J. R. Statist. Soc. B (2003)
65, *Part* 3, *pp.* 663–678

An empirical likelihood goodness-of-fit test for time series

Song Xi Chen,

National University of Singapore, Singapore

Wolfgang Härdle

Humboldt-Universität zu Berlin, Germany

and Ming Li

National University of Singapore, Singapore

[Received January 2001. Final revision January 2003]

Summary. Standard goodness-of-fit tests for a parametric regression model against a series of nonparametric alternatives are based on residuals arising from a fitted model. When a parametric regression model is compared with a nonparametric model, goodness-of-fit testing can be naturally approached by evaluating the likelihood of the parametric model within a nonparametric framework. We employ the empirical likelihood for an α-mixing process to formulate a test statistic that measures the goodness of fit of a parametric regression model. The technique is based on a comparison with kernel smoothing estimators. The empirical likelihood formulation of the test has two attractive features. One is its automatic consideration of the variation that is associated with the nonparametric fit due to empirical likelihood's ability to Studentize internally. The other is that the asymptotic distribution of the test statistic is free of unknown parameters, avoiding plug-in estimation. We apply the test to a discretized diffusion model which has recently been considered in financial market analysis.

Keywords: α-mixing; Empirical likelihood; Goodness-of-fit test; Nadaraya–Watson estimator; Parametric models; Power of test; Square-root processes; Weak dependence

1. Introduction

The analysis and prediction of time series is standard in statistics. The techniques that are employed usually rely on the actual model assumed to represent and generate the time series dynamics. Mismodelling might result in biased prediction and an incorrect parameter specification. The aim of this paper is to show how empirical likelihood (EL) (Owen, 1990) may be used to construct simple test procedures for the goodness of fit of standard time series models.

Suppose that $\{(X_i, Y_i)\}_{i=1}^n$ is a strictly stationary time series with $Y_i \in \mathbb{R}$ and $X_i \in \mathbb{R}^d$. Let $m(x) = E(Y|X = x)$ be the conditional mean function, f be the density of the design X and $\sigma^2(x) = \text{var}(Y|X = x)$ be the conditional variance of Y given $X = x \in S$, a set to be specified later. Suppose that $\{m_\theta|\theta \in \Theta\}$ is a parametric model for the mean function m and that $\hat{\theta}$ is an estimator of θ under this parametric model. The interest is to test the null hypothesis

Address for correspondence: Song Xi Chen, Department of Statistics and Applied Probability, National University of Singapore, Kent Ridge, 117543 Singapore.
E-mail: stacsx@nus.edu.sg

© 2003 Royal Statistical Society 1369–7412/03/65663

$H_0 : m(x) = m_\theta(x)$ for all $x \in S$ against a series of nonparametric alternatives $H_1 : m(x) = m_\theta(x) + c_n\, \Delta_n(x)$, where c_n is a non-random sequence tending to 0 as $n \to \infty$ and $\Delta_n(x)$ is a sequence of bounded functions.

The problem of testing a parametric mean regression against a nonparametric alternative is not new for an independent and identically distributed setting. Härdle and Mammen (1993) proposed a test statistic based on an L_2-distance between a nonparametric kernel estimator of the conditional mean and the hypothesized parametric function. This kind of goodness-of-fit test, comparing nonparametric with parametric fits, was also treated in Eubank and Spiegelman (1990), Hart (1997) and the references therein. Kreiss *et al.* (1998) have extended the Härdle–Mammen test to a time series context by implementing the wild bootstrap. Linearity tests based on local polynomial estimators were considered in Hjellvik *et al.* (1998). Koul and Stute (1999) proposed a test for time series based on the empirical processes constructed on the residuals of parametric fits under hypothesis H_0. The attractions of their test are

(a) its independence of a smoothing parameter and
(b) its capacity to test H_1 with c_n of order $n^{-1/2}$, which is smaller than $n^{-1/2}h^{-d/4}$ for the Härdle–Mammen test.

As they pointed out, however, the test is not applicable for $d > 1$ and it requires the conditional variance function $\sigma^2(\cdot)$ to be constant. Recently Horowitz and Spokoiny (2001) proposed a test for independent data that applies the Härdle–Mammen test simultaneously on a set of bandwidth values. They showed that the test is consistent for H_1 with $c_n = O(n^{-1/2}\sqrt{[\log\{\log(n)\}]})$. An adaption of a similar scheme for the EL test is possible, which would mean an improved resolution of the test that is proposed in the current paper.

The device that we use to formulate goodness-of-fit tests is EL (Owen, 1990, 2001), which is a computer-intensive nonparametric alternative to the bootstrap. It has been shown to share some key properties with parametric likelihood, e.g. Wilks's theorem and Bartlett correctability; see Hall and La Scala (1990), Qin and Lawless (1994) and Chen (1996). Kitamura (1997) considered blocked EL for parameters associated with weakly dependent processes. Monti (1997) constructed confidence regions for a parameter of a stationary time series via Whittle's estimation method. For independent data, EL has been employed for testing conditional moment restrictions (Tripathi and Kitamura, 2000) and for nonparametric functions based on a local approximation (Fan and Zhang, 2000).

The EL goodness-of-fit statistic that is proposed here is based on an EL ratio of the parametric model over a set within the domain of the design density f. The EL test is carried out by simulating a Gaussian random field with known mean and covariance function. The known covariance function is due to EL's ability to Studentize internally, leading to asymptotic pivotalness. The test proposed has the advantages of

(a) putting the goodness-of-fit measure within the context of its variation and
(b) avoiding secondary plug-in estimation, as the asymptotic distribution of the test statistic is free of unknown parameters.

The paper is structured as follows. In Section 2, we construct an EL ratio for $m(x)$, the basic building-block of the test proposed. In Section 3, we formulate the test statistic by integrating the EL ratio over the set S and establish the asymptotic equivalence of the test statistic with an integral of a squared Gaussian random field. The test procedure is proposed in Section 4 and is applied to test a parametric diffusion model in Section 5. Section 6 reports simulation results. All the technical details including assumptions are given in Appendix A.

2. Kernel estimator and empirical likelihood

We first introduce a nonparametric kernel estimator for m. Let $S = \{x \in \mathbb{R}^d | f(x) \geqslant \beta\}$ for some $\beta > 0$ be a compact set. Without loss of generality we assume that $S = [0, 1]^d$.

Let Λ be a univariate rth-order kernel which is compactly supported on $[-1, 1]$ such that

$$\int \Lambda(t)\,\mathrm{d}t = 1,$$

$$\int t^l \Lambda(t)\,\mathrm{d}t = 0 \qquad \text{if } 1 \leqslant l \leqslant r-1,$$

$$\int t^r \Lambda(t)\,\mathrm{d}t = \kappa_r \neq 0$$

for an integer $r \geqslant 2$, and let K be a d-dimensional product kernel of Λ, i.e.

$$K(t_1, \ldots, t_d) = \prod_{i=1}^{d} \Lambda(t_i).$$

Let h be a positive smoothing bandwidth which is used to smooth in every component of X, implying that the scale in each component is roughly the same. When the scales of the components are different, they can be standardized by using their standard deviations.

Let $K_h(u) = h^{-d} K(h^{-1}u)$. The nonparametric estimator of $m(x)$ considered is the Nadaraya–Watson estimator

$$\hat{m}(x) = \sum_{i=1}^{n} Y_i\, K_h(x - X_i) \Big/ \sum_{i=1}^{n} K_h(x - X_i). \tag{2.1}$$

Let

$$\tilde{m}_{\hat{\theta}}(x) = \sum_{i=1}^{n} K_h(x - X_i)\, m_{\hat{\theta}}(X_i) \Big/ \sum_{i=1}^{n} K_h(x - X_i)$$

be the smoothed parametric model. To avoid the issue of the bias that is associated with the nonparametric fit, the test statistic that we shall consider is based on the difference between $\tilde{m}_{\hat{\theta}}$ and $\hat{m}$, rather than between $\hat{m}$ and $m_{\hat{\theta}}$. The local linear estimator can be used to replace the Nadaraya–Watson estimator in estimating m because of its attractive bias properties. However, as we compare $\hat{m}$ with $\tilde{m}_{\hat{\theta}}$, the bias issue is unimportant.

We now introduce EL for a testing problem. At an arbitrary $x \in S$, let $p_i(x)$ be non-negative weights allocated to (X_i, Y_i). The EL for $\tilde{m}_{\hat{\theta}}(x)$ is

$$L\{\tilde{m}_{\hat{\theta}}(x)\} = \max\left\{\prod_{i=1}^{n} p_i(x)\right\}, \tag{2.2}$$

subject to

$$\sum_{i=1}^{n} p_i(x) = 1$$

and

$$\sum_{i=1}^{n} p_i(x)\, K\left(\frac{x - X_i}{h}\right)\{Y_i - \tilde{m}_{\hat{\theta}}(x)\} = 0.$$

By introducing Lagrange multipliers, the optimal weights are given as

$$p_i(x) = n^{-1}\left[1 + \lambda(x)\, K\left(\frac{x - X_i}{h}\right)\{Y_i - \tilde{m}_{\hat{\theta}}(x)\}\right]^{-1} \tag{2.3}$$

where $\lambda(x)$ is the root of

$$\sum_{i=1}^{n} \frac{K\{(x-X_i)/h\}\{Y_i - \tilde{m}_{\hat{\theta}}(x)\}}{1+\lambda(x)K\{(x-X_i)/h\}\{Y_i - \tilde{m}_{\hat{\theta}}(x)\}} = 0. \tag{2.4}$$

The maximum EL is achieved at $p_i(x) = n^{-1}$ corresponding to the Nadaraya–Watson estimator $\hat{m}(x)$. The log-EL ratio is $l\{\tilde{m}_{\hat{\theta}}(x)\} = -2 \log[L\{\tilde{m}_{\hat{\theta}}(x)\}n^n]$ which will be the basic building-block for the proposed EL goodness-of-fit statistic. We first evaluate $\lambda(x)$ in the following lemma.

Lemma 1. Under assumptions (a)–(g) given in Appendix A,

$$\sup_{x \in S} |\lambda(x)| = o_p\{(nh^d)^{-1/2} \log(n)\}.$$

The proof of lemma 1 is detailed in Chen *et al.* (2002). Let $\gamma(x)$ be a random process with $x \in S$. Throughout this paper the notation $\gamma(x) = \tilde{O}_p(\delta_n)$ means $\sup_{x \in S} |\gamma(x)| = O_p(\delta_n)$ for a sequence $\{\delta_n\}$, and similarly for $\gamma(x) = \tilde{o}_p(\delta_n)$.

Let

$$w_i(x) = K\left(\frac{x - X_i}{h}\right)\{Y_i - \tilde{m}_{\hat{\theta}}(x)\},$$

$$\bar{U}_j(x) = (nh^d)^{-1} \sum_{i=1}^{n} w_i(x)^j$$

for integers j. Expansion of expression (2.4) yields

$$\bar{U}_1(x) - \lambda(x)\, \bar{U}_2(x) + \lambda^2(x)(nh^d)^{-1} \sum_{i=1}^{n} \frac{w_i(x)^3}{1+\lambda(x)\, w_i(x)} + \tilde{o}_p\{(nh^d)^{-3/2} \log^3(n)\} = 0. \tag{2.5}$$

It may be shown by the technique that is used in the proof of lemma 1 that

$$(nh^d)^{-1} \sum_{i=1}^{n} \frac{w_i(x)^3}{1+\lambda(x)\, w_i(x)} = \tilde{O}_p(1).$$

Inverting equation (2.5), we have $\lambda(x) = \bar{U}_2^{-1}(x)\, \bar{U}_1(x) + \tilde{o}_p\{(nh^d)^{-1} \log^2(n)\}$. This, together with equation (2.3) and lemma 1, means that

$$\begin{aligned} l\{\tilde{m}_{\hat{\theta}}(x)\} &= 2 \sum_{i=1}^{n} \log\left[1 + \lambda(x)\, K\left(\frac{x - X_i}{h}\right)\{Y_j - \tilde{m}_{\hat{\theta}}(x)\}\right] \\ &= 2(nh^d)\, \lambda(x)\, \bar{U}_1(x) - (nh^d)\, \lambda^2(x)\, \bar{U}_2(x) + \tilde{o}_p\{(nh^d)^{-1/2} \log^3(n)\} \\ &= (nh^d)\, \bar{U}_2^{-1}(x)\, \bar{U}_1^2(x) + \tilde{o}_p\{(nh^d)^{-1/2} \log^3(n)\}. \end{aligned} \tag{2.6}$$

Let $v(x;h) = h^d \int K_h^2(x-y)\, f(y)\, \sigma^2(y)\, \mathrm{d}y$, $b(x;h) = h^d \int K_h(x-y)\, f(y)\, \mathrm{d}y$ and $V(x;h) = v(x;h)/b^2(x;h)$. Note that $V(x;h)/nh^d$ is the asymptotic variance of $\hat{m}(x)$ provided that $nh^d \to \infty$. It may be shown that $\sup_{x \in S} |\hat{m}(x) - \tilde{m}_\theta(x)| = o_p\{(nh^d)^{-1/2} \log(n)\}$ and $\sup_{x \in S} |\hat{f}(x) - b(x)| = o_p\{(nh^d)^{-1/2} \log(n)\}$. From the proof of lemma 1, we have $\bar{U}_1(x) = b(x;h)\{\hat{m}(x) - \tilde{m}_\theta(x)\} + \tilde{o}_p\{(nh^d)^{-1} \log^2(n)\}$ and $\sup_{x \in S} |\bar{U}_2(x) - v(x;h)| = o_p\{(nh^d)^{-1/2} \log(n) + h^2\}$. These and equation (2.6) mean that

$$l\{\tilde{m}_{\hat{\theta}}(x)\} = nh^d \frac{\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2}{V(x;h)} + \tilde{O}\{(nh^d)^{-1/2} \log^3(n) + h^2 \log^2(n)\}, \tag{2.7}$$

implying that $l\{\tilde{m}_{\hat{\theta}}(x)\}$ is asymptotically equivalent to a Studentized L_2-distance between $\tilde{m}_{\hat{\theta}}(x)$ and $\hat{m}(x)$.

3. Goodness-of-fit test statistic

On the basis of the property of $l\{\tilde{m}_{\hat{\theta}}(x)\}$ revealed in equation (2.7), the EL-based goodness-of-fit statistic proposed is

$$l_n(\tilde{m}_{\hat{\theta}}) = \int_{x \in S} l\{\tilde{m}_{\hat{\theta}}(x)\}\, \mathrm{d}x.$$

From equation (2.7),

$$l_n(\tilde{m}_{\hat{\theta}}) = nh^d \int \frac{\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2}{V(x;h)}\, \mathrm{d}x + O_p\{(nh^d)^{-1/2} \log^3(n) + h^2 \log^2(n)\}. \tag{3.1}$$

Härdle and Mammen (1993) proposed $T_n = nh^{d/2} \int\{\hat{m}(x) - \tilde{m}_{\hat{\theta}}(x)\}^2\, \pi(x)\, \mathrm{d}x$ as a measure of goodness of fit where $\pi(x)$ is a given weight function. Equation (3.1) indicates that $l_n(\tilde{m}_{\hat{\theta}})$ is asymptotically equivalent to $h^{d/2} T_n$ with $\pi(x) = V^{-1}(x;h)$ which is proportional to $f(x)/\sigma^2(x)$. The differences between the two test statistics are that

(a) the EL automatically Studentizes so there is no need to estimate $V(x;h)$ and
(b) the EL can capture features of data such as the skewness and kurtosis.

The test statistic proposed by Kreiss *et al.* (1998) is equivalent to the Härdle–Mammen statistic with $\pi(x) = f^2(x)$, which is designed to downweight low density areas. However, this may lead to a loss of power.

The test statistic proposed can be readily extended to testing a parametric specification, σ^2_θ, of the volatility function $\sigma^2(x)$. We need only to replace the second constraint in equation (2.2) with

$$\sum_{i=1}^{n} p_i(x)\, K\left(\frac{x - X_i}{h}\right) [\{Y_i - \hat{m}(X_i)\}^2 - \tilde{\sigma}^2_{\hat{\theta}}(x)] = 0,$$

where $\tilde{\sigma}^2_{\hat{\theta}}$ like $\tilde{m}_{\hat{\theta}}$ is a kernel smooth of $\sigma^2_{\hat{\theta}}$. The advantage of the internal Studentizing that is offered by the EL becomes more apparent in this case as explicit variance estimation of the kernel estimator of σ^2 involves fourth-order moments and so is more difficult.

The smoothing bandwidth h can be chosen by any bandwidth selector that produces h which minimizes the mean integrated square error of the curve estimation, e.g. those from the cross-validation or the plug-in methods. This is an area that has been intensively studied in nonparametric curve estimation.

Theorem 1. Under assumptions (a)–(g) in Appendix A, $l_n(\tilde{m}_{\hat{\theta}})$ and $\int_S \mathcal{N}^2(s)\, \mathrm{d}s$ have the same asymptotic distribution as $n \to \infty$, where $\mathcal{N}$ is a normal process on $S = [0,1]^d$ such that $E\{\mathcal{N}(s)\} = h^{d/4}\, \Delta_n(s)\, V^{-1/2}(s;h)$ and

$$\Omega(s,t) =: \operatorname{cov}\{\mathcal{N}(s), \mathcal{N}(t)\} = \sqrt{\left\{\frac{f(s)\, \sigma^2(s)}{f(t)\, \sigma^2(t)}\right\}}\, K^{(2)}(0)^{-1}\, K^{(2)}\left(\frac{s-t}{h}\right),$$

where $K^{(2)}$ is the convolution of K.

The proof of the theorem is given in Appendix A. As K is a compact kernel on $[-1,1]^d$, $\Omega(s,t) = 0$ if $|s-t| > 2h$, which means that $\mathcal{N}(s)$ and $\mathcal{N}(t)$ are independent if $|s-t| > 2h$. As $f(s)\, \sigma^2(s) = f(s)\, \sigma^2(t) + O(h)$ when $|s-t| \leqslant 2h$,

$$\Omega(s,t) = K^{(2)}(0)^{-2}\, K^{(2)}\left(\frac{s-t}{h}\right) + O(h). \tag{3.2}$$

Hence, the leading order term of the covariance function is completely known and will facilitate a test based on the simulation of the normal process. The proof of theorem 1 contains the following corollary.

Corollary 1. Under assumptions (a)–(f) in Appendix A, $h^{-d/2}\{l_n(\tilde{m}_{\hat{\theta}}) - \mu_0\} \rightarrow^{\mathrm{d}} N(0, \sigma_0^2)$ as $n \to \infty$ where $\sigma_0^2 = 2\, K^{(4)}(0)\, K^{(2)}(0)^{-2}$ and $\mu_0 = 1 + h^{d/2} \int V^{-1}(s)\, \Delta_n^2(s)\, \mathrm{d}s$.

The corollary indicates that $l_n(\tilde{m}_{\hat{\theta}})$ is $O_p(h^{d/2})$, vanishing as $n \to \infty$. However, for a given h, $l_n(\tilde{m}_{\hat{\theta}})$ is a monotone function of $h^{-d/2}\{l_n(\tilde{m}_{\hat{\theta}}) - 1\}$ which is $O_p(1)$. Hence, a test based on $l_n(\tilde{m}_{\hat{\theta}})$ is equivalent to a test based on the standardized statistic $h^{-d/2}\{l_n(\tilde{m}_{\hat{\theta}}) - 1\}$.

4. Goodness-of-fit test

An α-level test based on the asymptotic normality of $l_n(\tilde{m}_{\hat{\theta}})$ rejects hypothesis H_0 if

$$h^{-d/2}\{l_n(\tilde{m}_{\hat{\theta}}) - 1\} > z_\alpha \surd\{2K^{(4)}(0)\}/K^{(2)}(0)$$

where z_α is the $(1-\alpha)$-quantile of $N(0, 1)$. The corollary implies that the asymptotic power of the test under H_1 is

$$1 - \Phi\left[z_\alpha - \frac{K^{(2)}(0) \int V^{-1}(s)\, \Delta_n^2(s)\, \mathrm{d}s}{\surd\{2K^{(4)}(0)\}}\right]$$

which is sensitive to alternatives in all directions and converges to 1 if c_n is of a larger order than $n^{-1/2}h^{-d/4}$, hence implying that the test is consistent.

The asymptotic normal test may not work well for a finite sample owing to the many approximations that are used in establishing the asymptotic normality, which subsequently make it too far from the finite sample distribution of $l_n(\tilde{m}_{\hat{\theta}})$. Indeed, obtaining a better approximation to the distribution of the test statistic motivated the use of the wild bootstrap (Härdle and Mammen, 1993; Kreiss *et al.*, 1998; Hjellvik *et al.*, 1998).

Rather than resort to the wild bootstrap, we propose to use the distribution of $\int_S \mathcal{N}^2(s)\, \mathrm{d}s$ to approximate that of $l_n(\tilde{m}_{\hat{\theta}})$. This proposal, as well as the result contained in theorem 1, is motivated by the observation that the distribution of $\mathcal{N}(x)$ mimics that of

$$(nh^d)^{1/2}\, V^{-1}(x; h)\, \{\hat{m}(x) - \tilde{m}_{\hat{\theta}}(x)\}$$

at each x. In particular, let

$$\Omega_0(s, t) = K^{(2)}(0)^{-2}\, K^{(2)}\left(\frac{s-t}{h}\right)$$

which is the leading term of $\Omega(s, t)$, the covariance of $\mathcal{N}$, and let $\mathcal{N}_0$ be a Gaussian random field with zero mean function and Ω_0 as its covariance function. Since the law of $\mathcal{N}_0$ is completely known on given an h, the distribution of $\int_S \mathcal{N}_0^2(s)\, \mathrm{d}s$ can be obtained by simulating $\{\mathcal{N}_0(s) | s \in S\}$ over a lattice in S. We use an algorithm proposed by Wood and Chan (1994). Our numerical experience indicates that the best effect is achieved when the lattice used in discretizing $\int_{x \in S} l\{\tilde{m}_{\hat{\theta}}(x)\}\, \mathrm{d}x$ is also used in simulating $\{\mathcal{N}_0(s) | s \in S\}$. Let w_α be the simulated upper α-level quantile of $\int_S \mathcal{N}_0^2(s)\, \mathrm{d}s$. Then, the EL test proposed is to reject hypothesis H_0 if $l_n(\tilde{m}_{\hat{\theta}}) \geqslant w_\alpha$.

Fig. 1 contains the simulated null densities of $l_n(\tilde{m}_{\hat{\theta}})$ and $\int_S \mathcal{N}_0^2(s)\, \mathrm{d}s$ as well as two normal densities: the full scale asymptotic normal $N(1, h^d\sigma_0^2)$ and $N(1, \sigma_1^2)$ densities, where

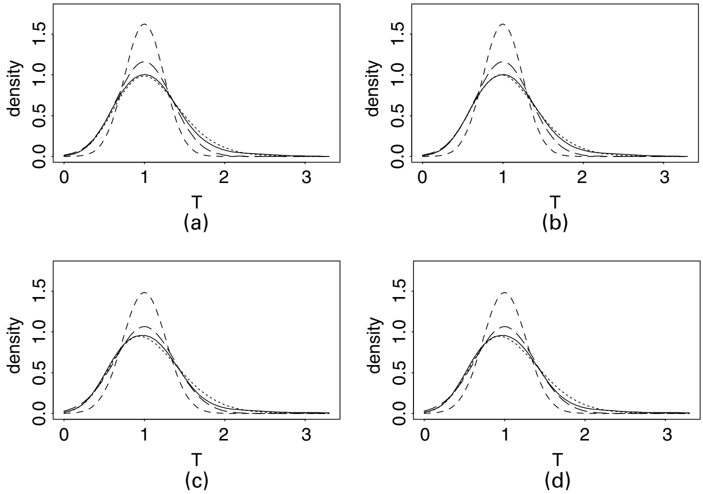

Fig. 1. Approximations to the density of $l_n(\tilde{m}_{\hat\theta})$ (———) by those of $\int_S \mathcal{N}_0^2(s)\,\mathrm{d}s$ (·······), $N(1, h^d\sigma_0^2)$ (-------) and $N(1, \sigma_1^2)$ (– – –) under models (6.1) and (6.2) with $n = 500$: (a) $h = 0.05$, $k_n = 300$; (b) $h = 0.05$, $k_n = 600$; (c) $h = 0.06$, $k_n = 300$; (d) $h = 0.06$, $k_n = 600$

$$\sigma_1^2 = K^{(2)}(0)^{-2} \int_S \int_S K^{(2)}\left(\frac{s-t}{h}\right)^2 \mathrm{d}s\,\mathrm{d}t$$

is the quantity that $h^d\sigma_0^2$ approximates. Let k_n be the number of lattice points used in discretizing both $\int_{x\in S} l\{\tilde{m}_{\hat\theta}(x)\}\,\mathrm{d}x$ and $\int_S \mathcal{N}_0^2(s)\,\mathrm{d}s$. Although Fig. 1 is for only two values of h and k_n as part of a comprehensive simulation study that is reported in Section 6, it shows typical results. The simulation of $\int_S \mathcal{N}_0^2(s)\,\mathrm{d}s$ provides quite a satisfactory approximation to the distribution of $l_n(\tilde{m}_{\hat\theta})$. Of the two normal approximations, $N(1, \sigma_1^2)$ is better than $N(1, h^d\sigma_0^2)$ owing to its better approximation to the variance. The simulated distribution nicely reflects the skewness of $l_n(\tilde{m}_{\hat\theta})$, which the two normal proxies cannot pick up. It is also worth noting that the test based on the normal densities will typically have a much larger size than the nominal level α. We see that the distributions of $l_n(\tilde{m}_{\hat\theta})$ and $\int_S \mathcal{N}_0^2(s)\,\mathrm{d}s$ are insensitive to the number of lattice points used.

5. Testing the Cox–Ingersoll–Ross model for Standard and Poors 500 data

In mathematical finance, interest rates, stocks and other financial products are modelled by diffusion processes with specific parametric assumptions on the drift and diffusion functions. A well-known model for modelling the dynamic of interest rates is the Cox–Ingersoll–Ross model (Cox *et al.*, 1985) given in equation (5.2) below. Specification tests for diffusion models have been considered by Aït-Sahalia (1996), Hong and Li (2001) and others.

In this section we apply the empirical likelihood test for a financial market model proposed by Platen (1999) on the Standard and Poors 500 share index data which contain the daily closing values of the index for 5479 trading days from December 31st, 1976, to December 31st, 1997. In Fig. 2(a), the index series shows an exponential trend which is estimated by using the method of Härdle *et al.* (2000). Fig. 2(a) also displays a residual process $X(t)$ at the bottom after removing the exponential trend. In mathematical finance, we assume a specific dynamic form for this $X(t)$ process. More precisely, Platen (1999) assumed the following model for an index process:

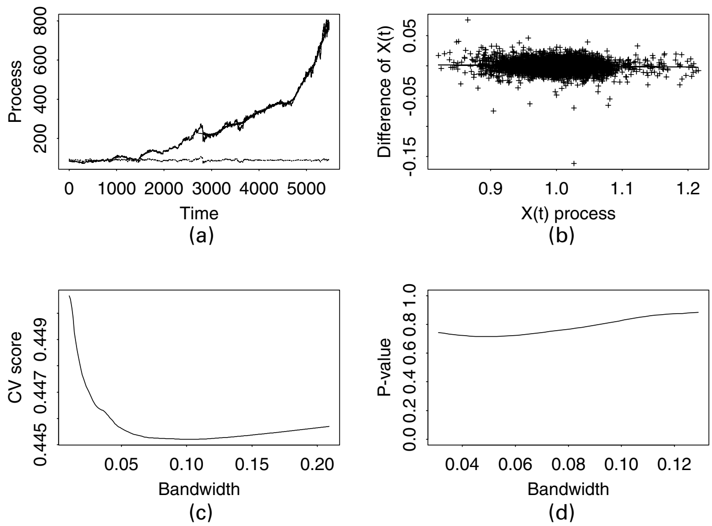

Fig. 2. (a) Raw Standard and Poors 500 data and their exponential trend (top) and the $X(t)$ process (bottom), (b) $X_i = X(i\Delta)$ *versus* $Y_i = X_{i+1} - X_i$ (+) and the parametric fit for the conditional mean (———), (c) cross-validation scores CV and (d) P-values of the EL test

$$S(t) = S(0)\, X(t)\, \exp\left\{\int_0^t \eta(s)\,\mathrm{d}s\right\}, \tag{5.1}$$

with a diffusion component $X(t)$ solving a stochastic differential equation

$$\mathrm{d}X(t) = \alpha\{1 - X(t)\}\,\mathrm{d}t + \sigma\, X^{1/2}(t)\,\mathrm{d}W(t), \tag{5.2}$$

where $W(t)$ is a Brownian motion and α and σ are parameters. Discretizing this series with sampling interval Δ leads to observations (X_i, Y_i) shown in Fig. 2(b) with $Y_i = X_{(i+1)\Delta} - X_{i\Delta}$ and $X_i = X_{i\Delta}$, which is α mixing, and fulfil all the other conditions assumed on the basis of the results of Genon-Catalot *et al.* (2000).

We applied the EL test to verify the parametric mean function $m(x) = a(1 - x)$ specified by the Cox–Ingersoll–Ross model. The process $X(t)$ was restored from the observed residuals by the approach that was introduced in Härdle *et al.* (2000). The estimate for the drift parameter a was $\hat{a} = 0.00968$ by using methods based on the marginal distribution and the autocorrelation structure of X. Cross-validation was used to find a suitable value of h. The cross-validation score function CV attained a minimum at $h = 0.107$ as shown in Fig. 2(c). As CV is known for its slow convergence to the optimal bandwidth, the prescribed h served as a reference only. Further investigation showed that an h-value larger than 0.06 oversmoothed the drift function whereas an h-value smaller than 0.03 undersmoothed. Therefore, the EL test was carried out for a set of h-values ranging from 0.02 to 0.13. The P-values plotted in Fig. 2(d) are all quite large. So, the data conform with the conditional mean structure specified by the Cox–Ingersoll–Ross model.

6. Simulation

In this section we report results from a simulation study that was designed to evaluate the performance of the EL test proposed. The test based on the wild bootstrap given in Kreiss *et al.* (1998) serves as a comparison. The simulation considers testing of time series models for $d = 1, 2$.

For $d = 1$, the model was

$$Y_i = \theta_1 X_i + \theta_2 X_i^2 + c_n \cos(8X_i) + \sigma_0 X_i \varepsilon_i, \tag{6.1}$$

$$X_i = \gamma X_{i-1} + \rho \eta_i, \qquad i = 1, 2, \ldots, n, \tag{6.2}$$

where $\theta = (\theta_1, \theta_2) = (0.3, 0.1)$ and $\sigma_0 = 0.5$; $\{\varepsilon_i\}_{i=1}^n$ and $\{\eta_i\}_{i=1}^n$ are mutually independent and identically distributed innovations, which are all independent of X_i. We fixed $\varepsilon_i \sim N(0, 1)$ and $\eta_i \sim \text{Unif}[-0.5, 0.5]$ with $\rho = 1$ and $\gamma = 0$. The choice of ρ was to make the density of X_i bounded away from 0 within $S = [-0.5, 0.5]$ which contains about 90% of observed X_i. Three c_n-values were used: $c_n = 0$, corresponding to hypothesis H_0, $c_n = 0.002$ and $c_n = 0.004$.

For $d = 2$, we considered an autoregressive conditional heteroscedastic ARCH (2) model

$$Y_i = \theta_1 Y_{i-1} + \theta_2 Y_{i-2} + c_n\, \Delta_n(Y_{i-1}, Y_{i-2}) + \sigma_0\, \sigma(Y_{i-1}, Y_{i-2}) \varepsilon_i \tag{6.3}$$

where $\theta = (\theta_1, \theta_2) = -(0.3, 0.3)$, $\varepsilon_i \overset{\text{IID}}{\sim} N(0, 1)$, $\sigma_0 = 0.7$ and $\sigma(x, y) = \Delta_n(x, y) = \surd(0.3x^2 + 0.2y^2 + 0.2)$. Three c_n-values, 0, 0.04 and 0.06, were used, and $S = [-0.5, 0.5]^2$.

In both models, the parameter θ was estimated by maximizing the conditional quasi-likelihood given the X_is. We fixed the nominal size of the tests at $\alpha = 5\%$, took $n = 500$ and $n = 1000$ and chose k_n, the number of lattice points used in approximating the integral $\int_{x \in S} l\{\tilde{m}_{\hat{\theta}}(x)\}\, \mathrm{d}x$, to be 300 for $d = 1$ and 900 for $d = 2$. The increase in k_n reflected the increase in the dimensionality of X_i. In the simulation of $\int_S \mathcal{N}_0^2(s)\, \mathrm{d}s$, the same lattice was used to ensure a better approximation.

The simulated power of the EL and the wild bootstrap tests for a set of h-values is summarized in Fig. 3 for $d = 1$ and in Fig. 4 for $d = 2$. Both Fig. 3 and Fig. 4 show that the power of the EL test was consistently higher than that of the bootstrap test, whereas the size of the EL test was in general the same as that of the bootstrap test. The performance of the bootstrap test for the univariate model was disappointing especially when $n = 500$. The power was improved for $d = 2$ with $n = 1000$, although the EL test was still noticeably better. The better power of the EL test was probably because it could obtain contributions of lack of fit from areas of low density where the wild bootstrap downweighted. For a given $c_n > 0$, the higher power that is observed

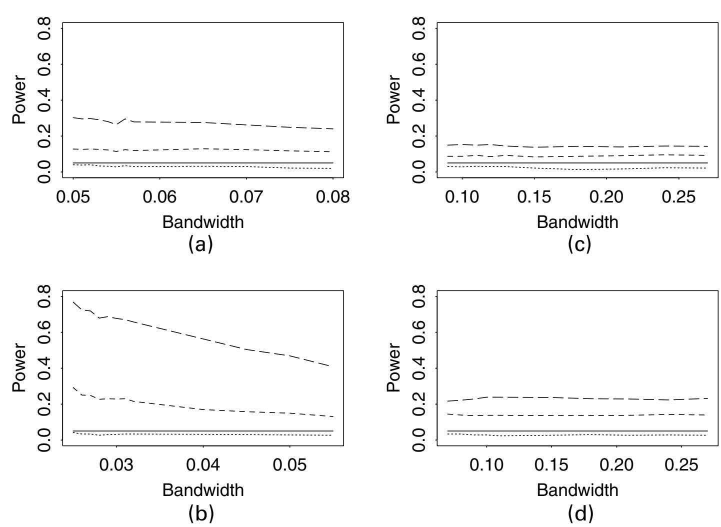

Fig. 3. Power of the EL test for (a) n = 500 and (b) n = 1000 and of the wild bootstrap test for (c) n = 500 and (d) n = 1000 under models (6.1) and (6.2): ·······, c_n = 0.0; -------, c_n = 0.02; – – –, c_n = 0.04; ———, nominal 5%

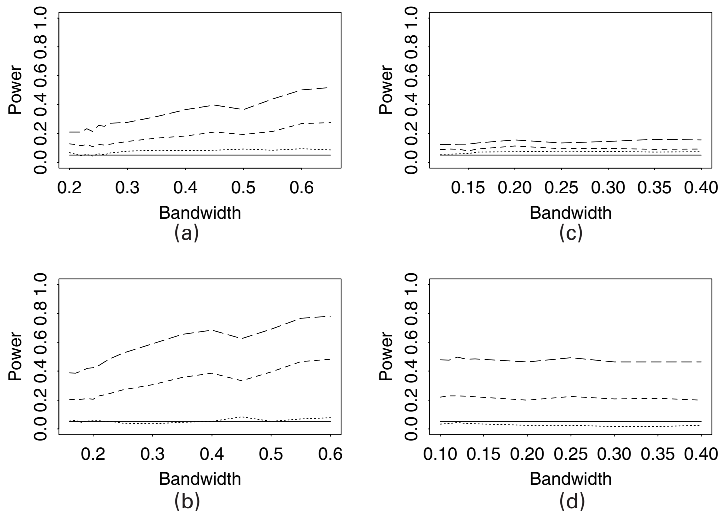

Fig. 4. Power of the EL test for (a) $n = 500$ and (b) $n = 1000$ and of the wild bootstrap test for (c) $n = 500$ and (d) $n = 1000$ under models (6.3): ·······, $c_n = 0.0$; -------, $c_n = 0.04$; – – –, $c_n = 0.06$; ———, nominal 5%

from $n = 500$ to $n = 1000$ was due to an increased distance between H_1 and H_0 although c_n was the same. One overall feature of the simulation was that the powers of both tests were not very sensitive to the value of h. The simulation was also conducted with real parameter values, giving the same pattern of results except that the sizes of both tests were closer to 5% and the power was slightly smaller.

Acknowledgements

The authors thank T. Kleinow, J. Gao, two referees, the Associate Editor and the Joint Editor for comments and suggestions which improved the presentation of the paper. The authors acknowledge support by the National University of Singapore and the Deutsch Forschungsgemeinschaft via Sonderforschungsbereich 373 'Quantifikation und Simulation ökonomischer Prozesse' at Humboldt-Universität zu Berlin.

Appendix A: Technical details

A.1. Assumptions

Let $\mathcal{F}_k^l$ be the σ-algebra of events generated by $\{(X_i, Y_i), k \leqslant i \leqslant l\}$ for $l \geqslant k$. The measure for dependence between a α-mixing time series is

$$\alpha(k) = \sup_{\mathcal{A}\in\mathcal{F}_1^i,\ \mathcal{B}\in\mathcal{F}_{i+k}^\infty} |P(AB) - P(A)\ P(B)|.$$

The assumptions that are required to establish the results given in the paper are the following:

(a) Λ is a univariate rth-order kernel which is compactly supported in $[-1, 1]$ and is Lipschitz continuous, $d < 4$; the smoothing bandwidth $h = O(n^{-1/(d+2r)})$.

(b) f, m and σ^2 have continuous derivatives up to the second order in S and both f and σ^2 are bounded below in S.

(c) $\hat{\theta}$ is a parametric estimator of θ within the family of the parametric model, and

$$\sup_{x\in S} |m_{\hat{\theta}}(x) - m_\theta(x)| = O_p(n^{-1/2}).$$

(d) $\Delta_n(x)$, the local shift in hypothesis H_1, is uniformly bounded with respect to x and n, and $c_n = n^{-1/2}h^{-d/4}$, which is the order of the difference between H_0 and H_1.
(e) $E[\exp\{a_0|Y_1 - m(X_1)|\}] < \infty$ for some $a_0 > 0$; $E(|Y_i|^k|X_i) < \infty$ for some $k > 1$; for all i, $E\{Y_i - m(X_i)|\Omega_{i-1}\} = 0$ where Ω_{i-1} is the σ-field generated by $\{(X_{j+1}, Y_j\}_{j=1}^{i-1}$.
(f) The conditional density of X given Y, $f_{X|Y} \leqslant A_1 < \infty$, the conditional joint density of (X_1, X_l) given (Y_1, Y_l) is bounded for all $l > 1$ and the joint density of $(X_1, Y_1, X_s, Y_s, X_t, Y_t)$ for $t > s > 1$ is continuous and bounded by a constant that is free of s and t.
(g) The process $\{(X_i, Y_i)\}$ is strictly stationary and α mixing, and $\alpha(k) \leqslant a\rho^k$ for some $a > 0$ and $\rho \in (0, 1)$.

Assumptions (a) and (b) on the kernel and bandwidth are standard in nonparametric curve estimation. The assumption of $d < 4$ is to make the bias in kernel variance estimation a smaller order of $h^{d/2}$. The kernel method will encounter the curse of dimension when $d \geqslant 4$ anyway. The bandwidth selected by either cross-validation or the plug-in method satisfies the order specified in assumption (a). Assumptions (c) and (d) are common in nonparametric goodness-of-fit tests, whereas assumptions (e), (f) and (g) are standard assumptions for dependent processes. In particular, assumption (g) means that the data are geometric α mixing. It can be seen from the proof that the geometric α-mixing condition can be weakened to $\alpha(k) \leqslant Ck^{-s(d)}$ where $s(d) > 2$ and is a monotone function of d. It is convenient technically to assume geometric α-mixing. For a univariate linear causal process $Y_t = \Sigma_{s=0}^{\infty} g_{t-s}\xi_s$ with independent and identically distributed innovation $\{\xi_s\}_{s=0}^{\infty}$, Gorodeskii (1977) showed that the linear process is α mixing under certain conditions and established the rate for the α-mixing coefficient. Pham and Tran (1985) showed that, if each coefficient g_t of the process is $O(\gamma^t)$, $0 < \gamma < 1$, then the process is geometric α mixing. For a Markov process $Y_i = m(X_i) + \sigma(X_i)\varepsilon_i$ where $X_i = (Y_{i-1}, \ldots, Y_{i-p})$ are lagged values and the ε_i are independent and identically distributed random variables, Masry and Tjøstheim (1995) provided conditions for geometric ergodicity and geometric α-mixing of the process.

Throughout the proof we shall use C to denote positive constants which may take different values at different places.

A.2. Proof of theorem 1

The proof of theorem 1 is carried out by proving that both $l_n(\tilde{m}_{\hat{\theta}})$ and $\int_S \mathcal{N}^2(s)\,\mathrm{d}s$ have the same limiting normal distribution. We first prove that $l_n(\tilde{m}_{\hat{\theta}})$ is asymptotically normally distributed.

Let $V(x) = R(K)\,\sigma^2(x)/f(x)$ such that $V(x; h) = V(x) + O(h^2)$. From equation (3.1)

$$l_n(\tilde{m}_{\hat{\theta}}) = S_n + o(h^{d/2})$$

where

$$S_n = nh^d \int_{x\in S} V^{-1}(x)\{\hat{m}(x) - \tilde{m}_\theta(x)\}^2\,\mathrm{d}x.$$

Let $H_{n1}(x) = n^{-1}\,\Sigma\,K_h(x - X_i)\varepsilon_i$ and $H_{n2}(x) = c_n n^{-1}\,\Sigma\,K_h(x - X_i)\,\Delta_n(x)$. Then,

$$S_n = nh^d \int_{x\in S} V^{-1}(x)\,f^{-2}(x)\{H_{n1}^2(x) + H_{n2}^2(x)\}\,\mathrm{d}x + 2A_n$$

where

$$\begin{aligned} A_n &= nh^d \int_{x\in S} V^{-1}(x)\,f^{-2}(x)\,H_{n1}(x)\,H_{n2}(x)\,\mathrm{d}x \\ &= n^{-1/2}h^{3d/4}\sum \varepsilon_i \int_{x\in S} V^{-1}(x)\,f^{-1}(x)\,\Delta_n(x)\,K_h(x - X_i)\,\mathrm{d}x\{1 + o_p(h^{d/2})\} \end{aligned}$$

since $H_{n2}(x) = \int K_h(x - y)\,\Delta_n(y)\,f(y)\,\mathrm{d}y + o_p\{(nh^d)^{-1/2}\,\log(n)\} = \Delta_n(x)\,f(x) + o_p(h^{d/2})$.

Let $s_0(X_i) = \int_{x\in S} K_h(x - X_i)\,V^{-1}(x)\,f^{-1}(x)\,\Delta_n(x)\,\mathrm{d}x$ and $W_{n0} = n^{-1}\,\Sigma\,\varepsilon_i\,s_0(X_i)$. Clearly, $A_n = n^{1/2}h^{3d/4}W_{n0}\{1 + o_p(h^{d/2})\}$. Since $E(W_{n0}) = 0$ and $\mathrm{var}(W_{n0}) = n^{-1}\,E\{\varepsilon_i^2\,s_0^2(X_i)\} \leqslant Cn^{-1}$, $A_n = O_p(h^{3d/4})$. Therefore,

$$l_n(\tilde{m}_{\hat{\theta}}) = S_{n1} + S_{n2} + S_{n3} + o_p(h^{d/2}) \tag{A.1}$$

where

$$S_{n1} = n^{-1}h^d \sum_{i\neq j} \varepsilon_i\varepsilon_j \int_{x\in S} V^{-1}(x)\, f^{-2}(x)\, K_h(x-X_i)\, K_h(x-X_j)\, \mathrm{d}x,$$

$$S_{n2} = n^{-1}h^d \sum \varepsilon_i^2 \int_{x\in S} V^{-1}(x)\, f^{-2}(x)\, K_h^2(x-X_i)\, \mathrm{d}x,$$

$$S_{n3} = h^{d/2} \int_{x\in S} V^{-1}(x)\, f^{-2}(x)\{n^{-1}\sum K_h(x-X_i)\, \Delta_n(X_i)\}^2\, \mathrm{d}x.$$

As $n^{-1}\, \Sigma\, K_h(x-X_i)\, \Delta_n(X_i) = \Delta_n(x)\, f(x) + \tilde{o}_p\{n^{-1/2}h^{-d/2}\, \log(n) + h^r\}$,

$$S_{n3} = h^{d/2} \int_{x\in S} V^{-1}(x)\, \Delta_n^2(x)\, \mathrm{d}x + o_p(h^{d/2}). \tag{A.2}$$

Note that $E(S_{n2}) = 1 + O(h^2)$ and $\mathrm{var}(S_{n2}) = O(n^{-1}h^d)$. Hence

$$S_{n2} = 1 + o_p(h^{d/2}). \tag{A.3}$$

Let $\phi_{ij} = \varepsilon_i\varepsilon_j \int_{x\in S} V^{-1}(x)\, f^{-2}(x)\, K_h(x-X_i)\, K_h(x-X_j)\, \mathrm{d}x$ and $S_{n1}^0 = \Sigma_{1\leqslant i<j\leqslant n}\phi_{ij}$ so that $S_{n1} = 2n^{-1}h^d S_{n1}^0$. Note that S_{n1}^0 is a degenerate U-statistic. A central limit theorem for degenerate U-statistics for absolutely regular processes has been established in Hjellvik *et al.* (1996). By reading their proof closely, we find that the assumption of the absolute regularity is only to use an inequality of Yoshihara (1976) which can be replaced by the Davydov inequality (Bosq (1998), page 19) for α-mixing processes; see Gao and King (2001) for an updated proof. This means that we can use the theorem in Hjellvik *et al.* (1996) for α-mixing processes.

Let $\sigma_{ij}^2 = \mathrm{var}(\phi_{ij})$ and $\sigma_n^2 = \Sigma_{1\leqslant i<j\leqslant n}\sigma_{ij}^2$. Let E_i be expectations with respect to $\xi_i =: (X_i, Y_i)$. From remark B to theorem A of Hjellvik *et al.* (1996), $\sigma_n^2 \to n^2\sigma_{n0}^2/2$ as $n\to\infty$ where

$$\begin{aligned}\sigma_{n0}^2 &= E_iE_j\bigg\{\varepsilon_i^2\varepsilon_j^2 \int\int V^{-1}(x)\, f^{-2}(x)\, V^{-1}(y)\, f^{-2}(y)\\ &\qquad \times K_h(x-X_i)\, K_h(x-X_j)\, K_h(y-X_i)\, K_h(y-X_j)\, \mathrm{d}x\, \mathrm{d}y\bigg\}\\ &= h^{-d}\, K^{(4)}(0)\, K^{(2)}(0)^{-2}\{1+o(1)\}.\end{aligned}$$

This means that

$$\mathrm{var}(S_{n1}) = 2h^d\, K^{(4)}(0)\, K^{(2)}(0)^{-2}\{1+o(1)\}. \tag{A.4}$$

Let $P(\xi_i)$, $P(\xi_i,\xi_j)$, $P(\xi_i,\xi_j,\xi_k)$ and $P(\xi_i,\xi_j,\xi_k,\xi_l)$ be the probability measures of ξ_i, (ξ_i,ξ_j), (ξ_i,ξ_j,ξ_k) and $(\xi_i,\xi_j,\xi_k,\xi_l)$ for different $i,j,k,l \in \{1,\ldots,n\}$ respectively. Define, for some constant $\delta>0$,

$$M_{n1} = \max_{1<i<j\leqslant n} \max\left\{E|\phi_{1j}\phi_{ij}|^{1+\delta},\ \int |\phi_{1j}\phi_{ij}|^{1+\delta}\, \mathrm{d}P(\xi_1)\, \mathrm{d}P(\xi_i,\xi_j)\right\},$$

$$\begin{aligned}M_{n2} = \max_{1<i<j\leqslant n} \max\bigg\{&E|\phi_{1j}\phi_{ij}|^{2(1+\delta)},\ \int |\phi_{1j}\phi_{ij}|^{2(1+\delta)}\, \mathrm{d}P(\xi_1)\, \mathrm{d}P(\xi_i,\xi_j),\\ &\int |\phi_{1j}\phi_{ij}|^{2(1+\delta)}\, \mathrm{d}P(\xi_1,\xi_i)\, \mathrm{d}P(\xi_j),\ \int |\phi_{1j}\phi_{ij}|^{2(1+\delta)}\, \mathrm{d}P(\xi_1)\, \mathrm{d}P(\xi_i)\, \mathrm{d}P(\xi_j)\bigg\},\end{aligned}$$

$$M_{n3} = \max_{1<i<j\leqslant n}(E|\phi_{1j}\phi_{ij}|^2),$$

$$M_{n4} = \max_{1<i,\, j,\, k\leqslant n}\left\{\max_P\left(\int |\phi_{1j}\phi_{jk}|^{2(1+\delta)}\, \mathrm{d}P\right)\right\},$$

$$M_{n5} = \max_{1<i<j} \max\left\{E\left|\int \phi_{1i}\phi_{1j}\, \mathrm{d}P(\xi_1)\right|^{2(1+\delta)},\ \int\left|\int \phi_{1i}\phi_{1j}\, \mathrm{d}P(\xi_1)\right|^{2(1+\delta)} \mathrm{d}P(\xi_i)\, \mathrm{d}P(\xi_j)\right\},$$

$$M_{n6} = \max_{1<i<j} \left\{ E \left| \int \phi_{1i}\phi_{1j} \, \mathrm{d}P(\xi_1) \right|^2 \right\},$$

where, in M_{n4}, i, j and k are mutually different and the maximization over P is taken over $P(\xi_1, \xi_i, \xi_j, \xi_k)$, $P(\xi_1)\ P(\xi_i, \xi_j, \xi_k)$, $P(\xi_1)\ P(\xi_i)\ P(\xi_j, \xi_k)$ and $P(\xi_1)\ P(\xi_i)\ P(\xi_j)\ P(\xi_k)$.

According to theorem A of Hjellvik *et al.* (1996), to show that $\sigma_n^{-1} S_{n1}^0$ is asymptotically standard normally distributed, it is sufficient to check that for some $\delta > 0$ and as $n \to \infty$

$$\max[\sigma_n^{-2}\{n^2(M_{n1}^{1/(1+\delta)} + M_{n5}^{1/2(1+\delta)} + M_{n6}^{1/2}), n^{3/2}(M_{n2}^{1/2(1+\delta)} + M_{n3}^{1/2} + M_{n4}^{1/2(1+\delta)})\}] \to 0. \tag{A.5}$$

Rather than evaluating all the M_{nl}-terms, we present here only the order of magnitude of M_{n1} and M_{n6} as the other terms can be evaluated similarly. Let $p, q > 1$ such that $p^{-1} + q^{-1} = 1$. From condition (b)

$$\begin{aligned} A_1 =: E|\phi_{1j}\phi_{ij}|^{1+\delta} &\leqslant Ch^{-2d(1+\delta)}\ E\left|\varepsilon_1\varepsilon_i\varepsilon_j^2\,|K|^{(2)}\left(\frac{X_1 - X_j}{h}\right)|K|^{(2)}\left(\frac{X_i - X_j}{h}\right)\right|^{1+\delta} \\ &\leqslant Ch^{-2d(1+\delta)}(E|\varepsilon_1\varepsilon_i\varepsilon_j^2|^{(1+\delta)p})^{1/p}\left[E\left\{|K|^{(2)}\left(\frac{X_1 - X_j}{h}\right)\ |K|^{(2)}\left(\frac{X_i - X_j}{h}\right)\right\}^{(1+\delta)q}\right]^{1/q}. \end{aligned}$$

Condition (e) implies that $E|\varepsilon_1\varepsilon_i\varepsilon_j^2|^{(1+\delta)p} < C$. Let $f_{1,i,j}$ be the joint density of (X_1, X_i, X_j). Condition (f) means that

$$\begin{aligned} &E\left\{|K|^{(2)}\left(\frac{X_1 - X_j}{h}\right)\ |K|^{(2)}\left(\frac{X_i - X_j}{h}\right)\right\}^{(1+\delta)q} \\ &= h^{2d}\int\int\int\{|K|^{(2)}(u)||K|^{(2)}(v)|\}^{(1+\delta)q}\ f_{1,i,j}(z - hu, z - hv, z)\,\mathrm{d}u\,\mathrm{d}v\,\mathrm{d}z \\ &\leqslant Ch^{2d}. \end{aligned} \tag{A.6}$$

Therefore, $A_1^{1/(1+\delta)} \leqslant Ch^{-2d\{1-1/q(1+\delta)\}} = o(h^{-d})$ if we choose q such that $1 < q < 2/(1+\delta)$. To evaluate the other term in M_{n1}, let $E_{i,j}$ be expectations with respect to (ξ_i, ξ_j) and

$$\begin{aligned} B_1 &=: E_1 E_{i,j}|\phi_{1j}\phi_{ij}|^{1+\delta} \\ &\leqslant Ch^{-2d(1+\delta)}\ E_1\left(|\varepsilon_1|^{1+\delta}\left[E_{i,j}\left\{|K|^{(2)}\left(\frac{X_1 - X_j}{h}\right)\ |K|^{(2)}\left(\frac{X_i - X_j}{h}\right)\right\}^{(1+\delta)q}\right]^{1/q}\right) \\ &\leqslant Ch^{-2d(1+\delta)+2d/q}. \end{aligned}$$

Hence $B_1^{1/(1+\delta)} \leqslant Ch^{-2d\{1-1/q(1+\delta)\}} = o(h^{-d})$ if we choose q such that $1 < q < 2/(1+\delta)$. Combining the results on A_1 and B_1, we have $\sigma_n^{-2} n^2 M_{n1}^{1/(1+\delta)} \to 0$ as $n \to \infty$.

Now let us consider M_{n6}. Let $A_6 = E_{i,j}(E_1|\phi_{1i}\phi_{1j}|)^2$. Similar to inequality (A.6),

$$\begin{aligned} E_1|\phi_{1i}\phi_{1j}| &\leqslant C|\varepsilon_i\varepsilon_j|h^{-2d}\ E_1\left\{\varepsilon_1^2|K|^{(2)}\left(\frac{X_1 - X_j}{h}\right)\ |K|^{(2)}\left(\frac{X_i - X_j}{h}\right)\right\} \\ &\leqslant C|\varepsilon_i\varepsilon_j|h^{-2d+d/q}\int\left\{|K|^{(2)}\left(w + \frac{X_i - X_j}{h}\right)\right\}^q f(w)\,\mathrm{d}w. \end{aligned} \tag{A.7}$$

Define another pair of $p_1, q_1 > 1$ such that $p_1^{-1} + q_1^{-1} = 1$. From inequality (A.7),

$$\begin{aligned} A_6 &\leqslant Ch^{-4d+2d/q}\ E_{i,j}\left(\varepsilon_i^2\varepsilon_j^2\left[\int\left\{|K|^{(2)}\left(w + \frac{X_i - X_j}{h}\right)\right\}^q f(w)\,\mathrm{d}w\right]^{2/q}\right) \\ &\leqslant Ch^{-4d+2d/q+d/q_1}. \end{aligned} \tag{A.8}$$

Hence $M_{n6}^{1/2} \leqslant Ch^{-d(2-1/q-1/2q_1)}$. By choosing q_1 (after choosing p and q as above) such that $0 < 2 - 1/q - 1/2q_1 < 1$, $M_{n6}^{1/2} = o(h^{-d})$, so $\sigma_n^{-2} n^2 M_{n6}^{1/2} = o(1)$. Hence we establish expression (A.5) and the asymptotic normality of S_{n1}^0.

In summary of equations (A.1)–(A.4) and the asymptotic normality of S_{n1}^0, we have

$$h^{-d/2}\left\{l_n(\tilde{m}_{\hat{\theta}}) - 1 - h^{d/2}\int_{x\in S} V^{-1}(x)\,\Delta^2(x)\,\mathrm{d}x\right\} \xrightarrow{\mathrm{d}} N(0,\sigma_0^2) \tag{A.9}$$

as $n\to\infty$, where $\sigma_0^2 = 2\,K^{(4)}(0)\,K^{(2)}(0)^{-2}$.

In what follows we shall prove the asymptotic normality of $T =: \int_S \mathcal{N}^2(s)\,\mathrm{d}s$. Let

$$\mathcal{N}_1(s) = \mathcal{N}(s) - h^{d/4}\,\Delta_n(s)\,f(s)/\sqrt{V(s)}.$$

Then $\mathcal{N}_1(s)$ is a Gaussian process with zero mean and covariance Ω. Split T as $T_1+T_2+T_3$ where

$$T_1 = \int_S \mathcal{N}_1^2(s)\,\mathrm{d}s,$$

$$T_2 = 2h^{d/4}\int_S V^{-1/2}(s)\,\Delta_n(s)\,\mathcal{N}_1(s)\,\mathrm{d}s,$$

$$T_3 = h^{d/2}\int_S V^{-1}(s)\,\Delta_n^2(s)\,\mathrm{d}s.$$

As Ω is bounded, $\int_S \Omega(t,t)\,\mathrm{d}t<\infty$. From results on stochastic integrals and equation (3.2),

$$E(T_1) = \int_S \Omega(s,s)\,\mathrm{d}s = 1,$$

$$\begin{aligned}\operatorname{var}(T_1) &= 2\int_S\int_S \Omega^2(s,t)\,\mathrm{d}s\,\mathrm{d}t\\ &= 2\,K^{(2)}(0)^{-2}\int_S\int_S K^{(2)}\{(s-t)/h\}^2\,\mathrm{d}s\,\mathrm{d}t\{1+o(1)\} &\text{(A.10)}\\ &= h^d\,K^{(4)}(0)K^{(2)}(0)^{-2}+o(h^d). &\text{(A.11)}\end{aligned}$$

Therefore, $\operatorname{var}(T_1) = 2h^d\,K^{(4)}(0)K^{(2)}(0)^{-2}+o(h^{2d})$. It is obvious that $E(T_2)=0$ and

$$\operatorname{var}(T_2) = 4h^{d/2}\int\int V^{-1/2}(s)\,\Delta_n(s)\,\Omega(s,t)\,V^{-1/2}(t)\,\Delta_n(t)\,\mathrm{d}s\,\mathrm{d}t.$$

As Δ_n and V^{-1} are bounded in S, there are constants C_1 and C_2 such that

$$\operatorname{var}(T_2) \leqslant C_1 h^{d/2}\int\int \Omega(s,t)\,\mathrm{d}s\,\mathrm{d}t \leqslant C_2 h^{3d/2}.$$

Hence $h^{-d/2}T_2 \to^{\mathrm{p}} 0$ as $n\to\infty$. As T_3 is non-random, we have

$$E(T) = 1 + h^{d/2}\int V^{-1}(s)\,\Delta_n^2(s)\,\mathrm{d}s + o(h^{d/2}), \tag{A.12}$$

$$\operatorname{var}(T) = 2h^d\,K^{(4)}(0)\,K^{(2)}(0)^{-2}+o(h^d). \tag{A.13}$$

It remains to prove the asymptotic normality of $h^{-d/2}\int_S\{N_0^2(s)-1\}\,\mathrm{d}s$. Let us first consider the case of $d=1$. Let δ_1 and δ_2 be sequences tending to 0 as $n\to\infty$, $\delta=\delta_1+\delta_2$, and $r=[1/\delta]$. In particular, we assume that $\delta_2=o(\delta_1)$, $h=o(\delta_2)$ and $\delta_2>2h$. For $i=1,\ldots,r$, let

$$V_i = \int_{(i-1)\delta}^{(i-1)\delta+\delta_1}\{\mathcal{N}_1^2(s)-1\}\,\mathrm{d}s,$$

$$V_i' = \int_{(i-1)\delta+\delta_1}^{i\delta}\{\mathcal{N}_1^2(s)-1\}\,\mathrm{d}s.$$

Since $\delta_2>2h$, the covariance function and the strict stationarity of $\mathcal{N}_1(s)$ mean that $\{V_i\}_{i=1}^r$ and $\{V_i'\}_{i=1}^r$ are independent and identically distributed.

A standard derivation shows that $\mathrm{var}(V_i) = \sigma_0^2 \delta_1 h\{1 + o(1)\}$ and $\mathrm{var}(V_i') = \sigma_0^2 \delta_2 h\{1 + o(1)\}$ where $\sigma_0^2 = 2\, K^{(4)}(0)\, K^{(2)}(0)^{-2}$. According to the central limit theorem, as $n \to \infty$,

$$\begin{aligned} &\sum_{i=1}^{r} V_i / \surd(r\delta_1 \sigma_0^2 h) \xrightarrow{\mathrm{d}} N(0, 1), \\ &\sum_{i=1}^{r} V_i' / \surd(r\delta_2 \sigma_0^2 h) \xrightarrow{\mathrm{d}} N(0, 1). \end{aligned} \tag{A.14}$$

Note that

$$\begin{aligned} (\sigma_0^2 h)^{-1/2} \int_0^1 \{\mathcal{N}_0^2(s) - 1\}\, \mathrm{d}s = (\sigma_0^2 h)^{-1/2} \sum_{i=1}^{r} V_i + (\sigma_0^2 h)^{-1/2} \sum_{i=1}^{r} V_i' \\ + (\sigma_0^2 h)^{-1/2} \int_{r\delta}^1 \{\mathcal{N}_0^2(s) - 1\}\, \mathrm{d}s. \end{aligned} \tag{A.15}$$

Since $r\delta_1 \to 1$ and $r\delta_2 \to 0$, results (A.14) mean that

$$(\sigma_0^2 h)^{-1/2} \sum_{i=1}^{r} V_i \xrightarrow{\mathrm{d}} N(0, 1),$$

$$(\sigma_0^2 h)^{-1/2} \sum_{i=1}^{r} V_i' \xrightarrow{\mathrm{p}} 0.$$

It can be easily shown that $(\sigma_0^2 h)^{-1/2} \int_{r\delta}^1 \{\mathcal{N}_0^2(s) - 1\}\, \mathrm{d}s \to^{\mathrm{p}} 0$ as $n \to \infty$. Thus, we establish the asymptotic normality for $d = 1$.

For the case of $d > 1$, let $S_{d-1} = [0, 1]^{d-1}$ and $T(s) = \int_{S_{d-1}} \{\mathcal{N}_1^2(s, t) - 1\}\, \mathrm{d}t$ where $s \in [0, 1]$. For $i = 1, \ldots, r$, define

$$V_i = \int_{(i-1)\delta}^{(i-1)\delta+\delta_1} T(s)\, \mathrm{d}s,$$

$$V_i' = \int_{(i-1)\delta+\delta_1}^{i\delta} T(s)\, \mathrm{d}s.$$

where δ_1, δ_2 and r are the same quantities defined earlier for $d = 1$. Because of the use of a product kernel and strict stationarity, $\{V_i\}_{i=1}^r$ and $\{V_i'\}_{i=1}^r$ are independent and identically distributed random variables. The asymptotic normality can be proved in a similar fashion.

References

Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Rev. Finan. Stud.*, **9**, 385–426.

Bosq, D. (1998) Nonparametric statistics for stochastic processes. *Lect. Notes Statist.*, **110**.

Chen, S. X. (1996) Empirical likelihood for nonparametric density estimation. *Biometrika*, **83**, 329–341.

Chen, S. X., Härdle, W. and Li, M. (2002) An empirical likelihood goodness-of-fit test for time series. *Research Report 3/2001*. Department of Statistics and Applied Probability, National University of Singapore. (Available from `http://www.stat.nus.edu.sg/`.)

Cox, J. C., Ingersoll, J. E. and Ross, S. A. (1985) A theory of term structure of interest rates. *Econometrica*, **53**, 385–407.

Eubank, R. L. and Spiegelman, C. H. (1990) Testing the goodness of fit of a linear model via nonparametric regression techniques. *J. Am. Statist. Ass.*, **85**, 387–392.

Fan, J. and Zhang, J. (2000) Sieve empirical likelihood ratio tests for nonparametric functions. *Research Report*. Department of Statistics, Chinese University of Hong Kong.

Gao, J. T. and King, M. (2001) Estimation and model specification testing in nonparametric and semiparametric time series econometric models. *Technical Report*. School of Mathematics and Statistics, University of Western Australia, Perth.

Genon-Catalot, V., Jeantheau, T. and Laredo, C. (2000) Stochastic Volatility models as hidden markov models and statistical applications. *Bernoulli*, **6**, 1051–1079.

Gorodeskii, V. V. (1977) On the strong mixing property for linear sequences. *Theory Probab. Applic.*, **22**, 411–413.

Hall, P. and La Scala, B. (1990) Methodology and algorithms of empirical likelihood. *Int. Statist. Rev.*, **58**, 109–127.

Härdle, W., Kleinow, T., Korostelev, A., Logeay, C. and Platen, E. (2000) Diffusion estimation and modeling of a stock market index. *Discussion Paper*. Humboldt-Universität zu Berlin, Berlin.
Härdle, W. and Mammen, E. (1993) Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926–1947.
Hart, J. (1997) *Nonparametric Smoothing and Lack-of-fit Tests*. Heidelberg: Springer.
Hjellvik, V., Yao, Q. and Tjøstheim, D. (1996) Linearity testing using local polynomial approximation. *Discussion Paper*. Humboldt-Universität zu Berlin, Berlin.
Hjellvik, V., Yao, Q. and Tjøstheim, D. (1998) Linearity testing using local polynomial approximation. *J. Statist. Planng. Inf.*, **68**, 295–321.
Hong, Y. and Li, H. (2001) Nonparametric specification testing for continuous-time models with application to spot interest rates. *Research Report*. Department of Economics, Cornell University, Ithaca.
Horowitz, J. L. and Spokoiny, V. G. (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**, 599–631.
Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, **25**, 2084–2102.
Koul, H. L. and Stute, W. (1999) Nonparametric model checks for time series. *Ann. Statist.*, **27**, 204–236.
Kreiss, J., Neumann, M. H. and Yao, Q. (1998) Bootstrap tests for simple structures in nonparametric time series regression. *Discussion Paper*. Humboldt-Universität zu Berlin, Berlin.
Masry, E. and Tjøstheim, D. (1995) Nonparametric estimation and identification of nonlinear arch time series. *Econometr. Theory*, **11**, 258–289.
Monti, A. C. (1997) Empirical likelihood confidence regions in time series models. *Biometrika*, **84**, 395–405.
Owen, A. B. (1990) Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.
Owen, A. B. (2001) *Empirical Likelihood*. London: Chapman and Hall.
Pham, T. D. and Tran, L. T. (1985) Some mixing properties of time series models. *Stoch. Process. Applic.*, **19**, 297–303.
Platen, E. (1999) Risk premia and financial modelling without measure transformation. *Preprint*. School of Finance and Economics, University of Technology, Sydney.
Qin, J. and Lawless, J. (1994) Empirical likelihood and general estimating functions. *Ann. Statist.*, **22**, 300–325.
Tripathi, G. and Kitamura, Y. (2000) On testing conditional moment restrictions: the “canonical case”. *Research Report*. Department of Economics, University of Wisconsin—Madison, Madison.
Wood, A. T. A. and Chan, G. (1994) Simulation of stationary Gaussian process in $[0,1]^d$. *J. Comput. Graph. Statist.*, **3**, 409–432.
Yoshihara, K. (1976) Limiting behaviour of U-Statistics for stationary absolutely regular processes. *Z. Wahrsch. Ver. Geb.*, **35**, 237–252.