Econometric Theory, **27**, 2011, 154–177.
doi:10.1017/S0266466610000150

EMPIRICAL-LIKELIHOOD-BASED CONFIDENCE INTERVALS FOR CONDITIONAL VARIANCE IN HETEROSKEDASTIC REGRESSION MODELS

NGAI HANG CHAN
The Chinese University of Hong Kong

LIANG PENG
Georgia Institute of Technology

DABAO ZHANG
Purdue University

Fan and Yao (1998) proposed an efficient method to estimate the conditional variance of heteroskedastic regression models. Chen, Cheng, and Peng (2009) applied variance reduction techniques to the estimator of Fan and Yao (1998) and proposed a new estimator for conditional variance to account for the skewness of financial data. In this paper, we apply empirical likelihood methods to construct confidence intervals for the conditional variance based on the estimator of Fan and Yao (1998) and the reduced variance modification of Chen et al. (2009). Simulation studies and data analysis demonstrate the advantage of the empirical likelihood method over the normal approximation method.

1. INTRODUCTION

Let $\{(Y_i, X_i)\}$ be a strictly stationary process with the same marginal distribution as (Y, X), where $Y \in R$ and $X \in R^d$. Let $m(x) = \mathrm{E}[Y|X = x]$ and $\sigma^2(x) = \mathrm{Var}(Y|X = x) \neq 0$ be the regression function and the conditional variance, respectively. Consider the nonparametric regression model

$$Y_i = m(X_i) + \sigma(X_i)\epsilon_i. \tag{1.1}$$

We are grateful to two referees and Prof. Kitamura, one of the co-editors, for their constructive comments. Chan's research was supported in part by HKSAR-RGC grants CUHK 4043/02P, 400305, and 400306; Peng's research was supported in part by NSF grants DMS-0403443, SES-0631608, and a Humboldt Research Fellowship; and Zhang's research was supported in part by the Purdue Research Foundation. Address correspondence to Ngai Hang Chan, Department of Statistics, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong; email: nhchan@sta.cuhk.edu.hk.

 © Cambridge University Press 2010

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

Here, $\mathrm{E}[\epsilon_i|X_i] = 0$ and $\mathrm{Var}(\epsilon_i|X_i) = 1$. There is extensive literature on estimating the nonparametric regression function $m(x)$ and nonparametrically estimating the conditional variance $\sigma^2(x)$; see Fan and Yao (2003, Ch. 8) and the references therein.

One of the motivations to study model (1.1) is that it encompasses interesting models of volatility. For example, by taking $X_i = (Y_{i-1}, \ldots, Y_{i-d})^{\mathrm{T}}$, model (1.1) can be construed as an autoregressive conditional heteroskedasticity (ARCH) model when

$$m(X_i) = 0, \qquad \sigma(X_i) = \sqrt{1 + \sum_{j=1}^{d} b_i Y_{i-j}^2}. \tag{1.2}$$

When

$$m(X_i) = \sum_{j=1}^{q} a_j Y_{i-j}, \qquad \sigma(X_i) = \sqrt{1 + \sum_{j=1}^{d} b_i Y_{i-j}^2}, \tag{1.3}$$

model (1.1) becomes an autoregressive (AR) process with ARCH errors, as termed by Borkovec (2001) and Borkovec and Klüppelberg (2001). It was also referred to as a double autoregressive model by Ling (2004, 2007). For model (1.3), Borkovec (2001) and Borkovec and Klüppelberg (2001) discuss its probabilistic properties; Ling (2004) and Chan and Peng (2005) discuss its statistical properties with $q = 1$ and $d = 1$; and Klüppelberg and Pergamenchtchikov (2004) and Ling (2007) study its general properties. Although many traditional models are parametric in nature, by putting them under the nonparametric context of model (1.1), we can make use of developed nonparametric techniques to estimate $m(x)$ and $\sigma(x)$. Nonparametric estimators of $m(x)$ and $\sigma(x)$ can be used to construct goodness-of-fit test statistics for a given parametric model. There is extensive literature in statistics and econometrics on testing specific parametric forms against nonparametric forms; see, for example, Härdle and Mammen (1993), Andrews (1997), Fan and Huang (2001), Fan, Zhang, and Zhang (2001), Horowitz and Spokoiny (2001), Chen and Gao (2007), and Wang and Van Keilegom (2007).

Another equally important reason to study model (1.1) is that many commonly used continuous-time stochastic models in financial econometrics satisfy the following generalized diffusion process:

$$dX_t = \mu(X_t, t)\,dt + \sigma(X_t, t)\,dW_t, \tag{1.4}$$

where W_t is a standard Brownian motion; see, for example, Vasicek (1977), Cox, Ingersoll, and Ross (1985), and Kloeden and Platen (1999). Diffusion process (1.4) is often discretized and approximated by model (1.1) through the Euler scheme, that is,

$$X_{t+\Delta} - X_t = \mu(X_t, t)\Delta + \sigma(X_t, t)\Delta^{1/2}\epsilon_t, \tag{1.5}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

where $\{\epsilon_t\}$ is a sequence of independent and identically distributed (i.i.d.) standard normal random variables. By taking $Y_t = X_{t+\Delta} - X_t$, $\mu(X_t, t)\Delta = m(X_t)$, and $\sigma(X_t, t)\Delta^{1/2} = \sigma(X_t)$, model (1.5) can be viewed as a special case of model (1.1). Although in practice one often assumes certain parametric forms for model (1.4), many situations in finance and economics constitute proper set-ups to estimate $\sigma(x)$ nonparametrically, as argued by Aït-Sahalia (1996). One of the reasons is that in practice, it is virtually impossible to identify a parametric functional form of the diffusion function $\sigma(x)$ in (1.4) a priori because the volatility of financial series is unobservable; see Aït-Sahalia for further articulations on this issue. As a result, it is desirable to obtain a nonparametric estimator for $\sigma(x)$. Aït-Sahalia and Fan, Jiang, Zhang, and Zhou (2003) studied nonparametric estimates of $\mu(x)$ and $\sigma(x)$. Fan (2005) gave a succinct and comprehensive review of the nonparametric estimation and statistical inference for model (1.4) and its discretized version (1.5).

For model (1.1) with $d = 1$, Fan and Yao (1998) proposed to estimate $m(x)$ by local linear techniques, i.e., $\hat{m}(x) = \hat{a}$, where

$$(\hat{a}, \hat{b}) = \text{argmin}_{(a,b)} \sum_{i=1}^{n} \{Y_i - a - b(X_i - x)\}^2 K\left(\frac{X_i - x}{h_2}\right), \tag{1.6}$$

K is a density function, and $h_2 > 0$ is a bandwidth. They then proposed to estimate $\sigma^2(x)$ by $\hat{\sigma}^2(x) = \hat{\alpha}$, where

$$(\hat{\alpha}, \hat{\beta}) = \text{argmin}_{(\alpha,\beta)} \sum_{i=1}^{n} \{\hat{r}_i - \alpha - \beta(X_i - x)\}^2 W\left(\frac{X_i - x}{h_1}\right), \tag{1.7}$$

$\hat{r}_i = \{Y_i - \hat{m}(X_i)\}^2$, W is a density function, and $h_1 > 0$ is a bandwidth. One drawback of this approach is that the resulting estimate is not always positive. Instead of $\sigma^2(x)$, Yu and Jones (2004) recently studied local likelihood expansion for $\log \sigma^2(x)$, and the resulting conditional variance estimate is always positive. However, Yu and Jones (2004) derived the asymptotic limit of this local likelihood estimator under restrictive conditions such as independence between (X_i, Y_i)'s and $\epsilon_i \sim \text{N}(0, 1)$. This assumption of independence precludes many useful time series models that are subsumed under (1.1).

In this paper, we explore the feasibility of constructing an empirical-likelihood-based interval of $\sigma^2(x)$ for serially correlated data such as those exhibited by model (1.1). Note that conditional variance is pivotal for statistical analysis. It can be used to construct confidence intervals and conduct statistical inference. In finance, conditional variance is also referred to as volatility, which is directly related to option pricing and risk measure quantification such as the notion of value at risk in risk management. Since Owen (1988, 1990) introduced empirical likelihood methods for constructing a confidence interval for the mean of a random variable, empirical likelihood methods have been extended in many directions. This is due to the attractiveness of empirical likelihood methods, i.e.,

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

Bartlett-correctable and automatic-shape-determined confidence regions. Empirical likelihood methods were first studied by Kitamura (1997) to analyze dependent data. Other developments of empirical likelihood methods include Kitamura (2001), Imbens (2002), Chen, Härdle, and Li (2003), Donald, Imbens, and Newey (2003), Kitamura, Tripathi, and Ahn (2004), Newey and Smith (2004), and Chen and Cui (2007) and the references therein. Recently, Chen and Qin (2000) applied the empirical likelihood method to construct confidence intervals for $m(x)$ with i.i.d. observations and showed that the resulting confidence intervals have no boundary effect in terms of coverage accuracy, whereas the confidence intervals based on normal approximation suffer from serious boundary effects. Based on the conditional variance estimator of Fan and Yao (1998) and its variance reduction version by Chen, Cheng, and Peng (2009), we generalize the idea of Chen and Qin (2000) to construct confidence intervals for $\sigma^2(x)$ for stationary time series. For further details on empirical likelihood methods, see Owen (2001).

The remainder of this paper is organized as follows: In Section 2 methodologies and main results are presented. Simulation studies and data analysis are reported in Section 3. Proofs are relegated to Section 4.

2. MAIN RESULTS

Throughout, we consider model (1.1) with $d = 1$. To evaluate $\hat{a}$ in (1.6), we solve equations for a and b,

$$\begin{cases} \dfrac{\partial}{\partial a}\displaystyle\sum_{i=1}^{n}\{Y_i - a - b(X_i - x)\}^2 K\left(\dfrac{x - X_i}{h_2}\right) = 0, \\ \dfrac{\partial}{\partial b}\displaystyle\sum_{i=1}^{n}\{Y_i - a - b(X_i - x)\}^2 K\left(\dfrac{x - X_i}{h_2}\right) = 0, \end{cases}$$

and obtain the solution $\hat{a}$ for a that has the expression

$$\hat{m}(x) = \hat{a} = \sum_{i=1}^{n} \Delta_i^*(x) Y_i \Big/ \sum_{i=1}^{n} \Delta_i^*(x),$$

where

$$\Delta_i^*(x) = K\left(\frac{x - X_i}{h_2}\right)\left\{s_{n,2}^*(x) - \frac{(x - X_i)s_{n,1}^*(x)}{h_2}\right\},$$

$$s_{n,l}^*(x) = (nh_2)^{-1}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_2}\right)\left(\frac{x - X_i}{h_2}\right)^l, \qquad l = 0, 1, 2.$$

Similarly, to evaluate $\hat{\alpha}$ in (1.7), we solve equations for α and β,

$$\begin{cases} \dfrac{\partial}{\partial \alpha}\displaystyle\sum_{i=1}^{n}\{\hat{r}_i - \alpha - \beta(X_i - x)\}^2 W\left(\dfrac{x - X_i}{h_1}\right) = 0, \\ \dfrac{\partial}{\partial \beta}\displaystyle\sum_{i=1}^{n}\{\hat{r}_i - \alpha - \beta(X_i - x)\}^2 W\left(\dfrac{x - X_i}{h_1}\right) = 0, \end{cases}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

and obtain the solution $\hat{\alpha}$ for α that has the explicit expression

$$\hat{\sigma}^2(x) = \hat{\alpha} = \sum_{i=1}^{n} \Delta_i(x)\hat{r}_i \Big/ \sum_{i=1}^{n} \Delta_i(x),$$

where $\hat{r}_i = (Y_i - \hat{m}(X_i))^2$,

$$\Delta_i(x) = W\left(\frac{x - X_i}{h_1}\right)\left\{s_{n,2}(x) - \frac{(x - X_i)s_{n,1}(x)}{h_1}\right\},$$

$$s_{n,l}(x) = (nh_1)^{-1}\sum_{i=1}^{n} W\left(\frac{x - X_i}{h_1}\right)\left(\frac{x - X_i}{h_1}\right)^l, \qquad l = 0, 1, 2.$$

For fixed x, let $\theta = \sigma^2(x)$. As $\mathrm{E}[\Delta_i(x)(\hat{r}_i - \theta)] \to 0$ according to the proof of Theorem 1 in Section 4, by letting $\mathrm{E}[\Delta_i(x)(\hat{r}_i - \theta)] = 0$ we can apply the empirical likelihood method to obtain a confidence interval for θ as follows: Let $p_1, \ldots, p_n$ be nonnegative numbers summing to 1. Then, the empirical likelihood for θ is

$$L(\theta) = \sup_{\sum p_i \Delta_i(x)(\hat{r}_i - \theta) = 0} \prod_{i=1}^{n} p_i.$$

Using the Lagrange multiplier method to obtain the optimal p_i's, the log empirical likelihood ratio becomes

$$\ell(\theta) = -2\log\{L(\theta)n^n\} = 2\sum_{i=1}^{n} \log\{1 + \lambda(\theta)\Delta_i(x)(\hat{r}_i - \theta)\},$$

where $\lambda(\theta)$ satisfies

$$\sum_{i=1}^{n} \Delta_i(x)(\hat{r}_i - \theta)\{1 + \lambda(\theta)\Delta_i(x)(\hat{r}_i - \theta)\}^{-1} = 0. \tag{2.1}$$

To derive the asymptotic limit of $l(\theta)$, the regularity conditions below are assumed. Let $p(\cdot)$ be the marginal density function of X.

Condition (C1). For a given point x, the functions $\mathrm{E}[Y^3|X = x]$, $\mathrm{E}[Y^4|X = x]$, and $p(x)$ are continuous at the point x, and $\ddot{m}(z) = \frac{d^2}{dz^2}m(z)$ and $\ddot{\sigma}^2(z) = \frac{d^2}{dz^2}\sigma^2(z)$ are uniformly continuous on an open set containing the point x. Further, assume that $p(x) > 0$.

Condition (C2). Let $\mathrm{E}[Y^{4(1+\delta)}] < \infty$ for some $\delta \in (0, 1)$.

Condition (C3). The kernel functions W and K are symmetric density functions, each having a bounded support in $(-\infty, \infty)$. Furthermore, there exists $M > 0$ such that $|W(x_1) - W(x_2)| \le M|x_1 - x_2|$ for all x_1 and x_2 in the support of W, and $|K(x_1) - K(x_2)| \le M|x_1 - x_2|$ for all x_1 and x_2 in the support of K.

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

Condition (C4). The strictly stationary process $\{(Y_i, X_i)\}$ is absolutely regular, i.e.,

$$\beta(j) := \sup_{i \geq 1} \mathrm{E}\left[\sup_{A \in \mathcal{F}_{i+j}^{\infty}} |P(A|\mathcal{F}_1^i) - P(A)|\right] \to 0, \qquad \text{as } j \to \infty,$$

where $\mathcal{F}_i^j$ is the σ-field generated by $\{(Y_k, X_k) : k = i, \ldots, j\}$, $(j \geq i)$. Further, for the δ given in Condition (C2), $\sum_{j=1}^{\infty} j^2 \beta^{\frac{\delta}{1+\delta}}(j) < \infty$.

Condition (C5). As $n \to \infty$, $0 < \liminf nh_1^3 \leq \limsup nh_1^3 < \infty$, and $0 < \liminf nh_2^3 \leq \limsup nh_2^3 < \infty$.

THEOREM 1. *Under the regularity conditions (C1) to (C5), we have* $\ell(\sigma^2(x)) \xrightarrow{d} \chi^2(1)$. *A level* α *confidence interval for* $\sigma^2(x)$ *is*

$$I_\alpha = \{\theta : \ell(\theta) \leq u_\alpha\},$$

where u_α *is the level* α *critical point of* $\chi^2(1)$.

Remark. The preceding regularity conditions, except Condition (C5), were employed by Fan and Yao (1998). As the empirical likelihood method is designed for $\theta = \mathrm{E}[\hat{\sigma}^2(x)]$ rather than $\sigma^2(x)$, Condition (C5) is required to ensure that the bias term is asymptotically negligible. It is similar to the construction of normal-approximation-based confidence intervals for a density function, whereby statisticians often prefer to first under-smooth so that the bias is asymptotically negligible; see Hall (1992) for detailed descriptions and comparisons. To choose the bandwidths in terms of coverage accuracy, we have to expand the coverage probability, which can be difficult and tricky. Herein we simply choose $h_1 = c_1 n^{-1/3}$ and $h_2 = c_2 n^{-1/3}$. Note that the order $h_1 = n^{-1/3}$ is the optimal rate in nonparametric smoothing estimation in terms of coverage accuracy; see Hall (1992) and Chen and Qin (2000).

Recently, Chen et al. (2009) applied variance reduction techniques to the conditional variance estimator $\hat{\sigma}^2(x)$. The idea of the variance reduction strategy is to construct a linear combination of $\hat{\sigma}^2(z)$ at three points around x such that the asymptotic bias remains unchanged. Specifically, for any given point x, let $\{\beta_{x,0}, \beta_{x,1}, \beta_{x,2}\}$ be three equally spaced grid points with bandwidth $\delta h_1 = \beta_{x,1} - \beta_{x,0}$, such that $x = \beta_{x,1} + r\delta h_1$ for some $r \in [-1, 1]$. Chen et al. (2009) defined the variance reduction estimator of $\sigma^2(x)$ as

$$\tilde{\sigma}^2(x, r) = \frac{r(r-1)}{2}\hat{\sigma}^2(\beta_{x,0}) + (1 - r^2)\hat{\sigma}^2(\beta_{x,1}) + \frac{r(r+1)}{2}\hat{\sigma}^2(\beta_{x,2}). \tag{2.2}$$

Moreover, they showed that the estimator $\tilde{\sigma}^2(x, r)$ has a smaller asymptotic variance and mean squared error than $\hat{\sigma}^2(x)$. By choosing $r = \pm 2^{-1/2}$, they showed

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

that maximum variance reduction can be achieved irrespective of the values of h_1, δ, and W.

Let $r_0 = 2^{-1/2}$. As both $\tilde{\sigma}^2(x, r_0)$ and $\tilde{\sigma}^2(x, -r_0)$ use more data from one side of x than the other, Chen et al. (2009) proposed an alternate estimator, i.e.,

$$\tilde{\sigma}^2(x) = \left\{\tilde{\sigma}^2(x, r_0) + \tilde{\sigma}^2(x, -r_0)\right\}/2,$$

to balance the finite sample bias effect. When the support of X is $[0, 1]$, we choose $\delta = \delta(x) = \min\left\{\delta_0, x/((1+2^{-1/2})h_1), (1-x)/((1+2^{-1/2})h_1)\right\}$ with a positive constant δ_0 to keep the points $\left\{\beta_{x,0}, \beta_{x,1}, \beta_{x,2}\right\}$ (with $r = \pm 2^{-1/2}$) within the data range $[0, 1]$. Under this consideration, we apply the empirical likelihood method to assess the alternate variance reduction estimator $\tilde{\sigma}^2(x)$. As before, the log empirical likelihood ratio is given by

$$\tilde{\ell}(\theta) = 2\sum_{i=1}^{n} \log\{1 + \tilde{\lambda}(\theta)\tilde{\Delta}_i(x)(\hat{r}_i - \theta)\},$$

where

$$\begin{aligned}\tilde{\Delta}_i(x) = \frac{1}{2}\Bigg\{ & \frac{1-2^{1/2}}{4}\Delta_i(x - (1+2^{-1/2})\delta h_1) + \frac{1}{2}\Delta_i(x - 2^{-1/2}\delta h_1) \\ & + \frac{1+2^{1/2}}{4}\Delta_i(x - (2^{-1/2}-1)\delta h_1) + \frac{1+2^{1/2}}{4}\Delta_i(x + (2^{-1/2}-1)\delta h_1) \\ & + \frac{1}{2}\Delta_i(x + 2^{-1/2}\delta h_1) + \frac{1-2^{1/2}}{4}\Delta_i(x + (2^{-1/2}+1)\delta h_1)\Bigg\},\end{aligned}$$

and $\tilde{\lambda}(\theta)$ satisfies

$$\sum_{i=1}^{n} \tilde{\Delta}_i(x)(\hat{r}_i - \theta)\{1 + \tilde{\lambda}(\theta)\tilde{\Delta}_i(x)(\hat{r}_i - \theta)\}^{-1} = 0.$$

THEOREM 2. *Under the conditions of Theorem 1,* $\tilde{\ell}(\sigma^2(x)) \xrightarrow{d} \chi^2(1)$. *Hence a level* α *confidence interval for* $\sigma^2(x)$ *is*

$$\tilde{I}_\alpha = \{\theta : \tilde{\ell}(\theta) \le u_\alpha\},$$

where u_α *is the level* α *critical point of* $\chi^2(1)$.

3. SIMULATIONS AND DATA ANALYSIS

3.1. Simulations

This simulation study is designed to investigate the performance of the empirical-likelihood-based confidence intervals I_α and $\tilde{I}_\alpha$ and compare them with the

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

normal-approximation-based confidence intervals I_α^*, which are constructed with $\hat{\sigma}^2(x)$, and $\tilde{I}_\alpha^*$, which are constructed with $\tilde{\sigma}^2(x)$, i.e.,

$$I_\alpha^* = \left(1 - \frac{z_\alpha}{\sqrt{nh_1}}\hat{A}(x)\int W^2(s)\,ds, \quad 1 + \frac{z_\alpha}{\sqrt{nh_1}}\hat{A}(x)\int W^2(s)\,ds\right) \times \hat{\sigma}^2(x)$$

and

$$\tilde{I}_\alpha^* = \left(1 - \frac{z_\alpha}{\sqrt{nh_1}}\hat{A}(x)\left\{\int W^2(s)\,ds - \frac{C(\delta)}{4} - \frac{D(\delta)}{2}\right\},\right.$$

$$\left. 1 + \frac{z_\alpha}{\sqrt{nh_1}}\hat{A}(x)\left\{\int W^2(s)\,ds - \frac{C(\delta)}{4} - \frac{D(\delta)}{2}\right\}\right) \times \tilde{\sigma}^2(x),$$

where $C(\delta)$ and $D(\delta)$ are defined in the proof of Theorem 2 (see Section 4), z_α denotes the αth quantile of a standard normal random variable, and $\hat{A}(x)$ is an estimator of $\{p^{-1}(x)\mathrm{E}[(\epsilon_1^2 - 1)^2|X_1 = x]\}^{1/2}$ that is constructed from estimating $p(x)$ by a kernel smoother and $\mathrm{E}[(\epsilon_1^2 - 1)^2|X_1 = x]$ by a local linear estimator. The kernel smoother of $p(x)$ uses a fixed bandwidth $h = 0.6n^{-1/5}$, and the local linear estimator for $\mathrm{E}[(\epsilon_1^2 - 1)^2|X_1 = x]$ uses the bandwidth h_1 in (1.7). In each of the following cases, we generate 10,000 data sets, and for each data set we calculate the four different confidence intervals using the Epanechnikov kernel and setting $\delta = 1$.

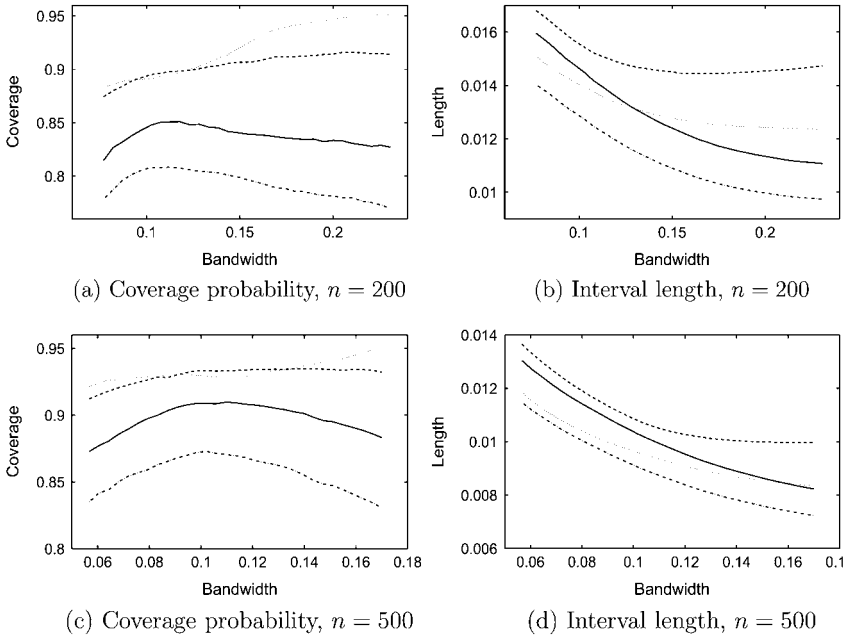
In line with Chen and Qin (2000), we first consider model (1.1) with

$$\begin{cases} m(x) = 0.3\exp\{-4(4x-1)^2\} + 0.7\exp\{-16(4x-3)^2\}, \\ \sigma^2(x) = 0.15\exp(-x), \end{cases} \tag{3.1}$$

and simulate X_i and ϵ_i independently from the uniform distribution on $[0, 1]$ and the standard normal distribution, respectively. In this example, $(X_1, Y_1), \ldots, (X_n, Y_n)$ are i.i.d. random vectors, which satisfy regularity conditions (C1)–(C4). We consider the point $x = 0.1$, which is near the boundary. For each simulated sample, 31 equally spaced levels of bandwidths $h_1 = h_2$, ranging from $0.45n^{-1/3}$ to $1.35n^{-1/3}$, are used to construct the four confidence intervals with the nominal level at 95%. The simulation results with sample sizes $n = 200$ and $n = 500$ are reported in Figure 1, which plots the average coverage probabilities and median lengths of the four confidence intervals against the bandwidth h_1.

As shown in Figure 1, the two empirical-likelihood-based confidence intervals I_α and $\tilde{I}_\alpha$ have better coverage accuracy than the normal-approximation-based confidence intervals I_α^* and $\tilde{I}_\alpha^*$. The uses of variance reduction clearly improve the coverage accuracy of the confidence intervals. Although for both empirical-likelihood-based and normal-approximation-based confidence intervals the uses of variance reduction lengthen the corresponding confidence intervals, the confidence interval constructed by applying variance reduction and the empirical

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press



(a) Coverage probability, $n = 200$ (b) Interval length, $n = 200$

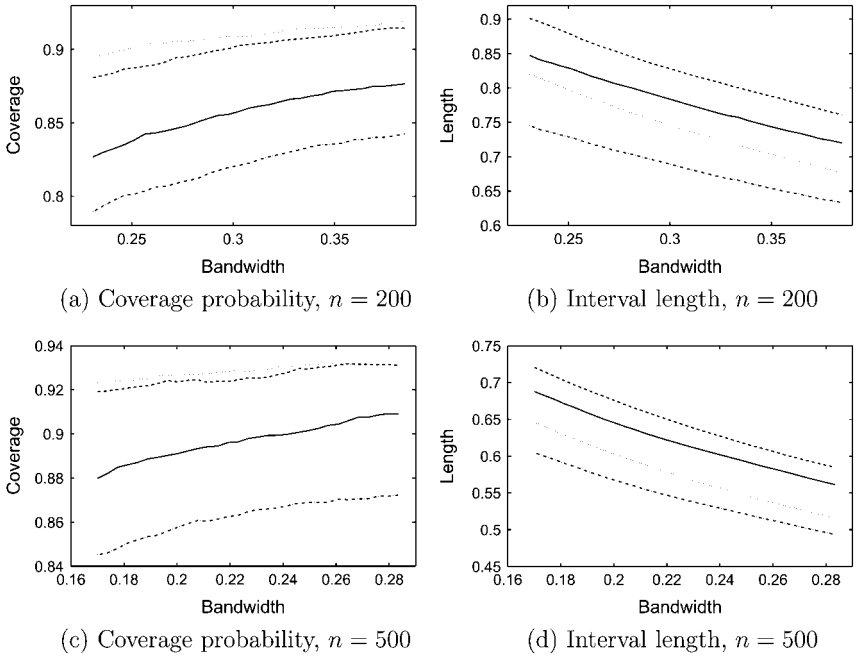
(c) Coverage probability, $n = 500$ (d) Interval length, $n = 500$

FIGURE 1. Model (3.1). The coverage probabilities and interval lengths of I_α^* (solid lines), $\tilde{I}_\alpha^*$ (dash-dotted lines), I_α (dashed lines), and $\tilde{I}_\alpha$ (dotted lines) at $x = 0.1$, plotted against bandwidth h_1.

likelihood method is clearly superior. We also observe that by increasing the sample sizes, the average accuracy of all four confidence intervals improves and their lengths shorten. It is interesting to observe that the coverage probability of all confidence intervals except $\tilde{I}_\alpha$ lies beneath the nominal 95% level, which cannot be attained by varying the bandwidths h_1 and h_2. The coverage probabilities of the two empirical-likelihood-based confidence intervals I_α and $\tilde{I}_\alpha$ are much closer to the nominal 95% level for all chosen bandwidths h_1 and h_2, which suggests that the normal-approximation-based confidence intervals may have a higher order of coverage probability error than the empirical-likelihood-based ones when x is near the boundary.

As a second example, consider model (1.1) with $m(x) = 0, \sigma(x) = \sqrt{1 + x^2/4}$, and $Y_i = X_{i+1}$. The errors $\{\epsilon_t\}$ are assumed to be i.i.d. standard normal. We consider the confidence intervals for $\sigma(0)$. Four confidence intervals with nominal level at 95% are constructed with the bandwidths $h_1 = h_2$ ranging from $1.35n^{-1/3}$ to $2.25n^{-1/3}$. The simulation results with sample sizes $n = 200$ and $n = 500$ are reported in Figure 2. Similar to the first example, both empirical-likelihood-based confidence intervals I_α and $\tilde{I}_\alpha$ have better coverage accuracy than the normal-approximation-based confidence intervals I_α^* and $\tilde{I}_\alpha^*$ (see Figure 2). The confidence interval constructed by applying both variance reduction and the empirical

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press



(a) Coverage probability, $n = 200$ (b) Interval length, $n = 200$

(c) Coverage probability, $n = 500$ (d) Interval length, $n = 500$

FIGURE 2. ARCH(1) Model. The coverage probabilities and interval lengths of I_α^* (solid lines), $\tilde{I}_\alpha^*$ (dash-dotted lines), I_α (dashed lines), and $\tilde{I}_\alpha$ (dotted lines) at $x = 0$, plotted against bandwidth h_1.

likelihood method has better coverage accuracy than the other three and is shorter than the others except $\tilde{I}_\alpha^*$.

We also consider the model (1.3),

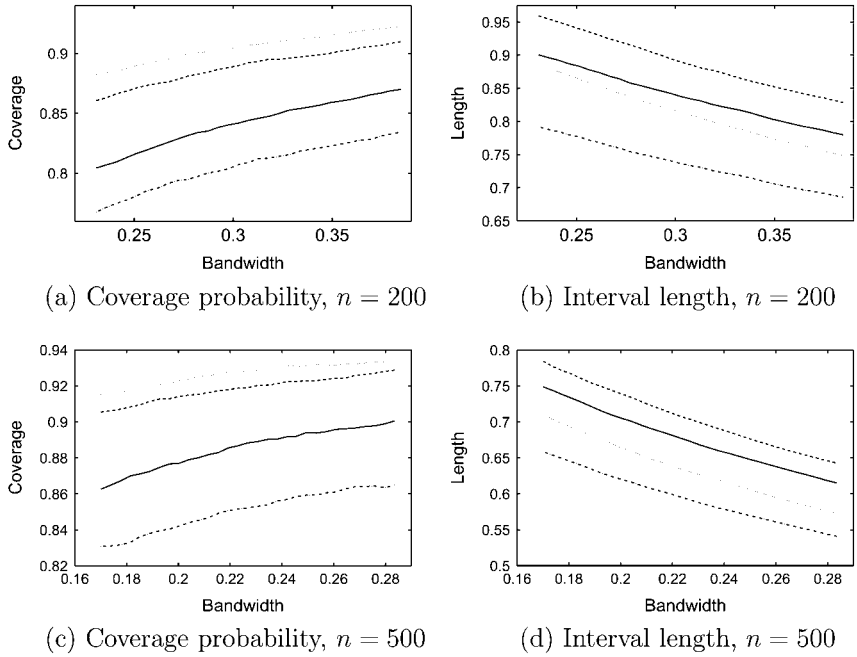
$$Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\epsilon_t,$$

where $m(x) = 0.5x$, $\sigma(x) = \sqrt{1 + 0.5x^2}$, and $\{\epsilon_t\}$ is a sequence of i.i.d. standard normal random variables. Using the same bandwidths as in the second example, we constructed four confidence intervals for $\sigma(0)$ with the nominal level at 95%. The simulation results with sample sizes $n = 200$ and $n = 500$ are reported in Figure 3, and the results are similar to those reported in Figure 2. The confidence interval that is constructed by applying both variance reduction and the empirical likelihood method is overwhelmingly better than the other three in terms of coverage accuracy.

3.2. Data Analysis

As an illustration, we examine the conditional variance of a financial data set. Consider the weekly interest rates of five-year U.S. Treasury notes from January

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press



(a) Coverage probability, $n = 200$ (b) Interval length, $n = 200$

(c) Coverage probability, $n = 500$ (d) Interval length, $n = 500$

FIGURE 3. AR(1) Process with ARCH(1) Errors. The coverage probabilities and interval lengths of I_α^* (solid lines), $\tilde{I}_\alpha^*$ (dash-dotted lines), I_α (dashed lines), and $\tilde{I}_\alpha$ (dotted lines) at $x = 0$, plotted against bandwidth h_1.

5, 1976, to March 26, 2004. Denote the weekly interest rates by X_i, $i = 1, \ldots, n$, and denote the weekly yields by $Y_i = X_{i+1} - X_i$. We fit model (1.1) to the data set (X_i, Y_i), see Figure 4(a), and construct confidence intervals for the conditional variance. Let x_L and x_U, respectively, denote the minimum and maximum values of the weekly interest rates in this data set. We use the same kernel and confidence level as in the simulation study and choose $h_1 = h_2 = 1.8(x_U - x_L)n^{-1/3}$ and $\delta(x) = \min\{\frac{1}{2}, \frac{x - x_L}{(1+2^{-1/2})h_1}, \frac{x_U - x}{(1+2^{-1/2})h_1}\}$. In Figure 4 we plot, against different values of x, $\hat{\sigma}^2(x)$ with its normal-approximation-based confidence interval $I_{0.95}^*$, $\tilde{\sigma}^2(x)$ with its normal-approximation-based confidence interval $\tilde{I}_{0.95}^*$, $\hat{\sigma}^2(x)$ with its empirical-likelihood-based confidence interval $I_{0.95}$, and $\tilde{\sigma}^2(x)$ with its empirical-likelihood-based confidence interval $\tilde{I}_{0.95}$. When calculating the normal-approximation-based confidence intervals, we fix the bandwidth for the kernel smoother of $p(x)$ at $h = 0.6(x_L - x_U)n^{-1/5}$. We observe that both empirical-likelihood-based confidence intervals, i.e., $I_{0.95}$ and $\tilde{I}_{0.95}$, are skewed to the right, which is consistent with the empirical finding that interest rates have an asymmetric distribution. In addition, both estimate and confidence intervals based on the variance reduction are smoother than those based on other methods.

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

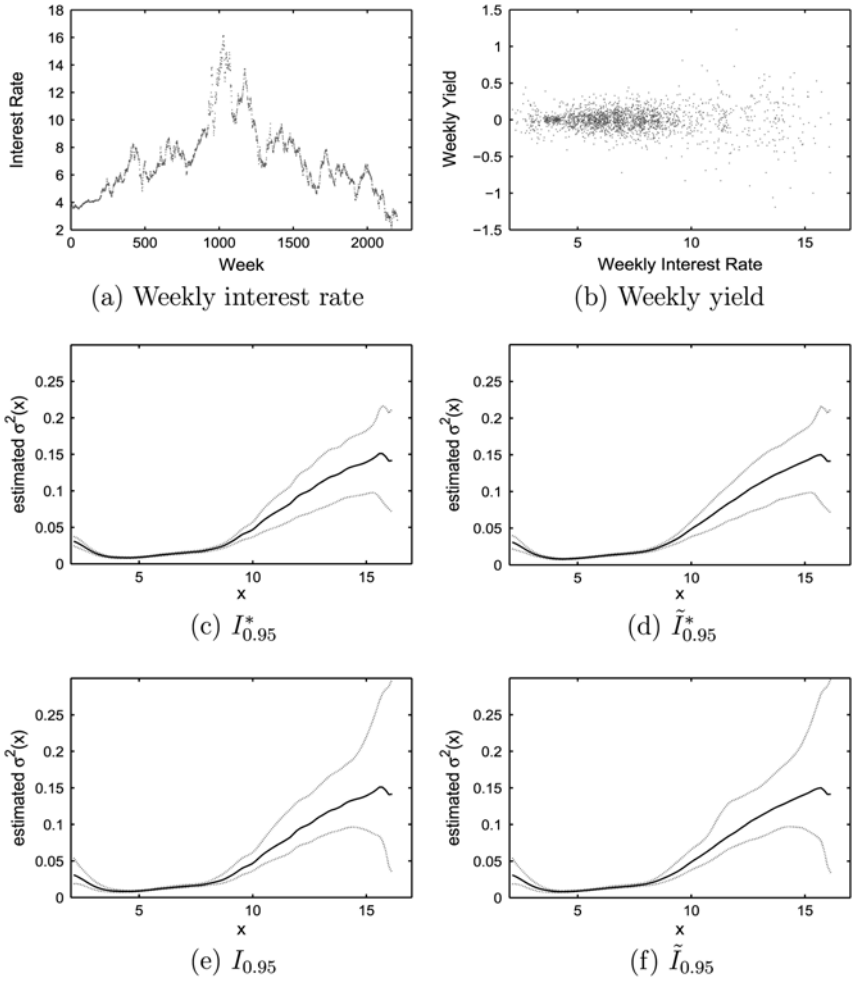


FIGURE 4. Five-year U.S. Treasury notes from January 5, 1976, to March 26, 2004, with the estimate (solid lines) and confidence interval (dotted lines) of the conditional variance of weekly yield.

4. PROOFS

Before proving the theorems, we need the following lemma, which is a modification of Theorem 1 of Fan and Yao (1998).

LEMMA 1. *Under the regularity conditions (C1) to (C5), we have*

$$\frac{1}{nh_1}\sum_{i=1}^{n} W\left(\frac{x-X_i}{h_1}\right)\{m(X_i)-\hat{m}(X_i)\}^2$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

$$= o_p\left(\frac{1}{n^{3/2}h_1h_2^2} + \frac{h_2}{nh_1} + \frac{1}{nh_2^2}\right) + O_p\left(\frac{h_2^3}{\sqrt{n}} + h_2^4\right) \tag{4.1}$$

$$= o_p(h_1 + h_2), \sqrt{nh_1}\{\hat{\sigma}^2(x) - \sigma^2(x)\}$$

$$= \frac{1}{p(x)\sqrt{nh_1}} \sum_{i=1}^{n} W\left(\frac{x - X_i}{h_1}\right) \sigma^2(X_i)(\epsilon_i^2 - 1) + o_p(1), \tag{4.2}$$

and

$$\sqrt{nh_1}\{\hat{\sigma}^2(x) - \sigma^2(x)\}$$

$$\xrightarrow{d} Z \sim N\left(0, p^{-1}(x)\sigma^4(x)\left\{\int W^2(s)\,ds\right\}\mathrm{E}[(\epsilon_1^2 - 1)^2 | X_1 = x]\right). \tag{4.3}$$

Proof. This lemma follows from the proof of Theorem 1 of Fan and Yao (1998) by using (C5) instead of $\liminf nh_1^4 > 0$ and $\liminf nh_2^4 > 0$. ■

Proof of Theorem 1. Without loss of generality, assume that the supports of K and W are $[-1, 1]$. It follows from Conditions (C1)–(C4) and the ergodic theorem that

$$\begin{cases} s_{n,0}(x) = \mathrm{E}\left[h_1^{-1} W\left(\dfrac{x - X_i}{h_1}\right)\right] + o_p(1) = p(x) + o_p(1) \\ s_{n,1}(x) = O_p(h_1) \\ s_{n,2}(x) = \mathrm{E}\left[h_1^{-3} W\left(\dfrac{x - X_i}{h_1}\right)(x - X_i)^2\right] + o_p(1) \\ \qquad\quad = p(x)\displaystyle\int s^2 W(s)\,ds + o_p(1). \end{cases} \tag{4.4}$$

Note that

$$\hat{r}_i = \sigma^2(X_i)\epsilon_i^2 + 2\sigma(X_i)\epsilon_i\{m(X_i) - \hat{m}(X_i)\} + \{m(X_i) - \hat{m}(X_i)\}^2$$

$$\le 2\sigma^2(X_i)\epsilon_i^2 + 2\{m(X_i) - \hat{m}(X_i)\}^2. \tag{4.5}$$

Put $Z_i = \Delta_i(x)\{\hat{r}_i - \sigma^2(x)\}$. By (4.5) we have

$$|Z_i| \le 2|\Delta_i(x)\sigma^2(X_i)\epsilon_i^2| + 2|\Delta_i(x)(m(X_i) - \hat{m}(X_i))^2| + |\Delta_i(x)\sigma^2(x)|$$

$$\le 2\left|W\left(\frac{x - X_i}{h_1}\right) s_{n,2}(x)\sigma^2(X_i)\epsilon_i^2\right|$$

$$+ 2\left|W\left(\frac{x - X_i}{h_1}\right)\frac{x - X_i}{h_1} s_{n,1}(x)\sigma^2(X_i)\epsilon_i^2\right|$$

$$+ 2\left|W\left(\frac{x - X_i}{h_1}\right) s_{n,2}(x)(m(X_i) - \hat{m}(X_i))^2\right|$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

$$
\begin{aligned}
&+2\left|W\left(\frac{x-X_i}{h_1}\right)\frac{x-X_i}{h_1}s_{n,1}(x)(m(X_i)-\hat{m}(X_i))^2\right| \\
&+\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,2}(x)\sigma^2(x)\right|+\left|W\left(\frac{x-X_i}{h_1}\right)\frac{x-X_i}{h_1}s_{n,1}(x)\sigma^2(x)\right| \\
&\le 2\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,2}(x)\sigma^2(X_i)\epsilon_i^2\right|+2\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,1}(x)\sigma^2(X_i)\epsilon_i^2\right| \\
&+2\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,2}(x)(m(X_i)-\hat{m}(X_i))^2\right| \\
&+2\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,1}(x)(m(X_i)-\hat{m}(X_i))^2\right| \\
&+\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,2}(x)\sigma^2(x)\right|+\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,1}(x)\sigma^2(x)\right|. \qquad \textbf{(4.6)}
\end{aligned}
$$

Using Condition (C2), we have, for any $b>0$,

$$
\begin{aligned}
&P\left(\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)\sigma^2(X_i)\epsilon_i^2\right|\ge b\sqrt{nh_1}\right) \\
&\le\sum_{i=1}^{n}P\left(\left|W\left(\frac{x-X_i}{h_1}\right)\sigma^2(X_i)\epsilon_i^2\right|\ge b\sqrt{nh_1}\right) \\
&\le\sum_{i=1}^{n}\mathrm{E}\left[\mathrm{E}\left[\left|W\left(\frac{x-X_i}{h_1}\right)\sigma^2(X_i)\epsilon_i^2\right|^{2(1+\delta)}\Big|X_i\right]\right]\left(b\sqrt{nh_1}\right)^{-2(1+\delta)} \\
&=O\left(nh_1/(nh_1)^{1+\delta}\right)=o(1). \qquad \textbf{(4.7)}
\end{aligned}
$$

It follows from (4.4) and (4.7) that

$$
\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,2}(x)\sigma^2(X_i)\epsilon_i^2\right|=o_p\left(\sqrt{nh_1}\right). \qquad \textbf{(4.8)}
$$

Similarly,

$$
\begin{cases}
\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,1}(x)\sigma^2(X_i)\epsilon_i^2\right|=o_p\left(\sqrt{nh_1}\right) \\
\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,2}(x)\sigma^2(x)\right|=o_p\left(\sqrt{nh_1}\right) \\
\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,1}(x)\sigma^2(x)\right|=o_p\left(\sqrt{nh_1}\right).
\end{cases} \qquad \textbf{(4.9)}
$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

It follows from Lemma 1 that

$$\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)(m(X_i)-\hat{m}(X_i))^2\right| \le \sum_{i=1}^{n} W\left(\frac{x-X_i}{h_1}\right)(m(X_i)-\hat{m}(X_i))^2$$
$$= o_p\left(nh_1(h_1+h_2)\right). \tag{4.10}$$

By (4.4), (4.10), and Condition (C5), we have

$$\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)s_{n,2}(x)(m(X_i)-\hat{m}(X_i))^2\right| = o_p\left(\sqrt{nh_1}\right). \tag{4.11}$$

Similarly

$$\max_{1\le i\le n}\left|W\left(\frac{x-X_i}{h_1}\right)\frac{x-X_i}{h_1}s_{n,1}(x)(m(X_i)-\hat{m}(X_i))^2\right| = o_p\left(\sqrt{nh_1}\right). \tag{4.12}$$

Therefore, by (4.6), (4.8), (4.9), (4.11), and (4.12), we have

$$\max_{1\le i\le n}|Z_i| = o_p\left(\sqrt{nh_1}\right). \tag{4.13}$$

As $1+\lambda(\theta)Z_i$ in $\ell(\theta)$ is nonnegative, we have

$$|1+\lambda(\theta)Z_i|^{-1} \ge \left\{1+|\lambda(\theta)|\max_{1\le i\le n}|Z_i|\right\}^{-1}.$$

By the definition of $\lambda(\theta)$ we have

$$0 = \frac{1}{nh_1}\left|\sum_{i=1}^{n}\left\{\frac{\lambda(\theta)Z_i^2}{1+\lambda(\theta)Z_i}-Z_i\right\}\right|$$
$$\ge \frac{1}{nh_1}\sum_{i=1}^{n}\frac{|\lambda(\theta)|Z_i^2}{|1+\lambda(\theta)Z_i|}-\left|\frac{1}{nh_1}\sum_{i=1}^{n}Z_i\right|$$
$$\ge \frac{1}{nh_1}\sum_{i=1}^{n}\frac{|\lambda(\theta)|Z_i^2}{1+|\lambda(\theta)|\max_{1\le i\le n}|Z_i|}-\left|\frac{1}{nh_1}\sum_{i=1}^{n}Z_i\right|,$$

which implies that

$$|\lambda(\theta)|\frac{1}{nh_1}\sum_{i=1}^{n}Z_i^2 \le \left|\frac{1}{nh_1}\sum_{i=1}^{n}Z_i\right|+|\lambda(\theta)|\max_{1\le i\le n}|Z_i|\left|\frac{1}{nh_1}\sum_{i=1}^{n}Z_i\right|,$$

i.e.,

$$|\lambda(\theta)|\left\{\frac{1}{nh_1}\sum_{i=1}^{n}Z_i^2-\max_{1\le i\le n}|Z_i|\left|\frac{1}{nh_1}\sum_{i=1}^{n}Z_i\right|\right\} \le \left|\frac{1}{nh_1}\sum_{i=1}^{n}Z_i\right|. \tag{4.14}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

It follows from Lemma 1 and (4.4) that

$$\begin{aligned}\frac{1}{\sqrt{nh_1}}\sum_{i=1}^{n} Z_i &= \frac{\sum_{j=1}^{n}\Delta_j(x)}{\sqrt{nh_1}}\left\{\hat{\sigma}^2(x)-\sigma^2(x)\right\}\\ &= \{s_{n,0}(x)s_{n,2}(x)-s_{n,1}^2(x)\}\sqrt{nh_1}\left\{\hat{\sigma}^2(x)-\sigma^2(x)\right\}\\ &\xrightarrow{d} p^2(x)\left\{\int s^2W(s)\,ds\right\}Z, \qquad \textbf{(4.15)}\end{aligned}$$

where Z is defined in Lemma 1. Write

$$\begin{aligned}\frac{1}{nh_1}\sum_{i=1}^{n} Z_i^2 &= \frac{1}{nh_1}\sum_{i=1}^{n}\Delta_i^2(x)\{\sigma^2(X_i)\epsilon_i^2-\sigma^2(x)\\ &\quad +2\sigma(X_i)\epsilon_i(m(X_i)-\hat{m}(X_i))+(m(X_i)-\hat{m}(X_i))^2\}^2\\ &= \frac{1}{nh_1}\sum_{i=1}^{n}\Delta_i^2(x)\{\sigma^2(X_i)\epsilon_i^2-\sigma^2(x)\}^2\\ &\quad +\frac{2}{nh_1}\sum_{i=1}^{n}\Delta_i^2(x)\{\sigma^2(X_i)\epsilon_i^2-\sigma^2(x)\}\{2\sigma(X_i)\epsilon_i(m(X_i)-\hat{m}(X_i))\\ &\quad +(m(X_i)-\hat{m}(X_i))^2\}\\ &\quad +\frac{1}{nh_1}\sum_{i=1}^{n}\Delta_i^2(x)\{2\sigma(X_i)\epsilon_i(m(X_i)-\hat{m}(X_i))+(m(X_i)-\hat{m}(X_i))^2\}^2\\ &= I_1+I_2+I_3\end{aligned}$$

and

$$\begin{aligned}I_1 &= \frac{s_{n,2}^2(x)}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)\{\sigma^2(X_i)\epsilon_i^2-\sigma^2(x)\}^2\\ &\quad -\frac{2s_{n,1}(x)s_{n,2}(x)}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)\frac{x-X_i}{h_1}\{\sigma^2(X_i)\epsilon_i^2-\sigma^2(x)\}^2\\ &\quad +\frac{s_{n,1}^2(x)}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)\frac{(x-X_i)^2}{h_1^2}\{\sigma^2(X_i)\epsilon_i^2-\sigma^2(x)\}^2\\ &= II_1+II_2+II_3.\end{aligned}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

By Conditions (C1)–(C4) and the ergodic theorem, we have

$$\begin{aligned}
&\frac{1}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)\{\sigma^2(X_i)\epsilon_i^2-\sigma^2(x)\}^2 \\
&\quad = h_1^{-1}\mathrm{E}[W^2\left(\frac{x-X_i}{h_1}\right)(\sigma^2(X_i)\epsilon_i^2-\sigma^2(x))^2]+o_p(1) \\
&\quad = \left\{\int W^2(s)\,ds\right\} p(x)\sigma^4(x)\mathrm{E}[(\epsilon_1^2-1)^2|X_1=x]+o_p(1). \qquad \textbf{(4.16)}
\end{aligned}$$

Hence, by (4.4) and (4.16),

$$II_1 \xrightarrow{p} p^2(x)\left\{\int s^2W(s)\,ds\right\}^2\left\{\int W^2(s)\,ds\right\} p(x)\sigma^4(x)\mathrm{E}[(\epsilon_1^2-1)^2|X_1=x].$$

Similarly, we can show that

$$II_2 = o_p(1) \quad \text{and} \quad II_3 = o_p(1)$$

by noting $s_{n,1}(x) = O_p(h_1)$ from (4.4). Hence,

$$I_1 \xrightarrow{p} p^2(x)\left\{\int s^2W(s)\,ds\right\}^2\left\{\int W^2(s)\,ds\right\} p(x)\sigma^4(x)\mathrm{E}[(\epsilon_1^2-1)^2|X_1=x]. \qquad \textbf{(4.17)}$$

By the boundedness of W and (4.1), we have

$$\frac{1}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)(m(X_i)-\hat{m}(X_i))^2 = o_p(h_1+h_2). \qquad \textbf{(4.18)}$$

By Condition (C5), (4.4), (4.1), and (4.18), we have

$$\begin{aligned}
I_3 &\le \frac{8}{nh_1}\sum_{i=1}^{n}\Delta_i^2(x)\sigma^2(X_i)\epsilon_i^2(m(X_i)-\hat{m}(X_i))^2 \\
&\quad + \frac{2}{nh_1}\sum_{i=1}^{n}\Delta_i^2(x)(m(X_i)-\hat{m}(X_i))^4 \\
&\le \frac{16}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)s_{n,2}^2(x)\sigma^2(X_i)\epsilon_i^2(m(X_i)-\hat{m}(X_i))^2 \\
&\quad + \frac{16}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)\frac{(x-X_i)^2}{h_1^2}s_{n,1}^2(x)\sigma^2(X_i)\epsilon_i^2(m(X_i)-\hat{m}(X_i))^2 \\
&\quad + \frac{2}{nh_1}\left\{\sum_{i=1}^{n}|\Delta_i(x)|(m(X_i)-\hat{m}(X_i))^2\right\}^2
\end{aligned}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

$$\le 16\left\{\frac{1}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right) s_{n,2}^4(x)\sigma^4(X_i)\epsilon_i^4\right\}^{1/2}$$

$$\times\left\{\frac{1}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)(m(X_i)-\hat{m}(X_i))^2\right\}^{1/2}$$

$$+16\left\{\frac{1}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)\frac{(x-X_i)^4}{h_1^4} s_{n,1}^4(x)\sigma^4(X_i)\epsilon_i^4\right\}^{1/2}\times$$

$$\times\left\{\frac{1}{nh_1}\sum_{i=1}^{n} W^2\left(\frac{x-X_i}{h_1}\right)(m(X_i)-\hat{m}(X_i))^2\right\}^{1/2}$$

$$+\frac{2}{nh_1}\left\{\sum_{i=1}^{n} W\left(\frac{x-X_i}{h_1}\right) s_{n,2}(x)(m(X_i)-\hat{m}(X_i))^2\right.$$

$$\left.+\sum_{i=1}^{n} W\left(\frac{x-X_i}{h_1}\right)\frac{|x-X_i|}{h_1}|s_{n,1}(x)|(m(X_i)-\hat{m}(X_i))^2\right\}^2$$

$$=32\{O_p(1)\}^{1/2}\{o_p(h_1+h_2)\}^{1/2}+\frac{2}{nh_1}\{(nh_1)o_p(h_1+h_2)\}^2$$

$$=o_p(1). \tag{4.19}$$

By (4.17) and (4.19), we have

$$|I_2|\le 2I_1^{1/2}I_3^{1/2}=o_p(1). \tag{4.20}$$

Thus, by (4.17), (4.19), and (4.20),

$$\frac{1}{nh_1}\sum_{i=1}^{n} Z_i^2 \xrightarrow{p} p^2(x)\left\{\int s^2W(s)\,ds\right\}^2\left\{\int W^2(s)\,ds\right\}p(x)\sigma^4(x)$$

$$\times \mathrm{E}[(\epsilon_1^2-1)^2|X_1=x]. \tag{4.21}$$

It follows from (4.13), (4.14), (4.15), and (4.21) that

$$|\lambda(\theta)|=O_p\left((nh_1)^{-1/2}\right). \tag{4.22}$$

Hence, by (4.13) and (4.22),

$$\max_{1\le i\le n}|\lambda(\theta)Z_i|=o_p(1). \tag{4.23}$$

Because of (4.23), expanding (2.1) and using (4.22) and (4.23) result in

$$\lambda(\theta)=\left(\frac{1}{nh_1}\sum_{i=1}^{n} Z_i\right)\Big/\left(\frac{1}{nh_1}\sum_{i=1}^{n} Z_i^2\right)+o_p\left((nh_1)^{-1/2}\right). \tag{4.24}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

By (4.23) we can expand

$$\log(1+\lambda(\theta)Z_i) = \lambda(\theta)Z_i - \frac{1}{2}\lambda^2(\theta)Z_i^2 + \eta_i,$$

where for some finite $B > 0$

$$P\{|\eta_i| \le B|\lambda(\theta)Z_i|^3, 1 \le i \le n\} \to 1 \tag{4.25}$$

as $n \to \infty$. Hence, by (4.24), (4.25), (4.15), and (4.21),

$$\begin{aligned}\ell(\theta) &= 2\sum_{i=1}^{n}\lambda(\theta)Z_i - \sum_{i=1}^{n}\lambda_i^2(\theta)Z_i^2 + 2\sum_{i=1}^{n}\eta_i \\ &= \frac{1}{nh_1}\left(\sum_{i=1}^{n}Z_i\right)^2\left\{\frac{1}{nh_1}\sum_{i=1}^{n}Z_i^2\right\}^{-1} + o_p(1) \\ &\xrightarrow{d} \chi^2(1).\end{aligned}$$

■

Proof of Theorem 2. Put $\tilde{Z}_i = \tilde{\Delta}_i(x)\{\hat{r}_i - \sigma^2(x)\}$. Using the exact arguments in proving (4.13), we have

$$\begin{cases}\max_{1\le i\le n}\left|\Delta_i\left(x-\left(1+2^{-1/2}\right)\delta h_1\right)\{\hat{r}_i-\sigma^2\left(x-\left(1+2^{-1/2}\right)\delta h_1\right)\}\right| = o_p\left(\sqrt{nh_1}\right)\\ \max_{1\le i\le n}\left|\Delta_i\left(x-2^{-1/2}\delta h_1\right)\{\hat{r}_i-\sigma^2\left(x-2^{-1/2}\delta h_1\right)\}\right| = o_p\left(\sqrt{nh_1}\right)\\ \max_{1\le i\le n}\left|\Delta_i\left(x-\left(2^{-1/2}-1\right)\delta h_1\right)\{\hat{r}_i-\sigma^2\left(x-\left(2^{-1/2}-1\right)\delta h_1\right)\}\right| = o_p\left(\sqrt{nh_1}\right)\\ \max_{1\le i\le n}\left|\Delta_i\left(x+\left(2^{-1/2}-1\right)\delta h_1\right)\{\hat{r}_i-\sigma^2\left(x+\left(2^{-1/2}-1\right)\delta h_1\right)\}\right| = o_p\left(\sqrt{nh_1}\right)\\ \max_{1\le i\le n}\left|\Delta_i\left(x+2^{-1/2}\delta h_1\right)\{\hat{r}_i-\sigma^2\left(x+2^{-1/2}\delta h_1\right)\}\right| = o_p\left(\sqrt{nh_1}\right)\\ \max_{1\le i\le n}\left|\Delta_i\left(x+\left(1+2^{-1/2}\right)\delta h_1\right)\{\hat{r}_i-\sigma^2\left(x+\left(1+2^{-1/2}\right)\delta h_1\right)\}\right| = o_p\left(\sqrt{nh_1}\right).\end{cases} \tag{4.26}$$

Hence, it follows from Condition (C4), (4.4), and (4.26) that

$$\max_{1\le i\le n}|\tilde{Z}_i| = o_p\left(\sqrt{nh_1}\right). \tag{4.27}$$

Like the proof of (4.14), we have

$$\left|\tilde{\lambda}(\theta)\right|\left\{\frac{1}{nh_1}\sum_{i=1}^{n}\tilde{Z}_i^2 - \max_{1\le i\le n}\left|\tilde{Z}_i\right|\left|\frac{1}{nh_1}\sum_{i=1}^{n}\tilde{Z}_i\right|\right\} \le \left|\frac{1}{nh_1}\sum_{i=1}^{n}\tilde{Z}_i\right|. \tag{4.28}$$

By (4.4) we have

$$\frac{1}{nh_1}\sum_{i=1}^{n}\Delta_i(x) = p^2(x)\int s^2W(s)\,ds + O_p(h_1). \tag{4.29}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

It follows from (4.2) and (4.29) that

$$\frac{p^{-2}(x)\left\{\int s^2 W(s)\,ds\right\}^{-1}}{\sqrt{nh_1}}\sum_{i=1}^{n}\Delta_i(x)\left(\hat{r}_i-\sigma^2(x)\right)$$
$$=\frac{1}{\sqrt{nh_1}}\sum_{i=1}^{n}W\left(\frac{x-X_i}{h_1}\right)\sigma^2(X_i)(\epsilon_i^2-1)+o_p(1). \tag{4.30}$$

Put $\alpha_{x,j}=x-(2^{-1/2}+1-j)\delta h_1$, $\beta_{x,j}=x-(-2^{-1/2}+1-j)\delta h_1$ for $j=0,1,2$. Denote $\gamma_0(r)=r(r-1)/2$, $\gamma_1(r)=(1-r^2)$, $\gamma_2(r)=r(r+1)/2$, $A_i(x)=W(\frac{x-X_i}{h_1})\sigma^2(X_i)(\epsilon_i^2-1)$, $C^*(a,b)=\int W(s+a)W(s+b)\,ds$, $C(s,t)=\int_{-\infty}^{\infty}W(u-st)W(u+st)\,du$, $\nu_{i,j}=\int_{-\infty}^{\infty}z^i W^j(z)\,dz$, $C(s)=\frac{3}{2}C(0,s)-2C(\frac{1}{2},s)+\frac{1}{2}C(1,s)$, and

$$D(\delta)=\nu_{0,2}-\frac{C(\delta)}{4}$$
$$-\frac{1}{16}\left\{4\left(1+\sqrt{2}\right)C\left(\sqrt{2}-1,\frac{\delta}{2}\right)+\left(3+2\sqrt{2}\right)C\left(2-\sqrt{2},\frac{\delta}{2}\right)\right.$$
$$+2C\left(\sqrt{2},\frac{\delta}{2}\right)+4\left(1-\sqrt{2}\right)C\left(\sqrt{2}+1,\frac{\delta}{2}\right)$$
$$\left.+\left(3-2\sqrt{2}\right)C\left(\sqrt{2}+2,\frac{\delta}{2}\right)\right\}.$$

Then

$$\mathrm{E}\left[\frac{1}{2}\sum_{j=0}^{2}(\gamma_j(2^{-1/2})A_i(\alpha_{x,j})+\gamma_j(-2^{-1/2})A_i(\beta_{x,j}))\right]^2$$
$$=\frac{1}{4}\sum_{j=0}^{2}\mathrm{E}\left[\gamma_j^2(2^{-1/2})A_i^2(\alpha_{x,j})+\gamma_j^2(-2^{-1/2})A_i^2(\beta_{x,j})\right.$$
$$\left.+\,2\gamma_j(2^{-1/2})\gamma_j(-2^{-1/2})A_i(\alpha_{x,j})A_i(\beta_{x,j})\right]$$
$$+\frac{1}{4}\sum_{j_1,j_2=0}^{2}\sum_{j_1\neq j_2}E\{\gamma_{j_1}(2^{-1/2})\gamma_{j_2}(2^{-1/2})A_i(\alpha_{x,j_1})A_i(\alpha_{x,j_2})$$
$$+\gamma_{j_1}(2^{-1/2})\gamma_{j_2}(-2^{-1/2})A_i(\alpha_{x,j_1})A_i(\beta_{x,j_2})$$
$$+\gamma_{j_1}(-2^{-1/2})\gamma_{j_2}(2^{-1/2})A_i(\beta_{x,j_1})A_i(\alpha_{x,j_2})$$
$$+\gamma_{j_1}(-2^{-1/2})\gamma_{j_2}(-2^{-1/2})A_i(\beta_{x,j_1})A_i(\beta_{x,j_2}) \tag{4.31}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

It is straightforward to check that

$$\begin{aligned}
&\mathrm{E}\left[A_i(\alpha_{x,j_1})A_i(\alpha_{x,j_2})\right] \\
&\quad = h_1\sigma^4(x)\mathrm{E}\left[(\epsilon_i^2-1)^2|X_i=x\right] \\
&\qquad \times p(x)C^*((2^{-1/2}+1-j_1)\delta,(2^{-1/2}+1-j_2)\delta)+o(h) \\
&\quad = h_1\sigma^4(x)\mathrm{E}\left[(\epsilon_i^2-1)^2|X_i=x\right]p(x)C^*(-j_1\delta,-j_2\delta)+o(h), \qquad \textbf{(4.32)}
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}\left[A_i(\beta_{x,j_1})A_i(\beta_{x,j_2})\right] \\
&\quad = h_1\sigma^4(x)\mathrm{E}\left[(\epsilon_i^2-1)^2|X_i=x\right] \\
&\qquad \times p(x)C^*((-2^{-1/2}+1-j_1)\delta,(-2^{-1/2}+1-j_2)\delta)+o(h) \\
&\quad = h_1\sigma^4(x)\mathrm{E}\left[(\epsilon_i^2-1)^2|X_i=x\right]p(x)C^*(-j_1\delta,-j_2\delta)+o(h), \qquad \textbf{(4.33)}
\end{aligned}$$

$$\begin{aligned}
&\mathrm{E}\left[A_i(\alpha_{x,j_1})A_i(\beta_{x,j_2})\right] \\
&\quad = h_1\sigma^4(x)E\{(\epsilon_i^2-1)^2|X_i=x\} \\
&\qquad \times p(x)C^*((2^{-1/2}+1-j_1)\delta,(-2^{-1/2}+1-j_2)\delta)+o(h) \\
&\quad = h_1\sigma^4(x)E\{(\epsilon_i^2-1)^2|X_i=x\} \\
&\qquad \times p(x)C^*((2^{-1/2}-j_1)\delta,(-2^{-1/2}-j_2)\delta)+o(h). \qquad \textbf{(4.34)}
\end{aligned}$$

We further have

$$C^*(-j_1\delta,-j_2\delta)=\begin{cases} C(0,\delta)=\nu_{0,2} & \text{if } |j_1-j_2|=0 \\ C(\frac{1}{2},\delta) & \text{if } |j_1-j_2|=1 \\ C(1,\delta) & \text{if } |j_1-j_2|=2, \end{cases} \qquad \textbf{(4.35)}$$

and

$$C^*((2^{-1/2}-j_1)\delta,(-2^{-1/2}-j_2)\delta)=\begin{cases} C(\sqrt{2},\delta/2) & \text{if } j_1-j_2=0 \\ C(\sqrt{2}-1,\delta/2) & \text{if } j_1-j_2=1 \\ C(2-\sqrt{2},\delta/2) & \text{if } j_1-j_2=2 \\ C(\sqrt{2}+1,\delta/2) & \text{if } j_1-j_2=-1 \\ C(\sqrt{2}+2,\delta/2) & \text{if } j_1-j_2=-2. \end{cases} \qquad \textbf{(4.36)}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

Hence, by (4.30)–(4.36),

$$\begin{aligned}
&\frac{p^{-2}(x)\{\int s^2W(s)\,ds\}^{-1}}{\sqrt{nh_1}}\sum_{i=1}^{n}\sum_{j=0}^{2}\frac{1}{2}\Big\{\gamma_j(2^{-1/2})\Delta_i(\alpha_{x,j})(\hat{r}_i-\sigma^2(\alpha_{x,j}))\\
&\qquad\qquad+\gamma_j(-2^{-1/2})\Delta_i(\beta_{x,j})(\hat{r}_i-\sigma^2(\beta_{x,j}))\Big\}\\
&=\frac{1}{\sqrt{nh_1}}\sum_{i=1}^{n}\sum_{j=0}^{2}\frac{1}{2}\Big\{\gamma_j(2^{-1/2})A_i(\alpha_{x,j})+\gamma_j(-2^{-1/2})A_i(\beta_{x,j})\Big\}+o_p(1)\\
&\xrightarrow{d} N\left(0,\left\{v_{0,2}-\frac{C(\delta)}{4}-\frac{D(\delta)}{2}\right\}p^{-1}(x)\sigma^4(x)\mathrm{E}\left[(\epsilon_1^2-1)^2|X_1=x\right]\right).
\end{aligned}\tag{4.37}$$

It follows from (4.29) that

$$\begin{aligned}
&\frac{1}{nh_1}\sum_{j=0}^{2}\sum_{i=1}^{n}\gamma_j\left(2^{-1/2}\right)\Delta_i(\alpha_{x,j})\left(\sigma^2(\alpha_{x,j})-\sigma^2(x)\right)\\
&=\sum_{j=0}^{2}\gamma_j(2^{-1/2})p^2(x)\left\{\int s^2W(s)\,ds\right\}(1+O_p(h))(\alpha_{x,j}-x)\\
&\quad\times\left\{\frac{d}{dx}\sigma^2(x)\right\}\{1+o(1)\}\\
&=\sum_{j=0}^{2}\gamma_j(2^{-1/2})(\alpha_{x,j}-x)(1+O_p(h_1)).
\end{aligned}\tag{4.38}$$

Note that

$$\sum_{j=0}^{2}\gamma_j(2^{-1/2})=1\quad\text{and}\quad\sum_{j=0}^{2}\gamma_j(2^{-1/2})(\alpha_{x,j}-x)=0.\tag{4.39}$$

By (4.38) and (4.39), we have

$$\frac{1}{nh_1}\sum_{j=0}^{2}\sum_{i=1}^{n}\gamma_j(2^{-1/2})\Delta_i(\alpha_{x,j})(\sigma^2(\alpha_{x,j})-\sigma^2(x))=O_p(h_1^2).\tag{4.40}$$

Similarly,

$$\frac{1}{nh_1}\sum_{j=0}^{2}\sum_{i=1}^{n}\gamma_j(-2^{-1/2})\Delta_i(\beta_{x,j})(\sigma^2(\beta_{x,j})-\sigma^2(x))=O_p(h_1^2).\tag{4.41}$$

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

Hence, it follows from (4.37), (4.39)–(4.41), and Condition (C4) that

$$\frac{p^{-2}(x)\{\int s^2 W(s)\,ds\}^{-1}}{\sqrt{nh_1}} \sum_{i=1}^{n} \tilde{Z}_i$$

$$\xrightarrow{d} N\left(0, \left\{\nu_{0,2} - \frac{C(\delta)}{4} - \frac{D(\delta)}{2}\right\} p^{-1}(x)\sigma^4(x)\mathrm{E}[(\epsilon_1^2 - 1)^2 | X_1 = x]\right). \tag{4.42}$$

Using the same arguments in proving (4.21) and (4.42), we can show that

$$\frac{1}{nh_1}\sum_{i=1}^{n} \tilde{Z}_i^2 \xrightarrow{p} p^2(x)\left\{\int s^2 W(s)\,ds\right\}^2 \left\{\nu_{0,2} - \frac{C(\delta)}{4} - \frac{D(\delta)}{2}\right\} p(x)\sigma^4(x)$$

$$\times \mathrm{E}[\mathrm{E}((\epsilon_1^2 - 1)^2 | X_1 = x]. \tag{4.43}$$

The rest of proof is the same as that of Theorem 1 by using (4.27), (4.28), (4.42), and (4.43). ■

REFERENCES

Aït-Sahalia, Y.A. (1996) Nonparametric pricing of interest rate derivative securities. *Econometrica* 64, 527–560.

Andrews, D.W.K. (1997) A conditional Kolmogorov test. *Econometrica* 65, 1097–1028.

Borkovec, M. (2001) Asymptotic behavior of the sample autocovariance and auto correlation function of the AR(1) process with ARCH(1) errors. *Bernoulli* 6, 847–872.

Borkovec, M. & C. Klüppelberg (2001) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. *Annals of Applied Probability* 11, 1220–1241.

Chan, N.H. & L. Peng (2005) Weighted least absolute deviations estimation for an AR(1) process with ARCH(1) errors. *Biometrika* 92, 477–484.

Chen, L., M. Cheng, & L. Peng (2009) Conditional variance estimation in heteroscedastik regression models. *Journal of Statistical Planning and Inference* 139, 236–245.

Chen, S.X. & H. Cui (2007) On the second order properties of empirical likelihood with moment restrictions. *Journal of Econometrics* 141, 492–516.

Chen, S.X. & J. Gao (2007) An adaptive empirical likelihood test for time series models. *Journal of Econometrics* 141, 950–972.

Chen, S.X., W. Härdle, & M. Li (2003) An empirical likelihood goodness-of-fit test for time series. *Journal of Royal Statistical Society B* 66, 63–78.

Chen, S.X. & Y. Qin (2000) Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* 87, 946–953.

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

Donald, S.G., G.W. Imbens, & W.K. Newey (2003) Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics* 117, 55–93.

Fan, J. (2005) A selective overview of nonparametric methods in financial econometrics with discussions. *Statistical Science* 20, 317–337.

Fan, J. & L. Huang (2001) Goodness-of-fit test for parametric regression models . *Journal of the American Statistical Association* 96, 640–652.

Fan, J., J. Jiang, C. Zhang, & Z. Zhou (2003) Time-dependent diffusion models for term structure dynamics. *Statistica Sinica* 13, 965–992.

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press

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

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag.

Fan, J., C. Zhang, & J. Zhang (2001) Generalized likelihood ratio statistics and Wilks phenomenon. *Annals of Statistics* 29, 153–193.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion.* Springer-Verlag.

Härdle, W. & E. Mammen (1993) Comparing nonparametric versus parametric regression fits. *Annals of Statistics* 21, 1926–1947.

Horowitz, J.L. & V.G. Spokoiny (2001) An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica* 69, 599–632.

Imbens, G.W. (2002) Generalized method of moments and empirical likelihood equations. *Journal of Business and Economic Statistics* 20, 493–506.

Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Annals of Statistics* 25, 2084–2102.

Kitamura, Y. (2001) Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 69, 1661–1672.

Kitamura, Y., G. Tripathi, & H. Ahn (2004) Empirical likelihood-based inference in conditional moment restriction models. *Econometrica* 72, 219–255.

Kloeden, P.E. & E. Platen (1999) *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag.

Klüppelberg, C. & S. Pergamenchtchikov (2004) The tail of the stationary distribution of a random coefficient AR(q) model. *Annals of Applied Probability* 14, 971–1005.

Ling, S. (2004) Estimation and testing stationarity for double autoregressive models. *Journal of Royal Statistical Society B* 66, 63–78.

Ling, S. (2007) A double AR(p) model: Structure and estimation. *Statistica Sinica* 17, 161–175.

Newey, W.K. & R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–255.

Owen, A.B. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237–249.

Owen, A.B. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90–120.

Owen, A.B. (2001) *Empirical Likelihood*. Chapman and Hall.

Vasicek, O.A. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.

Wang, L. & I. Van Keilegom (2007) Nonparametric test for the form of parametric regression with time series errors. *Statistica Sinica* 17, 369–386.

Yu, K. & M.C. Jones (2004) Likelihood-based local linear estimation of the conditional variance function. *Journal of American Statistical Association* 99, 139–144.

https://doi.org/10.1017/S0266466610000150 Published online by Cambridge University Press