Journal of Multivariate Analysis 130 (2014) 118–133

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Specification test for Markov models with measurement errors☆

Seonjin Kim [a], Zhibiao Zhao [b,*]

[a] Miami University, United States

[b] Penn State University, United States

ARTICLE INFO

Article history:
Received 13 October 2013
Available online 21 May 2014

AMS subject classifications:
62G10
62G08

Keywords:
Simultaneous confidence band
Markov model
Measurement errors
Time series
Nonparametric estimation
Specification testing

ABSTRACT

Most existing works on specification testing assume that we have direct observations from the model of interest. We study specification testing for Markov models based on contaminated observations. The evolving model dynamics of the unobservable Markov chain is implicitly coded into the conditional distribution of the observed process. To test whether the underlying Markov chain follows a parametric model, we propose measuring the deviation between nonparametric and parametric estimates of conditional regression functions of the observed process. Specifically, we construct a nonparametric simultaneous confidence band for conditional regression functions and check whether the parametric estimate is contained within the band.

1. Introduction

Let $\{X_i\}_{i\in\mathbb{N}}$ be a real-valued stationary time series of interest. In some applications, $\{X_i\}$ may not be directly observable and instead we observe a contaminated version $\{Y_i\}$:

$$Y_i = X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{1}$$

where $\{\varepsilon_i\}$ are independent and identically distributed (i.i.d.) measurement errors. For example, (1) has been proposed to explain the microstructure noise phenomenon observed in high-frequency financial data [3,29,21,24]. Another example is the widely used stochastic volatility model in financial econometrics:

$$Y_i = \sigma_i \varepsilon_i, \tag{2}$$

where $\{\sigma_i > 0\}$ is an unobservable volatility process. For example, Taylor [27] proposed an AR(1) model for $\log(\sigma_i^2)$: $\log(\sigma_i^2) = a\log(\sigma_{i-1}^2) + \eta_i$. In general, with the transformation $Y_i^* = \log(Y_i^2)$, $X_i = \log(\sigma_i^2)$, $\varepsilon_i^* = \log(\varepsilon_i^2)$, then (2) becomes the measurement errors model:

$$Y_i^* = X_i + \varepsilon_i^*. \tag{3}$$

☆ We are grateful to the editor and an anonymous referee for their constructive comments. Zhao acknowledges partial financial support from NSF grant DMS-1309213 and NIDA grant P50-DA10075. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NSF.

* Correspondence to: Department of Statistics, Penn State University, 326 Thomas Building, University Park, PA 16802, United States.
E-mail addresses: kims20@miamioh.edu (S. Kim), zuz13@stat.psu.edu (Z. Zhao).

http://dx.doi.org/10.1016/j.jmva.2014.05.008
0047-259X/

In the vast literature on errors-in-variables or measurement errors models, the central goal has been to study parameter estimation and inference of parametric regressions in the presence of measurement errors on the covariates; see the monographs Fuller [15] and Carroll et al. [6] and the recent survey paper Chen et al. [7] for an extensive account of related contributions. Unlike the aforementioned works, our focus is on inferring the model dynamics of the unobservable process $\{X_i\}$.

The main purpose of this article is to address specification testing regarding the underlying data-generating mechanism, denoted by $\mathcal{Q}$, that generates $\{X_i\}$ based on the contaminated observations $\{Y_i\}$. Specifically, we are interested in testing

$$H_0 : \mathcal{Q} = \mathcal{Q}_\theta, \text{ for a parametric specification } \mathcal{Q}_\theta \text{ with unknown parameter } \theta. \tag{4}$$

Parametric models can provide a parsimonious interpretation of the model dynamics, but a mis-specification of the underlying model may result in wrong conclusions. Therefore, it is necessary to validate the adequacy of the parametric model before employing it.

There is an extensive literature on specification testing but most existing works are concentrated on the case that data of interest are directly observable. Some representative works include pseudo-likelihood ratio test [4], square distance between parametric and nonparametric estimate [17], residuals-based tests [10,20], generalized likelihood ratio test [13], and density based approaches [1,16,19,2]. In the above works, direct observations from the model of interest are available, a feature unfortunately not shared by (1).

Due to the non-observability and dependence of $\{X_i\}$, the aforementioned methods are not applicable and it is a difficult task to address specification testing regarding $\mathcal{Q}$. To address this issue, we impose a Markovian assumption on $\{X_i\}$. Markov chains are used in a wide range of fields, ranging from quantitative fields such as econometrics and statistics to more applied fields such as biology and engineering. In econometrics, one important example is the nonlinear autoregressive conditional heteroscedastic model

$$X_i = \mu(X_{i-1}) + s(X_{i-1})\eta_i, \tag{5}$$

for i.i.d. errors $\{\eta_i\}_{i\in\mathbb{Z}}$. Given different specifications of (μ, s), (5) includes many popular models, such as threshold autoregressive models and autoregressive conditional heteroscedastic models. Another example is discrete samples from the diffusion model

$$dX_t = \mu(X_t)dt + s(X_t)dW_t, \quad t \geq 0, \tag{6}$$

where $\{W_t\}_{t\geq 0}$ is a standard Brownian motion. This model includes many widely used financial models, see Zhao [30] for a review.

In this article, we propose a conditional expectation generator based approach to address the specification testing problem (4). Our approach is motivated by three facts: (i) the evolving dynamics of the unobservable Markov chain $\{X_i\}$ is characterized by its transition density, denoted by $q_X(x'|x)$; (ii) the transition density $q_X(x'|x)$ of $\{X_i\}$ is implicitly coded into the conditional density, denoted by $q_Y(y'|y)$, of Y_i given Y_{i-1}; and (iii) furthermore, $q_Y(y'|y)$ is coded into the conditional expectation

$$\mathcal{G}_g(y) = \mathbb{E}[g(Y_i)|Y_{i-1} = y], \tag{7}$$

for proper transformations $g(\cdot)$. To address specification testing for hidden Markov models, Zhao [31] compared the parametric estimate of $q_Y(y'|y)$ to its nonparametric estimate. Using $\mathcal{G}_g(y)$ instead of $q_Y(y'|y)$ has some practical advantages. In terms of the sample size, in order to estimate the two-dimensional function $q_Y(y'|y)$ nonparametrically, Zhao [31] required a relatively large sample size in the order of thousands; by contrast, the proposed conditional expectation based method works reasonably well for a moderate sample size (for example, 200) in simulation studies. For bandwidth selection, it is a more challenging issue to choose the bandwidths in nonparametric conditional density estimation, while there are well-studied standard bandwidth selections for nonparametric mean regression; see, e.g., Li and Racine [23] for detailed discussions. Furthermore, it is practically more convenient to compare the univariate function $\mathcal{G}_g(y)$ than the bivariate function $q_Y(y'|y)$.

The main component of our methodology is the construction of a nonparametric simultaneous confidence band (SCB) for $\mathcal{G}_g(y)$. The constructed nonparametric SCB does not depend on any specific model structure and hence can serve as a true reference. To test (4), we then check whether the parametric estimate of $\mathcal{G}_g(y)$ under H_0 is contained within the nonparametric SCB. The problem of SCB construction has been studied previously for marginal density of independent data [5], nonparametric regression function for both independent data [22,8,12] and time series data [32]. For hidden Markov models, Zhao [31] studied SCB for conditional density function. Our development on SCB for $\mathcal{G}_g(y)$ under the Markov-chain measurement-error model involves novel technical developments. The main argument is to decompose summation of dependent variables into a leading summation of martingale differences and a negligible error term. Unlike the nonparametric kernel density estimation case where the summands are uniformly bounded, nonparametric kernel smoothing estimate of the regression function $\mathcal{G}_g(y)$ involves unbounded terms and is significantly more challenging to deal with.

Throughout, for a random variable Z, we write $Z \in \mathcal{L}^q$, $q > 0$, if $\|Z\|_q := [\mathbb{E}(|Z|^q)]^{1/q} < \infty$; for $z \in \mathbb{R}$, write $\lfloor z \rfloor$ as the integer part of z. Section 2 presents the main methodology. Section 3 contains simulation studies. Technical proofs are provided in Section 4.

2. Methodology

For convenience, we gather some notations below, which are used throughout the paper,

- $p_X(x, x')$: the joint density function of (X_{i-1}, X_i);
- $q_X(x'|x)$: the conditional density function of X_i given $X_{i-1} = x$;
- $p_Y(y, y')$: the joint density function of (Y_{i-1}, Y_i);
- $q_Y(y'|y)$: the conditional density function of Y_i given $Y_{i-1} = y$;
- $f_Y(\cdot)$: the density function of Y_i;
- $q_\varepsilon(\cdot)$: the density function of ε_i.

For the Markov chain $\{X_i\}$, its evolving dynamics is characterized by the joint density function $p_X(x, x')$ of (X_{i-1}, X_i). Since $\{X_i\}$ is unobservable, we propose extracting information about $p_X(x, x')$ from the observed chain $\{Y_i\}$. In (1), we assume that the measurement errors $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ are i.i.d. and independent of the Markov chain $\{X_i\}_{i\in\mathbb{Z}}$.

Recall the conditional expectation operator $\mathcal{G}_g$ defined in (7). Our method is motivated by the following result:

Proposition 1. *Let ε be a random sample from $q_\varepsilon(\cdot)$ independent of (X_{i-1}, X_i). Then*

$$\mathcal{G}_g(y) = \frac{\mathbb{E}[g(X_i + \varepsilon)q_\varepsilon(y - X_{i-1})]}{\mathbb{E}[q_\varepsilon(y - X_{i-1})]}. \tag{8}$$

By Proposition 1, $\mathcal{G}_g(y)$ contains rich information about the joint density of (X_{i-1}, X_i), and different choices of $g(\cdot)$ can extract different information. For example, $g_1(Y_i) = Y_i$ and $g_2(Y_i) = Y_i^2$ extract information from the first two conditional moments, and $g_t(Y_i) = \mathbf{1}_{Y_i\le t}$, $t \in \mathbb{R}$, extracts information from the conditional distribution; see Section 2.3 for more discussions.

Motivated by (8), we introduce our conditional expectation based approach to address the specification testing problem (4). First, we apply nonparametric kernel smoothing methods to construct a nonparametric estimate of $\mathcal{G}_g(y)$, denoted by $\hat{\mathcal{G}}_g(y)$. Without imposing any specific model structure, $\hat{\mathcal{G}}_g(y)$ is always a consistent estimate of $\mathcal{G}_g(y)$ and hence can be used as a reference quantity. Under H_0, we use the right hand side of (8) to construct a parametric estimate of $\mathcal{G}_g(y)$, denoted by $\mathcal{G}_g(y|\mathcal{Q}_{\hat\theta})$, where $\hat\theta$ is a consistent estimate of θ. To test H_0, we examine the distance between the parametric estimate $\mathcal{G}_g(y|\mathcal{Q}_{\hat\theta})$ and the nonparametric reference $\hat{\mathcal{G}}_g(y)$, with a large discrepancy indicating rejection of H_0.

To determine the critical value, we use the idea of simultaneous confidence band (SCB). For a significance level $\alpha \in (0, 1)$, we say that $[l_n(\cdot), u_n(\cdot)]$ is an asymptotic $(1 - \alpha)$ nonparametric SCB for $\mathcal{G}(y)$ on a given compact set $\mathcal{Y} \subset \mathbb{R}$ if

$$\lim_{n\to\infty} \mathbb{P}\{l_n(y) \le \mathcal{G}_g(y) \le u_n(y), \text{ for all } y \in \mathcal{Y}\} = 1 - \alpha. \tag{9}$$

Intuitively, the function $\mathcal{G}_g(\cdot)$ is contained within the nonparametric band $[l_n(\cdot), u_n(\cdot)]$ with asymptotic probability $(1-\alpha)$. As will be illustrated in Section 2.1, nonparametric SCB of $\mathcal{G}_g(\cdot)$ usually centers at a nonparametric estimate $\hat{\mathcal{G}}_g(y)$. Therefore, the band $[l_n(\cdot), u_n(\cdot)]$ with center $\hat{\mathcal{G}}_g(y)$ provides an acceptance region for H_0. If the parametric estimate $\mathcal{G}_g(y|\mathcal{Q}_{\hat\theta})$ under H_0 falls outside the band, then the deviation between $\hat{\mathcal{G}}_g(y)$ and $\mathcal{G}_g(y|\mathcal{Q}_{\hat\theta})$ is too large to be in favor of H_0. Clearly, the concept of SCB is an extension of the classical confidence interval for a one-dimensional parameter (e.g., the population mean) to a function.

We now summarize our nonparametric SCB based specification testing procedure:

(i) Apply nonparametric methods to construct a nonparametric estimate $\hat{\mathcal{G}}_g(y)$ of $\mathcal{G}_g(y)$, and then use $\hat{\mathcal{G}}_g(y)$ to build a $(1 - \alpha)$ nonparametric SCB for $\mathcal{G}_g(y)$, denoted by $[\ell_n(\cdot), u_n(\cdot)]$.

(ii) Under H_0, apply parametric methods to obtain an estimate $\hat\theta$ of θ, and further use the right hand side of (8) to obtain a parametric estimate $\mathcal{G}_g(y|\mathcal{Q}_{\hat\theta})$ of $\mathcal{G}_g(y)$.

(iii) Check whether $l_n(y) \le \mathcal{G}_g(y|\mathcal{Q}_{\hat\theta}) \le u_n(y)$ holds for all $y \in \mathcal{Y}$, or equivalently, whether $\mathcal{G}_g(y|\mathcal{Q}_{\hat\theta})$ is contained within the constructed SCB. If no, we reject H_0 at level α.

In Sections 2.1 and 2.2, we construct nonparametric SCB and parametric estimate of $\mathcal{G}_g(y)$, respectively; Section 2.3 discusses choices of $g(\cdot)$ and the Bonferroni correction.

Remark 1. Although the underlying process $\{X_i\}$ is a Markov chain, the observed process $\{Y_i\}$ is no longer a Markov chain and thus its distributional property cannot be completely characterized by the one-step conditional distribution of Y_i given Y_{i-1}. To incorporate more distributional information, in light of (7), one may consider $\mathcal{G}_g(y_1, \ldots, y_k) = \mathbb{E}[g(Y_i)|Y_{i-1} = y_1, \ldots, Y_{i-k} = y_k]$ for some given k. However, due to the well-known curse of dimensionality, it is practically infeasible to nonparametrically estimate the latter multivariate function. Thus, we shall not pursue this direction.

2.1. Nonparametric simultaneous confidence band

Consider the Nadaraya–Watson estimate of $\mathcal{G}_g(y)$:

$$\hat{\mathcal{G}}_g(y) = \frac{\sum_{i=1}^{n} g(Y_i) K_{b_n}(y - Y_{i-1})}{\sum_{i=1}^{n} K_{b_n}(y - Y_{i-1})}, \tag{10}$$

where and hereafter $K_{b_n}(u) = K(u/b_n)$ for a kernel function K satisfying $\int_{\mathbb{R}} K(u)du = 1$ and bandwidth $b_n > 0$. To study asymptotic properties of $\hat{\mathcal{G}}_g(y)$, we define the conditional variance function $\sigma_g^2(y)$ and impose Conditions 1–3 as follows:

$$\sigma_g^2(y) = \mathbb{E}\{[g(Y_i) - \mathcal{G}_g(Y_{i-1})]^2 | Y_{i-1} = y\}. \tag{11}$$

Condition 1 (*Kernel Assumption*)**.** The kernel K is bounded, symmetric, and has bounded derivative and support $[-\omega, \omega]$. Write $\varphi_K = \int_{-\omega}^{\omega} K^2(u)du$ and $\psi_K = \int_{-\omega}^{\omega} u^2 K(u)du$.

Condition 2 (*Regularity Assumption*)**.** Without loss of generality, let $\mathcal{Y} = [-T, T]$ for some $T > 0$. There exists some small $\epsilon > 0$ such that $f_Y(y) > 0$ and $\mathcal{G}_g(y)$ have bounded fourth order derivative on $\mathcal{Y}_\epsilon := [-T - \epsilon, T + \epsilon]$, and that $\sigma_g^2(y) > 0$ has bounded derivative on $\mathcal{Y}_\epsilon$. The measurement errors $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ are i.i.d. and independent of the Markov chain $\{X_i\}_{i\in\mathbb{Z}}$. The density function q_ε of ε_i is bounded and has bounded derivative on $\mathbb{R}$.

Condition 3 (*Dependence Assumption*)**.** The unobservable process $\{X_i\}$ is an α-mixing stationary Markov chain with α-mixing coefficients α_k, $k \in \mathbb{N}$. Assume that $g(Y_i) \in \mathcal{L}^\delta$ for some $\delta \geq 4$ and $\sum_{k=1}^{\infty} \alpha_k^{1-2/\delta} < \infty$.

Definition 1. Let $\tau_n \to 0$ and $m_n \to \infty$. We say that $\mathcal{Y}_n \subset \mathcal{Y}$ is a (τ_n, m_n) approximation of $\mathcal{Y}$ if: (i) $\mathcal{Y}_n$ contains m_n distinct points from $\mathcal{Y}$; (ii) the distance between any two points from $\mathcal{Y}_n$ is at least τ_n; and (iii) the distance between $\mathcal{Y}_n$ and $\mathcal{Y}$ goes to zero as $n \to \infty$.

Theorem 1 establishes a maximal deviation result for $\hat{\mathcal{G}}_g(y)$, which can be used to construct a nonparametric SCB for $\hat{\mathcal{G}}_g(y)$.

Theorem 1. *Assume that Conditions 1–3 hold and $nb_n^9 \log n + (nb_n^3)^{-1} \log n \to 0$. Then for any (τ_n, m_n) approximation $\mathcal{Y}_n$ of $\mathcal{Y}$ such that $b_n = o(\tau_n)$ and $(\log n)^3[(nb_n)^{-1} + b_n^2]m_n^2 \to 0$,*

$$\lim_{n\to\infty} \mathbb{P}\left\{ \sup_{y\in\mathcal{Y}_n} \left[\frac{nb_n f_Y(y)}{\varphi_K \sigma_g^2(y)}\right]^{1/2} \left|\hat{\mathcal{G}}_g(y) - \mathcal{G}_g(y) - \rho_g(y) b_n^2\right| \leq B_{m_n}(z) \right\} = e^{-2e^{-z}}, \tag{12}$$

for $z \in \mathbb{R}$, where $\sigma_g^2(y)$ is defined in (11), $\rho_g(y) = [f_Y'(y)\mathcal{G}_g'(y)/f_Y(y) + \mathcal{G}_g''(y)/2]\psi_K$, and

$$B_{m_n}(z) = \sqrt{2\log m_n} - \frac{1}{\sqrt{2\log m_n}}\left[\frac{1}{2}\log\log m_n + \log(2\sqrt{\pi}) - z\right].$$

Now we discuss estimation of $\sigma_g^2(y)$ and $f_Y(y)$. To estimate $f_Y(y)$, we use the nonparametric kernel density estimator:

$$\hat{f}_Y(y) = \frac{1}{nl_n}\sum_{i=1}^{n} K_{l_n}(y - Y_i), \tag{13}$$

for a bandwidth $l_n > 0$. Based on residuals $g(Y_i) - \hat{\mathcal{G}}_g(Y_{i-1})$, we propose the Nadaraya–Watson kernel smoothing estimate of $\sigma_g^2(y)$ in (11):

$$\hat{\sigma}_g^2(y) = \frac{\sum_{i=1}^{n}[g(Y_i) - \hat{\mathcal{G}}_g(Y_{i-1})]^2 K_{h_n}(y - Y_{i-1})}{\sum_{i=1}^{n} K_{h_n}(y - Y_{i-1})}, \tag{14}$$

where $K_{h_n}(u) = K(u/h_n)$ for another bandwidth $h_n > 0$.

Proposition 2. (i) *Under Conditions 1–3 and* $l_n^4 \log n + (nl_n)^{-1}(\log n)^2 \to 0$, *we have*

$$\sup_{y\in\mathcal{Y}} |\hat{f}_Y(y) - f_Y(y)| = o_p[(\log n)^{-1/2}].$$

(ii) *Assume that Conditions 1–3 hold. Further assume* $nb_n^8(\log n)^2 + (nb_n^4)^{-1}(\log n)^6 \to 0$ *and* $h_n^4 \log n + (nh_n^2)^{-1}(\log n)^3 \to 0$. *Then we have*

$$\sup_{y\in\mathcal{Y}} |\hat{\sigma}_g^2(y) - \sigma_g^2(y)| = o_p[(\log n)^{-1/2}].$$

We point out that, from the proof of Proposition 2, the bound $o_p[(\log n)^{-1/2}]$ can be substantially improved. For brevity, we present the loose bound $o_p[(\log n)^{-1/2}]$ since it is enough for our asymptotic results.

Due to the unknown derivatives $f_Y'(y)$, $\mathcal{G}_g'(y)$ and $\mathcal{G}_g''(y)$, it is generally difficult to estimate the bias $\rho_g(y)b_n^2$ in Theorem 1. To address this issue, we adopt a bias-correction procedure so that it is not necessary to estimate the second-order bias; see Section 3 for more details. Thus, combining Theorem 1 and Proposition 2, Corollary 1 provides an asymptotic $(1-\alpha)$ SCB for $\mathcal{G}_g(y)$.

Corollary 1. *Under the conditions in Theorem 1 and Proposition 2, an asymptotic* $(1-\alpha)$ *SCB for* $\mathcal{G}_g(y)$ *on the region* $\mathcal{Y}_n$ *with the bias-correction can be constructed as*

$$\hat{\mathcal{G}}_g(y) \pm \left[\frac{\varphi_K \hat{\sigma}_g^2(y)}{nb_n\hat{f}_Y(y)}\right]^{1/2} B_{m_n}(z_\alpha), \quad y \in \mathcal{Y}_n, \tag{15}$$

where $z_\alpha = -\log\log[(1-\alpha)^{-1/2}]$ *is the* $(1-\alpha)$ *quantile of the limiting distribution in (12).*

By Definition 1, $\mathcal{Y}_n$ becomes denser and denser in $\mathcal{Y}$ as $n \to \infty$. Thus, the constructed SCB on $\mathcal{Y}_n$ provides a good approximation to (9) for sufficiently large n. For any fixed $c > 0$, let $m_n = \lfloor 2T/[c(\log n)^2 b_n]\rfloor$ and

$$\mathcal{Y}_n = \{y_j = -T + c(\log n)^2 b_n j,\ j = 0, 1, \ldots, m_n - 1\}.$$

Then $\mathcal{Y}_n$ is a (τ_n, m_n) approximation of $\mathcal{Y}$ with $\tau_n = c(\log n)^2 b_n$. It is easily seen that, under $nb_n^8(\log n)^2 + (nb_n^4)^{-1}(\log n)^6 \to 0$ in Proposition 2, the conditions $b_n = o(\tau_n)$ and $(\log n)^3[(nb_n)^{-1} + b_n^2]m_n^2 \to 0$ in Theorem 1 automatically hold.

2.2. Parametric estimate under $H_0 : \mathcal{Q} = \mathcal{Q}_\theta$

In this section we develop a general procedure to construct parametric estimate of $\mathcal{G}_g(y)$ under $H_0 : \mathcal{Q} = \mathcal{Q}_\theta$. Without further assumptions, it is generally impossible to use (8) to construct a parametric estimate of $\mathcal{G}_g(y)$. For example, Fan [9] assumed that ε_i has an exactly known density function in order to study the nonparametric de-convolution problem of estimating the density of X_i based on noisy observations Y_i from (1). Here we assume that ε_i has the normal distribution $N(0, \sigma^2)$ for some unknown variance $\sigma^2 > 0$. Denote by $\phi(z)$ the standard normal density, and write $\phi_\sigma(z) = \sigma^{-1}\phi(z/\sigma)$. Then $q_\varepsilon(z) = \phi_\sigma(z)$.

In practice, there is generally no closed-form expression for the joint density of (X_{i-1}, X_i) for most time series models, and thus it is infeasible to evaluate the expectations on the right hand side of (8) directly. For example, even for the simplest threshold autoregressive model $X_i = a_1 X_{i-1}\mathbf{1}_{X_{i-1}\le 0} + a_2 X_{i-1}\mathbf{1}_{X_{i-1}>0} + \eta_i$ with i.i.d. errors $\eta_i \sim N(0, 1)$, the stationary joint density remains unknown. To solve this issue, we propose a Monte Carlo simulation based method below.

(i) Under $H_0 : \mathcal{Q} = \mathcal{Q}_\theta$, obtain consistent estimate of (θ, σ), denoted by $(\hat{\theta}, \hat{\sigma})$. Under the parametric specification, a natural parameter estimation method is the maximum likelihood estimator, which may be computationally expensive. In some cases, it is computationally appealing to use, for example, moments based methods.
(ii) Simulate sample path $\{X_i^*\}_{0\le i\le m}$ from the estimated null model $\mathcal{Q}_{\hat{\theta}}$.
(iii) Let ε_i be i.i.d. $N(0, \hat{\sigma}^2)$ variables. Using empirical version of (8), we propose

$$\hat{\mathcal{G}}_g(y|\mathcal{Q}_{\hat{\theta}}) = \frac{m^{-1}\sum_{i=1}^{m} g(X_i^* + \varepsilon_i)\phi_{\hat{\sigma}}(y - X_{i-1}^*)}{m^{-1}\sum_{i=1}^{m} \phi_{\hat{\sigma}}(y - X_{i-1}^*)}.$$

For large m, the numerator and the denominator of $\hat{\mathcal{G}}(y|\mathcal{Q}_{\hat{\theta}})$ approach their expectations.

As an illustration, we consider (5) with $\eta_i \sim N(0, 1)$. Let $\mathcal{Q} = (\mu, s)$ and $\mathcal{Q}_\theta = (\mu_\theta, s_\theta)$ for some parametric specification (μ_θ, s_θ). Then step (ii) above is implemented using

$$X_i^* = \mu_{\hat{\theta}}(X_{i-1}^*) + s_{\hat{\theta}}(X_{i-1}^*)\eta_i, \quad \eta_i \sim N(0, 1).$$

Clearly, the above proposed procedure can be readily extended to the case of non-Gaussian errors. We simply replace $\phi_{\hat{\sigma}}$ with another given parametric density with estimated parameters and draw ε_i from the latter density.

2.3. Choices of the transformation $g(\cdot)$ and Bonferroni correction

In (7), different choices of $g(\cdot)$ can extract different information about the underlying distribution. In many practical problems, the conditional mean and conditional variance are the two most important pieces of information researchers are interested in. For example, in model (5) with i.i.d. $\eta_i \sim N(0, 1)$, the conditional mean function $\mu(\cdot)$ and the conditional variance function $s^2(\cdot)$ fully determine the model structure. Similarly, in model (6), $\mu(\cdot)$ and $s^2(\cdot)$ represent the conditional drift (mean) function and conditional volatility (variance) function, respectively, and they fully characterize the underlying model. To study the underlying model dynamics in (5) and (6), we let $\mathcal{Q} = (\mu, s)$ and $H_0 : \mathcal{Q} = \mathcal{Q}_\theta = (\mu_\theta, s_\theta)$ for parametric specifications $(\mu_\theta, \sigma_\theta)$.

Motivated by the above discussion, we propose using two simple transformations $g_1(Y_i) = Y_i$ and $g_2(Y_i) = Y_i^2$. By combining the two transformations together, the test can detect deviations from the conditional mean and/or conditional variance in the underlying model. To combine the two corresponding tests together, we adopt the following procedure:

(**Bonferroni correction**): *Suppose the pre-specified significance level is α, then we construct $(1 - \alpha/2)$ SCBs, denoted by* SCB$_1$ *and* SCB$_2$, *for $\mathcal{G}_{g_1}(y)$ and $\mathcal{G}_{g_2}(y)$ separately, and reject H_0 if either* SCB$_1$ *or* SCB$_2$ *cannot cover the corresponding parametric estimates for $\mathcal{G}_{g_1}(y)$ or $\mathcal{G}_{g_2}(y)$.*

Theoretically speaking, we can combine tests across multiple transformations. For example, one natural choice is to combine multiple conditional moments, i.e., $g_k(Y_i) = Y_i^k,\ k = 1, \ldots, J$, for some $J \in \mathbb{N}$. However, we do not recommend this approach based on three considerations. First, the Bonferroni correction is well-known to be very conservative for multiple tests. Second, as discussed above, for the two most popular Markov models (5) and (6), the conditional mean and the conditional variance contain all important information and higher-order moments do not provide extra information. Third, using high-order moments would require high-order finite-moment assumptions, which may be too restrictive in practice.

2.4. Alternative approaches: conditional distribution or conditional characteristic function

Some alternative approaches are to use a class of transformations $g(\cdot)$ indexed by a continuous parameter. For example, $g_t(Y_i) = \mathbf{1}_{Y_i \le t}$ for $t \in \mathbb{R}$ corresponds to the conditional distribution function, and $g_t(Y_i) = \exp(\sqrt{-1}Y_i t)$ for $t \in \mathbb{R}$ corresponds to the conditional characteristic function. Under different contexts, Hong [18] and Pinkse [25] used empirical characteristic functions to test for serial dependence. In our SCB setting, using such choices of transformations involves studying maximum deviations of $\hat{\mathcal{G}}_{g_t}(y)$ over both $t \in \mathbb{R}$ and $y \in \mathbb{R}$. With $g_t(Y_i) = \mathbf{1}_{Y_i \le t}$, we expect that, after proper normalization, the process $\{\hat{\mathcal{G}}_{g_t}(y)\}_{t\in\mathbb{R}}$ with any fixed y converges in distribution in the Skorokhod space $D[-\infty, +\infty]$ to the process $\{\mathbb{B}(Q_Y(t|y))\}_{t\in\mathbb{R}}$, where $\mathbb{B}$ is the standard Brownian bridge and $Q_Y(t|y)$ is the conditional distribution function of Y_i given $Y_{i-1} = y$. Therefore, by the continuous mapping theorem, we can handle the supremum over $t \in \mathbb{R}$. Unfortunately, it is unclear how to deal with the supremum over $y \in \mathbb{R}$. Furthermore, in order to establish the latter functional convergence, we need to prove the tightness of the process. It seems that substantial theoretical developments are necessary and we leave them for future research.

3. Monte Carlo simulation study

In this section, we conduct a small simulation study to examine the empirical performance of the proposed specification test. First, we address some practical implementation issues.

(**Bias-correction**): We adopt a higher-order kernel to remove the second-order bias term $\rho_g(y)b_n^2$ in Theorem 1. Let $\phi(u)$ be the standard normal density function. In our numerical analysis, we use the kernel function $K(u) = 2\phi(u) - \phi(u/\sqrt{2})/\sqrt{2}$, which is symmetric and satisfies $\int_{\mathbb{R}} K(u)du = 1$ and $\psi_K = \int_{\mathbb{R}} u^2 K(u)du = 0$ so that $\rho_g(y) = 0$.

Remark 2. The higher-order kernel above can remove the second order bias, but the fourth order bias is still present. The bias issue is an intrinsic feature of any nonparametric regression methods, and there seems to be no satisfactory approach to address it. Our small simulation study shows that the above simple approach works reasonably well.

(**Bandwidth selection**): To select the bandwidth b_n in (10), we use the leave-one-out cross validation with the criterion of minimum mean squared error in [23, Chapter 2.2.2], which is implemented using the command `localpoly.reg` in R. To choose h_n in (14), first we obtain the residuals $g(Y_i) - \hat{\mathcal{G}}_g(Y_{i-1})$ using the selected bandwidth b_n. Then, the leave-one-out cross validation is employed again based on the nonparametric regression of $[g(Y_i) - \hat{\mathcal{G}}_g(Y_{i-1})]^2$ on Y_{i-1}. For l_n in (13), we use the rule-of-thumb nonparametric kernel density bandwidth selector in [26], which is implemented using the command `bw.nrd0` in R.

We compare the empirical performance of the proposed specification tests based on different transformations $g(\cdot)$:

(i) Test 1: using the single transformation $g_1(Y_i) = Y_i$;
(ii) Test 2: using the single transformation $g_2(Y_i) = Y_i^2$;
(iii) Test 3: combining the two transformations g_1 and g_2 with the Bonferroni correction.

Table 1
Empirical power: Test 1, Test 2, and Test 3 stand for the proposed specification tests based on SCB with $g_1(Y_i) = Y_i$, $g_2(Y_i) = Y_i^2$, and combining the two transformations together with the Bonferroni correction, respectively. Significance level is 5%.

	n		λ					
			0.0	0.2	0.4	0.6	0.8	1.0
Model 1	200	Test 1	0.039	0.090	0.278	0.525	0.695	0.838
		Test 2	0.038	0.021	0.038	0.054	0.054	0.068
		Test 3	0.038	0.042	0.181	0.358	0.558	0.738
	500	Test 1	0.029	0.159	0.532	0.866	0.965	0.997
		Test 2	0.052	0.037	0.043	0.075	0.113	0.090
		Test 3	0.037	0.098	0.379	0.752	0.921	0.985
	2000	Test 1	0.063	0.637	0.989	1.000	1.000	1.000
		Test 2	0.018	0.086	0.255	0.358	0.428	0.532
		Test 3	0.042	0.503	0.974	1.000	1.000	1.000
Model 2	200	Test 1	0.039	0.047	0.041	0.053	0.083	0.106
		Test 2	0.038	0.033	0.047	0.099	0.150	0.210
		Test 3	0.038	0.042	0.045	0.082	0.119	0.186
	500	Test 1	0.029	0.024	0.030	0.056	0.115	0.218
		Test 2	0.052	0.042	0.049	0.103	0.245	0.429
		Test 3	0.037	0.036	0.043	0.082	0.204	0.386
	2000	Test 1	0.063	0.085	0.110	0.152	0.235	0.321
		Test 2	0.018	0.049	0.170	0.447	0.784	0.946
		Test 3	0.042	0.069	0.147	0.367	0.716	0.917

In (1), we generate $\{X_i\}$ from the following true models:

$$\text{(Model 1) } X_i = 0.6[(1-\lambda)X_{i-1} + \lambda|X_{i-1}|] + \eta_i, \quad \lambda \in [0, 1],$$

$$\text{(Model 2) } X_i = 0.6X_{i-1} + \eta_i\sqrt{1 + 0.3\lambda X_{i-1}^2}, \quad \lambda \in [0, 1],$$

for i.i.d. noises $\eta_i \sim N(0, 1)$. We wish to test the null hypothesis $H_0 : X_i = \theta_1 X_{i-1} + \eta_i$ based on contaminated observations $\{Y_i\}$ from (1) with i.i.d. measurement errors $\varepsilon_i \sim N(0, 1)$. The parameter λ regulates the deviation from the null model. The case $\lambda = 0$ leads to the null model; as λ increases, Model 1 and Model 2 move further away from the null model. In particular, for $\lambda \neq 0$, Model 1 becomes the threshold autoregressive model, and Model 2 becomes the autoregressive conditional heteroscedastic model with a linear term. Model 1 and Model 2 are used to examine the sensitivity of the test to deviations in the conditional mean function and conditional variance function, respectively.

Under H_0, we need to estimate θ_1 and the variances of η_i and ε_i, denoted by θ_2^2 and θ_3^2, respectively. Then elementary calculations show that

$$\mathbb{E}(Y_i^2) = \frac{\theta_2^2}{1-\theta_1^2} + \theta_3^2, \qquad \mathrm{cov}(Y_{i-1}, Y_i) = \frac{\theta_1\theta_2^2}{1-\theta_1^2}, \qquad \mathrm{cov}(Y_{i-2}, Y_i) = \frac{\theta_1^2\theta_2^2}{1-\theta_1^2}.$$

Thus, we can estimate the parameters $(\theta_1, \theta_2, \theta_3)$ by the empirical versions of moments. We simulate 1000 realizations with sample size $n = 200, 500, 2000$ and significance level $\alpha = 0.05$. In (15), we need to select a set $\mathcal{Y}_n$ of grid points. For a realization $\{Y_i\}$, let $l_{0.15}$ and $l_{0.85}$ be their 15 and 85 percentiles, respectively. We take $\mathcal{Y}_n$ to be the set of 11 evenly spaced grid points $y_i = l_{0.15} + i(l_{0.85} - l_{0.15})/10$, $i = 0, \ldots, 10$ for $n = 2000$, and similarly we use the set of 5 and 7 evenly spaced grid points for $n = 200$ and $n = 500$, respectively.

The result is presented in Table 1. We see that, Test 1 based on the transformation g_1 is much more powerful in detecting deviations in the conditional mean (Model 1) than in detecting deviations in the conditional variance (Model 2). Similarly, Test 2 based on g_2 is more powerful in detecting deviations in the conditional variance. By contrast, Test 3 based on combining g_1 and g_2 through the Bonferroni correction can detect both deviations well. Moreover, the empirical size of Test 3 is quite close to the nominal size 0.05, and the power rises dramatically as the deviation parameter λ increases. This small simulation study demonstrates that the proposed test using the two transformations g_1 and g_2 with the Bonferroni correction works quite well.

4. Proofs

4.1. Proof of Proposition 1

Proof of Proposition 1. Recall the notations $p_X, q_X, p_Y, q_Y, f_Y, q_\varepsilon$ at the beginning of Section 2. Conditioning on (X_{i-1}, X_i), Y_{i-1} and Y_i are conditionally independent and have density $q_\varepsilon(y - X_{i-1})$ and $q_\varepsilon(y' - X_i)$, respectively. Thus, $p_Y(y, y') = \mathbb{E}[q_\varepsilon(y - X_{i-1})q_\varepsilon(y' - X_i)]$. By the latter identity and the independence between ε and (X_{i-1}, X_i),

$$\int_{\mathbb{R}} g(y')p_Y(y, y')dy' \quad = \quad \mathbb{E}\Big[q_\varepsilon(y - X_{i-1}) \int_{\mathbb{R}} g(y')q_\varepsilon(y' - X_i)dy'\Big]$$

$$\begin{aligned} &\overset{z=y'-X_i}{=} \mathbb{E}\Big[q_\varepsilon(y-X_{i-1})\int_{\mathbb{R}} g(X_i+z)q_\varepsilon(z)dz\Big] \\ &= \mathbb{E}\{q_\varepsilon(y-X_{i-1})\mathbb{E}[g(X_i+\varepsilon)|X_{i-1},X_i]\} \\ &= \mathbb{E}[g(X_i+\varepsilon)q_\varepsilon(y-X_{i-1})]. \end{aligned} \tag{16}$$

By conditioning on X_{i-1}, we can show $f_Y(y) = \mathbb{E}[q_\varepsilon(y-X_{i-1})]$. Therefore, the desired result follows from (16) and $\mathcal{G}_g(y) = \int_{\mathbb{R}} g(y')q_Y(y'|y)dy' = \int_{\mathbb{R}} g(y')p_Y(y,y')dy'/f_Y(y)$. $\diamondsuit$

4.2. Some preliminary facts of projection operator

For convenience, we recall some basic properties of conditional expectations. Let $Z \in \mathcal{L}^1$ be any integrable random variable and $\mathcal{F}$ a σ-algebra on the same probability space. Then

(C1) $\mathbb{E}(Z) = \mathbb{E}[\mathbb{E}(Z|\mathcal{F})]$.
(C2) Let $\mathcal{G}$ be another σ-algebra such that $\mathcal{F} \subset \mathcal{G}$. Then $\mathbb{E}(Z|\mathcal{F}) = \mathbb{E}[\mathbb{E}(Z|\mathcal{G})|\mathcal{F}]$.
(C3) If $Z \in \mathcal{L}^p$ for some $p \ge 1$, then $\|\mathbb{E}(|Z||\mathcal{F})\|_p \le \|Z\|_p$ and $(\mathbb{E}|Z|)^p \le \mathbb{E}(|Z|^p)$.

Recall that, in (1), $\{\varepsilon_i\}$ are i.i.d. and independent of the unobservable Markov chain $\{X_i\}$. Let $\mathcal{F}_i = \sigma(\varepsilon_j, X_{j+1} : j \le i)$ be the σ-algebra generated by $\varepsilon_j, X_{j+1}, j \le i$. Then $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$ is an increasing filtration. For $i \in \mathbb{Z}$, define the projection operator $\mathcal{P}_i$ by

$$\mathcal{P}_i Z = \mathbb{E}(Z|\mathcal{F}_i) - \mathbb{E}(Z|\mathcal{F}_{i-1}), \quad Z \in \mathcal{L}^1.$$

The projection operator $\mathcal{P}_i$ satisfies the following properties (in the statements below, $\{Z_i\}_{i\in\mathbb{Z}}$ is any sequence of random variables):

(C4) For any $\{Z_i \in \mathcal{L}^1\}_{i\in\mathbb{Z}}$, $\{\mathcal{P}_i Z_i\}_{i\in\mathbb{Z}}$ are martingale differences with respect to the increasing filtration $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$. Thus, $\sum_{i=1}^n \mathcal{P}_i Z_i$ is a martingale with respect to $\mathcal{F}_n$.
(C5) For any $\{Z_i \in \mathcal{L}^2\}_{i\in\mathbb{Z}}$, $\|\sum_{i=1}^n \mathcal{P}_i Z_i\|_2^2 = \sum_{i=1}^n \|\mathcal{P}_i Z_i\|_2^2$.
(C6) For any $Z \in \mathcal{L}^2$, $\|\mathcal{P}_i Z\|_2^2 \le \|\mathbb{E}(Z|\mathcal{F}_i)\|_2^2 \le \|Z\|_2^2$.
(C7) For any $Z \in \mathcal{L}^2$, $\mathbb{E}[(\mathcal{P}_i Z)^2|\mathcal{F}_{i-1}] = \mathbb{E}\{[\mathbb{E}(Z|\mathcal{F}_i)]^2|\mathcal{F}_{i-1}\} - [\mathbb{E}(Z|\mathcal{F}_{i-1})]^2$.
(C8) For any $Z \in \mathcal{L}^2$, $\mathbb{E}[(\mathcal{P}_i Z)^2|\mathcal{F}_{i-1}] \le \mathbb{E}(Z^2|\mathcal{F}_{i-1})$.

Proof. By definition, $\mathcal{P}_i Z_i$ is $\mathcal{F}_i$-measurable. Furthermore, by property (C2), $\mathbb{E}(\mathcal{P}_i Z_i|\mathcal{F}_{i-1}) = \mathbb{E}[\mathbb{E}(Z_i|\mathcal{F}_i)|\mathcal{F}_{i-1}] - \mathbb{E}(Z_i|\mathcal{F}_{i-1}) = 0$. Thus, (C4) holds. By (C4), (C5) follows from the orthogonality of martingale differences. To see (C6), let $Z^* = \mathbb{E}(Z|\mathcal{F}_i)\,\mathbb{E}(Z|\mathcal{F}_{i-1})$, by property (C2), $\mathbb{E}(Z^*|\mathcal{F}_{i-1}) = [\mathbb{E}(Z|\mathcal{F}_{i-1})]^2$. Thus, $\mathbb{E}(Z^*) = \mathbb{E}[\mathbb{E}(Z^*|\mathcal{F}_{i-1})] = \mathbb{E}\{[\mathbb{E}(Z|\mathcal{F}_{i-1})]^2\}$. Using the latter identity, we can show $\|\mathcal{P}_i Z\|_2^2 = \mathbb{E}\{[\mathbb{E}(Z|\mathcal{F}_i)]^2\} - \mathbb{E}\{[\mathbb{E}(Z|\mathcal{F}_{i-1})]^2\} \le \|\mathbb{E}(Z|\mathcal{F}_i)\|_2^2 \le \|Z\|_2^2$, where the last inequality follows from property (C3). By simple calculations, (C7) follows from the definition of $\mathcal{P}_i$ and property (C2). Finally, by (C7), (C8) follows from $\mathbb{E}[(\mathcal{P}_i Z)^2|\mathcal{F}_{i-1}] \le \mathbb{E}\{[\mathbb{E}(Z|\mathcal{F}_i)]^2|\mathcal{F}_{i-1}\} \le \mathbb{E}\{\mathbb{E}(Z^2|\mathcal{F}_i)|\mathcal{F}_{i-1}\} = \mathbb{E}(Z^2|\mathcal{F}_{i-1})$. $\diamondsuit$

By (C4), $\{\mathcal{P}_i\}_{i\in\mathbb{Z}}$ are martingale difference operators with respect to $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$. This idea of martingale construction serves as the building block for our technical arguments. See Wu [28] for more discussions.

4.3. Some preliminary results on mixing processes

In Condition 3, we impose α-mixing conditions on $\{X_i\}$. Lemmas 1–2 present some useful results for α-mixing processes.

Lemma 1 (*Proposition 2.5 in [11]*). *Let U and V be two random variables such that $U \in \mathcal{L}^p$ and $V \in \mathcal{L}^q$ for some $p > 1, q > 1$, and $1/p + 1/q < 1$. Then*

$$|\mathrm{cov}(U,V)| \le 8\alpha(U,V)^{1-1/p-1/q}\|U\|_p\|V\|_q.$$

Here $\alpha(U,V)$ is the α-mixing coefficient between the two σ-algebras generated by U and V.

Next, we present an important inequality regarding the supremum of any differentiable function $f(\cdot)$ on a given bounded interval $[a,b]$. Note that $|f(y)| = |f(a) + \int_a^y f'(z)dz| \le |f(a)| + \int_a^b |f'(z)|dz$ for all $y \in [a,b]$. Thus, using $(u+v)^2 \le 2(u^2+v^2)$ and the Cauchy–Schwarz inequality $[\int_a^b |f'(z)|dz]^2 \le (b-a)\int_a^b |f'(z)|^2dz$, we have the uniform bound:

$$\sup_{y\in[a,b]} |f(y)|^2 \le 2\left\{|f(a)|^2 + \left[\int_a^b |f'(z)|dz\right]^2\right\} \le 2\left\{|f(a)|^2 + (b-a)\int_a^b |f'(z)|^2dz\right\}. \tag{17}$$

Clearly, if $f(\cdot)$ is a random function, then taking expectation on both sides of (17) gives

$$\begin{aligned}\left\| \sup_{y\in[a,b]} |f(y)| \right\|_2^2 &\le 2\left\{ \|f(a)\|_2^2 + (b-a)\int_a^b \|f'(z)\|_2^2 dz \right\} \\ &\le 2\left\{ \|f(a)\|_2^2 + (b-a)^2 \sup_{y\in[a,b]} \|f'(y)\|_2^2 \right\}. \end{aligned} \tag{18}$$

In (18), while it is generally difficult to study the left hand side with "sup" inside $\|\cdot\|_2$, it is much easier to handle the right hand side with "sup" outside $\|\cdot\|_2$. Thus, (18) provides a useful inequality in bounding the supremum of random processes indexed by a continuous parameter. In particular, we can obtain the following useful result:

Lemma 2. *Let $\{X_i\}_{i\in\mathbb{N}}$ be an α-mixing stationary process with mixing coefficient α_k, $k \in \mathbb{N}$. For a bivariate measurable and differentiable function h, define*

$$H(y) = \sum_{i=1}^n \Big\{ h(y, X_i) - \mathbb{E}[h(y, X_i)] \Big\}.$$

Suppose there exists some $\delta > 2$ such that $c := \sup_{y\in\mathbb{R}}[\|h(y, X_1)\|_\delta + \|\partial h(y, X_1)/\partial y\|_\delta] < \infty$ and $\sum_{k=1}^\infty \alpha_k^{1-2/\delta} < \infty$. Let $[a, b]$ be any given bounded interval. Then

$$\mathbb{E}\left[\sup_{y\in[a,b]} |H(y)|^2 \right] = O(n). \tag{19}$$

Furthermore, if $b_n \to 0$ and $w(\cdot)$ is an integrable function with bounded support, then

$$\mathbb{E}\left[\sup_{y\in[a,b]} \left| \int_{\mathbb{R}} w(u) H(y - ub_n) du \right| \right]^2 = O(n). \tag{20}$$

Proof. Let $\gamma_k = \operatorname{cov}\{h(y, X_1), h(y, X_{k+1})\}$. Then $\gamma_0 \le \|h(y, X_1)\|_2^2 \le \|h(y, X_1)\|_\delta^2 \le c^2$. For $k \ge 1$, by Lemma 1, $|\gamma_k| \le 8\alpha_k^{1-2/\delta}\|h(y, X_1)\|_\delta^2 \le 8\alpha_k^{1-2/\delta} c^2$. Thus,

$$\|H(y)\|_2^2 = n\gamma_0 + 2\sum_{k=1}^{n-1}(n-k)\gamma_k \le n\left(\gamma_0 + 2\sum_{k=1}^n |\gamma_k| \right) \le nc^2\left(1 + 16\sum_{k=1}^\infty \alpha_k^{1-2/\delta} \right). \tag{21}$$

Similarly, using $H'(y) = \sum_{i=1}^n \{\partial h(y, X_i)/\partial y - \mathbb{E}[\partial h(y, X_i)/\partial y]\}$, we have

$$\|H'(y)\|_2^2 \le nc^2\left(1 + 16\sum_{k=1}^\infty \alpha_k^{1-2/\delta} \right). \tag{22}$$

The assertion (19) then follows by applying (21) and (22) to (18). To prove (20), by the bounded support of $w(\cdot)$ and $b_n \to 0$, for $y \in [a, b]$, we have $y - ub_n \in [a-1, b+1]$ for sufficiently large n. Thus,

$$\sup_{y\in[a,b]} \left| \int_{\mathbb{R}} w(u) H(y - ub_n) du \right| \le \sup_{z\in[a-1,b+1]} |H(z)| \int_{\mathbb{R}} |w(u)| du. \tag{23}$$

Taking square first and then taking expectation in (23), we can obtain (20) from (19). ◊

4.4. Proof of Theorem 1

Throughout our proofs, $c, c_1, c_2, \ldots$ are constants that may vary from places to places.

Proof of Theorem 1. Recall $\hat{\mathcal{G}}_g(y)$ in (10). Define

$$\tilde{f}_Y(y) = \frac{1}{nb_n}\sum_{i=1}^n K_{b_n}(y - Y_{i-1}), \tag{24}$$

$$\xi_i(y) = [g(Y_i) - \mathcal{G}_g(Y_{i-1})] K_{b_n}(y - Y_{i-1}). \tag{25}$$

By the definition of $\mathcal{G}_g$, we have $\mathbb{E}[\xi_i(y)] = \mathbb{E}\{\mathbb{E}[\xi_i(y)|Y_{i-1}]\} = 0$. Therefore, we can write

$$\hat{\mathcal{G}}_g(y) - \mathcal{G}_g(y) = \frac{\sum_{i=1}^{n}\{\xi_i(y) - \mathbb{E}[\xi_i(y)]\}}{nb_n\tilde{f}_Y(y)} + \frac{\sum_{i=1}^{n}[\mathcal{G}_g(Y_{i-1}) - \mathcal{G}_g(y)]K_{b_n}(y - Y_{i-1})}{nb_n\tilde{f}_Y(y)}$$
$$:= T_n(y) + U_n(y). \tag{26}$$

In (26), $T_n(y)$ is the stochastic component determining the asymptotic distribution of $\hat{\mathcal{G}}_g(y)$, and $U_n(y)$ is the bias component. By Lemma 4, $\tilde{f}_Y(y) = f_Y(y) + O[b_n^2 + (nb_n/\log n)^{-1/2}]$ uniformly in $y \in \mathcal{Y}$. Furthermore, by Lemma 5, $U_n(y) = \rho_g(y)b_n^2 + O_p[b_n^4 + (b_n \log n/n)^{1/2}] = \rho_g(y)b_n^2 + o_p[(nb_n \log n)^{-1/2}]$ uniformly in $y \in \mathcal{Y}$. By Slutsky's theorem, it suffices to establish a maximal deviation result for $\sum_{i=1}^{n}\{\xi_i(y) - \mathbb{E}[\xi_i(y)]\}$.

We use the projection operator $\mathcal{P}_i$ in Section 4.2 to write the decomposition

$$\begin{aligned}\sum_{i=1}^{n}\{\xi_i(y) - \mathbb{E}[\xi_i(y)]\} &= \sum_{i=1}^{n}\{\mathcal{P}_i\xi_i(y) + \mathcal{P}_{i-1}\xi_i(y) + \mathbb{E}[\xi_i(y)|\mathcal{F}_{i-2}] - \mathbb{E}[\xi_i(y)]\} \\ &= \sum_{i=1}^{n}[\mathcal{P}_i\xi_i(y) + \mathcal{P}_i\xi_{i+1}(y)] + \sum_{i=1}^{n}\{\mathbb{E}[\xi_i(y)|\mathcal{F}_{i-2}] - \mathbb{E}[\xi_i(y)]\} + [\mathcal{P}_0\xi_1(y) - \mathcal{P}_n\xi_{n+1}(y)] \\ &:= S_n(y) + R_n(y) + M_n(y).\end{aligned} \tag{27}$$

The decomposition (27) provides a convenient tool to study asymptotic properties. First, by property (C4) in Section 4.2, $S_n(y)$ is a martingale with respect to $\mathcal{F}_n$. To study asymptotic properties of $S_n(y)$, Lemma 6 studies its conditional variance. In Lemma 7, we use the obtained result to study the quadratic characteristic matrix of the multivariate martingale $[S_n(y_1), \dots, S_n(y_k)]^T$ for distinct $y_1, \dots, y_k$. Second, by Lemma 3, $\sup_{y\in\mathcal{Y}} |R_n(y)| = O_p(b_n\sqrt{n})$. Finally, it is easy to observe that $\sup_y |M_n(y)| = O_p(1)$. To see this, by the boundedness of $K(\cdot)$,

$$\sup_y |\mathcal{P}_0\xi_1(y)| \le \sup_u |K(u)| \Big\{\mathbb{E}[|g(Y_1) - \mathcal{G}_g(Y_0)||\mathcal{F}_0] + \mathbb{E}[|g(Y_1) - \mathcal{G}_g(Y_0)||\mathcal{F}_{-1}]\Big\}.$$

Thus, by property (C1) in Section 4.2,

$$\mathbb{E}\Big[\sup_y |\mathcal{P}_0\xi_1(y)|\Big] \le 2\sup_u |K(u)|[\mathbb{E}|g(Y_1)| + \mathbb{E}|\mathcal{G}_g(Y_0)|] = O(1),$$

where we have $\mathbb{E}|\mathcal{G}_g(Y_0)| = \mathbb{E}|\mathbb{E}[g(Y_1)|Y_0]| \le \mathbb{E}\{\mathbb{E}[|g(Y_1)||Y_0]\} = \mathbb{E}|g(Y_1)| = O(1)$. Similarly, $\mathbb{E}[\sup_y |\mathcal{P}_n\xi_{n+1}(y)|] = O(1)$. This proves $\sup_y |M_n(y)| = O_p(1)$. Finally, the desired result then follows from the maximal deviation of $S_n(y)$ in Lemma 7. ♢

Lemma 3. *For $R_n(y)$ in (27), we have $\sup_{y\in\mathcal{Y}} |R_n(y)| = O_p(b_n\sqrt{n})$.*

Proof. Recall $\xi_i(y)$ in (25). Write $\xi_i(y) = \xi_{i,1}(y) - \xi_{i,2}(y)$, where

$$\xi_{i,1}(y) = g(Y_i)K_{b_n}(y - Y_{i-1}) \quad \text{and} \quad \xi_{i,2}(y) = \mathcal{G}_g(Y_{i-1})K_{b_n}(y - Y_{i-1}). \tag{28}$$

Then it suffices to prove $\sup_{y\in\mathcal{Y}} |J_r(y)| = O_p(b_n\sqrt{n})$, where

$$J_r(y) = \sum_{i=1}^{n}\Big\{\mathbb{E}[\xi_{i,r}(y)|\mathcal{F}_{i-2}] - \mathbb{E}[\xi_{i,r}(y)]\Big\}, \quad r = 1, 2.$$

First, we consider $J_1(y)$. Since $\{\varepsilon_i\}_{i\in\mathbb{N}}$ are i.i.d. and independent of $\{X_i\}_{i\in\mathbb{N}}$, by writing $Y_{i-1} = X_{i-1} + \varepsilon_{i-1}$, we have

$$\begin{aligned}\mathbb{E}[\xi_{i,1}(y)|\mathcal{F}_{i-2}, X_i] &= \mathbb{E}[g(Y_i)K_{b_n}(y - X_{i-1} - \varepsilon_{i-1})|X_{i-1}, X_i] \\ &= \mathbb{E}\Big[g(Y_i)\int_{\mathbb{R}} K_{b_n}(y - X_{i-1} - v)q_\varepsilon(v)dv|X_{i-1}, X_i\Big] \\ &= b_n\int_{-\omega}^{\omega} K(u)e_i(u)du, \quad e_i(u) = \mathbb{E}[g(Y_i)q_\varepsilon(y - X_{i-1} - ub_n)|X_{i-1}, X_i].\end{aligned}$$

Here, the last equality follows from the transformation $u = (y - X_{i-1} - v)/b_n$. Note that the conditional expectation $e_i(u)$ is a function of X_{i-1}, X_i. Thus,

$$\begin{aligned}\mathbb{E}[\xi_{i,1}(y)|\mathcal{F}_{i-2}] &= \mathbb{E}\{\mathbb{E}[\xi_{i,1}(y)|\mathcal{F}_{i-2}, X_i]|\mathcal{F}_{i-2}\} \\ &= b_n \int_{-\omega}^{\omega} K(u)\mathbb{E}[e_i(u)|X_{i-1}]du \\ &= b_n \int_{-\omega}^{\omega} K(u)\mathbb{E}[g(Y_i)q_\varepsilon(y - X_{i-1} - ub_n)|X_{i-1}]du, \end{aligned} \tag{29}$$

where the first equality follows from property (C2) in Section 4.2, the second equality follows from the independence between $e_i(u)$ (which is a function of X_{i-1}, X_i) and $\{\varepsilon_i\}_{i\in\mathbb{N}}$ as well as the Markovian assumption on $\{X_i\}_{i\in\mathbb{N}}$, and the third equality follows from $\mathbb{E}[e_i(u)|X_{i-1}] = \mathbb{E}[g(Y_i)q_\varepsilon(y - X_{i-1} - ub_n)|X_{i-1}]$ (property (C2) in Section 4.2). Define $h(z, X_{i-1}) = \mathbb{E}[g(Y_i)q_\varepsilon(z - X_{i-1})|X_{i-1}]$. (After taking conditional expectation, it is a function of X_{i-1}.) Then, using $\mathbb{E}[\xi_{i,1}(y)] = \mathbb{E}\{\mathbb{E}[\xi_{i,1}(y)|\mathcal{F}_{i-2}]\}$ and by (29), we obtain

$$J_1(y) = b_n \int_{-\omega}^{\omega} K(u)H(y - ub_n)du, \quad \text{where } H(z) = \sum_{i=1}^{n}[h(z, X_{i-1}) - \mathbb{E}h(z, X_{i-1})].$$

Since $q_\varepsilon(\cdot)$ is bounded, by property (C3) in Section 4.2, we can easily see that $\|h(z, X_0)\|_\delta = O(1)\|\mathbb{E}[|g(Y_1)||X_0]\|_\delta \le O(1)\|g(Y_1)\|_\delta$ for all $z \in \mathbb{R}$. Similarly, using $\partial h(z, X_0)/\partial z = \mathbb{E}[g(Y_1)q'_\varepsilon(z - X_0)|X_0]$ and the boundedness of $q'_\varepsilon(\cdot)$, we have $\|\partial h(z, X_0)/\partial z\|_\delta \le O(1)\|g(Y_1)\|_\delta$ for all $z \in \mathbb{R}$. Thus, by (20) in Lemma 2, we conclude that $\sup_{y\in\mathcal{Y}} |J_1(y)| = O_p(b_n\sqrt{n})$.

Next, we consider $J_2(y)$. Using $Y_{i-1} = X_{i-1} + \varepsilon_{i-1}$, we obtain

$$\begin{aligned}\mathbb{E}[\xi_{i,2}(y)|\mathcal{F}_{i-2}] &= \int_{\mathbb{R}} \mathcal{G}_g(X_{i-1} + v)K_{b_n}(y - X_{i-1} - v)q_\varepsilon(v)dv \\ &= b_n \int_{-\omega}^{\omega} K(u)\mathcal{G}_g(y - ub_n)q_\varepsilon(y - X_{i-1} - ub_n)du. \end{aligned} \tag{30}$$

Thus, using $\mathbb{E}[\xi_{i,2}(y)] = \mathbb{E}\{\mathbb{E}[\xi_{i,2}(y)|\mathcal{F}_{i-2}]\}$, we have

$$J_2(y) = b_n \int_{-\omega}^{\omega} K(u)\mathcal{G}_g(y - ub_n)L(y - ub_n)du, \tag{31}$$

where $L(z) = \sum_{i=0}^{n-1}\{q_\varepsilon(z - X_i) - \mathbb{E}[q_\varepsilon(z - X_i)]\}$. Since $\mathcal{G}_g(\cdot)$ is bounded in the neighborhood $\mathcal{Y}_\epsilon$, the claim then follows from (20) in Lemma 2. ♢

Lemma 4. *For* $\tilde{f}_Y(y)$ *in* (24)*, we have* $\sup_{y\in\mathcal{Y}} |\tilde{f}_Y(y) - f_Y(y)| = O_p[b_n^2 + (nb_n/\log n)^{-1/2}]$.

Proof. Let $\gamma_i(y) = K_{b_n}(y - Y_i)$. Observe the decomposition

$$\tilde{f}_Y(y) = \frac{1}{nb_n}H_1(y) + \frac{1}{nb_n}H_2(y) + \frac{\mathbb{E}[\gamma_1(y)]}{b_n}, \tag{32}$$

where

$$H_1(y) = \sum_{i=0}^{n-1}\{\gamma_i(y) - E[\gamma_i(y)|\mathcal{F}_{i-1}]\} = \sum_{i=0}^{n-1}\mathcal{P}_i\gamma_i(y) \tag{33}$$

$$H_2(y) = \sum_{i=0}^{n-1}\{\mathbb{E}[\gamma_i(y)|\mathcal{F}_{i-1}] - \mathbb{E}[\gamma_i(y)]\}. \tag{34}$$

By the symmetry of $K(\cdot)$, we can show $b_n^{-1}\mathbb{E}[\gamma_1(y)] = f_Y(y) + O(b_n^2)$. To prove the desired result, it suffices to show $\sup_{y\in\mathcal{Y}} |H_1(y)| = O_p(\sqrt{nb_n \log n})$ and $\sup_{y\in\mathcal{Y}} |H_2(y)| = O_p(b_n\sqrt{n})$.

First, we consider $H_2(y)$. By the same argument in (30)–(31), $H_2(y) = b_n \int_{-\omega}^{\omega} K(u)L(y - ub_n)du$ with $L(\cdot)$ defined in (31). Thus, by (20) in Lemma 2, $\sup_{y\in\mathcal{Y}} |H_2(y)| = O_p(b_n\sqrt{n})$.

Next, we consider the martingale part $H_1(y)$. We shall adopt a chain argument to approximate $H_1(y)$, $y \in \mathcal{Y}$, on increasingly denser grid points. Let $N = n^2$ and $y_j = jT/N$, $j = -N, 1 - N, \dots, N - 1, N$. Then $y_{-N}, \dots, y_N$ partition $\mathcal{Y} = [-T, T]$ into $2N$ equally spaced intervals with length T/N. By the bounded derivative of $K(\cdot)$, there exists some constant c_1 such that, for all $y \in [y_j, y_{j+1}]$,

$$|\gamma_i(y) - \gamma_i(y_j)| + |\mathbb{E}[\gamma_i(y)|\mathcal{F}_{i-1}] - \mathbb{E}[\gamma_i(y_j)|\mathcal{F}_{i-1}]| \le c_1|y - y_j|/b_n \le c_1T/(Nb_n).$$

Thus, $\sup_{y\in[y_j,y_{j+1}]} |H_1(y) - H_1(y_j)| \le nc_1T/(Nb_n) = O[(nb_n)^{-1}]$, and consequently,

$$\sup_{y\in\mathcal{Y}} |H_1(y)| \le \max_{j=-N,\dots,N} |H_1(y_j)| + O[(nb_n)^{-1}]. \tag{35}$$

By property (C8) in Section 4.2, $\sum_{i=0}^{n-1}\mathbb{E}\{[\mathcal{P}_i\gamma_i(y_j)]^2|\mathcal{F}_{i-1}\} \le \sum_{i=0}^{n-1}\mathbb{E}[\gamma_i^2(y_j)|\mathcal{F}_{i-1}] \le c_2nb_n$ for some constant c_2. Let $c_3 = \sup_u |K(u)|$. Then $|\mathcal{P}_i\gamma_i(y_j)| \le 2c_3$. Thus, by Freedman's exponential inequality for bounded martingale differences [14], for any $c > 0$,

$$p_j := \mathbb{P}\left\{|H_1(y_j)| \ge c\sqrt{nb_n\log n}\right\} \le 2\exp\left[-\frac{c^2nb_n\log n}{2(2c_3c\sqrt{nb_n\log n}+c_2nb_n)}\right] = 2\exp(-\lambda_n\log n), \quad (36)$$

where

$$\lambda_n = \frac{c^2}{4c_3c\sqrt{\frac{\log n}{nb_n}}+2c_2}.$$

Since $(nb_n^3)^{-1}\log n \to 0$, $(nb_n)^{-1}\log n < 1$ for sufficiently large n. Thus, $\lambda_n > c^2/(4c_3c+2c_2) \ge 3$ by choosing a large enough c (for example, we may take $c = 12c_3+\sqrt{6c_2}$). Then

$$\mathbb{P}\left\{\max_{j=-N,\dots,N}|H_1(y_j)| \ge c\sqrt{nb_n\log n}\right\} \le \sum_{j=-N}^{N}p_j = O(Nn^{-\lambda_n}) = O(1/n) \to 0.$$

Therefore, $\max_{j=-N,\dots,N}|H_1(y_j)| = O_p(\sqrt{nb_n\log n})$. The result then follows from (35). ♢

Lemma 5. *Recall $\rho_g(y)$ in Theorem 1. Then*

$$\sum_{i=1}^{n}[\mathcal{G}_g(Y_{i-1})-\mathcal{G}_g(y)]K_{b_n}(y-Y_{i-1}) = nb_n^3f_Y(y)\{\rho_g(y)+O_p[b_n^2+(nb_n^3/\log n)^{-1/2}]\}.$$

Proof. We adopt the same argument in Lemma 4. Let $\eta_i(y) = [\mathcal{G}_g(Y_i)-\mathcal{G}_g(y)]K_{b_n}(y-Y_i)$. By the same decomposition in (32), we have

$$\sum_{i=0}^{n-1}\eta_i(y) = N_1(y)+N_2(y)+n\mathbb{E}[\eta_1(y)],$$

where $N_1(y) = \sum_{i=0}^{n-1}\mathcal{P}_i\eta_i(y)$ and $N_2(y) = \sum_{i=0}^{n-1}\{\mathbb{E}[\eta_i(y)|\mathcal{F}_{i-1}]-\mathbb{E}[\eta_i(y)]\}$. By the symmetry of $K(\cdot)$ and Taylor's expansion, we can show $\mathbb{E}[\eta_1(y)] = b_n^3f_Y(y)[\rho_g(y)+O(b_n^2)]$. For $N_2(y)$, by the same argument in (30)–(31), we can obtain

$$N_2(y) = b_n\int_{-\omega}^{\omega}K(u)[\mathcal{G}_g(y-ub_n)-\mathcal{G}_g(y)]L(y-ub_n)du, \quad (37)$$

where $L(\cdot)$ is defined in (31). Note that $|\mathcal{G}_g(y-ub_n)-\mathcal{G}_g(y)| = O(b_n)$. Thus, by (20) in Lemma 2, $\sup_{y\in\mathcal{Y}}|N_2(y)| = O_p(b_n^2\sqrt{n})$. For the martingale part $N_1(y)$, using $\mathcal{G}_g(Y_i)-\mathcal{G}_g(y) = O(b_n)$ for $y-Y_i = O(b_n)$, we have $\mathbb{E}[\eta_i^2(y)|\mathcal{F}_{i-1}] = O(b_n^2)\mathbb{E}[K_{b_n}^2(y-Y_i)|\mathcal{F}_{i-1}] = O(b_n^3)$. Thus, by property (C8) in Section 4.2, the conditional variance satisfies $\sum_{i=0}^{n-1}\mathbb{E}\{[\mathcal{P}_i\eta_i(y)]^2|\mathcal{F}_{i-1}\} \le \sum_{i=0}^{n-1}\mathbb{E}[\eta_i^2(y)|\mathcal{F}_{i-1}] = O(nb_n^3)$. By the same chain argument in the proof of $H_1(y)$ in Lemma 4, we can show $\sup_{y\in\mathcal{Y}}|N_1(y)| = O_p(\sqrt{nb_n^3\log n})$. ♢

Lemma 6. *Recall $\xi_i(y)$ in (25). Define*

$$d_i(y) = \frac{\mathcal{P}_i[\xi_i(y)+\xi_{i+1}(y)]}{\sigma_g(y)\sqrt{nb_n\varphi_Kf_Y(y)}}. \quad (38)$$

Then

$$\sup_{y\in\mathcal{Y}}\left\|\sum_{i=1}^{n}\mathbb{E}[d_i^2(y)|\mathcal{F}_{i-1}]-1\right\|_{3/2} = O[(nb_n)^{-1/2}+b_n^{2/3}]. \quad (39)$$

Proof. We drop the argument "y" and write $\xi_i = \xi_i(y)$. By property (C7), we can show

$$\begin{aligned}\mathbb{E}\{[\mathcal{P}_i(\xi_i+\xi_{i+1})]^2|\mathcal{F}_{i-1}\} &= 2\mathbb{E}(\xi_i\xi_{i+1}|\mathcal{F}_{i-1})-2\mathbb{E}(\xi_i|\mathcal{F}_{i-1})\mathbb{E}(\xi_{i+1}|\mathcal{F}_{i-1})-[\mathbb{E}(\xi_{i+1}|\mathcal{F}_{i-1})]^2\\ &\quad+\mathbb{E}(\xi_i^2|\mathcal{F}_{i-1})-[\mathbb{E}(\xi_i|\mathcal{F}_{i-1})]^2+\mathbb{E}\{[\mathbb{E}(\xi_{i+1}|\mathcal{F}_i)]^2|\mathcal{F}_{i-1}\}\\ &:= 2A_{i,1}-2A_{i,2}-A_{i,3}+A_{i,4}-A_{i,5}+A_{i,6}.\end{aligned} \quad (40)$$

Below we consider each of these six terms separately. For convenience, sometimes we give bounds for $\|\cdot\|_2$, which dominates $\|\cdot\|_{3/2}$.

($A_{i,5}$ *and* $A_{i,6}$ *terms*:) Let $\nu_i = [\mathbb{E}(\xi_{i+1}|\mathcal{F}_i)]^2$. Note that $A_{i+1,5} - A_{i,6} = \mathcal{P}_i \nu_i$. By consecutively using properties (C5), (C6) and (C3) in Section 4.2, we have

$$\left\| \sum_{i=1}^{n} (A_{i+1,5} - A_{i,6}) \right\|_2^2 = \sum_{i=1}^{n} \|\mathcal{P}_i \nu_i\|_2^2 \le \sum_{i=1}^{n} \|\nu_i\|_2^2 \le n\mathbb{E}(\xi_1^4) = O(nb_n). \tag{41}$$

Therefore, by the triangle inequality,

$$\left\| \sum_{i=1}^{n} (A_{i,5} - A_{i,6}) \right\|_2 = \left\| \nu_0 - \nu_n + \sum_{i=1}^{n} (A_{i+1,5} - A_{i,6}) \right\|_2 = O(\sqrt{nb_n}). \tag{42}$$

($A_{i,3}$ *terms*:) Recall $\xi_{i,1}(y)$ and $\xi_{i,2}(y)$ in (28). By $\xi_i = \xi_{i,1}(y) - \xi_{i,2}(y)$, (29) and (30), and the boundedness of $K(\cdot)$, $q_\varepsilon(\cdot)$ and $\mathcal{G}_g(y)$, $y \in \mathcal{Y}_\epsilon$, there exists a constant c_1 such that

$$|\mathbb{E}(\xi_{i+1}|\mathcal{F}_{i-1})| \le c_1 b_n \{1 + \mathbb{E}[|g(Y_{i+1})||X_i]\}, \quad \text{for all } y. \tag{43}$$

Using $\{1 + \mathbb{E}[|g(Y_{i+1})||X_i]\}^2 \le 2 + 2\{\mathbb{E}[|g(Y_{i+1})||X_i]\}^2 \le 2 + 2\mathbb{E}[g^2(Y_{i+1})|X_i]$, we obtain

$$\begin{aligned} \left\| \sum_{i=1}^{n} A_{i,3} \right\|_2 &\le 2c_1^2 b_n^2 \left\| \sum_{i=1}^{n} \{1 + \mathbb{E}[g^2(Y_{i+1})|X_i]\} \right\|_2 \\ &\le 2c_1^2 b_n^2 \sum_{i=1}^{n} \|1 + \mathbb{E}[g^2(Y_{i+1})|X_i]\|_2 = O(nb_n^2). \end{aligned} \tag{44}$$

Here, the last equality follows from $\|\mathbb{E}[g^2(Y_{i+1})|X_i]\|_2 \le \|g^2(Y_{i+1})\|_2 = \|g(Y_1)\|_4^2 < \infty$.

($A_{i,2}$ *terms*:) By the bounded support of $K(\cdot)$, it suffices to consider $|Y_{i-1} - y| = O(b_n)$ in ξ_i so that $|\mathcal{G}_g(Y_{i-1})| \le c_2$ for some constant c_2. Note that $\mathbb{E}[|g(Y_i)||\mathcal{F}_{i-1}] = \mathbb{E}[|g(Y_i)||X_i]$. Thus, $|\mathbb{E}(\xi_i|\mathcal{F}_{i-1})| \le |K_{b_n}(y - Y_{i-1})|\{c_2 + \mathbb{E}[|g(Y_i)||X_i]\}$. Combining this with (43) gives

$$|A_{i,2}| \le c_1 b_n |K_{b_n}(y - Y_{i-1})| Z_i, \qquad Z_i = \{c_2 + \mathbb{E}[|g(Y_i)||X_i]\}\{1 + \mathbb{E}[|g(Y_{i+1})||X_i]\}. \tag{45}$$

By the same argument in (30), $\mathbb{E}[|K_{b_n}(y - Y_{i-1})|^{3/2}|X_{i-1}, X_i] \le c_3 b_n$ for some constant c_3. Also, Z_i is a function of X_i and independent of ε_{i-1}. Thus, $\mathbb{E}[|K_{b_n}(y - Y_{i-1})Z_i|^{3/2}] = \mathbb{E}\{\mathbb{E}[|K_{b_n}(y - Y_{i-1})Z_i|^{3/2}|X_{i-1}, X_i]\} \le c_3 b_n \mathbb{E}(|Z_i|^{3/2})$. Now, by (45), we have

$$\left\| \sum_{i=1}^{n} A_{i,2} \right\|_{3/2} \le \sum_{i=1}^{n} \|A_{i,2}\|_{3/2} \le c_1 b_n \sum_{i=1}^{n} \|K_{b_n}(y - Y_{i-1}) Z_i\|_{3/2} = O(nb_n^{5/3}). \tag{46}$$

Here, we have used $\|Z_i\|_{3/2} \le \|Z_1\|_2 < \infty$ by the condition $g(Y_i) \in \mathcal{L}^\delta$ with $\delta \ge 4$.

($A_{i,1}$ *terms*:) In $\xi_i \xi_{i+1}$, thanks to the terms $K_{b_n}(y - Y_{i-1})$ and $K_{b_n}(y - Y_i)$, it suffices to consider $|Y_{i-1} - y| = O(b_n)$ and $|Y_i - y| = O(b_n)$. Thus, $|g(Y_i)| + |\mathcal{G}_g(Y_i)| + |\mathcal{G}_g(Y_{i-1})| \le c_4$ for some constant c_4, and consequently $|\xi_i| \le c_4 |K_{b_n}(y - Y_{i-1})|$. Also, conditioning on $\mathcal{F}_{i-1}$, the ε_i term in $Y_i = X_i + \varepsilon_i$ is independent of everything else, and thus the same argument in (30) shows that the term $K_{b_n}(y - Y_i)$ will result in an $O(b_n)$ factor. Thus,

$$|A_{i,1}| \le O(b_n) |K_{b_n}(y - Y_{i-1})| \{c_4 + \mathbb{E}[|g(Y_{i+1})||\mathcal{F}_{i-1}]\}.$$

By the independence between $\{\varepsilon_i\}$ and $\{X_i\}$ as well as the Markovian assumption on $\{X_i\}$, $\mathbb{E}[|g(Y_{i+1})||\mathcal{F}_{i-1}] = \mathbb{E}[|g(Y_{i+1})||X_i]$. The same argument in (46) then gives

$$\left\| \sum_{i=1}^{n} A_{i,1} \right\|_{3/2} \le O(b_n) \sum_{i=1}^{n} \|K_{b_n}(y - Y_{i-1})\{c_4 + \mathbb{E}[|g(Y_{i+1})||X_i]\}\|_{3/2} = O(nb_n^{5/3}). \tag{47}$$

($A_{i,4}$ *terms*:) Write $A_{i,4} - \mathbb{E}(\xi_i^2) = \mathcal{P}_{i-1}\xi_i^2 + [\mathbb{E}(\xi_i^2|\mathcal{F}_{i-2}) - \mathbb{E}(\xi_i^2)]$. Since $\{\mathcal{P}_{i-1}\xi_i^2\}_{i \in \mathbb{Z}}$ are martingale differences with respect to $\{\mathcal{F}_{i-1}\}_{i \in \mathbb{Z}}$, by the same argument in (41),

$$\left\| \sum_{i=1}^{n} \mathcal{P}_{i-1}\xi_i^2 \right\|_2 = O(\sqrt{nb_n}). \tag{48}$$

By the same argument in (29) and (30), it can be shown that

$$\mathbb{E}(\xi_i^2|\mathcal{F}_{i-2}) = b_n \int_{-\omega}^{\omega} K^2(u)\mathbb{E}\{[g(Y_i) - \mathcal{g}_g(y - ub_n)]^2 q_\varepsilon(y - X_{i-1} - ub_n)|X_{i-1}\}.$$

Thus, as in the proof of Lemma 3, an application of Lemma 2 gives $\| \sum_{i=1}^n [\mathbb{E}(\xi_i^2|\mathcal{F}_{i-2}) - \mathbb{E}(\xi_i^2)]\|_2 = O(b_n\sqrt{n})$. The latter bound along with (48) gives

$$\left\| \sum_{i=1}^n [A_{i,4} - \mathbb{E}(\xi_i^2)] \right\|_2 \le \left\| \sum_{i=1}^n \mathcal{P}_{i-1}\xi_i^2 \right\|_2 + \left\| \sum_{i=1}^n [\mathbb{E}(\xi_i^2|\mathcal{F}_{i-2}) - \mathbb{E}(\xi_i^2)] \right\|_2 = O(\sqrt{nb_n}). \tag{49}$$

Elementary calculation shows that $\mathbb{E}(\xi_i^2) = b_n\varphi_K f_Y(y)\sigma_g^2(y) + O(b_n^2)$. Finally, (39) then follows from (40), (42), (44), (46), (47) and (49) via the triangle inequality. ◊

Lemma 7. *Recall $S_n(y)$ in (27). Under the conditions and notations in Theorem 1,*

$$\lim_{n\to\infty} \mathbb{P}\left\{ \sup_{y\in\mathcal{Y}_n} \frac{|S_n(y)|}{\sigma_g(y)\sqrt{nb_n\varphi_K f_Y(y)}} \le B_{m_n}(z) \right\} = e^{-2e^{-z}}, \quad z \in \mathbb{R}. \tag{50}$$

Proof. Let $d_i(y)$ be defined in (38). Write

$$\tilde{S}_n(y) := \frac{S_n(y)}{\sigma_g(y)\sqrt{nb_n\varphi_K f_Y(y)}} = \sum_{i=1}^n d_i(y).$$

Write $\mathcal{Y}_n = \{y_1 < \cdots < y_{m_n}\}$. For fixed $k \in \mathbb{N}$ distinct integers $1 \le j_1, j_2, \ldots, j_k \le m_n$, define the k-dimensional column vectors

$$D_i = [d_i(y_{j_1}), \ldots, d_i(y_{j_k})]^T \quad \text{and} \quad S_{n,k} = \sum_{i=1}^n D_i = [\tilde{S}_n(y_{j_1}), \ldots, \tilde{S}_n(y_{j_k})]^T.$$

Then $\{D_i\}_{i\in\mathbb{Z}}$ are k-dimensional vectors of martingale differences with respect to $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$. Denote by Q_n the quadratic characteristic matrix of the martingale $S_{n,k}$, i.e.,

$$Q_n = \sum_{i=1}^n \mathbb{E}(D_i D_i^T|\mathcal{F}_{i-1}) := (q_{rs})_{1\le r,s\le k}.$$

Let $\tau_{rs} = \varphi_K\sigma_g(y_{j_r})\sigma_g(y_{j_s})\sqrt{f_Y(y_{j_r})f_Y(y_{j_s})}$. Then we can write q_{rs} as

$$\begin{aligned} q_{rs} &= \sum_{i=1}^n \mathbb{E}[d_i(y_{j_r})d_i(y_{j_s})|\mathcal{F}_{i-1}] \\ &= \frac{1}{nb_n\tau_{rs}} \sum_{i=1}^n \Big\{ \mathbb{E}[\mathcal{P}_i\xi_i(y_{j_r})\mathcal{P}_i\xi_i(y_{j_s})|\mathcal{F}_{i-1}] + \mathbb{E}[\mathcal{P}_i\xi_{i+1}(y_{j_r})\mathcal{P}_i\xi_{i+1}(y_{j_s})|\mathcal{F}_{i-1}] \\ &\quad + \mathbb{E}[\mathcal{P}_i\xi_i(y_{j_r})\mathcal{P}_i\xi_{i+1}(y_{j_s})|\mathcal{F}_{i-1}] + \mathbb{E}[\mathcal{P}_i\xi_{i+1}(y_{j_r})\mathcal{P}_i\xi_i(y_{j_s})|\mathcal{F}_{i-1}] \Big\}. \end{aligned} \tag{51}$$

For $r = s$, by Lemma 6, $\|q_{rr} - 1\|_{3/2} = O[(nb_n)^{-1/2} + b_n^{2/3}]$. For $r \ne s$, by the definition of $\mathcal{Y}_n$, since $|y_{j_r} - y_{j_s}| \ge \tau_n$, $b_n = o(\tau_n)$, and the kernel function $K(\cdot)$ has bounded support, we have $K_{b_n}(y_{j_r} - Y_{i-1})K_{b_n}(y_{j_s} - Y_{i-1}) = 0$ for large enough n. Thus, $\mathcal{P}_i\xi_i(y_{j_r})\mathcal{P}_i\xi_i(y_{j_s}) = 0$. For the other three terms on the right hand side of (51), their expansions of the form (40) involve terms of the form $A_{i,1}, A_{i,2}, A_{i,3}, A_{i,5}, A_{i,6}$ (the term $A_{i,4}$ vanishes, thanks to the choice of $\mathcal{Y}_n$ and the bounded support of $K(\cdot)$). Thus, we can use the same argument in Lemma 6 to show that their $\|\cdot\|_{3/2}$ norm can be bounded by $O[(nb_n)^{-1/2} + b_n^{2/3}]$. In summary, let I_{rs} be the (r, s)-element of the $k \times k$ identity matrix, then $\|q_{rs} - I_{rs}\|_{3/2} = O[(nb_n)^{-1/2} + b_n^{2/3}]$ uniformly over $1 \le r, s \le k$. It is easily seen that $\sum_{i=1}^n \mathbb{E}|d_i(y_{j_r})|^3 = O[(nb_n)^{-1/2}]$ uniformly over $1 \le r \le k$. Thus $\sum_{i=1}^n \mathbb{E}|d_i(y_{j_r})|^3 + \mathbb{E}(|q_{rs} - I_{rs}|^{3/2}) = O(\Omega_n)$ uniformly, where $\Omega_n = (nb_n)^{-1/2} + b_n$.

For $j = 1, \ldots, m_n$, define events $T_j = \{|\tilde{S}_n(y_j)| \ge B_{m_n}(z)\}$. Then $\mathbb{P}\{\sup_{y\in\mathcal{Y}_n} |\tilde{S}_n(y)| \ge B_{m_n}(z)\} = \mathbb{P}\{\cup_{j=1}^{m_n} T_j\}$. By the same argument in the proof of Theorem 3 in [31], we can show $\mathbb{P}\{\cup_{j=1}^{m_n} T_j\} \to 1 - \exp\{-2e^{-z}\}$. The details are omitted. ◊

4.5. Proof of Proposition 2

Lemma 8. *The uniform consistency holds:* $\sup_{y\in\mathcal{Y}}|\hat{\mathcal{G}}_g(y)-\mathcal{G}_g(y)|=O_p[b_n^2+(b_n\sqrt{n})^{-1}\log n]$.

Proof. By (26), (27), Lemmas 3–4, it suffices to prove $\sup_{y\in\mathcal{Y}}|S_n(y)|=O_p(\sqrt{n}\log n)$, where $S_n(y)=\sum_{i=1}^n\mathcal{P}_i[\xi_i(y)+\xi_{i+1}(y)]$ is defined in (27). Again, we adopt the chain argument in Lemma 4 to establish the uniform bound for the martingale $S_n(y)$.

Let $y_{-N},\ldots,y_N$ be the grid points defined in the proof of Lemma 4. By the same chain argument in Lemma 4, it suffices to prove $\max_{-N\le j\le N}|S_n(y_j)|=O_p(\sqrt{n}\log n)$. However, since the summands $\mathcal{P}_i[\xi_i(y)+\xi_{i+1}(y)]$ and their conditional variances are no longer bounded, we cannot directly use Freedman's exponential inequality for bounded martingale differences. To solve this issue, we adopt the following argument. Define

$$A_1=\max_{1\le i\le n}[\mathbb{E}(\zeta_i|\mathcal{F}_i)+\mathbb{E}(\zeta_i|\mathcal{F}_{i-1})],\qquad \zeta_i=|g(Y_i)|+|\mathcal{G}_g(Y_{i-1})|+|g(Y_{i+1})|+|\mathcal{G}_g(Y_i)|,$$

$$A_2=\sum_{i=1}^n\mathbb{E}[|g(Y_i)|^2+|\mathcal{G}_g(Y_{i-1})|^2+|g(Y_{i+1})|^2+|\mathcal{G}_g(Y_i)|^2|\mathcal{F}_{i-1}].$$

Let $c_1=\sup_u|K(u)|$. Then $|\xi_i(y)|+|\xi_{i+1}(y)|\le c_1\zeta_i$ uniformly in y, and consequently

$$|\mathcal{P}_i[\xi_i(y)+\xi_{i+1}(y)]|\le c_1A_1,\quad \text{uniformly in } i=1,\ldots,n,\ y\in\mathbb{R}. \tag{52}$$

Now, consider conditional variance. By (C8) in Section 4.2, $\mathbb{E}(\{\mathcal{P}_i[\xi_i(y)+\xi_{i+1}(y)]\}^2|\mathcal{F}_{i-1})\le\mathbb{E}\{[\xi_i(y)+\xi_{i+1}(y)]^2|\mathcal{F}_{i-1}\}\le c_1^2\mathbb{E}(\zeta_i^2|\mathcal{F}_{i-1})$. Thus, by the Cauchy–Schwarz inequality,

$$\sum_{i=1}^n\mathbb{E}(\{\mathcal{P}_i[\xi_i(y)+\xi_{i+1}(y)]\}^2|\mathcal{F}_{i-1})\le c_1^2\sum_{i=1}^n\mathbb{E}(\zeta_i^2|\mathcal{F}_{i-1})\le 4c_1^2A_2,\quad \text{uniformly in } y\in\mathbb{R}. \tag{53}$$

By (52) and (53), on the event $\{A_1\le n^{1/4}\log n, A_2\le n\log n\}$, the martingale differences $\mathcal{P}_i[\xi_i(y)+\xi_{i+1}(y)]$ are upper bounded by $c_1n^{1/4}\log n$ and the sum of conditional variances is upper bounded by $4c_1^2n\log n$. Therefore, as in (36), for any $j=-N,\ldots,N$ and $c>0$,

$$\begin{aligned}p_j&:=\mathbb{P}\left\{|S_n(y_j)|\ge c\sqrt{n}\log n, A_1\le n^{1/4}\log n, A_2\le n\log n\right\}\\&\le 2\exp\left\{-\frac{c^2n(\log n)^2}{2[(c_1n^{1/4}\log n)(c\sqrt{n}\log n)+4c_1^2n\log n]}\right\}\\&=2\exp(-\lambda_n\log n),\end{aligned} \tag{54}$$

where

$$\lambda_n=\frac{c^2}{2[cc_1n^{-1/4}(\log n)^2+4c_1^2]}.$$

For large enough c and n, we have $\lambda_n>3$. Thus,

$$\begin{aligned}\mathbb{P}\left\{\max_{-N\le j\le N}|S_n(y_j)|\ge c\sqrt{n}\log n\right\}&\le\mathbb{P}\left\{\max_{-N\le j\le N}|S_n(y_j)|\ge c\sqrt{n}\log n, A_1\le n^{1/4}\log n, A_2\le n\log n\right\}\\&\quad+\mathbb{P}\{A_1>n^{1/4}\log n\}+\mathbb{P}\{A_2>n\log n\}\\&\le\sum_{j=-N}^N p_j+\mathbb{P}\{A_1>n^{1/4}\log n\}+\mathbb{P}\{A_2>n\log n\}.\end{aligned} \tag{55}$$

By (54) and $N=n^2$, $\sum_{j=-N}^N p_j=O(1/n)\to 0$. Note that

$$\begin{aligned}\mathbb{E}(A_1^4)&\le\sum_{i=1}^n\mathbb{E}\{[\mathbb{E}(\zeta_i|\mathcal{F}_i)+\mathbb{E}(\zeta_i|\mathcal{F}_{i-1})]^4\}\\&\le 16\sum_{i=1}^n\mathbb{E}\{[\mathbb{E}(\zeta_i|\mathcal{F}_i)]^4+[\mathbb{E}(\zeta_i|\mathcal{F}_{i-1})]^4\}\le 32n\mathbb{E}(\zeta_1^4).\end{aligned}$$

Here the second "$\le$" follows from $(u+v)^4\le 16(u^4+v^4)$ and the third "$\le$" follows from property (C3) in Section 4.2. Thus, by Markov's inequality, $\mathbb{P}\{A_1>n^{1/4}\log n\}\le\mathbb{E}(A_1^4)/[n^{1/4}(\log n)]^4=O[(\log n)^{-4}]$. Another application of Markov's inequality gives $\mathbb{P}\{A_2>n\log n\}\le\mathbb{E}(A_2)/(n\log n)=O[(\log n)^{-1}]$. Thus, the right hand side of (55) goes to zero, and we conclude $\max_{-N\le j\le N}|S_n(y_j)|=O_p(\sqrt{n}\log n)$, completing the proof. $\diamond$

Proof of Proposition 2. (i) By the same argument in Lemma 4, $\sup_{y\in\mathcal{Y}}|\hat{f}_Y(y) - f_Y(y)| = O_p[l_n^2 + (nl_n/\log n)^{-1/2}] = o_p[(\log n)^{-1/2}]$.

(ii) Recall the definition of $\mathcal{Y}_\epsilon$ in Condition 2. Write $\Delta_1 = \sup_{y\in\mathcal{Y}_{\epsilon/2}}|\hat{\mathcal{G}}_g(y) - \mathcal{G}_g(y)|$. Clearly, the uniform bound in Lemma 8 also holds on $\mathcal{Y}_{\epsilon/2}$, i.e., $\Delta_1 = O_p[b_n^2 + (b_n\sqrt{n})^{-1}\log n]$. Define

$$\bar{\sigma}_g^2(y) = \frac{\sum_{i=1}^{n}[g(Y_i) - \mathcal{G}_g(Y_{i-1})]^2 K_{h_n}(y - Y_{i-1})}{\sum_{i=1}^{n} K_{h_n}(y - Y_{i-1})}.$$

By the bounded support of $K(\cdot)$, it suffices to consider Y_{i-1} in a neighborhood of $y \in \mathcal{Y}$ or consider Y_{i-1} in the neighborhood $\mathcal{Y}_{\epsilon/2}$ of $\mathcal{Y}$ so that $\max_{1\le i\le n}|\hat{\mathcal{G}}_g(Y_{i-1}) - \mathcal{G}_g(Y_{i-1})| \le \Delta_1$. Applying the inequality $|a^2 - b^2| \le x(x + 2|b|)$ for all $|a - b| \le x$, we have

$$|\hat{\sigma}_g^2(y) - \bar{\sigma}_g^2(y)| \le \Delta_1 \frac{\sum_{i=1}^{n}[\Delta_1 + 2|g(Y_i) - \mathcal{G}_g(Y_{i-1})|]|K_{h_n}(y - Y_{i-1})|}{\left|\sum_{i=1}^{n} K_{h_n}(y - Y_{i-1})\right|}. \quad (56)$$

Let $\Delta_2 = \max_{1\le i\le n}|g(Y_i) - \mathcal{G}_g(Y_{i-1})|$. Then $\mathbb{E}(\Delta_2^4) \le \sum_{i=1}^{n}\mathbb{E}[|g(Y_i) - \mathcal{G}_g(Y_{i-1})|^4] = O(n)$, which implies $\Delta_2 = O_p(n^{1/4})$. Thus, by (56) and Lemma 4, $\hat{\sigma}_g^2(y) - \bar{\sigma}_g^2(y) = O_p(\Delta_1^2 + \Delta_1\Delta_2) = o_p[(\log n)^{-1/2}]$, uniformly in $y \in \mathcal{Y}$. Finally, by the same argument in Lemma 8 (i.e., use a similar decomposition as in (27) along with Lemma 3, Lemma 4 and the argument in Lemma 8), we can show that $\sup_{y\in\mathcal{Y}}|\bar{\sigma}_g^2(y) - \sigma_g^2(y)| = O_p[h_n^2 + (h_n\sqrt{n})^{-1}\log n] = o_p[(\log n)^{-1/2}]$. This completes the proof. ◊

References

[1] Y. Aït-Sahalia, Testing continuous-time models of the spot interest rate, Rev. Financ. Stud. 9 (1996) 385–426.
[2] Y. Aït-Sahalia, J. Fan, H. Peng, Nonparametric transition-based tests for jump-diffusions, J. Amer. Statist. Assoc. 104 (2009) 1102–1116.
[3] Y. Aït-Sahalia, P.A. Mykland, L. Zhang, How often to sample a continuous-time process in the presence of market microstructure noise, Rev. Financ. Stud. 18 (2005) 351–416.
[4] A. Azzalini, A. Bowman, On the use of nonparametric regression for checking linear relationships, J. R. Stat. Soc. Ser. B 55 (1993) 549–557.
[5] P.J. Bickel, M. Rosenblatt, On some global measures of the deviations of density function estimates, Ann. Statist. 1 (1973) 1071–1095.
[6] R.J. Carroll, D. Ruppert, L.A. Stefanski, C.M. Crainiceanu, Measurement Error in Nonlinear Models: A Modern Perspective, CRC Press, 2006.
[7] X. Chen, H. Hong, D. Nekipelov, Nonlinear models of measurement errors, J. Econom. Lit. 49 (2011) 901–937.
[8] R.L. Eubank, P.L. Speckman, Confidence bands in nonparametric regression, J. Amer. Statist. Assoc. 88 (1993) 1287–1301.
[9] J. Fan, On the optimal rates of convergence for nonparametric deconvolution problems, Ann. Statist. 19 (1991) 1257–1272.
[10] Y. Fan, Q. Li, Consistent model specification tests: omitted variables and semiparametric functional forms, Econometrica 64 (1996) 865–890.
[11] J. Fan, Q. Yao, Nonlinear Time Series: Nonparametric and Parametric Methods, Springer, New York, 2003.
[12] J. Fan, W. Zhang, Simultaneous confidence bands and hypothesis testing in varying-coefficient models, Scand. J. Statist. 27 (2000) 715–731.
[13] J. Fan, C. Zhang, J. Zhang, Generalized likelihood ratio statistics and Wilks phenomenon, Ann. Statist. 29 (2001) 153–193.
[14] D.A. Freedman, On tail probabilities for martingales, Ann. Probab. 3 (1975) 100–118.
[15] W. Fuller, Measurement Error Models, John Wiley & Sons, New York, 1987.
[16] J. Gao, M. King, Adaptive testing in continuous-time diffusion models, Econometric Theory 20 (2004) 844–882.
[17] W. Härdle, E. Mammen, Comparing nonparametric versus parametric regression fits, Ann. Statist. 21 (1993) 1926–1947.
[18] Y. Hong, Hypothesis testing in time series via the empirical characteristic function: a generalized spectral density approach, J. Amer. Statist. Assoc. 94 (1999) 1201–1220.
[19] Y. Hong, H. Li, Nonparametric specification testing for continuous-time models with applications to term structure of interest rates, Rev. Financ. Stud. 18 (2005) 37–84.
[20] Y. Hong, H. White, Consistent specification testing via nonparametric series regression, Econometrica 63 (1995) 1133–1159.
[21] J. Jacod, Y. Li, P.A. Mykland, M. Podolskij, M. Vetter, Microstructure noise in the continuous case: the pre-averaging approach, Stochastic Process. Appl. 119 (2009) 2249–2276.
[22] G. Knafl, J. Sacks, D. Ylvisaker, Confidence bands for regression functions, J. Amer. Statist. Assoc. 80 (1985) 683–691.
[23] Q. Li, J. Racine, Nonparametric Econometrics, Princeton University Press, Princeton, New Jersey, 2007.
[24] Y. Li, Z. Zhang, X. Zheng, Volatility inference in the presence of both endogenous time and microstructure noise, Stochastic Process. Appl. 123 (2013) 2696–2727.
[25] J. Pinkse, A consistent nonparametric test for serial independence, J. Econometrics 84 (1998) 205–231.
[26] B.W. Silverman, Density Estimation, Chapman and Hall, London, 1986.
[27] S. Taylor, Modeling Financial Time Series, Wiley, Chichester, 1986.
[28] W.B. Wu, Nonlinear system theory: another look at dependence, Proc. Natl. Acad. Sci. USA 102 (2005) 14150–14154.
[29] L. Zhang, P.A. Mykland, Y. Aït-Sahalia, A tale of two time scales: determining integrated volatility with noisy high-frequency data, J. Amer. Statist. Assoc. 472 (2005) 1394–1411.
[30] Z. Zhao, Parametric and nonparametric models and methods in financial econometrics, Stat. Surv. 2 (2008) 1–42.
[31] Z. Zhao, Nonparametric model validations for hidden Markov models with applications in financial econometrics, J. Econometrics 162 (2011) 225–239.
[32] Z. Zhao, W.B. Wu, Confidence bands in nonparametric time series regression, Ann. Statist. 36 (2008) 1854–1878.