrews, D.W.K., 1987. Consistency in nonlinear econometric models: a generic uniform law of large numbers. Econometrica 55, 1465–1471.

Andrews, D.W.K., 1994. Asymptotics for semiparametric econometric models via stochastic equicontinuity. Econometrica 62, 43–72.

Bierens, H.J., 2005. Introduction to the Mathematical and Statistical Foundations of Econometrics. Cambridge University Press, Cambridge.

Fan, J., Gijbels, I., 1995. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. J. R. Statist. Soc. Ser. B 57, 371–394.

Fan, J., Yao, Q., 1998. Efficient estimation of conditional variance functions in stochastic regression. Biometrika 85, 645–660.

Hall, P., Kay, J.W., Titterington, D.M., 1990. Asymptotically optimal difference based estimation of variance in nonparametric regression. Biometrika 77, 521–528.

Hardle, W., Tsybakov, A., 1997. Local polynomial estimators of the volatility function in nonparametric autogression. J. Econom. 81, 223–242.

Levins, M., 2003. Variance estimation for nonparametric regression and its applications. Ph.D. Dissertation, University of Pennsylvania, 2003.

Parzen, E., 1962. On estimating of a probability density and mode. Ann. Math. Statist. 35, 1065–1076.

Peligrad, M., Utev, S., 1997. Central limit theorem for linear processes. Ann. Probab. 25, 443–456.

Rosenblatt, M., 1956. Remarks on some nonparametric estimates of a density function. Ann. Math. Statist. 27, 832–837.

Shiryaev, A., 1996. Probability, second ed. Springer, Berlin.

Simonoff, J.S., 1996. Smoothing Methods in Statistics. Springer, Berlin.