Specification test for Markov models with measurement errors

Seonjin Kim, Zhibiao Zhao

A R T I C L E I N F O

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

A B S T R A C T

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.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Let \( \{X_t\}_{t \in \mathbb{N}} \) be a real-valued stationary time series of interest. In some applications, \( \{X_t\} \) may not be directly observable and instead we observe a contaminated version \( \{Y_t\} \):

\[
Y_t = X_t + \varepsilon_t, \quad i = 1, 2, \ldots, n, \tag{1}
\]

where \( \{\varepsilon_t\} \) 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:

\[
Y_t = \sigma_t \varepsilon_t, \tag{2}
\]

where \( \{\sigma_t \} > 0 \) is an unobservable volatility process. For example, Taylor \([27]\) proposed an AR(1) model for \( \log(\sigma_t^2) \):

\[
\log(\sigma_t^2) = a \log(\sigma_{t-1}^2) + \eta_t. \tag{3}
\]

In general, with the transformation \( Y_t^* = \log(Y_t^2), X_t = \log(\sigma_t^2), \varepsilon_t^* = \log(\varepsilon_t^2) \), then (2) becomes the measurement errors model:

\[
Y_t^* = X_t + \varepsilon_t^*, \tag{3}
\]

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/© 2014 Elsevier Inc. All rights reserved.
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 \( \{X_t\} \).

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

\[
H_0 : \mathcal{Q} = \mathcal{Q}_\theta, \quad \text{for a parametric specification } \mathcal{Q}_\theta \text{ with unknown parameter } \theta.
\] (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 \( \{X_t\} \), 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 \( \{X_t\} \). 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

\[
X_t = \mu(X_{t-1}) + s(X_{t-1})\eta_t,
\] (5)

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

\[
dX_t = \mu(X_t)dt + s(X_t)dW_t, \quad t \geq 0,
\] (6)

where \( \{W_t\}_{t\geq0} \) 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 \( \{X_t\} \) is characterized by its transition density, denoted by \( q_X(x'|x) \); (ii) the transition density \( q_X(x'|x) \) of \( \{X_t\} \) is implicitly coded into the conditional density, denoted by \( q_Y(y'|y) \), of \( Y_t \) given \( Y_{t-1} \); and (iii) furthermore, \( q_Y(y'|y) \) is coded into the conditional expectation

\[
g_y(y) = \mathbb{E}[g(Y_t)\mid Y_{t-1} = y].
\] (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 \( g_y(y) \) 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 \( g_y(y) \) 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 \( g_y(y) \). 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 \( g_y(y) \) 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 \( g_y(y) \) 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 \( g_y(y) \) 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 \( \|Z\|_q := \left[\mathbb{E}(|Z|^q)\right]^{1/q} < \infty \); 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 \( (X_{t-1}, X_t) \);
- \( q_X(x'|x) \): the conditional density function of \( X_t \) given \( X_{t-1} = x \);
- \( p_Y(y, y') \): the joint density function of \( (Y_{t-1}, Y_t) \);
- \( q_Y(y'|y) \): the conditional density function of \( Y_t \) given \( Y_{t-1} = y \);
- \( f_t(\cdot) \): the density function of \( Y_t \);
- \( q_e(\cdot) \): the density function of \( \varepsilon_t \).

For the Markov chain \( \{X_t\} \), its evolving dynamics is characterized by the joint density function \( p_X(x, x') \) of \( (X_{t-1}, X_t) \). Since \( \{X_t\} \) is unobservable, we propose extracting information about \( p_X(x, x') \) from the observed chain \( \{Y_t\} \). In (1), we assume that the measurement errors \( \{\varepsilon_t\}_{t \in \mathbb{Z}} \) are i.i.d. and independent of the Markov chain \( \{X_t\}_{t \in \mathbb{Z}} \).

Recall the conditional expectation operator \( \hat{g}_\theta \) defined in (7). Our method is motivated by the following result:

**Proposition 1.** Let \( \varepsilon \) be a random sample from \( q_e(\cdot) \) independent of \( (X_{t-1}, X_t) \). Then

\[
\hat{g}_\theta(y) = \frac{\mathbb{E}[g(X_t + \varepsilon)q_e(y - X_{t-1})]}{\mathbb{E}[q_e(y - X_{t-1})]},
\]

(8)

By **Proposition 1**, \( \hat{g}_\theta(y) \) contains rich information about the joint density of \( (X_{t-1}, X_t) \), and different choices of \( g(\cdot) \) can extract different information. For example, \( g_1(Y_t) = Y_t \) and \( g_2(Y_t) = Y_t^2 \) extract information from the first two conditional moments, and \( g_3(Y_t) = 1_{Y_t \leq t}, t \in \mathbb{R} \), 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 \( \hat{g}_\theta(y) \), denoted by \( \hat{g}_\theta(y) \). Without imposing any specific model structure, \( \hat{g}_\theta(y) \) is always a consistent estimate of \( \hat{g}_\theta(y) \) 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 \( \hat{g}_\theta(y) \), denoted by \( \hat{g}_\theta(y|\hat{\theta}) \), where \( \hat{\theta} \) is a consistent estimate of \( \hat{\theta} \). To test \( H_0 \), we examine the distance between the parametric estimate \( \hat{g}_\theta(y|\hat{\theta}) \) and the nonparametric reference \( \hat{g}_\theta(y) \), 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 \( \hat{g}_\theta(y) \) on a given compact set \( \mathcal{Y} \subset \mathbb{R} \) if

\[
\lim_{n \to \infty} \mathbb{P}[l_n(y) \leq \hat{g}_\theta(y) \leq u_n(y), \text{ for all } y \in \mathcal{Y}] = 1 - \alpha.
\]

(9)

Intuitively, the function \( \hat{g}_\theta(\cdot) \) 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 \( \hat{g}_\theta(\cdot) \) usually centers at a nonparametric estimate \( \hat{g}_\theta(y) \). Therefore, the band \([l_n(\cdot), u_n(\cdot)]\) with center \( \hat{g}_\theta(y) \) provides an acceptance region for \( H_0 \). If the parametric estimate \( \hat{g}_\theta(y|\hat{\theta}) \) under \( H_0 \) falls outside the band, then the deviation between \( \hat{g}_\theta(y) \) and \( \hat{g}_\theta(y|\hat{\theta}) \) 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 \( \hat{g}_\theta(y) \) of \( \hat{g}_\theta(y) \), and then use \( \hat{g}_\theta(y) \) to build a \((1 - \alpha)\) nonparametric SCB for \( \hat{g}_\theta(y) \), denoted by \([l_n(\cdot), u_n(\cdot)]\).

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

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

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

**Remark 1.** Although the underlying process \( \{X_t\} \) is a Markov chain, the observed process \( \{Y_t\} \) is no longer a Markov chain and thus its distributional property cannot be completely characterized by the one-step conditional distribution of \( Y_t \) given \( Y_{t-1} \). To incorporate more distributional information, in light of (7), one may consider \( \hat{g}_\theta(y_1, \ldots, y_k) = \mathbb{E}[g(Y_t) | Y_{t-1} = y_1, \ldots, Y_{t-k} = y_k] \) 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 $\hat{g}_n(y)$:

$$\hat{g}_n(y) = \frac{\sum_{i=1}^{n} g(Y_i)K_{b_n}(y - Y_{i-1})}{\sum_{i=1}^{n} K_{b_n}(y - Y_{i-1})},$$  \hspace{1cm} (10)

where and hereafter $K_{b_n}(u) = K(u/b_n)$ for a kernel function $K$ satisfying $\int_0^1 K(u)du = 1$ and bandwidth $b_n > 0$. To study asymptotic properties of $\hat{g}_n(y)$, we define the conditional variance function $\sigma^2_g(y)$ and impose Conditions 1–3 as follows:

$$\sigma^2_g(y) = \mathbb{E}[(g(Y_i) - \hat{g}_n(Y_{i-1}))^2 | Y_{i-1} = y].$$  \hspace{1cm} (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 $f_\epsilon(y) > 0$ and $\hat{g}_n(y)$ have bounded fourth order derivative on $\mathcal{Y}_\epsilon := [-T - \epsilon, T + \epsilon]$, and that $\sigma^2_g(y) > 0$ 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 $q_\varepsilon$ of $\varepsilon_i$ is bounded and has bounded derivative on $\mathbb{R}$.

**Condition 3** (Dependence Assumption). The unobservable process $(X_i)$ is an $\alpha$-mixing stationary Markov chain with $\alpha$-mixing coefficients $\alpha_k, k \in \mathbb{N}$. Assume that $g(Y_i) \in L^3$ 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 $(\tau_n, m_n)$ 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 most $\tau_n$; 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 $\hat{g}_n(y)$, which can be used to construct a nonparametric SCB for $\hat{g}_n(y)$.

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

$$\lim_{n \to \infty} \mathbb{P} \left( \sup_{y \in \mathcal{Y}_n} \left[ \frac{nb^2_n f_\epsilon(y)}{\varphi_K \sigma^2_g(y)} \right]^{1/2} \left| \hat{g}_n(y) - g(y) - \rho_\varepsilon(y)b^2_n \right| \leq B_{m_n}(z) \right) = e^{-2e^{-z}},$$  \hspace{1cm} (12)

for $z \in \mathbb{R}$, where $\sigma^2_g(y)$ is defined in (11), $\rho_\varepsilon(y) = [f'_\epsilon(y)\hat{g}_n(y)/f_\epsilon(y) + \hat{g}_n(y)/2] \psi_K$, and

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

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

$$\hat{f}_\epsilon(y) = \frac{1}{nh} \sum_{i=1}^{n} K_{b_n}(y - Y_i),$$  \hspace{1cm} (13)

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

$$\hat{\sigma}^2_n(y) = \frac{\sum_{i=1}^{n} [g(Y_i) - \hat{g}_n(Y_{i-1})]^2 K_{b_n}(y - Y_{i-1})}{\sum_{i=1}^{n} K_{b_n}(y - Y_{i-1})},$$  \hspace{1cm} (14)

where $K_{b_n}(u) = K(u/h_n)$ for another bandwidth $h_n > 0$. 
Proposition 2. (i) Under Conditions 1–3 and \( \lambda_n^2 \log n + (nl_n)^{-1}(\log n) \to 0 \), we have
\[
\sup_{y \in \mathcal{Y}} |\hat{f}_n(y) - f_n(y)| = o_p((\log n)^{-1/2}).
\]

(ii) Assume that Conditions 1–3 hold. Further assume \( nb_n^2(\log n)^2 + (nb_n^2)^{-1}(\log n)^6 \to 0 \) and \( h_n^2 \log n + (nh_n)^{-1}(\log n)^3 \to 0 \). Then we have
\[
\sup_{y \in \mathcal{Y}} |\hat{\sigma}_n^2(y) - \sigma_n^2(y)| = o_p((\log n)^{-1/2}).
\]

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.

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 \( \hat{g}_n(y) \).

Corollary 1. Under the conditions in Theorem 1 and Proposition 2, an asymptotic \((1 - \alpha)\) SCB for \( \hat{g}_n(y) \) on the region \( \mathcal{Y}_n \) with the bias-correction can be constructed as
\[
\hat{g}_n(y) \pm \left[ \frac{\varphi_\epsilon \hat{\sigma}_n^2(y)}{nb_n f_n(y)} \right]^{1/2} B_{m_0}(z_\alpha), \quad y \in \mathcal{Y}_n,
\]
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 \( y \) 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 \( m_n = \lfloor 2T/\epsilon(c(\log n)^2 b_n) \rfloor \) and
\[
\mathcal{Y}_n = \{ y_j = -T + c(\log n)^2 b_n j, \quad j = 0, 1, \ldots, m_n - 1 \}.
\]
Then \( \mathcal{Y}_n \) is a \((\tau_n, m_n)\) approximation of \( \mathcal{Y} \) with \( \tau_n = c(\log n)^2 b_n \). It is easily seen that, under \( nb_n^2(\log n)^2 + (nb_n^2)^{-1}(\log n)^6 \to 0 \) 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 \( H_0: \mathcal{Q} = \mathcal{Q}_0 \)

In this section we develop a general procedure to construct parametric estimate of \( \hat{g}_n(y) \) under \( H_0: \mathcal{Q} = \mathcal{Q}_0 \). Without further assumptions, it is generally impossible to use (8) to construct a parametric estimate of \( \hat{g}_n(y) \). For example, Fan [9] assumed that \( \epsilon_i \) 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 \( \epsilon_i \) 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 \( \epsilon_i(z) = \phi_\sigma(z) \).

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 \( X_i = a_1 X_{i-1} 1_{X_{i-1} \leq 0} + a_2 X_{i-1} 1_{X_{i-1} > 0} + \eta_i \) 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 \( H_0: \mathcal{Q} = \mathcal{Q}_0 \), obtain consistent estimate of \((\theta, \sigma)\), 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 \( \{X^*_i\}_{0 \leq i \leq m} \) from the estimated null model \( \mathcal{Q}_0 \).

(iii) Let \( \epsilon_i \) be i.i.d. \( N(0, \sigma^2) \) variables. Using empirical version of (8), we propose
\[
\hat{g}_n(y|\mathcal{Q}_0) = \frac{m^{-1} \sum_{i=1}^m g(X^*_i + \epsilon_i) \phi_\sigma(y - X^*_i)}{m^{-1} \sum_{i=1}^m \phi_\sigma(y - X^*_i)}. \]

For large \( m \), the numerator and the denominator of \( \hat{g}_n(y|\mathcal{Q}_0) \) approach their expectations.

As an illustration, we consider (5) with \( \eta_i \sim N(0, 1) \). Let \( \mathcal{Q} = (\mu, s) \) and \( \mathcal{Q}_0 = (\mu_0, s_0) \) for some parametric specification \((\mu_0, s_0)\). Then step (ii) above is implemented using
\[
X^*_i = \mu_0(X^*_i) + s_0(X^*_i) \eta_i, \quad \eta_i \sim N(0, 1).
\]

Clearly, the above proposed procedure can be readily extended to the case of non-Gaussian errors. We simply replace \( \phi_\sigma \) with another given parametric density with estimated parameters and draw \( \epsilon_i \) 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 \( \sigma^2(\cdot) \) fully determine the model structure. Similarly, in model (6), \( \mu(\cdot) \) and \( \sigma^2(\cdot) \) represent the conditional 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 \( Q = (\mu, \sigma) \) and \( H_0 : Q = Q_0 = (\mu_0, \sigma_0) \) for parametric specifications \((\mu_0, \sigma_0)\).

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 \( \alpha \), then we construct \((1 - \alpha/2)\) SCBs, denoted by SCB_1 and SCB_2, for \( g_1(y) \) and \( g_2(y) \) separately, and reject \( H_0 \) if either SCB_1 or SCB_2 cannot cover the corresponding parametric estimates for \( g_1(Y_i) \) or \( g_2(Y_i) \).

Theoretically speaking, we can combine tests across multiple samples. 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, \( g_k(Y_i) = 1_{Y_i \leq t} \) for \( t \in \mathbb{R} \) corresponds to the conditional distribution function, and \( g_k(Y_i) = \exp(\sqrt{-1} \mathbb{Y}_t(t) \) 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 \( g_k(Y_i) \) over both \( t \in \mathbb{R} \) and \( y \in \mathbb{R} \). With \( g_1(Y_i) = 1_{Y_i \leq t} \), we expect that, after proper normalization, the process \( \{g_k(Y_i)\}_{t \in \mathbb{R}} \), with any fixed \( y \) converges in distribution in the Skorokhod space \( D[-\infty, +\infty] \) to the process \( \mathbb{B}(Q_Y(t|y)) \), where \( \mathbb{B} \) is the standard Brownian bridge and \( Q_Y(t|y) \) is the conditional distribution function of \( Y_t \) given \( Y_{t-} = y \). 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 \( \rho_b(y)b_n^2 \) 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 \( \int_{-\infty}^{\infty} K(u)du = 1 \) and \( \psi(u) = \int_{-\infty}^{\infty} u^2 K(u)du = 0 \) so that \( \rho_b(y) = 0 \). 

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 \( h_n \) 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.xreg in R. To choose \( h_n \) in (14), first we obtain the residuals \( g(Y_i) - \hat{g}_k(Y_{i-}) \) using the selected bandwidth \( h_n \). Then, the leave-one-out cross validation is employed again based on the nonparametric regression of \( \log(Y_i) - \hat{g}_k(Y_{i-})^2 \) on \( Y_{i-} \). For \( h_n \) in (13), we use the rule-of-thumb nonparametric kernel density bandwidth selector in [26], which is implemented using the command \bwnrd0 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.
The case calculations show that for i.i.d. noises with samples size \( n \) (gridpoints together with the Bonferroni correction, respectively. Significance level is 5%.

<table>
<thead>
<tr>
<th>Test</th>
<th>λ</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>200</td>
<td>Test 1</td>
<td>0.039</td>
<td>0.090</td>
<td>0.278</td>
<td>0.525</td>
<td>0.695</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 2</td>
<td>0.038</td>
<td>0.021</td>
<td>0.038</td>
<td>0.054</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 3</td>
<td>0.038</td>
<td>0.042</td>
<td>0.181</td>
<td>0.358</td>
<td>0.358</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>Test 1</td>
<td>0.029</td>
<td>0.159</td>
<td>0.532</td>
<td>0.866</td>
<td>0.866</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 2</td>
<td>0.052</td>
<td>0.037</td>
<td>0.043</td>
<td>0.075</td>
<td>0.113</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 3</td>
<td>0.037</td>
<td>0.098</td>
<td>0.379</td>
<td>0.752</td>
<td>0.921</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>Test 1</td>
<td>0.063</td>
<td>0.637</td>
<td>0.989</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 2</td>
<td>0.018</td>
<td>0.086</td>
<td>0.255</td>
<td>0.358</td>
<td>0.428</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 3</td>
<td>0.042</td>
<td>0.503</td>
<td>0.974</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>Model 2</td>
<td>200</td>
<td>Test 1</td>
<td>0.039</td>
<td>0.047</td>
<td>0.041</td>
<td>0.053</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 2</td>
<td>0.038</td>
<td>0.033</td>
<td>0.047</td>
<td>0.099</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 3</td>
<td>0.038</td>
<td>0.042</td>
<td>0.045</td>
<td>0.082</td>
<td>0.119</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>Test 1</td>
<td>0.029</td>
<td>0.024</td>
<td>0.030</td>
<td>0.056</td>
<td>0.115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 2</td>
<td>0.052</td>
<td>0.042</td>
<td>0.049</td>
<td>0.103</td>
<td>0.245</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 3</td>
<td>0.037</td>
<td>0.036</td>
<td>0.043</td>
<td>0.082</td>
<td>0.204</td>
</tr>
<tr>
<td></td>
<td>2000</td>
<td>Test 1</td>
<td>0.063</td>
<td>0.085</td>
<td>0.110</td>
<td>0.152</td>
<td>0.235</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 2</td>
<td>0.018</td>
<td>0.049</td>
<td>0.170</td>
<td>0.447</td>
<td>0.784</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Test 3</td>
<td>0.042</td>
<td>0.069</td>
<td>0.147</td>
<td>0.367</td>
<td>0.716</td>
</tr>
</tbody>
</table>

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

\[
\begin{align*}
\text{(Model 1)} & : X_i = 0.6(1 - \lambda)X_{i-1} + \lambda |X_{i-1}| + \eta_i, \quad \lambda \in [0, 1], \\
\text{(Model 2)} & : X_i = 0.6X_{i-1} + \eta_i \sqrt{1 + 0.3\lambda Y_{i-1}^2}, \quad \lambda \in [0, 1],
\end{align*}
\]

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 \( \lambda \) regulates the deviation from the null model. The case \( \lambda = 0 \) leads to the null model; as \( \lambda \) 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 \( \theta_1 \) and the variances of \( \eta_i \) and \( \varepsilon_i \), denoted by \( \theta_2^2 \) and \( \theta_3^2 \), respectively. Then elementary calculations show that

\[
\mathbb{E}(Y_i^2) = \frac{\theta_2^2}{1 - \theta_1^2} + \theta_3^2, \quad \text{cov}(Y_{i-1}, Y_i) = \frac{\theta_1 \theta_2^2}{1 - \theta_1^2}, \quad \text{cov}(Y_{i-2}, Y_i) = \frac{\theta_1^2 \theta_2^2}{1 - \theta_1^2}.
\]

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 \( Y_n \) 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 \( Y_n \) 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 \( \lambda \) 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 \( p_X, q_X, p_{YI}, q_{YI}, f_Y, q_e \) 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 \( q_e(y - X_{i-1}) \) and \( q_e(y' - X_i) \), respectively. Thus, \( p_Y(y, y') = \mathbb{E}[q_e(y - X_{i-1})q_e(y' - X_i)] \). By the latter identity and the independence between \( \varepsilon \) and \( (X_{i-1}, X_i) \),

\[
\int_R g(y') p_Y(y, y') dy' = \mathbb{E} \left[ q_e(y - X_{i-1}) \int_R g(y') q_e(y' - X_i) dy' \right]
\]
For any variable and

\[
\begin{align*}
\frac{z - y}{x_i} & \equiv \mathbb{E}\left[q_x(y - X_{i-1}) \int_R g(X_i + z)q_z(z)dz\right] \\
& = \mathbb{E}\{q_x(y - X_{i-1})\mathbb{E}[g(X_i + \varepsilon)|X_{i-1}, X_i]\} \\
& = \mathbb{E}[g(X_i + \varepsilon)q_x(y - X_{i-1})].
\end{align*}
\]

By conditioning on \(X_{i-1}\), we can show \(f(y) = \mathbb{E}\{q_x(y - X_{i-1})\}\). Therefore, the desired result follows from (16) and \(q_y(y) = \int_R g(y')q_y(y')dy' = \int_R g(y')p_y(y, y')dy'/f_y(y)\). □

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 \(\sigma\)-algebra on the same probability space. Then

(C1) \(\mathbb{E}(Z) = \mathbb{E}[\mathbb{E}(Z|\mathcal{F})]\).

(C2) Let \(\mathcal{G}\) be another \(\sigma\)-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 \(p \geq 1\), then \(\|\mathbb{E}(Z|\mathcal{F})\|_p \leq \|Z\|_p\) and \((\mathbb{E}(Z))^p \leq \mathbb{E}(Z^p)\).

Recall that, in (1), \((\varepsilon_t)\) are i.i.d. and independent of the unobservable Markov chain \(\{X_t\}\). Let \(\mathcal{F}_t = \sigma(\varepsilon_j, X_{j+1} : j \leq t)\) be the \(\sigma\)-algebra generated by \(\varepsilon_j, X_{j+1}, j \leq t\). Then \(\{\mathcal{F}_t\}_{t \in \mathbb{Z}}\) 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, \(\{Z_t\}_{t \in \mathbb{Z}}\) is any sequence of random variables):

(C4) For any \(\{Z_t \in \mathcal{L}^1\}_{t \in \mathbb{Z}}, \{\mathcal{P}_i Z_t\}_{t \in \mathbb{Z}}\) are martingale differences with respect to the increasing filtration \(\{\mathcal{F}_t\}_{t \in \mathbb{Z}}\). Thus, \(\sum_{i=1}^n \mathcal{P}_i Z_t\) is a martingale with respect to \(\mathcal{F}_n\).

(C5) For any \(\{Z_t \in \mathcal{L}^2\}_{t \in \mathbb{Z}}\), \(\sum_{i=1}^n \|\mathcal{P}_i Z_t\|_2^2 = \sum_{i=1}^n \|P_i Z_t\|_2^2\).

(C6) For any \(Z \in \mathcal{L}^2\), \(\|\mathcal{P}_i Z_i\|_2^2 \leq \mathbb{E}(Z|\mathcal{F}_i)\|_2^2 \leq \|Z\|_2^2\).

(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 \(Z \in \mathcal{L}^2\), \(\mathbb{E}(\mathcal{P}_i Z^2|\mathcal{F}_{i-1}) \leq \mathbb{E}(Z^2|\mathcal{F}_{i-1})\).

Proof. By definition, \(\mathcal{P}_i Z_t\) is \(\mathcal{F}_i\)-measurable. Furthermore, by property (C2), \(\mathbb{E}(\mathcal{P}_i Z_t|\mathcal{F}_{i-1}) = \mathbb{E}(\mathbb{E}(Z_t|\mathcal{F}_i)|\mathcal{F}_{i-1}) - \mathbb{E}(Z_t|\mathcal{F}_{i-1}) = 0\). Thus, (C4) holds. By (C4), (C5) follows from the orthogonality of martingale differences. To see (C6), let \(Z^* = \mathbb{E}(Z|\mathcal{F}_n)\), by property (C2), \(\mathbb{E}(Z^*|\mathcal{F}_{i-1}) = \mathbb{E}(Z|\mathcal{F}_{i-1})\). Thus, \(E(Z^*) = \mathbb{E}([E(Z|\mathcal{F}_i)^2 - \mathbb{E}(Z|\mathcal{F}_{i-1})^2] \leq \mathbb{E}(Z|\mathcal{F}_i)^2 \leq \mathbb{E}(Z|\mathcal{F}_{i-1})^2\) \(\|\mathcal{P}_i Z_i\|_2^2 \leq \mathbb{E}(Z|\mathcal{F}_i)\|_2^2 \leq \|Z\|_2^2\), 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 \(\mathbb{E}(\mathcal{P}_i Z^2|\mathcal{F}_{i-1}) \leq \mathbb{E}(\mathbb{E}(Z^2|\mathcal{F}_i)^2|\mathcal{F}_{i-1}) \leq \mathbb{E}(\mathbb{E}(Z^2|\mathcal{F}_i)|\mathcal{F}_{i-1}) = \mathbb{E}(Z^2|\mathcal{F}_{i-1})\). □

By (C4), \(\{\mathcal{F}_i\}_{t \in \mathbb{Z}}\) are martingale difference operators with respect to \(\{\mathcal{F}_t\}_{t \in \mathbb{Z}}\). 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 \(\alpha\)-mixing conditions on \(\{X_i\}\). Lemmas 1–2 present some useful results for \(\alpha\)-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

\[
|\text{cov}(U, V)| \leq 8\alpha(U, V)^{1-1/p-1/q}\|U\|_p\|V\|_q.
\]

Here \(\alpha(U, V)\) is the \(\alpha\)-mixing coefficient between the two \(\sigma\)-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 \(|f(y)| = |f(a)| + \int_a^y f'(z)dz| \leq |f(a)| + \int_a^b |f'(z)|dz|\) for all \(y \in [a, b]\). Thus, using \((u+v)^2 \leq 2(u^2+v^2)\) and the Cauchy–Schwarz inequality \(\int_a^b |f'(z)|dz| \leq (b-a) \int_a^b |f'(z)|^2dz\), we have the uniform bound:

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

(17)
Clearly, if \( f(\cdot) \) is a random function, then taking expectation on both sides of (17) gives
\[
\left\| \sup_{y \in [a,b]} f(y) \right\|_2^2 \leq 2 \left\{ \|f(a)\|_2^2 + (b - a) \int_a^b \|f'(z)\|_2^2 dz \right\}
\leq 2 \left\{ \|f(a)\|_2^2 + (b - a)^2 \sup_{y \in [a,b]} \|f'(y)\|_2^2 \right\}.
\]

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 \( \alpha \)-mixing stationary process with mixing coefficient \( \alpha_k, k \in \mathbb{N} \). For a bivariate measurable and differentiable function \( h \), define
\[
H(y) = \sum_{i=1}^n \left\{ h(y, X_i) - E[h(y, X_i)] \right\}.
\]
Suppose there exists some \( \delta > 2 \) such that \( c := \sup_{y \in \mathbb{R}} [\|h(y, X_1)\|_2 + \|\partial h(y, X_1)/\partial y\|_2] < \infty \) and \( \sum_{k=1}^\infty \alpha_k^{1-2/\delta} < \infty \). Let \( [a, b] \) be any given bounded interval. Then
\[
E \left[ \sup_{y \in [a,b]} \|H(y)\|^2 \right] = O(n).
\]
Furthermore, if \( b_n \to 0 \) and \( w(\cdot) \) is an integrable function with bounded support, then
\[
E \left[ \sup_{y \in [a,b]} \left| \int_{\mathbb{R}} w(u)H(y - ub_n)du \right|^2 \right] = O(n).
\]

**Proof.** Let \( \gamma_k = \text{cov}(h(y, X_i), h(y, X_{i+1})). \) Then \( \gamma_0 \leq \|h(y, X_1)\|_2^2 \leq \|h(y, X_1)\|_2^2 \leq c^2. \) For \( k \geq 1 \), by Lemma 1, \( |\gamma_k| \leq 8\alpha_k^{1-2/\delta} \|h(y, X_1)\|_2^2 \leq 8\alpha_k^{1-2/\delta} c^2. \) Thus,
\[
\|H(y)\|_2^2 = n\gamma_0 + 2 \sum_{k=1}^{n-1} (n - k)\gamma_k \leq n \left( \gamma_0 + 2 \sum_{k=1}^n |\gamma_k| \right) \leq nc^2 \left( 1 + 16 \sum_{k=1}^{\infty} \alpha_k^{1-2/\delta} \right).
\]

Similarly, using \( H'(y) = \sum_{i=1}^n \left\{ \partial h(y, X_i)/\partial y - E[\partial h(y, X_i)/\partial y] \right\} \), we have
\[
\|H'(y)\|_2^2 \leq nc^2 \left( 1 + 16 \sum_{k=1}^{\infty} \alpha_k^{1-2/\delta} \right).
\]

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 \( y - ub_n \in [a - b + 1] \) for sufficiently large \( n \). Thus,
\[
\sup_{y \in [a,b]} \left| \int_{\mathbb{R}} w(u)H(y - ub_n)du \right| \leq \sup_{z \in [a-1, b+1]} |H(z)| \int_{\mathbb{R}} |w(u)|du.
\]
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 \( \tilde{g}_n(y) \) in (10). Define
\[
\tilde{f}_y(y) = \frac{1}{nb_n} \sum_{i=1}^n K_{b_n}(y - Y_i - 1),
\]
\[
\tilde{g}_i(y) = [g(Y_i) - g(Y_{i-1})] K_{b_n}(y - Y_{i-1}).
\]
By the definition of \( \hat{g}_S \), we have \( \mathbb{E}[\xi_i(y)] = \mathbb{E}[\mathbb{E}[\xi_i(y)|Y_{i-1}]] = 0 \). Therefore, we can write

\[
\hat{g}_S(y) - g_S(y) = \frac{\sum_{i=1}^{n}[\xi_i(y) - \mathbb{E}[\xi_i(y)]]}{nb_n f(y)} + \frac{\sum_{i=1}^{n}[g_S(Y_{i-1}) - \hat{g}_S(y)]K_0(y - Y_{i-1})}{nb_n f(y)} := T_n(y) + U_n(y).
\]

In (26), \( T_n(y) \) is the stochastic component determining the asymptotic distribution of \( \hat{g}_S(y) \), and \( U_n(y) \) is the bias component. By Lemma 4, \( f(y) = f_1(y) + O(b_n^2 + (nb_n)/\log n)^{-1/2} \) uniformly in \( y \in \mathcal{Y} \). Furthermore, by Lemma 5, \( U_n(y) = \rho_S(y)b_n^2 + \varphi_n(y)b_n^2 + O_p((nb_n)/\log n)^{-1/2} \) 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 \( \mathcal{P}_1 \) in Section 4.2 to write the decomposition

\[
\sum_{i=1}^{n}[\xi_i(y) - \mathbb{E}[\xi_i(y)]] = \sum_{i=1}^{n}[\mathcal{P}_1 \xi_i(y) + \mathcal{P}_{i-1} \xi_i(y) + \mathbb{E}[\xi_i(y)|\mathcal{F}_{i-2}] - \mathbb{E}[\xi_i(y)]]
\]

\[
= \sum_{i=1}^{n}[\mathcal{P}_1 \xi_i(y) + \mathcal{P}_{i+1} \xi_i(y)] + \sum_{i=1}^{n}[\mathbb{E}[\xi_i(y)|\mathcal{F}_{i-2}] - \mathbb{E}[\xi_i(y)]] + [\mathcal{P}_0 \xi_1(y) - \mathcal{P}_n \xi_{n+1}(y)]
\]

\[
:= S_n(y) + R_n(y) + M_n(y).
\]

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 \( [S_n(y_1), \ldots, S_n(y_k)]^T \) for distinct \( y_1, \ldots, y_k \). 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 \( K(\cdot) \),

\[
\sup_y |\mathcal{P}_0 \xi_1(y)| \leq \sup_u |K(u)| \left[ \mathbb{E}[|g(Y_1) - \hat{g}_S(Y_0)||\mathcal{F}_0] + \mathbb{E}[|g(Y_1) - \hat{g}_S(Y_0)||\mathcal{F}_{-1}] \right].
\]

Thus, by property (C1) in Section 4.2,

\[
\mathbb{E}\left[ \sup_y |\mathcal{P}_0 \xi_1(y)| \right] \leq 2 \sup_u |K(u)| \left[ \mathbb{E}[|g(Y_1)|] + \mathbb{E}[|g(Y_0)|] \right] = O(1),
\]

where we have \( \mathbb{E}[|g(Y_1)|] = \mathbb{E}[\mathbb{E}[|g(Y_1)||Y_0]] \leq \mathbb{E}[\mathbb{E}[|g(Y_1)|]|Y_0] = \mathbb{E}[g(Y_1)] = O(1) \). Similarly, \( \mathbb{E}[\sup_y |\mathcal{P}_0 \xi_{n+1}(y)|] = O(1) \). 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. \hfill \dagger

**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

\[
\xi_{i,1}(y) = g(Y_i)K_0(y - Y_{i-1}) \quad \text{and} \quad \xi_{i,2}(y) = \hat{g}_S(Y_{i-1})K_0(y - Y_{i-1}).
\]

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

\[
J_i(y) = \sum_{r=1}^{2} \left\{ \mathbb{E}[\xi_{i,r}(y)|\mathcal{F}_{i-2}] - \mathbb{E}[\xi_{i,r}(y)] \right\}, \quad r = 1, 2.
\]

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

\[
\mathbb{E}[\xi_{i,1}(y)|\mathcal{F}_{i-2}, X_{i-1}] = \mathbb{E}[g(Y_i)K_0(y - X_{i-1} - \varepsilon_{i-1})|X_{i-1}, X_i]
\]

\[
= \mathbb{E}\left[ g(Y_i) \int_{\mathbb{R}} K_0(y - X_{i-1} - v)q_i(v)dv|X_{i-1}, X_i \right]
\]

\[
= b_n \int_{-\infty}^{\infty} K(u)e_i(u)du, \quad e_i(u) = \mathbb{E}[g(Y_i)q_i(y - X_{i-1} - ub_i)|X_{i-1}, X_i].
\]

Here, the last equality follows from the transformation \( u = (y - X_{i-1} - v)/b_i \). Note that the conditional expectation \( e_i(u) \) is a function of \( X_{i-1}, X_i \). Thus,
Thus, using 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 \( \{ e_i \}_{i \in \mathbb{N}} \) as well as the Markovian assumption on \( \{ X_i \}_{i \in \mathbb{N}} \), and the third equality follows from \( \mathbb{E}[e_i(u)|X_{i-1}] = \mathbb{E}[g(Y_i)q_r(y - X_{i-1} - ub_n)|X_{i-1}] \) (property (C2) in Section 4.2). Define \( h(z, X_{i-1}) = \mathbb{E}[g(Y_i)q_r(z - X_{i-1})|X_{i-1}] \). (After taking conditional expectation, it is a function of \( X_{i-1} \).) Then, using \( \mathbb{E}[\xi_{1,1}(y)|\mathcal{F}_{i-2}] \) and by (29), we obtain

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

Since \( q_r(\cdot) \) is bounded, by property (C3) in Section 4.2, we can easily see that \( \|h(z, X_0)\|_8 = O(1)\|\mathbb{E}[g(Y_1)|X_0]\|_8 \leq O(1)\|g(Y_1)\|_8 \) for all \( z \in \mathbb{R} \). Similarly, using \( \partial h(z, X_0)/\partial z = \mathbb{E}[g(Y_1)q_r'(z - X_0)|X_0] \) and the boundedness of \( q_r'(\cdot) \), we have \( \|\partial h(z, X_0)/\partial z\|_8 \leq O(1)\|g(Y_1)\|_8 \) 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 \( Y_{i-1} = X_{i-1} + e_i \), we obtain

\[
\mathbb{E}[\xi_{1,2}(y)|\mathcal{F}_{i-2}] = \int_{\mathbb{R}} g_i(X_{i-1} + v)K_n(y - X_{i-1} - v)q_r(v)dv
\]

\[
\mathbb{E}[\xi_{1,2}(y)|\mathcal{F}_{i-2}] = b_n \int_{-\infty}^{\infty} K(u)g_i(y - ub_n)q_r(y - X_{i-1} - ub_n)du.
\]

Thus, using \( \mathbb{E}[\xi_{1,2}(y)] = \mathbb{E}[\mathbb{E}[\xi_{1,2}(y)|\mathcal{F}_{i-2}]] \), we have

\[
J_2(y) = b_n \int_{-\infty}^{\infty} K(u)g_i(y - ub_n)L(y - ub_n)du,
\]

where \( L(z) = \sum_{i=0}^{n-1}[q_r(z - X_i) - \mathbb{E}q_r(z - X_i)] \). Since \( g_i(\cdot) \) is bounded in the neighborhood \( \mathcal{Y} \), the claim then follows from (20) in Lemma 2. \( \diamond \)

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

**Proof.** Let \( \gamma(y) = K_n(y - Y_i) \). Observe the decomposition

\[
\bar{f}(y) = \frac{1}{nb_n}H_1(y) + \frac{1}{nb_n}H_2(y) + \frac{\mathbb{E}[\gamma(y)]}{b_n}.
\]

where

\[
H_1(y) = \sum_{i=0}^{n-1}[\gamma(y) - \mathbb{E}[\gamma(y)|\mathcal{F}_{i-1}]] = \sum_{i=0}^{n-1} \mathbb{E}\gamma_i(y)
\]

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

By the symmetry of \( K(\cdot) \), we can show \( b_n^{-1}\mathbb{E}[\gamma_i(y)] = f_i(y) + O(b_n^2) \). To prove the desired result, it suffices to show \( \sup_{y \in \mathcal{Y}} |H_1(y)| = O_p(\sqrt{\log n}b_n) \) 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), \( H_2(y) = b_n \int_{-\infty}^{\infty} K(u)q_r(y - ub_n)du \) 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 \( y_j = jT/N, j = -N, -N+1, \ldots, N-1, N \). Then \( y_{-N}, \ldots, y_N \) 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 \( y \in [y_j, y_{j+1}] \),

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

Thus, \( \sup_{y \in \mathcal{Y}} |H_1(y)| \leq \max_{j=-N}^{N} |H_1(y_j)| + O((nb_n)^{-1}) \), and consequently,

\[
\sup_{y \in \mathcal{Y}} |H_1(y)| \leq \max_{j=-N}^{N} |H_1(y_j)| + O((nb_n)^{-1}).
\]
By property (C8) in Section 4.2, \( \sum_{i=0}^{n-1} \mathbb{E}[\left| \mathcal{P}_i \gamma_i(y_i) \right|^2 | \mathcal{F}_{i-1}] \leq \sum_{i=0}^{n-1} \mathbb{E}[\gamma_i^2(y_i) | \mathcal{F}_{i-1}] \leq c_2 nb_n \) for some constant \( c_2 \). Let \( c_1 = \sup_y |K(u)| \). Then \( |\mathcal{P}_i \gamma_i(y_i)| \leq 2c_3 \). Thus, by Freedman’s exponential inequality for bounded martingale differences [14], for any \( \epsilon > 0 \),

\[
p_j := \mathbb{P}\left( |H_1(y_i)| \geq c \sqrt{nb_n \log n} \right) \leq 2 \exp\left( - \frac{c^2 nb_n \log n}{2(2c_3c \sqrt{\log n} + c_2 nb_n)} \right) = 2 \exp(-\lambda_n \log n),
\]

where

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

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

\[
\mathbb{P}\left( \max_{j=-N,-N+1} \cdots N \left| H_1(y_j) \right| \geq c \sqrt{nb_n \log n} \right) \leq \sum_{j=-N}^{N} p_j = O(Nn^{-\lambda_n}) = O(1/n) \to 0.
\]

Therefore, \( \max_{j=-N,-N+1} \cdots N \left| H_1(y_j) \right| = O_p(\sqrt{nb_n \log n}) \). The result then follows from (35).

\[\]

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

\[
\sum_{i=1}^{n} \left[ g_y(Y_{i-1}) - g_y(y) \right] K_{n_b}(y - Y_{i-1}) = nb_n^2 f_y(y) \{ \rho_y(y) + O_p[2 + (nb_n^2 \log n)^{-1/2}] \}.
\]

**Proof.** We adopt the same argument in Lemma 4. Let \( \eta_i(y) = \left[ g_y(Y_{i-1}) - g_y(y) \right] K_{n_b}(y - Y_{i-1}) \). By the same decomposition in (32), we have

\[
\sum_{i=0}^{n-1} \eta_i(y) = N_1(y) + N_2(y) + n \epsilon_1[\eta_1(y)].
\]

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

\[
N_2(y) = b_n \int_{0}^{b_n} K(u) [g_y(y - ub_n) - g_y(y)] L(y - ub_n) du,
\]

where \( L(\cdot) \) is defined in (31). Note that \( \left| g_y(y - ub_n) - g_y(y) \right| = O(b_n) \). 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 \( g_y(Y_i) - g_y(y) = O(b_n) \) for \( y - Y_i = O(b_n) \), we have \( \mathbb{E}[\eta_1^2(y) | \mathcal{F}_{i-1}] = O(b_n^2) \mathbb{E}[ K_{n_b}^2(y - Y_i) | \mathcal{F}_{i-1}] = O(b_n^4) \). Thus, by property (C8) in Section 4.2, the conditional variance satisfies \( \sum_{i=0}^{n-1} \mathbb{E}[\eta_i^2(y) | \mathcal{F}_{i-1}] \leq \sum_{i=0}^{n-1} \mathbb{E}[\eta_i^2(y) | \mathcal{F}_{i-1}] = O(nb_n^2) \). By the same chain argument in the proof of \( H_1(y) \) in Lemma 4, we can show \( \sup_{y \in \mathcal{Y}} |N_1(y)| = O_p(\sqrt{nb_n^2 \log n}) \).

\[\]

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

\[
d_i(y) = \frac{\mathbb{E}[\xi_i(y) + \xi_{i+1}(y)]}{\sigma_y(y) \sqrt{nb_n^2 f_y(y)}}
\]

Then

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

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

\[
\mathbb{E}[\mathcal{P}_i (\xi_i + \xi_{i+1})^2 | \mathcal{F}_{i-1}] = 2 \mathbb{E}(\xi_i \xi_{i+1} | \mathcal{F}_{i-1}) - 2 \mathbb{E}(\xi_i | \mathcal{F}_{i-1}) \mathbb{E}(\xi_{i+1} | \mathcal{F}_{i-1}) - \mathbb{E}(\xi_{i+1} | \mathcal{F}_{i-1})^2 + \mathbb{E}(\xi_i^2 | \mathcal{F}_{i-1}) - \mathbb{E}(\xi_i | \mathcal{F}_{i-1})^2 + \mathbb{E}(\mathbb{E}(\xi_{i+1} | \mathcal{F}_{i}) | \mathcal{F}_{i-1})^2 | \mathcal{F}_{i-1})
\]

\[
\stackrel{\text{def}}{=} 2A_i - 2A_{i-2} - A_{i+3} + A_{i-4} - A_{i-5} + A_{i-6}.
\]
Below we consider each of these six terms separately. For convenience, sometimes we give bounds for \(\|\cdot\|_2\), which dominates \(\|\cdot\|_{3/2}\).

(A1,5 and A1,6 terms:) Let \(v_i = [E(\xi_{i+1}| \mathcal{F}_i)]^2\). Note that \(A_{i+1.5} - A_{i.6} = \mathcal{P}_i v_i\). By consecutively using properties (C5), (C6) and (C3) in Section 4.2, we have

\[
\left\| \sum_{i=1}^{n} (A_{i+1.5} - A_{i.6}) \right\|_2^2 = \sum_{i=1}^{n} \left\| \mathcal{P}_i v_i \right\|_2^2 \leq \sum_{i=1}^{n} \left\| v_i \right\|_2^2 \leq n E(\xi_i^4) = O(nb_n). \tag{41}
\]

Therefore, by the triangle inequality,

\[
\left\| \sum_{i=1}^{n} (A_{i.5} - A_{i.6}) \right\|_2 = \left\| v_0 - v_n + \sum_{i=1}^{n} (A_{i+1.5} - A_{i.6}) \right\|_2 \leq O(\sqrt{nb_n}). \tag{42}
\]

(A1,3 terms:) Recall \(\xi_{1,1}(y)\) and \(\xi_{1,2}(y)\) in (28). By \(\xi_{1,1} = \xi_{1,2}(y)\), (29) and (30), and the boundedness of \(K(\cdot), q_\epsilon(\cdot)\) and \(g_\delta(y), y \in \mathcal{Y}_i\), there exists a constant \(c_1\) such that

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

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

\[
\left\| \sum_{i=1}^{n} A_{1.3} \right\|_2 \leq 2 c_1^2 b_n^2 \sum_{i=1}^{n} \{1 + E[|g(Y_{i+1})|]\}^2 \leq 2 c_1^2 b_n^2 \sum_{i=1}^{n} \left\| 1 + E[|g(Y_{i+1})|]\}^2 \right\|_2 = O(nb_n^2). \tag{44}
\]

Here, the last equality follows from \(\|E[|g(Y_{i+1})|]\}^2 \leq \|g(Y_{i+1})\}^2 \leq \|g(Y)\}^2 < \infty\).

(A1,2 terms:) By the bounded support of \(K(\cdot)\), it suffices to consider \(|Y_{i-1} - y| = O(b_n)\) in \(\xi_i\) so that \(g_\delta(y_{i-1}) \leq c_2\) for some constant \(c_2\). Note that \(E[|g(Y)|]\}^2 \leq E[|g(Y)|]\}^2 \leq K_{b_n}(y - Y_{i-1})|c_2 + E||g(Y)|\}^2\}. Combining this with (43) gives

\[
|A_{1.2}| \leq c_1 b_n K_{b_n}(y - Y_{i-1})|Z_i|, \quad Z_i = \{c_2 + E[|g(Y)|]\}^2 \{1 + E[|g(Y)|]\}^2 \}]. \tag{45}
\]

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

\[
\left\| \sum_{i=1}^{n} A_{1.2} \right\|_{3/2} \leq \sum_{i=1}^{n} \left\| A_{1.2} \right\|_{3/2} \leq c_1 b_n \sum_{i=1}^{n} \left\| K_{b_n}(y - Y_{i-1})Z_i \right\|_{3/2} = O(nb_n^{5/3}). \tag{46}
\]

Here, we have used \(\|Z_i\|_{3/2} \leq \|Z_i\|_2 < \infty\) by the condition \(g(Y) \in L^{\delta}\) with \(\delta > 4\).

(A1,1 terms:) In \(\xi_i\), 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, \(g_\delta(y_{i-1}) + g_\delta(y_i) + |g_\delta(Y_{i-1})| \leq c_4\) for some constant \(c_4\), and consequently \(|\xi_i| \leq c_4|K_{b_n}(y - Y_{i-1})|\}. Also, conditioning on \(\mathcal{F}_{i-1}\), the \(\xi_i\) term in \(Y_i = X_i + \xi_i\) 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,

\[
|A_{1.1}| \leq O(b_n)|K_{b_n}(y - Y_{i-1})|\}c_4 + E||g(Y_{i+1})|]\}F_{i-1}|\}. \tag{47}
\]

By the independence between \(\{\xi_i\} \) and \(\{X_i\}\) as well as the Markovian assumption on \(\{X_i\}\), \(E[|g(Y_{i+1})|]\}F_{i-1} = E[|g(Y_{i+1})|]\}X_i\].

The same argument in (46) then gives

\[
\left\| \sum_{i=1}^{n} A_{1.1} \right\|_{3/2} \leq O(b_n) \sum_{i=1}^{n} \left\| K_{b_n}(y - Y_{i-1})\}c_4 + E||g(Y_{i+1})|]\}X_i\} \right\|_{3/2} = O(nb_n^{5/3}). \tag{48}
\]

(A4 terms:) Write \(A_{i.4} = E(\xi^2_i) = \mathcal{P}_{i-1}\xi^2_i + E(\xi^2_i|\mathcal{F}_{i-2}) - E(\xi^2_i)\). Since \(\mathcal{P}_{i-1}\xi^2_i \) are martingale differences with respect to \(\{\mathcal{F}_{i-1}\}_{i \in \mathbb{Z}}\), by the same argument in (41),

\[
\left\| \sum_{i=1}^{n} \mathcal{P}_{i-1}\xi^2_i \right\|_2 = O(\sqrt{nb_n}). \tag{49}
\]
By the same argument in (29) and (30), it can be shown that
\[
\mathbb{E}(\xi_i^2|\mathcal{F}_{i-2}) = b_n \int_{-\infty}^{\infty} K^2(u) \mathbb{E}[(g(Y_i) - g_\delta(y - ub_{n})]^2 q_{\delta}(y - X_{i-1} - ub_{n})|X_{i-1}).
\]
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
\[
\left\| \sum_{i=1}^{n} [A_{i,4} - \mathbb{E}(\xi_i^2)] \right\|_2 \leq \left\| \sum_{i=1}^{n} \mathcal{P}_{i-1}\xi_i^2 \right\|_2 + \left\| \sum_{i=1}^{n} [\mathbb{E}(\xi_i^2|\mathcal{F}_{i-2}) - \mathbb{E}(\xi_i^2)] \right\|_2 = O(\sqrt{nb_n}).
\]

Elementary calculation shows that \( \mathbb{E}(\xi_i^2) = b_n \varphi_{k}\lambda(y)\sigma_\delta(y) + O(b_n^2). \) 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,
\[
\lim_{n \to \infty} \mathbb{P} \left\{ \sup_{y \in \mathcal{Y}_n} \frac{|S_n(y)|}{\sigma_\delta(y)\sqrt{nb_n}\varphi_{k}\lambda(y)} \leq B_{m_n}(z) \right\} = e^{-2e^{-2}}, \quad z \in \mathbb{R}.
\]

**Proof.** Let \( d_j(y) \) be defined in (38). Write
\[
\tilde{S}_n(y) := \frac{S_n(y)}{\sigma_\delta(y)\sqrt{nb_n}\varphi_{k}\lambda(y)} = \sum_{i=1}^{n} d_i(y).
\]
Write \( y_{\mathcal{Y}} = \{y_1 < \cdots < y_{m_n}\}. \) For fixed \( k \in \mathcal{N} \) distinct integers \( 1 \leq j_1, j_2, \ldots, j_k \leq m_n, \) 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.,
\[
Q_n = \sum_{i=1}^{n} \mathbb{E}(D_iD_i^T|\mathcal{F}_{i-1}) := (q_{n,s})_{1 \leq r, s \leq k}.
\]

Let \( \tau_n = \varphi_{k}\sigma_\delta(y_{j_1})\sigma_\delta(y_{j_k})\sqrt{\lambda(y_{j_1})\lambda(y_{j_k})}. \) Then we can write \( q_{n,s} \) as
\[
q_{n,s} = \sum_{i=1}^{n} \mathbb{E}[d_i(y_{j_1})d_i(y_{j_k})|\mathcal{F}_{i-1}]
\]
\[
= \frac{1}{nb_n \tau_n} \sum_{i=1}^{n} \left[ \mathbb{E}[\mathcal{P}_{i}\xi_i(y_{j_1})\mathcal{P}_{i}\xi_i(y_{j_k})|\mathcal{F}_{i-1}] + \mathbb{E}[\mathcal{P}_{i}\xi_{i+1}(y_{j_1})\mathcal{P}_{i}\xi_i(y_{j_k})|\mathcal{F}_{i-1}]
\]
\[
+ \mathbb{E}[\mathcal{P}_{i}\xi_i(y_{j_1})\mathcal{P}_{i}\xi_{i+1}(y_{j_k})|\mathcal{F}_{i-1}] + \mathbb{E}[\mathcal{P}_{i}\xi_{i+1}(y_{j_1})\mathcal{P}_{i}\xi_{i+1}(y_{j_k})|\mathcal{F}_{i-1}] \right].
\]

For \( r = s, \) by Lemma 6, \( \|q_{n,r} - I|_{3/2} = O((nb_n)^{-1/2} + b_n^{2/3}). \) For \( r \neq s, \) by the definition of \( y_{\mathcal{Y}}, \) since \( |y_{\mathcal{Y}}| \geq \tau_n, b_n = o(\tau_n), \) and the kernel function \( K(\cdot) \) has bounded support, we have \( K_{b_n}(y_{\mathcal{Y}} - Y_{i-1} - Y_{i-1} = 0 \) for large enough \( n. \) Thus, \( P_{\xi_1}(y_{j_1})P_{\xi_1}(y_{j_k}) = 0. \) 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 \( y_{\mathcal{Y}} \) 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 \( I_{n,s} \) be the \((r,s)\)-element of the \( k \times k \) identity matrix, then \( \|q_{n,s} - I_{n,s}\|_{3/2} = O((nb_n)^{-1/2} + b_n^{2/3}) \) uniformly over \( 1 \leq r, s \leq k. \) It is easily seen that \( \sum_{i=1}^{n} \mathbb{E}[d_i(y_{j_1})]^2 = O((nb_n)^{-1/2}) \) uniformly over \( 1 \leq r, s \leq k. \) Thus \( \sum_{i=1}^{n} \mathbb{E}[d_i(y_{j_1})]^2 = O(\tau_n) \) uniformly, where \( \tau_n = (nb_n)^{-1/2} + b_n. \)

For \( j = 1, \ldots, m_n, \) define events \( T_j = \{ |\tilde{S}_n(y_{j})| \geq B_{m_n}(z) \}. \) Then \( \mathbb{P}\left( \sup_{y \in \mathcal{Y}_n} |\tilde{S}_n(y)| \geq B_{m_n}(z) \right) = \mathbb{P}(\cup_{j=1}^{m_n} T_j). \) By the same argument in the proof of Theorem 3 in [31], we can show \( \mathbb{P}(\cup_{j=1}^{m_n} T_j) \rightarrow 1 - \exp\{-2e^{-2}\}. \) The details are omitted. ⊓⊔
4.5. Proof of Proposition 2

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

**Proof.** By (26), (27), Lemmas 3–4, it suffices to prove $\sup_{y \in Y} |S_n(y)| = O_p(\sqrt{n} \log n)$, 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 $y, \ldots, y_N$ be the grid points defined in the proof of Lemma 4. By the same chain argument in Lemma 4, it suffices to prove $\max_{-N \leq j \leq N} |S_n(y_j)| = O_p(\sqrt{n} \log n)$. 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

$$A_1 = \max_{1 \leq i \leq n} \{\mathbb{E}(\xi_i|\mathcal{F}_i) + \mathbb{E}(\xi_i|\mathcal{F}_{i-1})\}, \quad \xi_i = |g(Y_i)| + |\hat{g}_S(Y_{i-1})| + |g(Y_{i+1})| + |\hat{g}_S(Y_i)|,$$

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

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

$$|\mathcal{P}_i[\xi_i(y) + \xi_{i+1}(y)]| \leq c_1 A_1, \quad \text{uniformly in } i = 1, \ldots, n, \ y \in \mathbb{R}. \quad (52)$$

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

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

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

$$p_j := \mathbb{P}\left\{ |S_n(y_j)| \geq c \sqrt{n} \log n, A_1 \leq n^{1/4} \log n, A_2 \leq n \log n \right\} \leq 2 \exp\left\{ -\frac{c^2 n(\log n)^2}{2[(c_1 n^{1/4} \log n)(c \sqrt{n} \log n) + 4c_1^2 n \log n]} \right\} \leq 2 \exp(-\lambda_n \log n), \quad (54)$$

where

$$\lambda_n = \frac{c^2}{2[(c_1 n^{-1/4} \log n)^2 + 4c_1^2]}. \quad \text{For large enough } c \text{ and } n, \text{ we have } \lambda_n > 3. \text{ Thus,}$$

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

By (54) and $N = n^{3/2}, \sum_{j=-N}^{N} p_j = O(1/n) \to 0$. Note that

$$\mathbb{E}(A_1^4) \leq \sum_{i=1}^n \mathbb{E}[(\mathbb{E}(\xi_i|\mathcal{F}_i) + \mathbb{E}(\xi_i|\mathcal{F}_{i-1}))^4] \leq 16 \sum_{i=1}^n \mathbb{E}[(\mathbb{E}(\xi_i|\mathcal{F}_i))^4 + (\mathbb{E}(\xi_i|\mathcal{F}_{i-1}))^4] \leq 32n\mathbb{E}(\xi_i^4).$$

Here the second “≤” follows from $(u + v)^4 \leq 16(u^4 + v^4)$ and the third “≤” follows from property (C3) in Section 4.2. Thus, by Markov’s inequality, $\mathbb{P}\{A_1 > n^{1/4} \log n\} \leq \mathbb{E}(A_1^4)/(n^{1/4} \log n)^4 = O((\log n)^{-4})$. Another application of Markov’s inequality gives $\mathbb{P}\{A_2 > n \log n\} \leq \mathbb{E}(A_2)/(n \log n) = O((\log n)^{-1})$. Thus, the right hand side of (55) goes to zero, and we conclude $\max_{-N \leq j \leq N} |S_n(y_j)| = O_p(\sqrt{n} \log n)$, completing the proof. \hfill \Box
Proof of Proposition 2. (i) By the same argument in Lemma 4, \( \sup_{y \in \mathcal{Y}} |\hat{f}_T(y) - f_T(y)| = O_p[\ell_n^2 + (nl_n / \log n)^{-1/2}] = o_p[(\log n)^{-1/2}] \).

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

\[
\hat{\sigma}_k^2(y) = \frac{\sum_{i=1}^n [g(Y_i) - g_k(Y_{i-1})]^2 K_{b_n}(y - Y_{i-1})}{\sum_{i=1}^n K_{b_n}(y - Y_{i-1})}.
\]

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 \( \mathcal{Y}_{1/2} \) of \( y \) so that \( \sup_{y \in \mathcal{Y}} |\hat{g}_k(y) - g_k(Y_{i-1})| \leq \Delta_1 \). Applying the inequality \( |a^2 - b^2| \leq x(x + 2|b|) \) for all \( |a - b| \leq x \), we have

\[
|\hat{\sigma}_k^2(y) - \sigma_k^2(y)| \leq \Delta_1 \frac{\sum_{i=1}^n [\Delta_1 + 2|g(Y_i) - g_k(Y_{i-1})|]K_{b_n}(y - Y_{i-1})}{\sum_{i=1}^n K_{b_n}(y - Y_{i-1})}.
\]

(56)

Let \( \Delta_2 = \max_{1 \leq i \leq n} |g(Y_i) - g_k(Y_{i-1})| \). Then \( \mathbb{E}(\Delta_2^4) \leq \sum_{i=1}^n \mathbb{E}[|g(Y_i) - g_k(Y_{i-1})|^4] = O(n) \), which implies \( \Delta_2 = O_p(n^{1/4}) \).

Thus, by (56) and Lemma 4, \( \hat{\sigma}_k^2(y) - \sigma_k^2(y) = O_p[(\Delta_1^2 + \Delta_1 \Delta_2)] = o_p[(\log n)^{-1/2}] \), 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 \( \sup_{y \in \mathcal{Y}} |\hat{\sigma}_k^2(y) - \sigma_k^2(y)| = O_p[\ell_n^2 + (b_n \sqrt{n})^{-1} \log n] = o_p[(\log n)^{-1/2}] \). This completes the proof. \( \diamondsuit \)

References


