Econometric Theory, **31**, 2015, 1078–1101.
doi:10.1017/S0266466614000681

PARAMETRIC SPECIFICATION TEST FOR NONLINEAR AUTOREGRESSIVE MODELS

KUN HO KIM
Hanyang University

TING ZHANG
Boston University

WEI BIAO WU
University of Chicago

The paper considers testing parametric assumptions on the conditional mean and variance functions for nonlinear autoregressive models. To this end, we compare the kernel density estimate of the marginal density of the process with a convolution-type density estimate. It is shown that, interestingly, the latter estimate has a parametric ($\sqrt{n}$) rate of convergence, thus substantially improving the classical kernel density estimates whose rates of convergence are much inferior. Our results are confirmed by a simulation study for threshold autoregressive processes and autoregressive conditional heteroskedastic processes.

1. INTRODUCTION

Consider the nonlinear autoregressive model

$$X_i = \mu(X_{i-1}) + \sigma(X_{i-1})\epsilon_i, \quad i \in \mathbb{Z}, \tag{1}$$

where $\mu : \mathbb{R} \to \mathbb{R}$ and $\sigma : \mathbb{R} \to \mathbb{R}$ are unknown conditional mean and variance functions, respectively, and $\epsilon_k, k \in \mathbb{Z}$, are independent and identically distributed (iid) innovations. Assuming certain parametric forms on $\mu(\cdot)$ and $\sigma(\cdot)$, model (1) covers many popular time series models as special cases including the linear autoregressive (AR) process $X_i = aX_{i-1} + \epsilon_i$, the threshold autoregressive (TAR) process $X_i = a\max(X_{i-1}, 0) + b\min(X_{i-1}, 0) + \epsilon_i$ by Tong (1990), the autoregressive conditional heteroscedastic (ARCH) process $X_i = \sqrt{a^2 + b^2 X_{i-1}^2}\epsilon_i$ by Engle (1982), and the exponential autoregressive (EAR) process $X_i = \{a + b\exp(-cX_{i-1})\}X_{i-1} + \epsilon_i$ by Jones (1976) among others. Although the aforementioned parametric models have been widely used in the literature due to their ease of

We thank the Editor, the Co-Editor, and two anonymous referees for their helpful comments and suggestions. Kim gratefully acknowledges funding from Hanyang University (HY-2013-G). Ting is grateful for the support of NSF grant DMS-1461796. Wu is grateful for the support of NSF grants DMS-0906073 and DMS-1106790. Address correspondence to Kun Ho Kim, Department of Economics and Finance, 222 Wangsimni-ro, Seongdong-gu Seoul 133-791, South Korea; e-mail: kunhokim8@gmail.com.

 © Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

interpretation and prediction, they may suffer from misspecification leading to erroneous conclusions. The primary goal of the paper is to develop a rigorous test for parametric assumptions on the conditional mean and variance functions in (1), namely we are interested in testing the null hypothesis

$$H_0 : \{\mu(\cdot), \sigma(\cdot)\} = \{\mu_\theta(\cdot), \sigma_\theta(\cdot)\} \text{ holds for some } \theta \in \Theta, \tag{2}$$

where $\mu_\theta(\cdot)$ and $\sigma_\theta(\cdot)$ have a known parametric form, θ is a finite dimensional parameter vector, and Θ is a parameter space. Throughout the paper we denote by θ_0 the true parameter and assume that the process (X_i) is stationary.

Aït-Sahalia (1996) proposed a parametric specification test by comparing the nonparametric kernel density estimate of the marginal density of X_i with its closed form solution under the null hypothesis. Specifically, suppose we observe $X_1, \ldots, X_n$, then the marginal density of X_i can be estimated nonparametrically by

$$\hat{f}(x) = \frac{1}{nb_n} \sum_{i=1}^{n} K\left(\frac{x - X_i}{b_n}\right), \tag{3}$$

where K is a kernel function and b_n is a bandwidth sequence satisfying $b_n \to 0$ and $nb_n \to \infty$. Nonparametric density estimation has been extensively studied in statistics; see for example, Robinson (1983), Silverman (1986), Scott (1992), and Fan and Yao (2005) among others. We remark that the estimate $\hat{f}(x)$ in (3) does not require specific structural assumptions of the process (X_i). On the other hand, however, if $\mu(\cdot)$ and $\sigma(\cdot)$ in (1) fall into certain parametric forms so that the marginal density of X_i has a closed form solution $f_\theta(\cdot)$, then one can plug in the parametric estimate $\hat{\theta}$ of the unknown parameter θ_0 and obtain a $\sqrt{n}$-consistent density estimate $f_{\hat{\theta}}(\cdot)$. Aït-Sahalia (1996) designed a testing procedure for the null hypothesis (2) by comparing $f_{\hat{\theta}}(\cdot)$ with the kernel density estimate $\hat{f}(\cdot)$; see also Stanton (1997), Pritsker (1998), Chapman and Pearson (2000), Gao and King (2004), and Aït-Sahalia, Fan, and Peng (2009) among others.

The test of Aït-Sahalia (1996) relies critically on the existence of a closed form marginal density under the null hypothesis. Nevertheless, it is not always guaranteed that such a closed form expression exists for the nonlinear autoregressive process (1), especially when the innovations are non-Gaussian. This makes the testing procedure of Aït-Sahalia (1996) not directly applicable, and it seems desirable if we can have an alternative testing procedure to handle the case where the marginal density does not have a closed form solution under the null. Here we shall utilize the structure in (1) and the independence of X_{i-1} and ϵ_i, and propose a density estimate which can play the role of the closed form marginal density as in Aït-Sahalia's (1996) setting. Hence, the test of the null hypothesis (2) can be similarly carried out. To illustrate the idea, consider the special case of (1) with $\sigma(\cdot) \equiv 1$:

$$X_i = m_\theta(X_{i-1}) + \epsilon_i, \tag{4}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

where $\theta \in \mathbb{R}^d$ is a d-dimensional parameter vector, $m_\theta(\cdot)$ is a measurable function, and ϵ_k, $k \in \mathbb{Z}$, are iid innovations with density function $h(\cdot)$. Assume at the outset that the innovation density h is known and $Z_i = m_{\theta_0}(X_i)$ are observed. Since Z_{i-1} and ϵ_i are independent, we can use the following convolution estimate

$$\check{f}(x) = \int_{\mathbb{R}} h(x-y)\hat{g}(y)dy, \quad \text{where } \hat{g}(y) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left(\frac{y - Z_{i-1}}{b_n}\right). \tag{5}$$

Note that $\hat{g}(\cdot)$ is the usual kernel density estimate of $g(\cdot)$, the density of Z_i. Various types of convolution estimates have been discussed in the literature; see for example, Bickel and Ritov (1988), Frees (1994), Saavedra and Cao (2000), Schick and Wefelmeyer (2004, 2007), and Giné and Mason (2007), where it is shown that convolution estimates can have parametric, namely $\sqrt{n}$, rate of convergence. In a more realistic situation, however, both the parameter θ and the innovation density f are *unknown*. Assume that $\hat{\theta}$ is a consistent estimate of θ_0. Let

$$\hat{h}_\theta(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left[\frac{x - \{X_i - m_\theta(X_{i-1})\}}{b_n}\right] \tag{6}$$

$$= \frac{1}{nb_n}\sum_{i=1}^{n} K\left\{\frac{x - \epsilon_i + m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1})}{b_n}\right\}. \tag{7}$$

Then $\hat{h}_{\hat{\theta}}(x)$ provides a natural nonparametric estimate of h based on the estimated residuals $\hat{\epsilon}_i = X_i - m_{\hat{\theta}}(X_{i-1}) = X_i - \hat{Z}_{i-1}$, $1 \le i \le n$. As in (5), we can estimate f by

$$\tilde{f}(x) = \int_R \hat{h}_{\hat{\theta}}(x-y)\hat{g}_{\hat{\theta}}(y)dy, \text{ where } \hat{g}_\theta(y) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left\{\frac{y - m_\theta(X_{i-1})}{b_n}\right\}. \tag{8}$$

Under suitable conditions, we expect that $\tilde{f}$ has a $\sqrt{n}$-convergence rate. Note that the construction of $\tilde{f}$ is based on the independence of $m_\theta(X_{i-1})$ and ϵ_i in (4). Hence we can test the hypothesis $H_0 : \mu(\cdot) = m_\theta(\cdot)$ for some $\theta \in \Theta$ by comparing the kernel density estimate (3) and the convolution estimate (8). The latter plays the role of the closed form marginal density as in Aït-Sahalia's (1996) setting.

The rest of the paper is organized as follows. Section 2 presents asymptotic properties of the convolution estimates. We prove that they achieve $\sqrt{n}$-consistency. In Section 3 we shall apply the convolution estimate and devise a model specification test. Simulation studies are provided in Section 5. Proofs are given in the Appendix.

2. ASYMPTOTICS OF CONVOLUTION ESTIMATES

If not otherwise specified, throughout this section we shall work with the process (4) and consider estimating the marginal density of (X_i). Note that

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

(X_i) can be nonlinear. The case of linear processes is covered in Schick and Wefelmeyer (2007), where a similar convolution-type estimate is proposed. Due to the nonlinearity, the involved derivations are much more complicated and the techniques developed for linear processes are no longer directly useful. Hence we need to resort to other theoretical frameworks and technical tools. It turns out that, by using the dependence concept introduced in Wu (2005), we can show that under proper regularity conditions the estimate $\tilde{f}(x)$ can have a $\sqrt{n}$-convergence rate. A central limit theorem for $\sqrt{n}[\tilde{f}(x) - f(x)]$ can also be obtained. Such $\sqrt{n}$-consistent density estimates can be used for specification test for inferring the underlying data-generating mechanisms for time series whose marginal densities may not have a closed form expression; see Section 3. Section 2.1 deals with the case that the innovation density h in (4) is known, while Section 2.2 is for unknown h.

Now we introduce some notation. For a random variable W, write $W \in \mathcal{L}^p$, $p > 0$, if $\|W\|_p := [\mathbb{E}(|W|^p)]^{1/p} < \infty$, and write $\|W\| = \|W\|_2$. We define the projection operator $\mathcal{P}$ as $\mathcal{P}_i \cdot \equiv \mathbb{E}[\cdot|\mathcal{F}_i] - \mathbb{E}[\cdot|\mathcal{F}_{i-1}]$, where $\mathcal{F}_i = (\epsilon_i, \epsilon_{i-1}, \ldots)$.

2.1. Case 1: Innovation Density Is Known

If h, the density of ϵ_i, is known, based on $X_0, X_1, \ldots, X_n$, we can estimate f by

$$\breve{f}(x) = \int_{\mathbb{R}} h(x-y)\hat{g}_{\hat{\theta}}(y)dy, \tag{9}$$

where $\hat{\theta}$ is an estimate of θ_0 and

$$\hat{g}_\theta(y) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left\{\frac{y - m_\theta(X_{i-1})}{b_n}\right\}. \tag{10}$$

Note that $\hat{g}_{\hat{\theta}}(\cdot)$ estimates the density of $m_{\theta_0}(X_{i-1})$. The following assumptions are needed for the derivation of asymptotic properties of $\breve{f}$.

Assumption 1. Let the kernel function K be bounded, symmetric, with bounded support $[-A, A]$, $K \in \mathcal{C}^1[-A, A]$, $K(\pm A) = 0$, and $\sup_u |K'(u)| < \infty$.

Assumption 2. $\rho := \sup_{x \neq x'} |m_{\theta_0}(x) - m_{\theta_0}(x')|/|x - x'| < 1$, and $\epsilon_i \in \mathcal{L}^p$, $p > 0$.

Assumption 3. $A_0 := \sup_x [h(x) + |h'(x)| + |h''(x)|] < \infty$, and as $|x| \to \infty$, $h(x) = O(|x|^{-\beta})$ for some $\beta > 0$.

Assumption 4. $\dot{m}_\theta(x) = \partial m_\theta(x)/\partial\theta$ exists, $\mathbb{E}[|\dot{m}_{\theta_0}(X_0)|] < \infty$, and $\dot{m}_\theta(X_0)$ is stochastic continuous in the sense that $\mathbb{E}[\sup_{|\theta-\theta_0|\le\delta} |\dot{m}_\theta(X_0) - \dot{m}_{\theta_0}(X_0)|] \to 0$ as $\delta \to 0$.

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Assumption 5. $\hat{\theta}_n$ is an estimate of θ_0 such that

$$\hat{\theta}_n - \theta_0 = n^{-1} \sum_{i=1}^{n} R_i + o_{\mathbb{P}}(n^{-1/2}), \tag{11}$$

where $R_i = R(\epsilon_i, \epsilon_{i-1}, \ldots)$ satisfies the short-range dependence condition

$$\sum_{i=0}^{\infty} \|\mathcal{P}_0 R_i\| < \infty. \tag{12}$$

Comments on these regularity assumptions are as follows. Assumption 1 allows many popular kernels including Parzen and Epanechnikov kernels among others. For many nonlinear time series models, the innovation densities satisfy Assumption 3. For example, it holds for the normal and t-distributions. Assumption 4 is not the weakest possible. However, for the sake of presentational clarity, we decide to use the current form. Assumptions 2 and 5 are discussed in Remarks 1 and 2.

Remark 1. Assumption 2 is a contraction condition and it ensures that the process (X_i) defined by (4) has a stationary and ergodic solution of the form $X_i = G(\mathcal{F}_i)$, where $\mathcal{F}_i = (\epsilon_i, \epsilon_{i-1}, \ldots)$. So we can assume that (X_i) is a stationary and ergodic process, and it is also causal. Moreover, as argued in Wu and Shao (2004), (X_i) satisfies the following geometric moment contraction property: let ϵ_0' and $\epsilon_i, i \in \mathbb{Z}$, be iid, then

$$\|X_i - X_i^*\|_p = O(\rho^i), \text{ where } X_i^* = G(\mathcal{F}_i^*), \mathcal{F}_i^* = (\epsilon_i, \ldots, \epsilon_1, \epsilon_0', \epsilon_{-1}, \ldots). \tag{13}$$

Here X_i^* is a coupled process of X_i with ϵ_0 replaced by an iid copy ϵ_0'.

Remark 2. To obtain a central limit theorem for an estimate $\hat{\theta}_n$, (11) is usually an important intermediate step. In certain situations (11) is called Bahadur representation (Bahadur, 1966). In the context of ARCH and GARCH processes, Hall and Yao (2003) obtained a similar expansion for quasi-maximum likelihood estimates; see equation (5.9) in the latter paper. See also Hall and Heyde (1980), Amemiya (1985), Klimko and Nelson (1978), and Section 5.5 in Tong (1990) for similar expansions for method of moment estimators, least squares estimators, and maximum likelihood estimators of nonlinear time series models.

THEOREM 1. *Assume* $\sqrt{n} b_n^2 \to 0$ *and* $n b_n \to \infty$. *(i) Under Assumptions 1–5, we have*

$$\sqrt{n}[\breve{f}(x) - f(x)] \Rightarrow N[0, \|D_0(x)\|^2],$$

where, letting $c_0(x) = \mathbb{E}[h'(x - m_{\theta_0}(X_0))\dot{m}_{\theta_0}(X_0)]$,

$$D_0(x) = \sum_{i=0}^{\infty} \mathcal{P}_0[h(x - m_{\theta_0}(X_i)) - c_0(x) R_i].$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

(ii) If additionally $\sup_x |h'''(x)| < \infty$*, then there exists a mean zero Gaussian process* $\{W(x)\}$ *with covariance function* $\operatorname{cov}[W(x), W(x')] = \mathbb{E}[D_0(x)D_0(x')]$ *such that, for any compact interval* $\mathcal{I}$*,*

$$\left\{\sqrt{n}[\check{f}(x) - f(x)], x \in \mathcal{I}\right\} \Rightarrow \{W(x), x \in \mathcal{I}\}.$$

Remark 3. For the classical kernel density estimate $\hat{f}(\cdot)$, it is well known that, at different points $x_1, \ldots, x_k$, $\sqrt{nb_n}[\hat{f}(x_i) - \mathbb{E}\hat{f}(x_i)]$, $1 \le i \le n$, are asymptotically independent. Namely, the limiting distribution of $\{\sqrt{nb_n}[\hat{f}(x) - \mathbb{E}\hat{f}(x)], x \in \mathbb{R}\}$ is a white noise process. Besides the faster $\sqrt{n}$-convergence, the convolution type estimate $\check{f}(x)$ has another appealing property that the limiting process $W(x)$ is differentiable since $D_0'(x)$ exists. Thanks to the smoothness, the convolution type estimate is visually attractive.

Remark 4. If θ_0 is also known, then we can use the estimate $\check{f}^{\circ}(x) = \int_{\mathbb{R}} h(x-y)\hat{g}_{\theta_0}(y)dy$, which is obtained by replacing $\hat{\theta}$ in (9) by the true value θ_0. Following the argument in the proof of Theorem 1, we have the central limit theorem

$$\sqrt{n}[\check{f}^{\circ}(x) - f(x)] \Rightarrow N[0, \|D_0^{\circ}(x)\|^2],$$

where $D_0^{\circ}(x) = \sum_{i=0}^{\infty} \mathcal{P}_0 h\{x - m_{\theta_0}(X_i)\}$. Interestingly, it could happen that $\|D_0^{\circ}(x)\| > \|D_0(x)\|$, indicating that the asymptotic variance of $\check{f}$ can be reduced if we use the estimated parameters instead of the true ones.

2.2. Case 2: Innovation Density Is Unknown

Theorem 1 requires that h is known. It is considerably more challenging to establish a $\sqrt{n}$-norming central limit theorem for $\tilde{f}(x)$ defined in (8) for h with an *unknown* form. In this case h itself needs to be estimated nonparametrically. Extra conditions are needed so that we can have the $\sqrt{n}$-consistency for $\tilde{f}(x)$.

Assumption 6. (i) $Z_i = m_{\theta_0}(X_i) \in \mathcal{L}^p$, $p > 1$; (ii) Z_i has density $g \in \mathcal{C}^2(\mathbb{R})$ satisfying $\sup_z |g''(z)| < \infty$ and $\int_{\mathbb{R}} |g''(z)|dz < \infty$; and (iii) let $g(z|\mathcal{F}_i)$ be the conditional density of Z_{i+1} at z given $\mathcal{F}_i$. Assume that $\sum_{i=0}^{\infty} \int_{\mathbb{R}} \|\mathcal{P}_0 g(z|\mathcal{F}_i)\| dz < \infty$.

In Assumption 6(iii), following Wu (2005), the quantity $\int_{\mathbb{R}} \|\mathcal{P}_0 g(z|\mathcal{F}_i)\|dz$ measures the contribution of ϵ_0 in predicting Z_{i+1}. Hence Assumption 6(iii) suggests that the overall contribution of ϵ_0 in predicting future values $(Z_{i+1})_{i \ge 0}$ is finite.

THEOREM 2. *Let the bandwidth sequence* b_n *satisfy* $(\log n)^2 = o(nb_n^3)$ *and* $b_n = o(n^{-1/4})$*. (i) Under Assumptions 1–6, we have*

$$\sqrt{n}[\tilde{f}(x) - f(x)] \Rightarrow N[0, \|D_0^{\star}(x)\|^2], \tag{14}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

where

$$D_0^\star(x) = \sum_{i=0}^{\infty} \mathcal{P}_0\{h(x - m_{\theta_0}(X_i)) + g(x - \epsilon_i) + \gamma_0(x) R_i\}$$

and $\gamma_0(x) = -\mathrm{cov}[h'(x - m_{\theta_0}(X_i)), \dot{m}_{\theta_0}(X_i)]$. *(ii) If additionally* $\sup_x(|g'''(x)| + |h'''(x)|) < \infty$, *then for any compact interval* $\mathcal{I}$ *we have the functional convergence*

$$\{\sqrt{n}[\tilde{f}(x) - f(x)], x \in \mathcal{I}\} \Rightarrow \left\{W^\star(x), x \in \mathcal{I}\right\},$$

where $W^\star$ *is a mean zero Gaussian process with covariance function* $\mathbb{E}[W^\star(x)W^\star(x')] = \mathbb{E}[D_0^\star(x)D_0^\star(x')]$.

3. MODEL SPECIFICATION TEST

In this section we shall apply the convolution-type estimate and devise a test for the null hypothesis (2). Section 3.1 deals with the relatively easy case in which $\sigma(\cdot)$ is a constant, while the case for nonconstant variance functions is treated in Section 3.2. Aït-Sahalia (1996) used the $\mathcal{L}^2$ distance of $\hat{f}(x)$ in (3) and $f_{\hat{\theta}}(\cdot)$ when a closed form marginal density is obtainable. Here we shall use the normalized $\mathcal{L}^\infty$ or maximal distance (cf. (17) and (23)). Theorems 3 and 4 suggest that our test statistics are asymptotically pivotal in the sense that the limiting distribution does not depend on the unknown densities. This interesting property allows us to design a simulation-based procedure which can have a very good finite-sample performance; see Section 5.2 for a more detailed description. On the other hand, Pritsker (1998) remarked that Aït-Sahalia's (1996) $\mathcal{L}^2$ based test has a very slow convergence rate and thousands of years of continuously sampled data are needed. In comparison, with only about 500 data points, our testing procedure shows very accurate size; see Section 5.2.

For our testing procedures, a distinguished feature and important advantage is that we allow that the innovation density can be unknown. In the study of model specification test, it is usually assumed that the density function under the null can be computed with an arbitrary accuracy; see for example, Aït-Sahalia (1999), Chen, Gao, and Tang (2008), Hong and Li (2005), and Bhardwaj, Corradi, and Swanson (2008) for tests based on transition densities. Bai (2003) and Corradi and Swanson (2006) considered specification tests for cases in which the innovation density is known, and Zhao (2011) considered the problem of nonparametric model validation for hidden Markov models. Nevertheless, none of the techniques in the aforementioned papers can be directly applicable here since we allow unknown innovation densities.

3.1. Constant Volatility Functions

If $\sigma(\cdot)$ is a constant, without loss of generality we assume that it is 1 and hence we have the time series model

$$X_i = \mu(X_{i-1}) + \epsilon_i, \tag{15}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

where μ is an unknown function and ϵ_k, $k \in \mathbb{Z}$, are iid innovations. We want to test whether the process (15) is of form (4). In other words, we test the hypothesis

$$H_0 : \mu(\cdot) = m_\theta(\cdot) \text{ holds for some } \theta \in \Theta, \tag{16}$$

where Θ is a parametric space. Recall (3) for $\hat{f}(x)$ and (8) for $\tilde{f}(x)$. For an interval $\mathcal{I} = [l, u]$, where $l < u$ are fixed, let

$$\Xi_n = \sup_{l \le x \le u} \frac{\sqrt{nb_n}|\hat{f}(x) - \tilde{f}(x)|}{\sqrt{\tilde{f}(x)\int K^2(u)du}}. \tag{17}$$

THEOREM 3. *Suppose H_0 is true. Then under conditions of Theorem 2, we have the Gumbel convergence that for every $z \in \mathbb{R}$,*

$$\mathbb{P}((2\log \bar{b}_n^{-1})^{1/2}\Xi_n - 2\log \bar{b}_n^{-1} - C_K \le z) \to e^{-2e^{-z}}, \tag{18}$$

where $C_K = \frac{1}{2}\log\{\int_{-A}^{A}[K'(u)]^2du / \int_{-A}^{A} K^2(u)du\} - \log(2\pi)$ and $\bar{b}_n = b_n/(u-l)$.

The convergence in (18) can be quite slow. However, since Ξ_n is asymptotically pivotal, its distribution can be approximated by extensive simulations of Ξ_n; see Section 5.2.

Remark 5. A careful check of the proofs of Theorems 3 and 4 suggests that Assumption 5 on $\hat{\theta}_n$ can actually be replaced by the weaker one: $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_{\mathbb{P}}(1)$. We emphasize that in many situations one obtains asymptotic normality for $\hat{\theta}_n$ via the representation (11) and hence the condition $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_{\mathbb{P}}(1)$ is necessarily satisfied.

3.2. Nonconstant Volatility Functions

If $\sigma(\cdot)$ in model (1) is a nonconstant function, then the two components $\mu(X_{i-1})$ and $\sigma(X_{i-1})\epsilon_i$ are no longer independent. Note that the conditional density of $X_i = \mu(X_{i-1}) + \sigma(X_{i-1})\epsilon_i$ given $\mathcal{F}_{i-1}$ is $h[\{\cdot - \mu(X_{i-1})\}/\sigma(X_{i-1})]/\sigma(X_{i-1})$. So we can estimate f by

$$f^{\sharp}_{\hat{\theta}}(y) = \int_{\mathbb{R}} \hat{h}[\{y - \mu_{\hat{\theta}}(x)\}/\sigma_{\hat{\theta}}(x)]\hat{f}(x)/\sigma_{\hat{\theta}}(x)dx,$$

where $\hat{h}$ is an estimate of h and $\hat{f}$ is given in (3). However, it turns out that it is quite tedious and intractable to derive an asymptotic theory for $f^{\sharp}_{\hat{\theta}}$. Hence we consider the modified version of estimating the density of

$$Y_i := \frac{X_i}{\sigma(X_{i-1})} = \frac{\mu(X_{i-1})}{\sigma(X_{i-1})} + \epsilon_i = m(X_{i-1}) + \epsilon_i. \tag{19}$$

Here $m(x) = \mu(x)/\sigma(x)$. In this subsection we let $f^{\dagger}$ be the density of $Y_i = X_i/\sigma(X_{i-1})$. Under the null hypothesis (2), let θ_0 be the true parameter and let $\hat{\theta}$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

be a $\sqrt{n}$-consistent estimate. Note that $f^{\dagger}_{\theta_0}$, the density of $X_i/\sigma_{\theta_0}(X_{i-1})$, can be estimated by $\hat{f}^{\dagger}_{\hat{\theta}}(\cdot)$, where

$$\hat{f}^{\dagger}_{\theta}(x) = \frac{1}{nb_n}\sum_{i=1}^{n} K\left\{\frac{x - X_i/\sigma_{\theta}(X_{i-1})}{b_n}\right\}. \tag{20}$$

Let $m_\theta(x) = \mu_\theta(x)/\sigma_\theta(x)$. Denote by $\hat{g}_\theta(x)$ (cf. (10)) the kernel density estimate for the density of $m_\theta(X_{i-1}) = \mu_\theta(X_{i-1})/\sigma_\theta(X_{i-1})$. Similarly as (6), let

$$\hat{h}_\theta(x) = \frac{1}{nb_n}\sum_{i=1}^{n} W_i(x,\theta), \text{ where } W_i(x,\theta) = K\left[\frac{x - \{X_i - \mu_\theta(X_{i-1})\}/\sigma_\theta(X_{i-1})}{b_n}\right]. \tag{21}$$

Then $\hat{h}_{\hat{\theta}}(\cdot)$ is an estimate of h. Since the density for Y_i is the convolution of the densities of $m(X_{i-1})$ and ϵ_i, we can also estimate $f^{\dagger}$ by

$$\tilde{f}^{\dagger}_{\hat{\theta}}(x) = \int_{\mathbb{R}} \hat{h}_{\hat{\theta}}(x-y)\hat{g}_{\hat{\theta}}(y)dy. \tag{22}$$

As in Section 3.1, we can test hypothesis (2) by comparing $\hat{f}^{\dagger}$ and $\tilde{f}^{\dagger}$.

For the process (1) to have a stationary ergodic solution, we need to replace Assumption 2 by the following (Wu and Shao, 2004) Assumption 7, which implies (13):

Assumption 7. Both $\mu_{\theta_0}(\cdot), \sigma_{\theta_0}(\cdot) \in \mathcal{C}(\mathbb{R})$, $\epsilon_i \in \mathcal{L}^p$, $p > 0$, and $\rho := \sup_{x\neq x'} \|[\mu_{\theta_0}(x) + \epsilon_0\sigma_{\theta_0}(x)] - [\mu_{\theta_0}(x') + \epsilon_0\sigma_{\theta_0}(x')]\|_p/|x - x'| < 1$.

Assumption 8. All of $\mu_\theta(\cdot)$, $\sigma_\theta(\cdot)$, and $m_\theta(\cdot) = \mu_\theta(\cdot)/\sigma_\theta(\cdot)$ satisfy Assumption 4, and there exists $\delta_0 > 0$ such that $\inf_x \inf_{|\theta-\theta_0|\leq\delta} \sigma_\theta(x) > 0$, and $\mathbb{E}\{[\dot{\sigma}_*(\delta_0, X_i)/\sigma_{\theta_0}(X_i)]^2\} < \infty$, where the local maximal function $\dot{\sigma}_*(\delta, x) = \sup_{|\theta-\theta_0|\leq\delta} |\dot{\sigma}_\theta(x)|$.

THEOREM 4. *Assume* $(\log n)^2 = o(nb_n^3)$ *and* $b_n = o(n^{-1/4})$. *Under Assumptions 1, 3, and 5–8, we have (18) with* Ξ_n *therein replaced by*

$$\Xi^{\dagger}_n = \sup_{l\leq x\leq u} \frac{\sqrt{nb_n}|\hat{f}^{\dagger}_{\hat{\theta}}(x) - \tilde{f}^{\dagger}_{\hat{\theta}}(x)|}{\sqrt{\tilde{f}^{\dagger}_{\hat{\theta}}(x)\int K^2(u)du}}. \tag{23}$$

Example 1 Consider the AR(1)-ARCH(1) model

$$X_t = \theta_0 X_{t-1} + \sqrt{\theta_1^2 + \theta_2^2 X_{t-1}^2}\epsilon_t, \tag{24}$$

where $\theta = (\theta_0, \theta_1, \theta_2)$ is the unknown parameter vector and ϵ_t are iid with density h satisfying Assumption 3. Such models were discussed in Borkovec (2000) and

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Weiss (1986) among others. Assume $\theta_1 > 0$ and $\sup_{-1\le u\le 1} \|\theta_0 + \theta_2 u\epsilon_t\|_p < 1$ for some $p > 0$. Then Assumption 7 holds. Elementary calculations show that Assumption 8 holds as well. By Ling (2004, 2007) and Chan and Peng (2005), both the QMLE and the least absolute deviation estimate satisfy the central limit theorem. Hence our Theorem 4 is applicable. A simple variant of (24) is the threshold AR-ARCH model with $\theta_0 X_{t-1}$ therein replaced by $\theta_0|X_{t-1}|$; see Cline and Pu (2004) for stability properties of such processes. For such processes, the regularity conditions can be similarly verified.

4. EXTENSION TO CONTINUOUS-TIME DIFFUSION MODELS

In econometrics, the nonlinear autoregressive model (1) can be viewed as the discrete-time version of the diffusion model

$$dX_t = \mu(X_t)dt + \sigma(X_t)dIB(t), \quad t \ge 0, \tag{25}$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are the drift and diffusion functions respectively, and $IB(\cdot)$ is the standard Brownian motion. For the continuous-time model (25), various parametric forms on $\mu(\cdot)$ and $\sigma(\cdot)$ have been proposed in the literature to explain the underlying data-generating mechanism. For example, Black and Scholes (1973) assumed that $\{\mu(x), \sigma(x)\} = (ax, bx)$, while Vasicek (1977) suggested using $\{\mu(x), \sigma(x)\} = (a + bx, c)$; see also Cox, Ingersoll, and Ross (1985), Courtadon (1982), Chan, Karolyi, Longstaff, and Sanders (1992), Constantinides (1992), Aït-Sahalia (1996), Duffie and Kan (1996), and Zhao (2008) for other parametric specifications. Bandi and Phillips (2003) considered nonparametric estimation of the drift and diffusion functions in (25) for stationary and nonstationary recurrent processes, while a parametric estimation method is given by Bandi and Phillips (2007); see also the survey paper by Bandi and Phillips (2009). We shall here focus on stationary processes and consider a rigorous specification test.

For the continuous-time diffusion model (25), since $IB(\cdot)$ is a standard Brownian motion, by the forward Kolmogorov equation, the marginal density of (25) has a closed form expression

$$f(x) = \frac{C}{\sigma^2(x)} \exp\left\{\int_{x_0}^{x} \frac{2\mu(u)}{\sigma^2(u)} du\right\}, \tag{26}$$

where x_0 is a point in the state space of X_t and $C = C(\mu, \sigma, x_0)$ is a normalizing constant such that $\int f(x)dx = 1$. Under the null hypothesis (2), assume that based on the sample $X_1, \ldots, X_n$ one can have a $\sqrt{n}$-consistent estimate $\hat{\theta}$ for the unknown parameter θ_0. Then in (26) we plug in $\mu(\cdot)$ and $\sigma(\cdot)$ by $\mu_{\hat{\theta}}(\cdot)$ and $\sigma_{\hat{\theta}}(\cdot)$, respectively, then the resulting density, denoted by $f_{\hat{\theta}}(\cdot)$, is also $\sqrt{n}$-consistent. Therefore, when forming the test statistic (17), we can simply replace the convolution estimate $\tilde{f}(\cdot)$ given in (8) by $f_{\hat{\theta}}(\cdot)$. A careful check of the proof reveals that Theorem 3 will continue to hold which can be used to test parametric assumptions for continuous-time diffusion models (25). Note that the current approach is $\mathcal{L}^\infty$-based, while the one by Aït-Sahalia (1996) is $\mathcal{L}^2$-based.

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

5. A SIMULATION STUDY

In Section 5.1 we shall study the numerical performance of the traditional kernel and convolution-type density estimators. A simulation study for model specification test is given in Section 5.2.

5.1. Comparison of Estimators

Recall (3) for the kernel density estimator $\hat{f}$, (5) for the convolution estimator $\check{f}$ with known θ, (9) for the feasible convolution estimator $\check{f}$ with θ estimated, and (8) for the nonparametric convolution estimator $\tilde{f}$ with both f and θ estimated. The superior performance of the convolution estimators in relation to the classical kernel estimator in this section can be explained by the asymptotic results that we showed earlier.

We consider the TAR processes (Tong, 1990):

$$X_i = \theta|X_{i-1}| + \sqrt{1-\theta^2}\epsilon_i, \quad i = 1, \ldots, n, \tag{27}$$

where the parameter $\theta \in (-1, 1)$ and $\epsilon_i, i \in \mathbb{Z}$, are iid random variables. If ϵ_i are standard normal, then interestingly the marginal density of X_i has a closed form with $f(u) = 2\phi(u)\Phi(\delta u)$, where $\delta = \theta/\sqrt{1-\theta^2}$ and $\phi = \Phi'$ is the standard normal density function; see Andel, Netuka, and Svara (1984).

Throughout this section, we use the Epanechnikov kernel $K(u) = 0.75(1-u^2)$, $u \in [-1, 1]$. Choose $\theta = 0.6$ and $n = 500$. For an estimator $\hat{f}(\cdot)$, our criterion function is the mean integrated squared errors (MISE) defined as $\mathbb{E}\int_{\mathbb{R}} |\hat{f}(u) - f(u)|^2 du$. In our simulation we approximate it by averaging 1,000 realizations of the integrated squared errors $\int_{\mathbb{R}} |\hat{f}(u) - f(u)|^2 du$. The parameter θ in (27) is estimated by the Least Squares Estimate (LSE) $\hat{\theta}_n = \sum_{i=1}^n X_i|X_{i-1}| / \sum_{i=1}^n X_{i-1}^2$. Two types of innovations are considered: $\epsilon_i \sim N(0, 1)$ and $\epsilon_i \sim t_5$. In the latter case, since the true marginal density does not have a closed form, we approximate it by the kernel estimator $\hat{f}$ with $n = 10^5$, and expect it to be very close to the true density. The results are summarized in Table 1. We also provide in Figure 1 plots of convolution and kernel estimators along with true densities. Observe that convolution estimates with known error densities ($\check{f}$ and $\check{f}$) are uniformly better than traditional kernel estimates $\hat{f}$. If the marginal density of the innovation is unknown, convolution estimators also outperform traditional kernel estimators if the bandwidth $b \leq 0.25$ is used. Given the conditions in Theorem 2, a smaller bandwidth is preferred for the convolution density estimator due to the necessity of reducing the bias term. In practice, we recommend using $b_n = cn^{-0.3}$, where $c > 0$ is a constant, which meets the conditions $(\log n)^2 = o(nb_n^3)$ and $b_n = o(n^{-1/4})$ in Theorem 2. Note that the optimal bandwidth for the traditional kernel density estimate is in the form of $b_n = cn^{-0.2}$. As a rule of thumb, we suggest using the automatic bandwidth selector for the traditional kernel density estimate and denote the selected bandwidth as $\hat{b}_n$, then we use $\tilde{b}_n = (n^{-0.3}/n^{-0.2})\hat{b}_n$ for the convolution density estimate.

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

TABLE 1. Simulated MISE ($\times 1000$) for various estimators with different bandwidths b

	Gaussian				t_5			
b	$\hat{f}_X$	$\check{f}_X$	$\check{f}_X$	$\tilde{f}_X$	$\hat{f}_X$	$\check{f}_X$	$\check{f}_X$	$\tilde{f}_X$
0.05	9.70	0.12	0.48	1.58	9.92	0.23	0.53	1.51
0.10	5.00	0.12	0.46	1.36	5.16	0.21	0.50	1.26
0.15	3.34	0.16	0.51	1.34	3.46	0.19	0.49	1.11
0.20	2.58	0.26	0.60	1.59	2.51	0.14	0.43	1.09
0.25	2.26	0.47	0.80	2.21	2.02	0.17	0.45	1.21
0.30	2.29	0.85	1.17	3.37	1.83	0.32	0.58	1.85
0.35	2.62	1.43	1.73	5.12	1.80	0.59	0.83	2.71
0.40	3.22	2.28	2.57	7.52	1.93	0.90	1.15	3.94

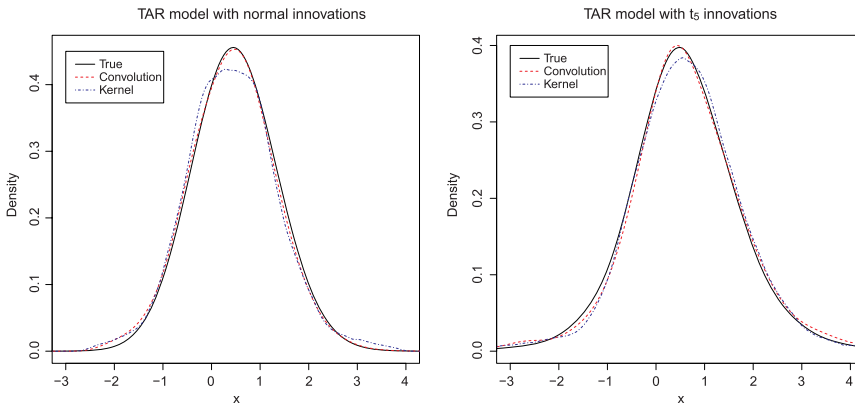


FIGURE 1. Density estimates for the TAR process (27) with both normal and t_5 innovations. In both graphs, solid (respectively dashed or dotted dashed) curves stand for true densities (respectively convolution estimates or the classical kernel density estimators). The true density in the right hand panel is approximated by the kernel estimator with $n = 10^5$.

5.2. Model Specification Tests

We now consider the hypothesis testing problem discussed in Section 3. Consider model (27) with $n = 500$, $\theta = 0.6$, and $(l, u) = (-0.5, 1.5)$, and ϵ_i can be standard normal or t_5 innovations. We want to test the hypothesis $H_0 : \mu(x) = \theta|x|$ for some $\theta \in \mathbb{R}$. Note that the Gumbel convergence in Theorem 3 can be quite slow. We shall here adopt a simulation-based procedure that can significantly improve the finite sample performance. Let $\epsilon_i^\star$, $i \in \mathbb{Z}$, be iid random variables with density h, we form $X_i^\star = m_\theta(X_{i-1}^\star) + \epsilon_i^\star$ and, as in (17), compute the corresponding

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

test statistic $\Xi_n^\star$. By Theorem 3, Ξ_n and $\Xi_n^\star$ have the same asymptotic distribution. Thus, the cutoff value, $(1-\alpha)$-th quantile of Ξ_n, can be approximated by the sample $(1-\alpha)$-th quantile of many simulated $\Xi_n^\star$. Note that $f(x)$ can be estimated by $\hat{f}_n(x,\hat{\theta})$. Let $\alpha = 10\%$ or 5%. For each run, we simulate 5,000 realizations and calculate their corresponding test statistics. For each realization, 5,000 $\Xi_n^\star$'s are generated and we compute their $(1-\alpha)$-th quantile, denoted by $\hat{q}_{1-\alpha}$. We reject the null hypothesis if $\Xi_n > \hat{q}_{1-\alpha}$. The results are summarized in Table 2. We can see that the empirical acceptance percentages are quite close to their nominal levels 90% or 95%. Pritsker (1998) commented that Aït-Sahalia's (1996) test procedure does not perform well unless one has an astronomically large sample. Table 2 suggests that our simulation-based procedure has a very satisfactory finite sample performance.

As another numerical example, we consider the TAR-ARCH (Threshold AR with ARCH errors) model

$$X_i = \theta_0|X_{i-1}| + \sigma_\theta(X_{i-1})\epsilon_i, \quad i = 1,\ldots,n, \tag{28}$$

where $\sigma_\theta(x) = (\theta_1 + \theta_2 x^2)^{1/2}$. Note that model (27) has a constant volatility function. Choose $n = 500$, $\theta = (\theta_0,\theta_1,\theta_2) = (1/2, 1/4, 1/4)$, and $(l,u) = (-1, 1.5)$. Also two types of innovations are considered: $\epsilon_i \sim N(0,1)$ and $\epsilon_i \sim (3/5)^{1/2} t_5$. For both cases $\text{var}(\epsilon_i) = 1$. The parameter vector θ can be estimated by the least squares method. Following the procedure above, we report the empirical levels in Table 2. We can see that they are also quite close to their nominal levels for a wide range of bandwidths.

5.3. Application to Continuous-Time Diffusion Models

As discussed in Section 4, the proposed test can be extended to continuous-time diffusion models, in which case the convolution density estimate is replaced by

TABLE 2. Empirical acceptance percentages with different bandwidths b and error distributions for model (27) and (28)

	TAR				TAR-ARCH			
	Gaussian		t_5		Gaussian		t_5	
b	90%	95%	90%	95%	90%	95%	90%	95%
0.05	89.2	94.8	90.0	95.2	90.8	95.3	90.4	95.1
0.10	89.5	94.6	90.6	95.5	90.9	95.7	90.7	95.6
0.15	90.5	94.9	90.3	95.2	90.3	94.9	91.0	95.9
0.20	89.5	94.7	87.8	93.6	89.8	94.7	89.4	94.7
0.25	88.5	94.0	91.3	95.7	90.3	95.5	89.9	94.2
0.30	88.6	93.9	90.2	95.0	90.0	94.8	90.0	94.6
0.35	86.5	93.0	90.4	95.2	89.1	94.2	89.2	94.1
0.40	90.2	95.3	88.1	93.9	90.2	95.2	90.5	95.2

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

TABLE 3. Empirical acceptance percentages for (i) the proposed simulation-based testing procedure and (ii) the test by Aït-Sahalia (1996) with different choices of bandwidth b and sampling frequency for continuous-time diffusion models

	Daily sampling				*Weekly sampling*				*Monthly sampling*			
	(i)		(ii)		(i)		(ii)		(i)		(ii)	
b	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
0.05	90.6	94.7	100.0	100.0	89.8	95.4	100.0	100.0	90.3	95.0	100.0	100.0
0.10	90.8	95.5	99.9	100.0	90.6	95.3	100.0	100.0	90.2	94.7	100.0	100.0
0.15	90.3	95.2	99.9	100.0	89.8	95.5	100.0	100.0	90.8	95.1	100.0	100.0
0.20	90.4	95.2	100.0	100.0	90.5	95.0	100.0	100.0	90.2	95.0	100.0	100.0
0.25	90.5	95.5	99.9	100.0	90.2	95.1	100.0	100.0	90.3	95.4	100.0	100.0
0.30	89.7	95.0	99.9	100.0	89.0	94.2	100.0	100.0	89.3	95.1	100.0	100.0
0.35	90.5	95.6	100.0	100.0	89.6	94.4	99.9	99.9	89.2	94.6	99.9	100.0
0.40	87.8	94.7	99.7	99.9	90.1	95.5	99.9	100.0	90.3	95.5	99.9	100.0

the parametric density estimate. Aït-Sahalia (1996) considered an $\mathcal{L}^2$-based specification test for diffusion processes, while we shall here apply the proposed $\mathcal{L}^\infty$-based test. For the diffusion process (25), we consider testing the null hypothesis $H_0: \mu(x) = 0.5(\gamma - x)$ and $\sigma(x) = \sigma$ for some $\gamma \in \mathbb{R}$ and $\sigma > 0$. Given a discrete sample X_i, $i = 1, \ldots, n$, the marginal distribution under the null hypothesis is $N(\gamma, \sigma^2)$ and thus the parameters can be naturally estimated by $\hat{\gamma}_n = \bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ and $\hat{\sigma}_n^2 = (n-1)^{-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$. To improve the finite sample performance, we propose a simulation-based testing procedure similar to the one discussed in Section 5.2 for nonlinear autoregressive models. In particular, the cut-off values are obtained by simulating diffusion processes using the specification under the null with parameter estimates $(\hat{\gamma}, \hat{\sigma}^2)$ plugged in. The results are summarized in Table 3 for different choices of bandwidth and sampling frequency. It can be seen from Table 3 that the proposed simulation-based testing procedure performs reasonably well as the empirical acceptance probabilities are quite close to their nominal levels (90% and 95%). In contrast, the test by Aït-Sahalia (1996) does not have a good finite sample performance due to its slow convergence rate; see also the discussion in Section 3.

REFERENCES

Aït-Sahalia, Y. (1996) Testing continuous-time models of the spot interest rate. *Review of Financial Studies* 9, 385–426.

Aït-Sahalia, Y. (1999) Transition densities for interest rate and other nonlinear diffusions. *Journal of Finance* 54, 1361–1395.

Aït-Sahalia, Y., J. Fan, & H. Peng (2009) Nonparametric transition-based tests for jump-diffusions. *Journal of the American Statistical Association* 104, 1102–1116.

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Amemiya, T. (1985) *Advanced Econometrics*. Harvard University Press.
Andel, J., I. Netuka, & K. Svara (1984) On threshold autoregressive processes. *Kibernetika* 20, 89–106.
Bahadur, R.R. (1966) A note on quantiles in large samples. *Annals of Mathematical Statistics* 37, 577–580.
Bai, J. (2003) Testing parametric conditional distribution of dynamic models. *Review of Economics and Statistics* 85, 531–549.
Bandi, F.M. & P.C.B. Phillips (2003) Fully nonparametric estimation of scalar diffusion models. *Econometrica* 71, 241–283.
Bandi, F.M. & P.C.B. Phillips (2007) A simple approach to the parametric estimation of potentially nonstationary diffusions. *Journal of Econometrics* 137, 354–395.
Bandi, F.M. & P.C.B. Phillips (2009) Nonstationary continuous-time processes. In Y. Aït-Sahalia & L.P. Hansen (eds.), *Handbook of Financial Econometrics*, pp. 139–202. North Holland.
Bhardwaj, G., V. Corradi, & N.R. Swanson (2008) A simulation-based specification test for diffusion processes. *Journal of Business and Economic Statistics* 26, 176–193.
Bickel, P. & Y. Ritov (1988) Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhya Series A* 50, 381–393.
Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.
Black, F. & M. Scholes (1973) The pricing of options and corporate liabilities. *Journal of Political Economy* 81, 637–654.
Borkovec, M. (2000) Extremal behavior of the autoregressive process with ARCH(1) errors. *Stochastic Processes and their Applications* 85, 189–207.
Chan, K.C., A.G. Karolyi, F.A. Longstaff, & A.B. Sanders (1992) An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance* 47, 1209–1227.
Chan, N.H. & L. Peng (2005) Weighted least absolute deviation estimation for an AR(1) process with ARCH(1) errors. *Biometrika* 92, 477–484.
Chapman, D.A. & N.D. Pearson (2000) Is the short rate drift actually nonlinear? *Journal of Finance* 55, 355–388.
Chen, S.X., J.T. Gao, & C.Y. Tang (2008) A test for model specification of diffusion processes. *Annals of Statistics* 36, 167–198.
Cline, D.B.H. & H.H. Pu (2004) Stability and the Lyapounov exponent of threshold AR-ARCH models. *Annals of Applied Probability* 14, 1920–1949.
Constantinides, G.M. (1992) A theory of the nominal term structure of interest rates. *Review of Financial Studies* 5, 531–552.
Corradi, V. & N.R. Swanson (2006) Bootstrap conditional distribution tests in the presence of dynamic misspecification. *Journal of Econometrics* 133, 779–806.
Courtadon, G. (1982) The pricing of options on default-free bonds. *Journal of Finance and Quantitative Analysis* 17, 75–100.
Cox, J.C., J.E. Ingersoll, & S.A. Ross (1985) A theory of the term structure of interest rates. *Econometrica* 53, 385–403.
Duffie, D. & R. Kan (1996) A yield factor model of interest rates. *Mathematical Finance* 6, 379–406.
Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
Fan, J.Q. & Q.W. Yao (2005) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.
Freedman, D.A. (1975) On tail probabilities for martingales. *Annals of Probability* 3, 100–118.
Frees, E.W. (1994) Estimating densities of functions of observations. *Journal of the American Statistical Association* 89, 517–525.
Gao, J. & M. King (2004) Adaptive testing in continuous-time diffusion models. *Econometric Theory* 20, 844–882.
Giné, E. & D.M. Mason (2007) On local U-statistic processes and the estimation of densities of functions of several sample variables. *Annals of Statistics* 35, 1105–1145.
Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Application*. Academic Press.

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Hall, P. & Q.W. Yao (2003) Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71, 285–317.
Hannan, E.J. (1973) Central limit theorems for time series regression. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 26, 157–170.
Hong, Y. & H. Li (2005) Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies* 18, 37–84.
Jones, D.A. (1976) Nonlinear autoregressive processes. *Proceedings of the Royal Society A* 360, 71–95.
Klimko, L.A. & P.I. Nelson (1978) On conditional least squares estimation for stochastic processes. *Annals of Statistics* 6, 629–642.
Ling, S. (2004) Estimation and testing stationarity for double-autoregressive models. *Journal of the Royal Statistical Society, Series B* 66, 63–78.
Ling, S. (2007) A double AR(p) model: Structure and estimation. *Statistica Sinica* 17, 161–175.
Liu, W.D. & W.B. Wu (2010) Simultaneous nonparametric inference of time series. *Annals of Statistics* 38, 2388–2421.
Pritsker, M. (1998) Nonparametric density estimation and tests of continuous time interest rate models. *Review of Financial Studies* 11, 449–487.
Robinson, P.M. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
Saavedra, A. & R. Cao (2000) On the estimation of the marginal density of a moving average process. *Canadian Journal of Statistics* 28, 799–815.
Schick, A. & W. Wefelmeyer (2004) Root n consistent density estimators for sums of independent random variables. *Journal of Nonparametric Statistics* 16, 925–935.
Schick, A. & W. Wefelmeyer (2007) Uniformly root-n consistent density estimators for weakly dependent invertible linear processes. *Annals of Statistics* 35, 815–843.
Scott, D.W. (1992) *Multivariate Density Estimation. Theory, Practice and Visualization*. Wiley.
Shao, X.F. & W.B. Wu (2007) Asymptotic spectral theory for nonlinear time series. *Annals of Statistics* 35, 1773–1801.
Silverman, B.W. (1986) *Density Estimation*. Chapman and Hall.
Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk. *Journal of Finance* 52, 1973–2002.
Tong, H. (1990) *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press.
Vasicek, O. (1977) An equilibrium characterization of the term structure. *Journal of Financial Economics* 5, 177–188.
Weiss, A.A. (1986) Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory* 2, 107–131.
Wu, W.B. (2005) Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* 102, 14150–14154.
Wu, W.B. & X.F. Shao (2004) Limit theorems for iterated random functions. *Journal of Applied Probability* 41, 425–436.
Zhao, Z.B. (2008) Parametric and nonparametric models and methods in financial econometrics. *Statistics Surveys* 2, 1–42.
Zhao, Z.B. (2011) Nonparametric model validations for hidden Markov models with applications in financial econometrics. *Journal of Econometrics* 162, 225–239.

APPENDIX

In this section we shall provide proofs for results in Sections 2 and 3. A careful check of the proofs implies that our asymptotic results also hold for the nonlinear AR(p) process

$$X_i = R(X_{i-1}, \ldots, X_{i-p}; \epsilon_i) := \mu(X_{i-1}, \ldots, X_{i-p}) + \sigma(X_{i-1}, \ldots, X_{i-p})\epsilon_i,$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

with the straightforward modification that the function $m_\theta(X_{i-1})$ in Assumptions 4 and 8 should now be replaced by $m_\theta(X_{i-1},\ldots,X_{i-p}) = \mu(X_{i-1},\ldots,X_{i-p})/\sigma(X_{i-1},\ldots,X_{i-p})$, and Assumption 7 is replaced by the following: for all $y_1,\ldots,y_p,y_1',\ldots,y_p' \in \mathbb{R}$,

$$\|R(y_1,\ldots,y_p;\epsilon_i) - R(y_1',\ldots,y_p';\epsilon_i)\|_\alpha \le \sum_{i=1}^{p} a_i |y_i - y_i'|^{\alpha'},$$

where a_i are nonnegative constants with $\sum_{i=1}^p a_i < 1$, $\epsilon_i \in \mathcal{L}^\alpha$, $\alpha > 0$, and $\alpha' = \min(1,\alpha)$. Shao and Wu (2007) showed that the above condition also implies the geometric moment contraction property (13). In our proofs for ease of reading we consider AR(1) models.

Proof. (Theorem 1) Let $\hat{S}_n(x) = \sum_{i=1}^n h(x - m_{\hat{\theta}}(X_{i-1}))$ and $S_n(x) = \sum_{i=1}^n h(x - m_{\theta_0}(X_{i-1}))$. By Assumption 3, $\sup_x |\hat{S}_n''(x)| = O(n)$. By Taylor's expansion, since $\int_{\mathbb{R}} uK(u)du = 0$,

$$\begin{aligned} |n\check{f}(x) - \hat{S}_n(x)| &= \left| \int_{\mathbb{R}} K(u)\hat{S}_n(x - ub_n)du - \hat{S}_n(x) \right| \\ &\le \int_{\mathbb{R}} |K(u)||\hat{S}_n(x - ub_n) - \hat{S}_n(x) + ub_n\hat{S}_n'(x)|du \\ &= \int_{\mathbb{R}} |K(u)|(ub_n)^2 O(n)du = O(nb_n^2). \end{aligned} \tag{29}$$

So $\sup_x |n\check{f}(x) - \hat{S}_n(x)| = o(\sqrt{n})$ since $\sqrt{n}b_n^2 \to 0$. Let

$$T_n(x,\theta) = \sum_{i=1}^{n} h'(x - m_\theta(X_{i-1}))\dot{m}_\theta(X_{i-1}).$$

By Assumptions 3 and 4 and the Lebesgue Dominated Convergence Theorem,

$$\lim_{\delta\to 0} \mathbb{E}\left[\sup_{|\theta-\theta_0|\le\delta} |h'(x - m_\theta(X_{i-1})) - h'(x - m_{\theta_0}(X_{i-1}))||\dot{m}_{\theta_0}(X_{i-1})| \right] = 0.$$

Therefore, again by Assumptions 3 and 4,

$$\lim_{\delta\to 0} \frac{1}{n}\mathbb{E}\left[\sup_{|\theta-\theta_0|\le\delta} |T_n(x,\theta) - T_n(x,\theta_0)| \right] = 0. \tag{30}$$

By Assumption 5, $\hat{\theta}_n - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$. By Taylor's expansion and (30),

$$\sup_x |\hat{S}_n(x) - S_n(x) - (\hat{\theta}_n - \theta_0)T_n(x,\theta_0)| = |\hat{\theta}_n - \theta_0| o_{\mathbb{P}}(n) = o_{\mathbb{P}}(n^{1/2}). \tag{31}$$

By the ergodic theorem, $\lim_{n\to\infty} T_n(x,\theta_0)/n \to c_0(x)$ almost surely. The latter convergence can be made uniform over $x \in \mathbb{R}$ since $c_0(\cdot)$ is uniformly continuous. By (29)–(31),

$$\sup_x \left| n\check{f}(x) - S_n(x) + nc_0(x)\sum_{i=1}^{n} R_i \right| = o_{\mathbb{P}}(\sqrt{n}). \tag{32}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Let $p' = \min(1, p)$. By Assumptions 2 and 3, and Remark 1,

$$\begin{aligned}\|\mathcal{P}_0 h(x - m_{\theta_0}(X_i))\| &= \|\mathbb{E}[h(x - m_{\theta_0}(X_i)) - h(x - m_{\theta_0}(X_i^*))|\mathcal{F}_0]\| \\ &\le \|h(x - m_{\theta_0}(X_i)) - h(x - m_{\theta_0}(X_i^*))\| \\ &\le \|A_0 \min[1, |m_{\theta_0}(X_i) - m_{\theta_0}(X_i^*)|]\| \\ &\le A_0[\mathbb{E}(|X_i - X_i^*|^{p'})]^{1/2} \\ &\le A_0 \|X_i - X_i^*\|_p^{p'/2} = O(\rho^{ip'/2}). \end{aligned} \tag{33}$$

Since $\sum_{i=0}^{\infty} \|\mathcal{P}_0 R_i\| < \infty$, we have

$$\sum_{i=0}^{\infty} \|\mathcal{P}_0[h(x - m_\theta(X_i)) - c_0(x) R_i]\| < \infty, \tag{34}$$

by Theorem 1 in Hannan (1973), (i) follows.

To prove (ii), since the finite dimensional convergence easily follows from (34), it remains to verify the tightness of $W_n(x) := \sqrt{n}[\check{f}(x) - f(x)]$; see Billingsley (1968). Since $\mathcal{I}$ is compact, by (32), it suffices to show that $\mathbb{E}[\sup_{x\in\mathcal{I}} |S_n'(x) - nf'(x)|] = O(\sqrt{n})$. Let $U_n(x) = [S_n'(x) - nf'(x)]/\sqrt{n}$. By Remark 1 and Assumption 3, as (33), we have

$$\sup_x \|\mathcal{P}_0 h''(x - m_{\theta_0}(X_i))\| = O(\rho^{ip'/2}).$$

So $\sup_x \|U_n'(x)\| \le \sum_{i=0}^{\infty} \sup_x \|\mathcal{P}_0 h''(x - m_{\theta_0}(X_i))\| = O(1)$, and $\mathbb{E}[\sup_{x\in\mathcal{I}} |U_n(x) - U_n(0)|] \le \int_{\mathcal{T}} \mathbb{E}|U_n'(x)|dx = O(1)$. Thus the tightness follows. ■

LEMMA 1. *Recall (8) for $\hat{g}_\theta(y)$. Assume $nb_n^2 \to \infty$ and $b_n \to 0$. Under Assumptions 1 and 4–6, we have*

$$\int_{\mathbb{R}} |\hat{g}_{\hat{\theta}}(y) - g(y)|dy = O_{\mathbb{P}}((\sqrt{n}b_n)^{-1} + b_n^2). \tag{35}$$

Proof. For $\delta > 0$ let $\dot{m}_*(\delta, x) = \sup_{|\theta-\theta_0|\le\delta} |\dot{m}_\theta(x)|$. By Assumption 4, there exists $\delta_0 > 0$ such that $\mathbb{E}[\dot{m}_*(\delta_0, X_0)] < \infty$. Choose $k_0 > 0$ such that $\int_{\mathbb{R}} |K(u) - K(u+\delta)|du \le |\delta|k_0$. Then

$$\begin{aligned}\int_{\mathbb{R}} |\hat{g}_{\hat{\theta}}(y) - \hat{g}_{\theta_0}(y)|dy &\le \frac{1}{nb_n} \sum_{i=1}^{n} \int_{\mathbb{R}} \left| K\left(\frac{y - m_{\hat{\theta}}(X_{i-1})}{b_n}\right) - K\left(\frac{y - m_{\theta_0}(X_{i-1})}{b_n}\right)\right| dy \\ &\le \frac{1}{nb_n} \sum_{i=1}^{n} |m_{\hat{\theta}}(X_{i-1}) - m_{\theta_0}(X_{i-1})|k_0 = O_{\mathbb{P}}((\sqrt{n}b_n)^{-1}). \end{aligned} \tag{36}$$

In the last step, we have applied the inequality $|m_\theta(x) - m_{\theta_0}(x)| \le |\theta - \theta_0|\dot{m}_*(|\theta - \theta_0|, x)$, the ergodic theorem, and the fact that $\hat{\theta} - \theta_0 = O_{\mathbb{P}}(n^{-1/2})$.

Without loss of generality let $1 < p \le 2$. Let $\lambda(y) = (1 + |y|)^p$. Since $Z_i \in \mathcal{L}^p$,

$$\begin{aligned}\int_{\mathbb{R}} \|K[(y - Z_0)/b_n]\|^2 \lambda(y)dy &= b_n \int_{\mathbb{R}} K^2(u)du \int_{\mathbb{R}} g(y - b_n u)\lambda(y)dy \\ &\le b_n \int_{\mathbb{R}} K^2(u)(1 + |b_n u|)^p du \int_{\mathbb{R}} g(y)\lambda(y)dy \\ &= O(b_n). \end{aligned}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Let

$$\hat{g}_n^0(y) = (nb_n)^{-1} \sum_{i=1}^{n} \mathbb{E}\{K[(y - Z_{i-1})/b_n] | \mathcal{F}_{i-2}\}.$$

By the orthogonality of martingale differences and Schwarz's inequality,

$$\begin{aligned}
\mathbb{E} \int_{\mathbb{R}} |\hat{g}_{\theta_0}(y) - \hat{g}_n^0(y)| dy &\le \int_{\mathbb{R}} \|\hat{g}_{\theta_0}(y) - \hat{g}_n^0(y)\| dy \\
&\le (nb_n)^{-1} \int_{\mathbb{R}} \sqrt{n} \|K[(y - Z_0)/b_n]\| dy \\
&\le \frac{1}{\sqrt{n} b_n} \left[\int_{\mathbb{R}} \|K[(y - Z_0)/b_n]\|^2 \lambda(y) dy \right]^{1/2} \left[\int_{\mathbb{R}} \lambda^{-1}(y) dy \right]^{1/2} \\
&= O((\sqrt{nb_n})^{-1}). \qquad (37)
\end{aligned}$$

Note that $\mathbb{E}\{K[(y - Z_{i-1})/b_n] | \mathcal{F}_{i-2}\} = b_n \int_{\mathbb{R}} K(u) f_Z(y - b_n u | \mathcal{F}_{i-2}) du$. Let

$$H_n(u) = \sum_{i=1}^{n} [f_Z(u | \mathcal{F}_{i-2}) - f_Z(u)].$$

By Theorem 2 in Wu (2005), $\|H_n(u)\| \le \sqrt{n} \sum_{i=0}^{\infty} \|\mathcal{P}_0 f_Z(u | \mathcal{F}_i)\|$, and Assumption 6(iii),

$$\begin{aligned}
\mathbb{E} \int_{\mathbb{R}} |\hat{g}_n^0(y) - \mathbb{E}[\hat{g}_n^0(y)]| dy &\le \frac{1}{n} \int_{\mathbb{R}} \int_{\mathbb{R}} |K(u)| \|H_n(y - b_n u)\| du dy \\
&\le \frac{1}{n} \int_{\mathbb{R}} \int_{\mathbb{R}} |K(u)| \sqrt{n} \sum_{i=0}^{\infty} \|\mathcal{P}_0 f_Z(y - b_n u | \mathcal{F}_i)\| du dy \\
&= O(n^{-1/2}). \qquad (38)
\end{aligned}$$

By the inequality $\int_{\mathbb{R}} |g(y + t) - g(y) - t g'(y)| dy \le (t^2/2) \int_{\mathbb{R}} |g''(y)| dy$, we have

$$\begin{aligned}
\int_{\mathbb{R}} |\mathbb{E}[\hat{g}_n^0(y)] - g(y)| dy &\le \int_{\mathbb{R}} \int_{\mathbb{R}} |K(u)| |g(y - b_n u) - g(y) + b_n u g'(y)| du dy \\
&= O(b_n^2) \qquad (39)
\end{aligned}$$

by Assumption 6(ii). By (36)–(39), the lemma follows. ■

LEMMA 2. *Assume* $(\log n)^2 = o(nb_n)$ *and* $b_n \to 0$. *Under Assumptions 1–5, we have*

$$\sup_{x \in \mathbb{R}} |\hat{h}_{\hat{\theta}}(x) - h(x)| = O_{\mathbb{P}}((nb_n)^{-1/2} \log n + b_n^2). \qquad (40)$$

Proof. Let $\delta_n \to 0$ be a positive sequence such that $\sqrt{n}\delta_n \to \infty$; let

$$\begin{aligned}
\hat{h}_\theta^0(x) &= \frac{1}{nb_n} \sum_{i=1}^{n} \mathbb{E}\{K[(x + m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1}) - \epsilon_i)/b_n] | \mathcal{F}_{i-1}\} \\
&= \frac{1}{n} \sum_{i=1}^{n} \int_{\mathbb{R}} K(u) h(x + m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1}) - b_n u) du.
\end{aligned}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Let $W_i(x,\theta) = K[(x + m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1}) - \epsilon_i)/b_n]$, $h_* = \sup_x h(x)$, $K_* = \sup_u |K(u)|$, and $K_2 = \int K^2(u)du$. Then

$$\mathbb{E}[W_i^2(x,\theta)|\mathcal{F}_{i-1}] = b_n \int_{\mathbb{R}} K^2(u)h(x - b_n u + m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1}))du$$
$$\leq b_n h_* K_2.$$

Note that $W_i(x,\theta) - \mathbb{E}[W_i(x,\theta)|\mathcal{F}_{i-1}]$, $i = 1, 2, \ldots$, form (triangular) stationary martingale differences bounded by $2K_*$. Using Freedman's (1975) exponential inequality for bounded martingale differences,

$$\mathbb{P}\left[|\hat{h}_\theta(x) - \hat{h}_\theta^0(x)| \geq \frac{\log n}{(nb_n)^{1/2}}\right] \leq 2\exp\left[-\frac{nb_n \log^2 n}{4K_*(nb_n)^{1/2}\log n + 4nb_n h_* K_2}\right]. \tag{41}$$

As in the proof of Lemma 1, let $\dot{m}_*(\delta, x) = \sup_{|\theta-\theta_0|\leq\delta} |\dot{m}_\theta(x)|$ and choose $\delta_0 > 0$ such that $\mathbb{E}[\dot{m}_*(\delta_0, X_0)] < \infty$. Let $N = n^4$ and define $\lfloor x \rfloor_N = \lfloor xN \rfloor/N$ and, for the d-dimensional vector $\theta = (\theta_1, \ldots, \theta_d)$, $\lfloor \theta \rfloor_N = (\lfloor \theta_1 N \rfloor/N, \ldots, \lfloor \theta_d N \rfloor/N)$. By Assumptions 1 and 4,

$$|\hat{h}_\theta(x) - \hat{h}_{\lfloor\theta\rfloor_N}(\lfloor x \rfloor_N)| \leq \sum_{i=1}^n \frac{k_0}{b_n}[N^{-1} + N^{-1}\dot{m}_*(\delta_0, X_{i-1})]$$

if $|\theta - \theta_0| \leq \delta_0$. Here $k_0 = \sup_u |K'(u)|$. Hence

$$\mathbb{E}\left[\sup_{x\in\mathbb{R}} \sup_{|\theta-\theta_0|\leq\delta_0} |\hat{h}_\theta(x) - \hat{h}_{\lfloor\theta\rfloor_N}(\lfloor x \rfloor_N)|\right] = o(n^{-2}). \tag{42}$$

Clearly the above relation (42) also holds if we replace $\hat{h}_\theta(x)$ therein by $\hat{h}_\theta^0(x)$. By Assumptions 2 and 4, there exists a constant $\tau > 0$ such that $\sup_{|\theta-\theta_0|\leq\delta} |m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1}) - \epsilon_i| \in \mathcal{L}^\tau$. Recall Assumption 3 for β and let the interval $\mathcal{I}_n = [-M, M]$, where $M = n^{8/\beta+8/\tau+8}$. Note that the cardinality $\#\{(\lfloor\theta\rfloor_N, \lfloor x \rfloor_N) : x \in \mathcal{I}_n, |\theta - \theta_0| \leq \delta_0\} = O(2MN^{1+d})$. In (41), since $(\log n)^2 = o(nb_n)$, we obtain

$$\mathbb{P}\left[\sup_{x\in\mathcal{I}_n} \sup_{|\theta-\theta_0|\leq\delta_0} |\hat{h}_{\lfloor\theta\rfloor_N}(\lfloor x \rfloor_N) - \hat{h}^0_{\lfloor\theta\rfloor_N}(\lfloor x \rfloor_N)| \geq \frac{\log n}{(nb_n)^{1/2}}\right]$$
$$= O\left(2MN^{1+d}\right)\exp\left[-\frac{nb_n \log^2 n}{4K_*(nb_n)^{1/2}\log n + 2nb_n h_* K_2}\right] = o(n^{-1}). \tag{43}$$

By (42) and (43), we obtain

$$\sup_{x\in\mathcal{I}_n} \sup_{|\theta-\theta_0|\leq\delta_n} |\hat{h}_\theta(x) - \hat{h}_\theta^0(x)| = O_{\mathbb{P}}((nb_n)^{-1/2}\log n). \tag{44}$$

By the choice of M and since K has bounded support, we have $\sup_{|x|\geq M} \sup_{|\theta-\theta_0|\leq\delta} |\hat{h}_\theta(x)| = 0$ in probability. By Assumption 3, $\sup_{|x|\geq M} h(x) = o(n^{-1})$. Hence

$$\sup_{|x|\geq M} \sup_{|\theta-\theta_0|\leq\delta_n} |\hat{h}_\theta(x) - h(x)| = o_{\mathbb{P}}(n^{-1}). \tag{45}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

On the other hand, let

$$I_\theta(x) = I_{\theta,n}(x) = \sum_{i=1}^{n} h(x + m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1})). \tag{46}$$

As (36), by the ergodic theorem, $\sup_x \sup_{|\theta-\theta_0|\le\delta_n} |I_{\theta,n}(x)/n - h(x)| = O_{\mathbb{P}}(\delta_n)$. Hence

$$\sup_{x\in\mathbb{R}} \sup_{|\theta-\theta_0|\le\delta_n} |\hat{h}^0_\theta(x) - \int_{\mathbb{R}} K(u)h(x-b_nu)du| = O_{\mathbb{P}}(\delta_n). \tag{47}$$

Since the speed $\sqrt{n}\delta_n \to \infty$ can be made arbitrarily slow, by (44) and (47), we have

$$\sup_{x\in\mathcal{I}_n} \left| \hat{h}_{\hat\theta}(x) - \int_{\mathbb{R}} K(u)h(x-b_nu)du \right| = O_{\mathbb{P}}((nb_n)^{-1/2}\log n). \tag{48}$$

Since $\sup_x |h''(x)| < \infty$, $\sup_x |\int_{\mathbb{R}} K(u)[h(x-b_nu) - h(x) + b_nuh'(x)]du| = O(b_n^2)$. Note that $\int_{\mathbb{R}} uK(u)du = 0$, Lemma 2 then follows from (45) and (48). ■

Proof. (Theorem 2) (i) Since $(\log n)^2 = o(nb_n^3)$ and $b_n = o(n^{-1/4})$, simple calculations show that

$$\left[(\sqrt{n}b_n)^{-1} + b_n^2\right]\left[(nb_n)^{-1/2}\log n + b_n^2\right] = o(n^{-1/2}).$$

By Lemmas 1 and 2, we have

$$\sup_x \int_{\mathbb{R}} |[\hat{h}_{\hat\theta}(x-y) - h(x-y)][\hat{g}_{\hat\theta}(y) - g(y)]|dy = o_{\mathbb{P}}(n^{-1/2}). \tag{49}$$

Let

$$\hat{W}_n(x) = \sum_{i=1}^{n} [h(x - m_{\hat\theta}(X_{i-1})) + g(x - \hat\epsilon_i)].$$

Since $\sup_x [|h''(x)| + |g''(x)|] < \infty$,

$$\hat{W}_n(x - b_nu) - 2nf(x - b_nu) = \left\{\hat{W}_n(x) - 2nf(x)\right\} - b_nu\left\{\hat{W}_n'(x) - 2nf_X'(x)\right\} + b_n^2u^2O(n).$$

Therefore, by Assumption 1, we have $\int_{\mathbb{R}} K(u)f(x-b_nu) = f(x) + O(b_n^2)$ and

$$\begin{aligned}
\Delta_n(x) &:= \int_{\mathbb{R}} [\hat{h}_{\hat\theta}(x-y) - h(x-y)][\hat{g}_{\hat\theta}(y) - g(y)]dy \\
&= \tilde{f}(x) + f(x) - \int_{\mathbb{R}} [\hat{h}_{\hat\theta}(x-y)g(y) + h(x-y)\hat{g}_{\hat\theta}(y)]dy \\
&= \tilde{f}(x) + f(x) - n^{-1}\int_{\mathbb{R}} K(u)\hat{W}_n(x-b_nu)du \\
&= \tilde{f}(x) + f(x) - 2\int_{\mathbb{R}} K(u)f(x-b_nu)du - \frac{\hat{W}_n(x) - 2nf(x)}{n} + O(b_n^2) \\
&= \tilde{f}(x) - f(x) - \frac{\hat{W}_n(x) - 2nf(x)}{n} + O(b_n^2).
\end{aligned}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Let $W_n(x) = \sum_{i=1}^n [h(x - m_{\theta_0}(X_{i-1})) + g(x - \epsilon_i)]$ and

$$V_n(x,\theta) = \sum_{i=1}^n \dot{m}_\theta(X_{i-1})[g'(x - \epsilon_i + m_\theta(X_{i-1}) - m_{\theta_0}(X_{i-1})) - h'(x - m_\theta(X_{i-1}))].$$

Similarly as (31), we have

$$\sup_x |\hat{W}_n(x) - W_n(x) - (\hat{\theta}_n - \theta_0)V_n(x,\theta_0)| = o_{\mathbb{P}}(n^{1/2}).$$

Note that $\mathbb{E}\{\dot{m}_{\theta_0}(X_0)[g'(x - \epsilon_1) - h'(x - m_{\theta_0}(X_0))]\} = \gamma_0(x)$. By the ergodic theorem, since $\sup_x[|h''(x)| + |g''(x)|] < \infty$, for any compact interval $\mathcal{I}$, we have

$$\lim_{n\to\infty} \sup_{x\in\mathcal{I}} \mathbb{E}\left|\frac{V_n(x,\theta_0)}{n} - \gamma_0(x)\right| = 0.$$

Since $\sqrt{n}b_n^2 \to 0$, we have

$$\sup_{x\in\mathcal{I}} |\tilde{f}_X(x) - f(x) - n^{-1}W_n(x) - (\hat{\theta}_n - \theta_0)\gamma_0(x)| = o_{\mathbb{P}}(n^{-1/2}). \tag{50}$$

Then similarly as the argument in (34), we have (14).

(ii) It can be similarly proved by using the argument in the proof of Theorem 1(ii). ■

Proof. (Theorem 3) We shall apply Theorem 2.1 in Liu and Wu (2010). The bandwidth condition $(\log n)^2 = o(nb_n^3)$ and $b_n = o(n^{-1/4})$ in our Theorem 2 implies Condition (C1) of Theorem 2.1 in Liu and Wu (2010). Condition (C3) in the latter paper is just our Assumption 3, and (C2) follows from the geometric contraction (13). Since $b_n = o(n^{-1/4})$, $\sqrt{nb_n}|\mathbb{E}\hat{f}(x) - f(x)| = O(n^{-1/8})$. By Theorem 2.1 in Liu and Wu (2010), (18) holds for Ξ_n° with

$$\Xi_n^\circ = \sup_{l\le x\le u} \frac{\sqrt{nb_n}|\hat{f}(x) - f(x)|}{\sqrt{f(x)\int K^2(u)du}}. \tag{51}$$

Then (18) results from Theorem 2 in view of $\sup_{l\le x\le u}|\tilde{f}(x) - f(x)| = O_{\mathbb{P}}(n^{-1/2})$. ■

Proof. (Theorem 4) Under Assumption 7, by Wu and Shao (2004), the process (X_i) satisfies the geometric moment contraction property (13). By Assumption 8, with elementary manipulations, the process $\mu_{\theta_0}(X_{i-1})/\sigma_{\theta_0}(X_{i-1})$ also satisfies (13). Since $X_i/\sigma_{\theta_0}(X_{i-1}) = \epsilon_i + \mu_{\theta_0}(X_{i-1})/\sigma_{\theta_0}(X_{i-1})$, Theorem 2.1 in Liu and Wu (2010) entails (18) with

$$\Xi_n^\diamond := \sup_{l\le x\le u} \frac{\sqrt{nb_n}|\hat{f}_{\theta_0}^\dagger(x) - f_{\theta_0}^\dagger(x)|}{\sqrt{f_{\theta_0}^\dagger(x)\int K^2(u)du}}. \tag{52}$$

Then it remains to show that $\hat{f}_{\hat{\theta}}^\dagger(x)$ (respectively $\tilde{f}_{\hat{\theta}}^\dagger(x)$) is close to $\hat{f}_{\theta_0}^\dagger(x)$ (respectively $f_{\theta_0}^\dagger(x)$) in an appropriate sense. We shall first deal with the difference $\tilde{f}_{\hat{\theta}}^\dagger(x) - f_{\theta_0}^\dagger(x)$. By Lemma 1, under Assumptions 1, 5, 6, and 8, we have (35). We now show that (40) holds

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

for the estimate $\hat{h}_{\hat{\theta}}(\cdot)$ defined in (21). Let $|\theta-\theta_0| \le \delta_n$. Note that the conditional density of $X_i = \mu_{\theta_0}(X_{i-1})+\epsilon_i\sigma_{\theta_0}(X_{i-1})$ given $\mathcal{F}_{i-1}$ is $h[(\cdot-\mu_{\theta_0}(X_{i-1}))/\sigma_{\theta_0}(X_{i-1})]/\sigma_{\theta_0}(X_{i-1})$. Then there exists a constant λ, independent of x, θ, and n, such that $\mathbb{E}[W_i^2(x,\theta)|\mathcal{F}_{i-1}] \le b_n\lambda$. Then following the proof of Lemma 2, we obtain (40), by noting that the function $I_\theta(x)$ in (46) shall now be replaced by

$$I_\theta^\dagger(x) = \sum_{i=1}^n h\left(\frac{x\sigma_\theta(X_{i-1})+\mu_\theta(X_{i-1})-\mu_{\theta_0}(X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}\right)\frac{\sigma_\theta(X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}.$$

Hence, with (35) and (40), by the argument in the proof of Theorem 2, we obtain a similar version of (50), which implies that

$$\sup_{l\le x\le u}|\tilde{f}_{\hat{\theta}}^\dagger(x) - f_{\theta_0}^\dagger(x)| = O_{\mathbb{P}}(n^{-1/2}). \tag{53}$$

We shall now establish a similar bound for $\hat{f}_\theta^\dagger(x) - \hat{f}_{\theta_0}^\dagger(x)$. Let $\mathcal{I} = [l,u]$; let $(\delta_n)_{n\ge0}$ be a decreasing sequence satisfying $\sqrt{n}\delta_n \to \infty$ and $\sqrt{n}\delta_n = o(\log n)$. Let

$$\begin{aligned} W_i^\dagger(x,\theta) &= K\left(\frac{x - X_i/\sigma_\theta(X_{i-1})}{b_n}\right) \\ &= K\left(\frac{x\sigma_\theta(X_{i-1})-\mu_{\theta_0}(X_{i-1})-\epsilon_i\sigma_{\theta_0}(X_{i-1})}{b_n\sigma_\theta(X_{i-1})}\right). \end{aligned}$$

As in the proof of Lemma 2, let $h_* = \sup_x h(x)$, $K_* = \sup_u |K(u)|$, and $K_2 = \int K^2(u)du$. By Assumption 1, there exists a constant C_K, only depending on the kernel K, such that

$$\mathbb{E}[|W_i^\dagger(x,\theta) - W_i^\dagger(x,\theta_0)||\mathcal{F}_{i-1}] \le C_K(|x|+b_n)\left|1-\frac{\sigma_\theta(X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}\right|\frac{\sigma_\theta(X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}h_*. \tag{54}$$

By Assumption 8, $|\sigma_\theta(X_{i-1})-\sigma_{\theta_0}(X_{i-1})| \le \delta_n\dot{\sigma}_*(\delta_0, X_{i-1})$. Hence, for

$$V_n := \sum_{i=1}^n \frac{\dot{\sigma}_*(\delta_0, X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}\left[1+\delta_0\frac{\dot{\sigma}_*(\delta_0, X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}\right],$$

we have by (54) that, for $\lambda_0 = 2K_*C_K(|l|+|u|+\sup_n b_n)h_*$, we have

$$\sum_{i=1}^n \mathbb{E}\left[|W_i^\dagger(x,\theta) - W_i^\dagger(x,\theta_0)|^2|\mathcal{F}_{i-1}\right] \le \lambda_0\delta_n V_n. \tag{55}$$

By Assumption 8, $\upsilon_1 = \mathbb{E}V_1 < \infty$. Let

$$G_n(x,\theta) = \sum_{i=1}^n\{W_i^\dagger(x,\theta) - \mathbb{E}[W_i^\dagger(x,\theta)|\mathcal{F}_{i-1}]\}.$$

Note that the summands of $G_n(x,\theta)$ are martingale differences bounded by $2K_*$. Again by Freedman's (1975) inequality and (55), let $t_n = \sqrt{n\delta_n}\log n$, we have

$$\mathbb{P}\left[|G_n(x,\theta) - G_n(x,\theta_0)| \ge t_n,\, V_n \le 2n\upsilon_1\right] \le 2\exp\left[-\frac{t_n^2}{4K_*t_n + 4\lambda_0\delta_n n\upsilon_1}\right]. \tag{56}$$

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press

Since $\sqrt{n}\delta_n \to \infty$ and $\sqrt{n}\delta_n = o(\log n)$, letting $N = n^4$, as in (43), we have

$$\mathbb{P}\left[\sup_{x\in\mathcal{I}}\sup_{|\theta-\theta_0|\le\delta_0}|G_n(\lfloor x\rfloor_N,\lfloor\theta\rfloor_N)-G_n(\lfloor x\rfloor_N,\theta_0)|\ge t_n,\ V_n\le 2nv_1\right]=o(n^{-1}),$$

which, by the argument in (42) and the fact that $\mathbb{P}(V_n > 2nv_1) \to 0$, implies

$$\sup_{x\in\mathcal{I}}\sup_{|\theta-\theta_0|\le\delta_0}|G_n(x,\theta)-G_n(x,\theta_0)|=O_{\mathbb{P}}(t_n). \tag{57}$$

We shall next deal with $\sum_{i=1}^n \mathbb{E}[W_i^\dagger(x,\theta)|\mathcal{F}_{i-1}]$. Analogous to (46), we now define

$$I_\theta(x)=\sum_{i=1}^n h\left(x\frac{\sigma_\theta(X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}-\frac{\mu_\theta(X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}\right)\frac{\sigma_\theta(X_{i-1})}{\sigma_{\theta_0}(X_{i-1})}. \tag{58}$$

Then

$$\sum_{i=1}^n \mathbb{E}[W_i^\dagger(x,\theta)-W_i^\dagger(x,\theta_0)|\mathcal{F}_{i-1}]=b_n\int_{\mathbb{R}}K(v)[I_\theta(x-b_nv)-I_{\theta_0}(x-b_nv)]dv.$$

By the ergodic theorem and Assumption 8, we have $\mathbb{E}[\sup_{x\in\mathcal{I}}|\dot{I}_\theta(x)|]=O(n)$. Then

$$\sup_{x\in\mathcal{I}}\sup_{|\theta-\theta_0|\le\delta_n}\left|\sum_{i=1}^n \mathbb{E}[W_i^\dagger(x,\theta)-W_i^\dagger(x,\theta_0)|\mathcal{F}_{i-1}]\right|=b_n\delta_n O_{\mathbb{P}}(n). \tag{59}$$

By (57) and (59),

$$\sup_{x\in\mathcal{I}}\sup_{|\theta-\theta_0|\le\delta_n}|\hat{f}_\theta^\dagger(x)-\hat{f}_{\theta_0}^\dagger(x)|=\frac{O_{\mathbb{P}}(t_n+nb_n\delta_n)}{nb_n}=\frac{O_{\mathbb{P}}(1)}{(nb_n)^{1/2}\log^2 n} \tag{60}$$

since $b_n = o(n^{-1/4})$, $\sqrt{n}\delta_n \to \infty$ and $\sqrt{n}\delta_n = o(\log n)$. Hence $\Xi_n^\dagger - \Xi_n^\diamond = O(\log^{-2} n)$ and the theorem follows. ■

https://doi.org/10.1017/S0266466614000681 Published online by Cambridge University Press