Published in final edited form as:

J Econom. 2011 June ; 162(2): 225–239. doi:10.1016/j.jeconom.2011.01.002.

Nonparametric model validations for hidden Markov models with applications in financial econometrics

Zhibiao Zhao*

Department of Statistics, Penn State University, University Park, PA 16802, United States

Abstract

We address the nonparametric model validation problem for hidden Markov models with partially observable variables and hidden states. We achieve this goal by constructing a nonparametric simultaneous confidence envelope for transition density function of the observable variables and checking whether the parametric density estimate is contained within such an envelope. Our specification test procedure is motivated by a functional connection between the transition density of the observable variables and the Markov transition kernel of the hidden states. Our approach is applicable for continuous time diffusion models, stochastic volatility models, nonlinear time series models, and models with market microstructure noise.

Keywords

Confidence envelope; Diffusion model; Hidden Markov model; Market microstructure noise; Model validation; Nonlinear time series; Transition density; Stochastic volatility

1. Introduction

Let $\{X_t\}_{t \in \mathcal{T}}$ be a stationary process with time index $t \in \mathcal{T} = \{0, 1, 2, \dots\}$ in discrete-time setting or the interval $\mathcal{T} = [0, \infty)$ in continuous-time setting. Examples include stock prices, interest rates, temperature series, rainfall measurements, and unemployment rates over a certain period of time among others. The data-generating mechanism underlying the process $\{X_t\}_{t \in \mathcal{T}}$ usually involves some unknown parameter θ which could be finite dimensional real-valued parameters in parametric settings or nonparametric functions in nonparametric inference problems. To conduct statistical inference about θ , one often needs to impose a certain dependence structure on the underlying process $\{X_t\}_{t \in \mathcal{T}}$. For the latter purpose, Markov chains are widely used in virtually every scientific subjects, such as biology, engineering, queueing theory, physics, finance, econometrics, and statistics.

In finance, an important Markov chain example is the continuous-time diffusion model

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \quad (1)$$

where $\{W_t\}_{t \geq 0}$ is a standard Brownian motion, and μ and σ are the drift and volatility functions. Then $\{X_t\}_{t \geq 0}$ is a Markov chain, and so is the discrete sample $\{X_{i\Delta}\}_{i \in \mathbb{N}}$ with Δ being the sampling interval. In nonparametric setting, no parametric forms are imposed on

(μ, σ) , and we may take $\vartheta = (\mu, \sigma)$ to be a vector of two functions. On the other hand, if we know $(\mu, \sigma) = (\mu(\cdot; \theta), \sigma(\cdot; \theta))$ for some known parametric forms $\mu(\cdot; \theta)$ and $\sigma(\cdot; \theta)$ with unknown parameter θ , then we may take $\vartheta = \theta$. The survey paper Zhao (2008) reviewed different specifications of (1). In econometrics, another useful example is the discrete-time version of (1): $X_i = \mu(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i$, where $\varepsilon_i, i \in \mathbb{Z}$, are independent and identically distributed (i.i.d.) random variables. See Section 3.4 for different specifications.

Despite the popularity of Markov chains, the Markovian assumption seems too restrictive in many situations. One distinctive example violating the Markovian assumption is the class of stochastic volatility models. Stochastic volatility model has emerged as a useful alternative to the traditional deterministic volatility model (e.g. Hull and White, 1987; Taylor, 1994; Kim et al., 1998; Ball and Torous, 1999). Consider the continuous-time stochastic volatility model

$$d\log(S_t) = \sigma_t dW_1(t) \quad \text{and} \quad d\sigma_t^2 = r(\sigma_t^2)dt + s(\sigma_t^2)dW_2(t), \quad (2)$$

where $\{W_1(t)\}_{t \geq 0}$ and $\{W_2(t)\}_{t \geq 0}$ are two independent standard Brownian motions. Then the return series $\{X_i = \log(S_{i\Delta}) - \log(S_{(i-1)\Delta})\}_{i \in \mathbb{N}}$ is not a Markov chain. Similarly, for the discrete time stochastic volatility model

$$X_i = \sigma_i \varepsilon_i \quad \text{and} \quad \sigma_i^2 = r(\sigma_{i-1}^2) + s(\sigma_{i-1}^2)\eta_i, \quad (3)$$

where $\{\varepsilon_i\}$ and $\{\eta_i\}$ are two independent i.i.d. sequences. Then $\{X_i\}$ from (3) is not a Markov chain. Volatility plays an important role in risk analysis and options pricing. To study the evolving dynamics of volatilities, we can let $\vartheta = (r, s)$ be the parameter of interest in (2) and (3). Due to the unobservable volatilities, it is more challenging to conduct statistical inferences for (2) and (3) than (1); see Broto and Ruiz (2004).

In this article we study inferences for hidden Markov models (HMM). Given discrete observations $\{X_i\}$ from a stationary process $\{X_i\}_{i \in \mathbb{Z}}$, we assume that $\{(X_i, Y_i)\}$ is a HMM with observable variables $\{X_i\}$ and some hidden or unobservable states $\{Y_i\}$, where $\{Y_i\}$ is a Markov chain; see the definition in Section 2. While Markov chain can be viewed as a special example of HMM, HMM includes many models that otherwise cannot be studied under the Markov chain framework. For example, while the returns $\{X_i = \log(S_{i\Delta}) - \log(S_{(i-1)\Delta})\}_{i \in \mathbb{N}}$ in (2) are non-Markovian, they form HMM with the hidden states $\{Y_i =$

$(\sigma_{s+(i-1)\Delta}, s \in [0, \Delta])\}_{i \in \mathbb{N}}$ or $\left\{Y_i = \left(\sigma_{i\Delta}, \int_{(i-1)\Delta}^{i\Delta} \sigma_s^2 ds\right)\right\}_{i \in \mathbb{N}}$; see Section 3.2. Similarly, for (3), $\{X_i\}$ is not a Markov chain but a HMM with the hidden states $\{Y_i = \sigma_i\}$; see Section 3.5. For a third example, consider

$$X_i = Y_i + \varepsilon_i, \quad (4)$$

where $\{X_i\}$ is the observation sequence, $\{Y_i\}$ is the underlying true but unknown process, and $\{\varepsilon_i\}$ is the measurement or contamination error independent of $\{Y_i\}$. Several researchers have used (4) to model financial markets in the presence of market microstructure noise (e.g. Aït-Sahalia et al., 2005; Zhang et al., 2005). If $\{Y_i\}$ is a Markov chain, then $\{X_i\}$ is a HMM with hidden chain $\{Y_i\}$; see Section 3.6. For other applications of HMM, see the monograph

by MacDonald and Zucchini (1997). Due to the unobservable states, statistical inference for HMM is more challenging than that for Markov chains.

Given observations $\{X_i\}$, researchers want to draw statistical inferences about $\mathcal{Q}$ which generates the HMM with unobservable states $\{Y_i\}$. If we know an *a priori* parametric family $\{\mathcal{Q}_\theta, \theta \in \Theta\}$ for $\mathcal{Q}$, then a parametric setting would be reasonable and the main focus becomes the estimation of parameter θ . In many situations, however, researchers have no or little prior information, and a mis-specification of the underlying model could lead to wrong conclusions. For example, specifying the correct model of the price process of the underlying assets plays a key role in the pricing of derivatives. In such circumstances, it is essential to test the null hypothesis $H_0: \mathcal{Q} = \mathcal{Q}_\theta$ for some unknown parameter $\theta \in \Theta$ before using any parametric model $\mathcal{Q}_\theta$. In different contexts in the literature, the latter hypothesis testing problem is often called model validation, model checking, goodness-of-fit, or specification testing. This is the primary goal of this article.

Nonparametric model validation under dependence has been an important yet difficult problem. A few model-specific model validation approaches have been proposed. Azzalini and Bowman (1993) studied model checking by using pseudo-likelihood ratio test. Härdle and Mammen (1993) proposed measuring the discrepancy between parametric and nonparametric estimates of the mean regression function. For residuals based tests, see Fan and Li (1996) and Hong and White (1995). Fan and Yao (2003) dealt with model validation problem for time series data by using generalized likelihood ratio test in Fan et al. (2001), which has been developed for independent data. Zhao and Wu (2008) studied model validations for time series models through simultaneous confidence bands. Most aforementioned approaches rely on nonparametric regression estimation. Recently, there has been considerable interest in density based specification tests. Aït-Sahalia (1996) introduced a density based test by comparing the parametric and nonparametric density estimates for model (1); see also Hong and Li (2005), Bosq (1998), and Gao and King (2004) for density based approaches. For Markov models, Aït-Sahalia et al. (2009) proposed specification tests based on transition densities. Aït-Sahalia et al. (2010) also used transition densities to test the Markov hypothesis.

In this article we consider model validations for HMM $\{(X_i, Y_i)\}$ with observable variables $\{X_i\}$ and unobservable states $\{Y_i\}$, without imposing any specific model structure. We achieve this goal by constructing a nonparametric simultaneous confidence envelope (SCE) for the transition density or conditional density function of X_i given X_{i-1} . Our specification test procedure is motivated by a functional connection between the transition density of the observable variables $\{X_i\}$ and the Markov transition kernel of the unobservable states $\{Y_i\}$; see Section 2.1 for more discussions. The proposed method constructs nonparametric SCE for transition density and checks whether the parametrically implied density estimate is contained within such an envelope. As demonstrated in Section 3, the proposed method works for a variety of models widely used in financial econometrics, including continuous-time diffusion models, stochastic volatility models, nonlinear time series models, and models with market microstructure noise among others.

The rest of this article is organized as follows. In Section 2, we first discuss the motivation of our method and then address the model validation problem by constructing nonparametric SCE for transition density function. In Section 3, we demonstrate the applicability of our methods for several widely used models. The finite sample performance is studied in Section 4. We defer the proofs to Section 5.

2. Transition density specification test for HMM

Given discrete samples $\{X_i\}$ from a stationary process $\{X_t\}_{t \in \mathcal{T}}$ whose data-generating mechanism involves some unobservable states $\{Y_i\}$ and unknown characteristics $\mathcal{Q}$, we are interested in the hypothesis testing problem $H_0: \mathcal{Q} = \mathcal{Q}_\theta$, where $\mathcal{Q}_\theta$ is some parametric form, $\theta \in \Theta$ is an unknown parameter, and Θ is the parameter space. We shall address this model validation problem for hidden Markov models (HMM).

First we give a formal definition of HMM by following Bickel and Ritov (1996) with minor modifications. For a set $\mathcal{S}$, we denote its Borel set by $\mathcal{B}(\mathcal{S})$.

Definition 1—A real-valued stochastic process $\{X_i\}_{i \in \mathbb{Z}}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a hidden Markov model with the hidden chain or states $\{Y_i\}_{i \in \mathbb{Z}}$ on a polish space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ if

- i. $\{Y_i\}_{i \in \mathbb{Z}}$ is a strictly stationary Markov chain.
- ii. for all i , given $\{Y_j\}_{j \leq i}$, $\{X_j\}_{j \leq i}$ are conditionally independent, and the conditional distribution of X_i depends only on Y_i .
- iii. The conditional distribution of X_i given Y_i does not depend on i .

In the definition, the observable process $\{X_i\}$ is real-valued whereas the hidden states process $\{Y_i\}$ can take values in a general polish space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$. For example, as we will show in Section 3.2, for the continuous-time stochastic volatility model (2), Y_i itself is a stochastic process with continuous sample path and $\mathcal{Y}$ is the set of continuous functions. If $\{X_i\}$ is a stationary Markov chain, then it is also a HMM with $\{Y_i = X_i\}$. In Section 3, we show that many continuous and discrete time models are special examples of HMM.

2.1. Transition density: connecting the observable $\{X_i\}$ and unobservable $\{Y_i\}$

To motivate the idea, we first focus on the simple case that the unobservable stationary Markov chain $\{Y_i\}$ is real-valued, and we will deal with the general case at the end of this section. For the Markov chain $\{Y_i\}$, the transition density $q_Y(y'|y)$ of Y_i at y' given $Y_{i-1} = y$ plays a key role in characterizing the probabilistic properties of $\{Y_i\}$. In many applications, $q_Y(y'|y)$ can fully characterize the underlying process. For example, consider $Y_i = \mu(Y_{i-1}) + \sigma(Y_{i-1})\varepsilon_i$, where ε_i are i.i.d. standard normal random variables. Denote by φ the standard normal density function. Then $q_Y(y'|y) = \varphi\{[y' - \mu(y)]/\sigma(y)\}/\sigma(y)$. It can be easily shown that, if two specifications (μ, σ) and (μ^*, σ^*) result in the same transition density $q_Y(y'|y)$, then $(\mu, \sigma) = (\mu^*, \sigma^*)$, and hence $q_Y(y'|y)$ completely determines the underlying data-generating mechanism. In fact, several researchers have used marginal and transition densities to address specification testing problems for Markov models (e.g. Aït-Sahalia, 1996; Aït-Sahalia et al., 2009). Aït-Sahalia et al. (2010) also used transition densities to test the Markov hypothesis.

Unfortunately, since $\{Y_i\}$ is not observable, $q_Y(y'|y)$ is not directly applicable, and we need to seek a reasonable “proxy” of it based on observations $\{X_i\}$. Let $\pi_X(x)$, $p_X(x, x')$ and $q_X(x'|x)$ be the marginal density of X_i , the joint density function of (X_{i-1}, X_i) at point (x, x') , and the transition density or conditional density of X_i at x' given $X_{i-1} = x$, respectively. Similarly, we define $\pi_Y(y)$ and $p_Y(y, y')$ as the marginal density of Y_i and joint density of (Y_{i-1}, Y_i) , respectively. Denote by $q_{X|Y}(x|y)$ the conditional density of X_i at x given $Y_i = y$. By the definition of HMM, conditioning on $\{Y_j\}_{j \leq i}$, X_{i-1} and X_i are independent and their conditional distributions depend only on Y_{i-1} and Y_i , respectively. So, by the conditioning argument, we have

$$p_X(x, x') = \iint q_{X|Y}(x|y)q_{X|Y}(x'|y')p_Y(y, y')dydy'. \quad (5)$$

Conditioning on Y_i , we have

$$\pi_X(x) = \int q_{X|Y}(x|y)\pi_Y(y)dy. \quad (6)$$

Since $q_X(x'|x) = p_X(x, x')/\pi_X(x)$ and $q_Y(y'|y) = p_Y(y, y')/\pi_Y(y)$, by (5) and (6),

$$q_X(x'|x) = \frac{\iint q_{X|Y}(x|y)q_{X|Y}(x'|y')\pi_Y(y)q_Y(y'|y)dydy'}{\int q_{X|Y}(x|y)\pi_Y(y)dy}. \quad (7)$$

The identity (7) establishes a functional connection between $q_Y(y'|y)$ of the hidden states $\{Y_i\}$ and $q_X(x'|x)$ of the observable variables $\{X_i\}$. In particular, $q_X(x'|x)$ is uniquely determined by $q_Y(y'|y)$ together with $q_{X|Y}(x|y)$ and $\pi_Y(y)$. On the other hand, $\{q_X(x'|x)\}_{x, x' \in \mathbb{R}}$ also contains rich information about $q_Y(y'|y)$ through the functional relationship (7). Therefore, we can view the process from the observable $q_X(x'|x)$ to the unobservable $q_Y(y'|y)$ as a functional inverse problem. In practice, $q_{X|Y}$ is often specified through model assumptions. For example, for the HMM $\{(X_i, Y_i) = (X_i, \sigma_i)\}$ in (3), $q_{X|Y}$ is simply the normal density if we assume that ε_i therein are normal errors. In summary, (7) motivates us to use $q_X(x'|x)$ as a “proxy” for $q_Y(y'|y)$.

Now we consider the case of a general polish space $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$. The probabilistic properties of $\{Y_i\}$ are characterized by the invariant probability measure $\Pi_Y(\mathcal{S}) = \mathbb{P}(Y_i \in \mathcal{S})$ with $\mathcal{S} \in \mathcal{B}(\mathcal{Y})$ and the Markov transition kernel or generator $Q_Y(\mathcal{S}|y) = \mathbb{P}(Y_i \in \mathcal{S} | Y_{i-1} = y)$ with $y \in \mathcal{Y}$ and $\mathcal{S} \in \mathcal{B}(\mathcal{Y})$. The identity (7) then becomes

$$q_X(x'|x) = \frac{\int_{\mathcal{Y}} \int_{\mathcal{Y}} q_{X|Y}(x|y)q_{X|Y}(x'|y')\Pi_Y(dy)Q_Y(dy'|y)}{\int_{\mathcal{Y}} q_{X|Y}(x|y)\Pi_Y(dy)}. \quad (8)$$

In some applications, it is more convenient to express (8) through expectations

$$q_X(x'|x) = \frac{\mathbb{E}[q_{X|Y}(x|Y_{i-1})q_{X|Y}(x'|Y_i)]}{\mathbb{E}[q_{X|Y}(x|Y_{i-1})]}. \quad (9)$$

In particular, the expectation motivates us to estimate the numerator and denominator by their corresponding empirical versions; see Section 2.4.

2.2. Confidence envelope and specification testing

In this section we propose a specification testing procedure based on $q_X(x'|x)$. We achieve this by constructing a simultaneous confidence envelope (SCE) for $q_X(x'|x)$ over a compact

set $\mathcal{X} \subset \mathbb{R}^2$. For a significance level $\alpha \in (0, 1)$ and a pair of bivariate functions $l_n(\cdot, \cdot)$ and $u_n(\cdot, \cdot)$ based on observations, we say that $[l_n(\cdot, \cdot), u_n(\cdot, \cdot)]$ is an asymptotic $(1 - \alpha)$ nonparametric SCE for $q_X(x'|x)$ on a compact set $\mathcal{X} \subset \mathbb{R}^2$ if

$$\lim_{n \rightarrow \infty} \mathbb{P}\{l_n(x, x') \leq q_X(x'|x) \leq u_n(x, x') \text{ for all } (x, x') \in \mathcal{X}\} = 1 - \alpha. \quad (10)$$

With asymptotic probability $(1 - \alpha)$, the unknown true density $q_X(x'|x)$ is contained within the envelope $[l_n(\cdot, \cdot), u_n(\cdot, \cdot)]$. The concept of confidence envelope for a bivariate function is a natural extension of the confidence interval for one-dimensional parameter.

Now we relate the idea of nonparametric SCE to our model validation problem. Consider the popular nonparametric kernel density estimate for the transition density $q_X(x'|x)$

$$\widehat{q}_X(x'|x) = \frac{\widehat{p}_X(x, x')}{\widehat{\pi}_X(x)}, \quad (11)$$

where

$$\widehat{p}_X(x, x') = \frac{1}{nb_n^2} \sum_{i=1}^n K_{b_n}(x - X_{i-1}) K_{b_n}(x' - X_i), \quad (12)$$

$$\widehat{\pi}_X(x) = \frac{1}{nb_n} \sum_{i=1}^n K_{b_n}(x - X_{i-1}). \quad (13)$$

Here 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 $b_n > 0$. Under mild conditions, $\widehat{q}_X(x'|x)$ is a consistent estimate of $q_X(x'|x)$, regardless of the underlying model structure, and hence it can be used as the “true” reference density. Under $H_0: \mathcal{Q} = \mathcal{Q}_\theta$, denote by $\widehat{q}_X(x'|x; \mathcal{Q}_\theta)$ a parametric estimate of $q_X(x'|x)$, where $\hat{\theta}$ is an estimator of θ . To test H_0 , we can measure the discrepancy between the parametric estimate $\widehat{q}_X(x'|x; \mathcal{Q}_\theta)$ under H_0 and the nonparametric estimate $\widehat{q}_X(x'|x)$. A general test statistic has the form $T_n = d(\widehat{q}_X(\cdot|x; \mathcal{Q}_\theta), \widehat{q}_X(\cdot|x))$ with a proper choice of the distance measure $d(\cdot, \cdot)$. Aït-Sahalia (1996) studied model (1) based on square distance.

Following Bickel and Rosenblatt (1973), we consider maximal deviation:

$$T_n = \max_{(x, x') \in \mathcal{X}} \frac{|\widehat{q}_X(x'|x) - \widehat{q}_X(x'|x; \mathcal{Q}_{\hat{\theta}})|}{\omega(x, x')}, \quad (14)$$

where $\omega(\cdot, \cdot)$ is a weight function that may depend on observations. Bickel and Rosenblatt (1973) considered marginal density for i.i.d. observations. Classical result asserts the

asymptotic normality of $\sqrt{nb_n^2}[\hat{q}_X(x'|x) - q_X(x'|x)]$. On the other hand, under H_0 , the parametric density estimate $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ is usually $\sqrt{n}$ -consistent. Therefore, under H_0 : $\mathcal{Q} = \mathcal{Q}_\theta$, as $b_n \rightarrow 0$, $\hat{q}_X(x'|x; \mathcal{Q}_\theta) - q_X(x'|x)$ is negligible compared to $\hat{q}_X(x'|x) - q_X(x'|x)$, and we have

$$\begin{aligned} T_n &= \max_{(x, x') \in \mathcal{X}} \frac{|\hat{q}_X(x'|x) - q_X(x'|x)| - |\hat{q}_X(x'|x; \mathcal{Q}_\theta) - q_X(x'|x)|}{\omega(x, x')} \\ &\approx \max_{(x, x') \in \mathcal{X}} \frac{|\hat{q}_X(x'|x) - q_X(x'|x)|}{\omega(x, x')} := T_n^*. \end{aligned} \quad (15)$$

The quantity T_n^* is closely related to nonparametric SCE for $q_X(x'|x)$. To see this, we assume that there exist normalizing sequences (γ_n, β_n) such that $\gamma_n T_n^* - \beta_n$ has a limiting distribution $\mathbb{G}$. Denote by $\mathbb{G}_\alpha$ the $(1 - \alpha)$ -quantile of $\mathbb{G}$. Based on T_n^* , a nonparametric SCE for $q_X(x'|x)$ can be constructed as $\hat{q}_X(x'|x) \pm \gamma_n^{-1}(\mathbb{G}_\alpha + \beta_n)\omega(x, x')$ in view of

$$\begin{aligned} \mathbb{P}\{|\hat{q}_X(x'|x) - q_X(x'|x)| \leq \gamma_n^{-1}(\mathbb{G}_\alpha + \beta_n)\omega(x, x'), (x, x') \in \mathcal{X}\} \\ = \mathbb{P}\{\gamma_n T_n^* - \beta_n \leq \mathbb{G}_\alpha\}. \end{aligned}$$

The above argument has a number of implications. First, the test statistic T_n is equivalent to constructing nonparametric SCE for $q_X(x'|x)$. Second, under H_0 , the parametric estimate $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ is contained within the nonparametric SCE with asymptotic probability $(1 - \alpha)$. Third, it is essential to construct SCE based on the nonparametric estimate $\hat{q}_X(x'|x)$, otherwise the method might fail. Consider, for example, the case where the SCE were constructed based on the parametric estimate $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ under H_0 . Due to the $\sqrt{n}$ -consistency of $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$, the constructed SCE would have a width proportional to $O(n^{-1/2})$

and therefore cannot cover the more volatile $\sqrt{nb_n^2}$ -consistent nonparametric estimate $\hat{q}_X(x'|x)$. We now summarize our specification testing procedure:

- i. Construct $(1 - \alpha)$ nonparametric SCE for $q_X(x'|x)$: $[\ell_n(\cdot, \cdot), u_n(\cdot, \cdot)]$.
- ii. Under H_0 , apply parametric methods to obtain an estimate $\hat{\theta}$ of θ .
- iii. Under H_0 , obtain a parametric estimate $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ of $q_X(x'|x)$.
- iv. Check whether $\ell_n(x, x') \leq \hat{q}_X(x'|x; \mathcal{Q}_\theta) \leq u_n(x, x')$ holds for all $(x, x') \in \mathcal{X}$. If so, then we accept H_0 at level α . Otherwise H_0 is rejected.

To implement the above idea, two tasks remain. The first is to establish some maximal deviation result of the form (15) for the nonparametric density estimate $\hat{q}_X(x'|x)$. The second is to construct parametrically implied density estimate $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ of $q_X(x'|x)$ under H_0 . We shall address these two issues in Sections 2.3 and 2.4.

2.3. Construct confidence envelope for transition density

For maximal deviations of nonparametric density estimate, Bickel and Rosenblatt (1973) dealt with i.i.d. data. Under the independence assumption, the problem of constructing simultaneous confidence bands (SCB, the term “band” is used for univariate function in contrast to “envelope” for bivariate function) have been studied previously under various

settings (e.g. Johnston, 1982; Knafl et al., 1985; Eubank and Speckman, 1993; Fan and Zhang, 2000). Zhao and Wu (2008) considered SCB construction for time series models.

Here we shall extend Bickel and Rosenblatt (1973)'s result to the transition density of hidden Markov models (HMM). Recall that $\mathfrak{X} \subset \mathbb{R}^2$ is a two-dimensional compact set.

Condition 1 (Dependence Assumption)—Suppose that $\{X_i\}$ is a HMM with respect to a stationary hidden chain $\{Y_i\}$. Denote by $\mathcal{G}_i^j$ the sigma-field generated by $\{Y_t, i \leq t \leq j\}$. Define the α -mixing coefficient of $\{Y_i\}$ by

$$\alpha(k) = \sup\{|\mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A \cap B)|, A \in \mathcal{G}_{-\infty}^0, B \in \mathcal{G}_k^\infty\}, \quad k \in \mathbb{N}.$$

Assume that $\sum_{k=1}^{\infty} \alpha(k) < \infty$.

Condition 2 (Kernel Assumption)—Assume that the kernel K is bounded, symmetric, and has bounded derivative and bounded support $[-\omega, \omega]$. Further assume that K is a 4-th order kernel in the sense that $\int_{-\omega}^{\omega} u^j K(u) du = 0, j = 1, 2, 3$. Let $\phi_K = \int_{-\omega}^{\omega} K^2(u) du$.

Condition 3 (Regularity Assumption)—Without loss of generality write $\mathfrak{X} = [-T, T] \times [-T, T]$ for some $T > 0$. Denote by $\mathfrak{X}^\varepsilon = [-T - \varepsilon, T + \varepsilon] \times [-T - \varepsilon, T + \varepsilon]$ the ε -neighborhood of $\mathfrak{X}$. Assume that there exists some $\varepsilon > 0$ such that $\pi_X(x)$ has bounded fourth order derivatives on $[-T - \varepsilon, T + \varepsilon]$ and $q_X(x'|x)$ has bounded fourth order derivatives with respect to both x and x' on $\mathfrak{X}^\varepsilon$, $\inf_{x \in [-T - \varepsilon, T + \varepsilon]} \pi_X(x) > 0$, $\inf_{(x, x') \in \mathfrak{X}^\varepsilon} q_X(x'|x) > 0$, and $\sup_{x \in [-T - \varepsilon, T + \varepsilon], y \in \mathbb{Y}} [|q_{X|Y}(x|y)| + |\partial q_{X|Y}(x|y)/\partial x|] < \infty$.

We briefly comment on Conditions 1–3. Condition 1 is frequently used in statistical inferences involving dependent data. Condition 2 is a standard assumption on the kernel function in nonparametric inference problems. Condition 3 imposes smoothness assumptions.

Theorem 1—Recall $\hat{q}_X(x'|x)$ in (11). Assume that Conditions 1–3 hold. Further assume that $nb_n^{10} \rightarrow 0, nb_n^2 \rightarrow \infty$. Then for every $(x, x') \in \mathfrak{X}$, as $n \rightarrow \infty$,

$$Z_n(x, x') := \frac{b_n \sqrt{n} \hat{q}_X(x'|x) - q_X(x'|x)}{\phi_K \sqrt{q_X(x'|x)/\pi_X(x)}} \Rightarrow N(0, 1). \quad (16)$$

Moreover, for distinct points

$$(x_1, x'_1), \dots, (x_k, x'_k) \in \mathfrak{X}, Z_n(x_1, x'_1), \dots, Z_n(x_k, x'_k)$$

are asymptotically independent.

Theorem 1 can be used to construct a point-wise confidence envelope for $q_X(x'|x)$. Theorem 2 provides a maximal deviation result for $\hat{q}_X(x'|x)$.

Theorem 2—Let $\hat{q}_X(x'|x)$ be as in (11). Recall that, in Condition 2, K has support $[-\omega, \omega]$ and $\mathfrak{X}$ is a compact subset of $\mathbb{R}^2$. For $m_n \rightarrow \infty$, let

$$\mathcal{X}_n = \{(x_j, x'_j) \in \mathcal{X}, j = 1, \dots, m_n\}$$

be a subset of $\mathcal{X}$ consisting of m_n points such that $\max\{|x_j - x_{j'}|, |x'_j - x'_{j'}|\} \geq 2\omega b_n$ for all $1 \leq j \neq j' \leq m_n$. Assume that Conditions 1–3 hold. Further assume that

$$nb_n^{10} \log n + m_n^2 (\log m_n)^3 [b_n^3 + (nb_n^2)^{-1}] \rightarrow 0. \quad (17)$$

For $k \geq 2$ define

$$B_k(z) = \sqrt{2 \log k} - \frac{1}{\sqrt{2 \log k}} \left[\frac{1}{2} \log \log k + \log(2 \sqrt{\pi}) - z \right], k \in \mathbb{N}, z \in \mathbb{R}. \quad (18)$$

Then for every $z \in \mathbb{R}$,

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{(x, x') \in \mathcal{X}_n} \frac{b_n \sqrt{n}}{\varphi_K} \frac{|\hat{q}_X(x'|x) - q_X(x'|x)|}{\sqrt{q_X(x'|x)/\pi_X(x)}} \leq B_{m_n}(z) \right\} \\ = e^{-2e^{-z}}. \end{aligned} \quad (19)$$

In (17), the first term $nb_n^{10} \log n \rightarrow 0$ is needed to control the bias while the second term $m_n^2 (\log m_n)^3 [b_n^3 + (nb_n^2)^{-1}] \rightarrow 0$ ensures the validity of the moderate deviation (cf. Theorem 3). In particular, if $m_n \asymp 1/b_n$, then (17) holds if $b_n = n^{-\beta}$ for $\beta \in (1/10, 1/4)$.

We can use Theorem 2 to construct asymptotic $(1 - \alpha)$ SCE for $q_X(x'|x)$ as

$$\begin{aligned} \widehat{q}_X(x'|x) \pm \frac{\phi_K B_{m_n}(z_\alpha)}{b_n \sqrt{n}} \sqrt{\widehat{q}_X(x'|x)/\widehat{\pi}_X(x)}, \\ z_\alpha = -\log \log \left(\frac{1}{\sqrt{1-\alpha}} \right), \end{aligned} \quad (20)$$

in the following approximated version of (10) (note that z_α is the $(1 - \alpha)$ -quantile of the limiting distribution on the right-hand side of (19)):

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{P}\{l_n(x, x') \leq q_X(x'|x) \\ \leq u_n(x, x') \text{ for all } (x, x') \in \mathcal{X}_n\} = 1 - \alpha. \end{aligned} \quad (21)$$

As $m_n \rightarrow \infty$, we can let $\mathcal{X}_n$ become asymptotically dense in $\mathcal{X}$. Therefore, for smooth function $q_X(x'|x)$, (21) provides a good approximation to (10) for large n .

Remark 1—In Condition 3, we have assumed that the conditional density $q_{X|Y}$ of X_i given Y_i exists. If $\{X_i\}$ itself is a Markov chain, then we can let $\{Y_i = X_i\}$ be the observable states.

However, the conditional density $q_{X|Y}$ does not exist. A careful examination of the proof reveals that, Theorems 1 and 2 still hold if we replace $q_{X|Y}$ in Condition 3 by the conditional density of X_i given X_{i-1} and apply the new filtration $\mathcal{F}_i = \sigma(X_j, j \leq i)$ in the proofs. Thus, the theoretical results developed also apply to ordinary Markov chains.

2.4. Construct parametric transition density estimate

After constructing nonparametric SCE for $q_X(x'|x)$, in order to test $H_0: \mathcal{Q} = \mathcal{Q}_\theta$, we need to obtain a parametric density estimate $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ of $q_X(x'|x)$ under H_0 and check whether it is contained within SCE. Parametric estimation of $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ involves estimating parameter θ through various parametric methods, such as least-squares, maximum likelihood, generalized method of moments, and M -estimation methods.

Under $H_0: \mathcal{Q} = \mathcal{Q}_\theta$, let us assume at the outset that θ is known. In practice, the conditional density $q_{X|Y}$ of X_i given Y_i is often specified through model assumptions. For example, for the HMM $\{(X_i, Y_i) = (X_i, \sigma_i)\}$ in (3), $q_{X|Y}$ is the normal density if we assume that ε_i therein are normal errors. Therefore, under $H_0: \mathcal{Q} = \mathcal{Q}_\theta$ with known θ , the underlying model is completely specified. Theoretically speaking, depending on whether $\{Y_i\}$ takes value in $\mathbb{R}$ or a general polish space $\mathcal{Y}$, one can use either (7) or (8) to obtain theoretical parametric density $q_X(x'|x; \mathcal{Q}_\theta)$ by replacing (π_Y, q_Y) and (Π_Y, Q_Y) therein by their theoretical expressions.

In many applications, however, the above naive method fails. First, there is no closed-form (π_Y, q_Y) in (7) or (Π_Y, Q_Y) in (8). Even for the simplest ARCH(1) model $Y_i = (a^2 + b^2 Y_{i-1}^2)^{1/2} \varepsilon_i$, there is no closed-form stationary density. For (2) with hidden states $\{Y_i = (\sigma_{s+(i-1)\mathcal{A}})_{s \in [0, \mathcal{A}]}\}_{i \geq 1}$, $\mathcal{Y}$ is the set of continuous function on $[0, \mathcal{A}]$, and it is infeasible to compute (Π_Y, Q_Y) . Second, it is not easy to evaluate the integrals in (7) and (8), especially when $\{Y_i\}$ takes value in a general polish space $\mathcal{Y}$.

To attenuate these issues, we propose a Markov Chain Monte Carlo simulation based method. Under $H_0: \mathcal{Q} = \mathcal{Q}_\theta$ with known θ so that the model structure is completely specified, we can simulate the hidden states $\{Y_j^*\}_{1 \leq j \leq m}$ from the underlying model $\mathcal{Q}_\theta$. In (9), replacing the numerator and denominator by their corresponding empirical versions, we propose the following estimate of $q_X(x'|x)$:

$$\hat{q}_X(x'|x; \mathcal{Q}_\theta) = \frac{m^{-1} \sum_{i=1}^m q_{X|Y}(x|Y_{i-1}^*; \theta) q_{X|Y}(x'|Y_i^*; \theta)}{m^{-1} \sum_{i=1}^m q_{X|Y}(x|Y_{i-1}^*; \theta)}. \quad (22)$$

Here we include θ to mean that the estimate is based on samples from $\mathcal{Q}_\theta$. Under mild conditions, for example mixing condition, by the Strong Law of Large Numbers for both the numerator and denominator in (22), $\hat{q}_X(x'|x; \mathcal{Q}_\theta) \rightarrow q_X(x'|x; \mathcal{Q}_\theta)$ as $m \rightarrow \infty$.

To implement the above idea, we need a consistent estimate $\hat{\theta}$ of θ . Parameter estimation for HMM is usually a difficult task. Under parametric specification $\mathcal{Q}_\theta$, a natural choice is the likelihood method. By the conditional independence, the likelihood for observations $\mathbf{x} := \{X_i = x_i\}_{1 \leq i \leq n}$ is

$$L(\mathbf{x}; \theta) = \int_{\mathcal{Y}} \cdots \int_{\mathcal{Y}} q_{X|Y}(x_1|y_1) \cdots q_{X|Y}(x_n|y_n) \times P_Y(dy_1, \dots, dy_n), \quad (23)$$

where $P_Y(s_1, \dots, s_n)$ is the joint probability measure of $\{Y_i\}_{1 \leq i \leq n}$. Therefore, we can in principle obtain an estimate $\hat{\theta}$ of θ by maximizing the likelihood $L(\mathbf{x}; \theta)$. However, $L(\mathbf{x}; \theta)$ is not directly computable even for simple models. A possible solution is to use the same idea in (22) by approximating $L(\mathbf{x}; \theta)$ as

$$\widehat{L}(\mathbf{x}; \theta) = \frac{1}{m} \sum_{i=1}^m q_{X|Y}(x_1|Y_1^{(i)}) \cdots q_{X|Y}(x_n|Y_n^{(i)}), \quad (24)$$

where $Y_1^{(i)}, \dots, Y_n^{(i)}$ is the i th sample path. This method requires a large number of sample paths (m sample paths of size n each) and is computationally expensive. In general, there are no universally efficient parameter estimation methods for HMM. In next section we discuss related references for some specific models.

Now we summarize our procedure to obtain parametric density estimate $\hat{q}_X(x'|x; \mathcal{Q}_s)$.

- i. Under H_0 : $\mathcal{Q} = \mathcal{Q}_\theta$, obtain an estimate $\hat{\theta}$.
- ii. Under the estimated parametric model $\mathcal{Q}_s$, simulate $Y_1^*, \dots, Y_m^*$.
- iii. Plug $\hat{\theta}$ and simulated samples $Y_1^*, \dots, Y_m^*$ into (22) to obtain

$$\hat{q}_X(x'|x; \mathcal{Q}_{\hat{\theta}}) = \frac{m^{-1} \sum_{i=1}^m q_{X|Y}(x|Y_{i-1}^*; \hat{\theta}) q_{X|Y}(x'|Y_i^*; \hat{\theta})}{m^{-1} \sum_{i=1}^m q_{X|Y}(x|Y_{i-1}^*; \hat{\theta})}. \quad (25)$$

3. Examples

In this section, we show that many popular continuous-time and discrete-time models in financial econometrics can be viewed as HMMs with properly chosen $\{(X_i, Y_i)\}$. Under H_0 : $\mathcal{Q} = \mathcal{Q}_\theta$, since various parameter estimation methods, such as maximum likelihood estimate (MLE), generalized least-squares method, M-estimation, and generalized method of moments (GMM), are available in a vast literature, we only briefly discuss some related references and focus our attention on constructing parametric density estimate under H_0 .

Throughout the rest of this article, we denote by $\varphi(x)$ and $\Phi(x)$ the standard normal density and distribution functions, respectively. For $p > 0$ and a random variable X write $X \in \mathcal{L}^p$ if $\|X\|_p := [\mathbb{E}(|X|^p)]^{1/p} < \infty$.

3.1. Continuous-time diffusion models

Let $X_i \equiv X_{i\Delta}$ be discrete samples from the diffusion model (1). Here $\Delta > 0$ is a small but fixed number representing sampling interval. In practice, for daily or weekly data, Δ is one

day or one week, respectively. The process $\{X_t\}$ is a Markov chain. Denote by π_X be the marginal density function of the stationary solution X_t on $D = (D_l, D_u)$ with $-\infty \leq D_l < D_u \leq +\infty$. By Aït-Sahalia (1996),

$$\pi_X(x) = \frac{c(x_0)}{\sigma^2(x)} \exp \left\{ \int_{x_0}^x \frac{2\mu(y)}{\sigma^2(y)} dy \right\}, \quad (26)$$

where the choice of the lower bound point $x_0 \in D$ is irrelevant, and $c(x_0)$ is a normalizing constant. Similar versions of Condition 4 below have been discussed in Hansen and Scheinkman (1995), Aït-Sahalia (1996), and Genon-Catalot et al. (2000).

Condition 4—Assume that: (i) Both μ and σ are twice continuously differentiable and $\sigma^2 > 0$ on D ; (ii) The integral of $m(x) = \exp\{\Lambda(x)\}/\sigma^2(x)$ converges at both boundaries of D , and the integral of $s(x) = \exp\{-\Lambda(x)\}$ diverges at both boundaries of D , where $\Lambda(x) = \int_{x_0}^x [2\mu(y)/\sigma^2(y)] dy$; (iii) $\lim_{x \rightarrow D_l, D_u} \sigma(x)\pi_X(x) = 0$ and $\lim_{x \rightarrow D_l, D_u} \sigma(x)/|2\mu(x) - \sigma(x)\sigma'(x)| < \infty$.

Proposition 1—Under Condition 4, Condition 1 holds with $Y_i = X_i$.

Proof—Under Condition 4, Hansen and Scheinkman (1995) proved that, the operator $\mathcal{H}_t$ defined by $\mathcal{H}_t g(y) = \mathbb{E}[g(X_t)|X_0 = y]$ for $g(X_0) \in \mathcal{L}^2$ is a strong contraction in the sense that there exists $\lambda \in (0, 1)$ such that

$$\begin{aligned} \|\mathcal{H}_t g(X_0)\|_2 &\leq \lambda^t \|g(X_0)\|_2 \quad \text{for } g(X_0) \in \mathcal{L}^2 \quad \text{and} \\ \mathbb{E}[g(X_0)] &= 0. \end{aligned} \quad (27)$$

Therefore, for $g_1(X_0), g_2(X_0) \in \mathcal{L}^2$ satisfying $\mathbb{E}[g_1(X_0)] = \mathbb{E}[g_2(X_0)] = 0$, by the Cauchy–Schwarz inequality,

$$\begin{aligned} |\mathbb{E}[g_1(X_0)g_2(X_t)]| &= |\mathbb{E}\{g_1(X_0)\mathbb{E}[g_2(X_t)|X_0]\}| = |\mathbb{E}[g_1(X_0)\mathcal{H}_t g_2(X_0)]| \\ &\leq \|g_1(X_0)\|_2 \|\mathcal{H}_t[g_2(X_0)]\|_2 \leq \|g_1(X_0)\|_2 \|g_2(X_0)\|_2 \lambda^t. \end{aligned} \quad (28)$$

Denote by $\mathcal{G}_t$ and $\mathcal{G}'_t$ the sigma fields generated by $\{X_s\}_{s \leq t}$ and $\{X_s\}_{s \geq t}$, respectively. The ρ -mixing coefficient of $\{X_t\}_{t \geq 0}$ is defined as

$$\rho_t = \sup \rho(\mathcal{G}_s, \mathcal{G}_s^{s+t}), \quad \text{where } \rho(\mathcal{G}, \mathcal{G}') = \sup \left\{ \frac{|\text{Cov}(X, X')|}{\|X\|_2 \|X'\|_2} \right\},$$

where the supremum is taken over all $\mathcal{G}$ (respectively $\mathcal{G}'$)-measurable random variable X (respectively X') satisfying $X, X' \in \mathcal{L}^2$, $\mathbb{E}(X) = \mathbb{E}(X') = 0$. Since $\{X_t\}_{t \geq 0}$ is a stationary Markov process, by Theorem 4.1 in Bradley (1986) and (28),

$$\rho_t = \sup \left\{ \frac{|\text{Cov}[g_1(X_0), g_2(X_t)]|}{\|g_1(X_0)\|_2 \|g_2(X_t)\|_2} \right\} = O(\lambda^t).$$

Hence, $\{X_t\}_{t \geq 0}$ is ρ -mixing with mixing coefficient $\rho(t) = O(\lambda^t)$, which completes the proof since α -mixing coefficient is less than the corresponding ρ -mixing coefficient.

After we construct SCE for $q_X(x'|x)$ as in Section 2.3, we need to obtain a parametric estimate $\hat{q}_X(x'|x; \mathcal{Q}_\theta)$ of $q_X(x'|x)$ under H_0 : $\mathcal{Q} = \mathcal{Q}_\theta = (\mu(\cdot; \theta), \sigma(\cdot; \theta))$ for a parametric specification $\mathcal{Q}_\theta$. To estimate θ under H_0 , a natural choice is the maximum likelihood estimate (MLE). Consider the following Euler discretization scheme:

$$X_{t+\Delta} - X_t = \mu(X_t)\Delta + \sigma(X_t)(W_{t+\Delta} - W_t). \quad (29)$$

Then θ can be estimated by maximizing the approximate conditional log-likelihood

$$\hat{\theta} = \underset{\theta}{\text{argmax}} \sum_{i=1}^n \log \times \left[\frac{1}{\sigma(X_{i-1}; \theta) \sqrt{\Delta}} \varphi \left\{ \frac{X_i - X_{i-1} - \mu(X_{i-1}; \theta)\Delta}{\sigma(X_{i-1}; \theta) \sqrt{\Delta}} \right\} \right]. \quad (30)$$

And the transition density is given by

$$\hat{q}_X(x'|x; \mathcal{Q}_{\hat{\theta}}) = \frac{1}{\sigma(x; \hat{\theta}) \sqrt{\Delta}} \phi \left\{ \frac{x' - x - \mu(x; \hat{\theta})\Delta}{\sigma(x; \hat{\theta}) \sqrt{\Delta}} \right\}. \quad (31)$$

3.2. Continuous-time stochastic volatility models

Consider the continuous-time stochastic volatility model (2). Let $X_i = \log(S_{i\Delta}) - \log(S_{(i-1)\Delta})$ be the log returns during time period $[(i-1)\Delta, i\Delta]$. The unobserved volatility process $\{\sigma_t^2\}_{t \geq 0}$ is a stationary Markov process. Volatility plays an important role in risk analysis and options pricing. To study volatilities, we can take $\mathcal{Q} = (r, s)$. Part (i) of Proposition 2 is adopted from Genon-Catalot et al. (2000).

Proposition 2—Assume that (r, s) in model (2) governing σ_t^2 satisfy Condition 4. Then

- i. $\{X_i\}$ is a HMM with the hidden chain $\{Y_i^{(1)} = (\sigma_{t+(i-1)\Delta}^2)_{t \in [0, \Delta]}\}$ or $\{Y_i^{(2)} = (\sigma_{i\Delta}^2, \int_{(i-1)\Delta}^{i\Delta} \sigma_t^2 dt)\}$;
Condition 1 holds;
- ii. Condition 3 holds provided that with probability one $c_1 < \inf_{t \geq 0} \sigma_t^2 \leq \sup_{t \geq 0} \sigma_t^2 < c_2$ for some constants $0 < c_1 < c_2 < \infty$.

Proof

- i. We shall only consider $\{Y_i^{(1)}\}$; see Genon-Catalot et al. (2000) for the proof of $\{Y_i^{(2)}\}$. Let $\mathcal{G}_t$ be the sigma field generated by $\{\sigma_s^2\}_{s \leq t}$. By the independence of $\{W_1(t)\}$ and $\{W_2(t)\}$, conditional on $\mathcal{G}_{i\Delta}$, $X_j = \int_{(j-1)\Delta}^{j\Delta} \sigma_t dW_1(t)$, $j \leq i$, are independent normal random variables with zero mean and variance $\sum_i^2 = \int_{(i-1)\Delta}^{i\Delta} \sigma_t^2 dt$. Notice that the random variables $\sum_j^2$, $j \leq i$, are measurable with respect to the sigma field

$$\sigma(Y_j^{(1)}, j \leq i) \subset \mathcal{G}_{i\Delta}$$

. Therefore, we have for all $\lambda_1, \dots, \lambda_i \in \mathbb{R}$,

$$\begin{aligned} & \mathbb{E}[e^{\lambda_1 X_1 + \dots + \lambda_i X_i} | Y_1^{(1)}, \dots, Y_i^{(1)}] \\ &= \mathbb{E}[\mathbb{E}[e^{\lambda_1 X_1 + \dots + \lambda_i X_i} | \mathcal{G}_{i\Delta}] | Y_1^{(1)}, \dots, Y_i^{(1)}] \\ &= \exp\{(\lambda_1^2 \sum_1^2 + \dots + \lambda_i^2 \sum_i^2)/2\}, \end{aligned}$$

entailing the conditional independence of X_j , $j \leq i$. The Markovian property of $\{\sigma_t^2\}_{t>0}$ implies that $\{Y_i^{(1)}\}_{i>0}$ is Markovian, completing the proof.

- ii. See Proposition 1.
- iii. Let $Y_i = Y_i^{(1)}$. Then the conditional density $q_{X|Y}$ of X_i given Y_i is uniformly bounded in view of

$$q_{X|Y}(x|Y_i) = \varphi(x / \sum_i) / \sum_i = (2\pi \sum_i^2)^{-1/2} \times \exp[-x^2 / (2 \sum_i^2)] \leq (2\pi c\Delta)^{-1/2}, \quad (32)$$

and $\sum_i^2 = \int_{(i-1)\Delta}^{i\Delta} \sigma_t^2 dt > c\Delta$. Similarly, $|\partial q_{X|Y}(x|Y_i)|$ is uniformly bounded. Clearly, $q_{X|Y}(x|Y_i)$ is also uniformly bounded away from zero on any compact set. By (9) and the Dominated Convergence Theorem, it is easy to verify that $q_X(x'|x)$ has bounded derivatives of all orders on any compact set. So, Condition 3 holds.

We now discuss parametric density estimate under and $H_0: \mathcal{Q} = \mathcal{Q}_\theta = (r(\cdot; \theta), s(\cdot; \theta))$. First, we need to estimate θ . Under parametric specification $\mathcal{Q}_\theta$, a natural choice is the likelihood method. Let $f_\Sigma(\sum_0^2, \dots, \sum_n^2; \theta)$ be the joint density of $\{\sum_i^2\}_{0 \leq i \leq n}$ under $H_0: \mathcal{Q} = \mathcal{Q}_\theta$. By the argument in the proof of Proposition 2, the likelihood for given observations $\mathbf{x} = \{X_i = x_i\}_{0 \leq i \leq n}$ from (2) is

$$L(\mathbf{x}; \theta) = \int \prod_{i=0}^n \sum_i^{-1} \varphi(x_i / \sum_i) f_\Sigma(\sum_0^2, \dots, \sum_n^2; \theta) d\sum_0^2 \cdots d\sum_n^2. \quad (33)$$

Under H_0 , f_Σ is completely determined up to unknown parameter θ . Therefore, we can in principle obtain an estimate $\hat{\theta}$ of θ by maximizing the likelihood $L(\mathbf{x}; \theta)$. However, $L(\mathbf{x}; \theta)$ is

not directly computable even for very simple models. Since

$$L(\mathbf{x};\theta)=\mathbb{E}\left[\prod_{i=0}^n\sum_i^{-1}\varphi(x_i/\sum_i)\right],$$
 a possible solution is Markov chain Monte Carlo method

by simulating a large number of sample paths of $\{\sigma_t^2\}_{t\geq 0}$ from the parametric model (2), which, unfortunately, involves extensive computation. Possible alternatives have been proposed in, for example, Kim et al. (1998) for Bayesian method, and Andersen and Sørensen (1996) for moments based method. For other contributions, see the survey paper by Broto and Ruiz (2004). We point out that, most existing estimation methods deal with simple (say, lognormal autoregressive) stochastic volatility models. It still remains open how to develop efficient estimation techniques for general stochastic volatility models. In our simulation studies, we use a moment based method.

Recall $q_{X|Y}$ in (32). Once we have a consistent estimate of θ , we can apply (25) to estimate $q_X(x'|x)$ parametrically as

$$\hat{q}_X(x'|x; \mathcal{Q}_{\hat{\theta}}) = \frac{m^{-1} \sum_{i=1}^m \phi(x/\Sigma_{i-1}^*) \phi(x'/\Sigma_i^*) / (\Sigma_{i-1}^* \Sigma_i^*)}{m^{-1} \sum_{i=1}^m \phi(x/\Sigma_{i-1}^*) / \Sigma_{i-1}^*}, \quad (34)$$

where $\sum_i^{*2} = \int_{(i-1)\Delta}^{i\Delta} \sigma_t^{*2} dt$ and $\{\sigma_t^{*2}\}$ is a simulated realization from the estimated null model $d\sigma_t^{*2} = r(\sigma_t^{*2}; \hat{\theta})dt + s(\sigma_t^{*2}; \hat{\theta})dW_t$.

3.3. Stochastic volatility models driven by stable Lévy process

Model (2) can be extended to non-Gaussian stable Lévy processes. A process $\{Z_t\}$ is said to be a α -stable Lévy process if it has independent and stationary increments, and Z_1 has a stable distribution with index $\alpha \in (0, 2]$. The special case of $\alpha = 2$ corresponds to Brownian motion. Consider

$$\begin{aligned} d\log(S_t) &= \sigma_t dZ(t) \quad \text{and} \\ d\sigma_t^\alpha &= r(\sigma_t^\alpha)dt + s(\sigma_t^\alpha)dW(t), \quad t \geq 0, \end{aligned}$$

where $\{Z(t)\}_{t\geq 0}$ is a α -stable Lévy process with index $\alpha \in (0, 2]$ independent of the Brownian motion $\{W(t)\}_{t\geq 0}$. Using the scaling property of stable Lévy process, the same argument in Section 3.2 shows that Proposition 2 still holds with the hidden chain

$$\{Y_i^{(1)} = (\sigma_{t+(i-1)\Delta}^\alpha)_{t \in [0, \Delta]}\} \text{ or } \{Y_i^{(2)} = (\sigma_{i\Delta}^\alpha, \int_{(i-1)\Delta}^{i\Delta} \sigma_t^\alpha dt)\}.$$
 We omit the details.

3.4. Nonlinear time series

Consider the nonlinear autoregressive conditional heteroscedastic model

$$X_i = \mu(X_{i-1}) + \sigma(X_{i-1})\varepsilon_i, \quad (35)$$

where ε_i are i.i.d. random variables. Special cases of (35) include linear AR $X_i = aX_{i-1} + \varepsilon_i$, ARCH (Engle, 1982) $X_i = (a^2 + b^2 X_{i-1}^2)^{1/2} \varepsilon_i$, TAR (Tong, 1990) $X_i = a \max(X_i, 0) + b \min(X_i, 0) + \varepsilon_i$, and EAR (Haggan and Ozaki, 1981) $X_i = [a + b \exp(-cX_{i-1})]X_{i-1} + \varepsilon_i$ among others. Clearly, $\{X_i\}$ is a Markov chain. We refer the reader to Bradley (2005) for discussions on mixing conditions for Markov chains.

Let $\mathcal{Q} = (\mu, \sigma)$. Under H_0 : $\mathcal{Q} = \mathcal{Q}_\theta$ for a specification $\mathcal{Q}_\theta = (\mu(\cdot; \theta), \sigma(\cdot; \theta))$, Zhao (2010) considered the following estimate

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^n \left\{ \left[\frac{X_i - \mu(X_{i-1}; \theta)}{\sigma(X_{i-1}; \theta)} \right]^2 + 2 \log \sigma(X_{i-1}; \theta) \right\}. \quad (36)$$

Under mild conditions, Zhao (2010) derived a Bahadur representation for $\hat{\theta}$ and established its $\sqrt{n}$ -consistency. Assume that ε_i are standard normal random variables. By the plug-in method, the parametrically estimated transition density is

$$\hat{q}_X(x'; \mathcal{Q}_{\hat{\theta}}) = \frac{1}{\sigma(x; \hat{\theta})} \phi \left\{ \frac{x' - \mu(x; \hat{\theta})}{\sigma(x; \hat{\theta})} \right\}. \quad (37)$$

3.5. Discrete-time stochastic volatility models

Let $\{X_i\}$ be samples from the discrete-time stochastic volatility model (3). Special examples have been studied in Ruiz (1994), Jacquier et al. (1994) and Kim et al. (1998). If we are interested in the data-generating mechanism of the unobservable volatility process $\{\sigma_i\}$, then we can let $\mathcal{Q} = (r, s)$. Clearly, $\{X_i\}$ is a HMM with the hidden chain $\{Y_i = \sigma_i\}$.

Now we consider parametric density construction under H_0 : $\mathcal{Q} = \mathcal{Q}_\theta = (r(\cdot; \theta), s(\cdot; \theta))$. Let $\hat{\theta}$ be an estimate of θ ; see Section 3.2 for various estimation methods. Assume that $\{\varepsilon_i\}$ and $\{\eta_i\}$ are i.i.d. standard normal random variables. Then the conditional density $q_{X|Y}$ of X_i given $Y_i = y$ is $q_{X|Y}(x|y) = y^{-1} \phi(x/y)$. By (25), we propose

$$\hat{q}_X(x'; \mathcal{Q}_{\hat{\theta}}) = \frac{m^{-1} \sum_{i=1}^m \phi(x/\sigma_{i-1}^*) \phi(x'/\sigma_i^*) / (\sigma_{i-1}^* \sigma_i^*)}{m^{-1} \sum_{i=1}^m \phi(x/\sigma_{i-1}^*) / \sigma_{i-1}^*},$$

where $\{\sigma_i^{*2}\}$ are simulated samples from the estimated null model
 $\sigma_i^{*2} = r(\sigma_{i-1}^{*2}; \hat{\theta}) + s(\sigma_{i-1}^{*2}; \hat{\theta}) \eta_i$.

3.6. Models with market microstructure noise

Let $\{Y_t\}_{t \in \mathcal{T}}$ be the true process. In practice, we often do not observe Y_t directly but a contaminated version X_t of it. For example, if Y_t is the actual stock price at the t -th sampling time point, then we may only observe $X_t = Y_t + \varepsilon_t$ with i.i.d. errors ε_t . Assume that the errors

$\{\varepsilon_i\}$ are independent of $\{Y_i\}$. This framework has been proposed as models with market microstructure noise (Aït-Sahalia et al., 2005; Zhang et al., 2005).

If $\{Y_i\}$ is a Markov chain, then $\{X_i\}$ is a HMM with the unobservable chain $\{Y_i\}$. To construct parametric density estimate, assume that $\{Y_i\}$ is a Markov chain with data-generating mechanism $\mathcal{Q}$ of interest. Further assume that the contamination errors ε_i are centered normal random variables with variance σ^2 . Denote by $q_{X|Y}$ the conditional density of X_i given Y_i . Then $q_{X|Y}(x|y) = \sigma^{-1} \phi\{(x - y)/\sigma\}$. Under $H_0: \mathcal{Q} = \mathcal{Q}_\theta$, let $(\hat{\sigma}, \hat{\theta})$ be a consistent estimate of (σ, θ) . By (25), we can obtain the parametric density estimate

$$\begin{aligned} \hat{q}_X(x'|x; \mathcal{Q}_{\hat{\theta}}) \\ = \frac{m^{-1} \sum_{i=1}^m \hat{\sigma}^{-1} \phi\{(x - Y_{i-1}^*)/\hat{\sigma}\} \hat{\sigma}^{-1} \phi\{(x' - Y_i^*)/\hat{\sigma}\}}{m^{-1} \sum_{i=1}^m \hat{\sigma}^{-1} \phi\{(x - Y_{i-1}^*)/\hat{\sigma}\}}, \end{aligned} \quad (38)$$

where $\{Y_i^*\}$ are simulated from estimated null model $\mathcal{Q}_{\hat{\theta}}$.

3.7. Extension to higher order Markov models

The proposed transition density based test can be extended to deal with higher order Markov models. Consider

$$X_i = \mu(X_{i-1}, \dots, X_{i-p}) + \sigma(X_{i-1}, \dots, X_{i-p})\varepsilon_i, \quad (39)$$

where μ and σ are p -dimensional functions. Let $\mathcal{Q} = (\mu, \sigma)$. Suppose that we wish to test $H_0: \mathcal{Q} = \mathcal{Q}_\theta$ for a specification $\mathcal{Q}_\theta$. For $p \geq 3$, due to the ‘‘curse of dimensionality’’, it is practically infeasible to obtain a nonparametric estimate of the conditional density of X_i given $X_{i-1}, \dots, X_{i-p}$. Here we shall give a partial solution based on transition density.

In Section 2, our specification testing procedure is based on one-step transition density $q_X(x'|x)$. Here, due to the p th order Markovian structure in (39), we need to consider k -step transition densities, $k = 1, 2, \dots, p$, simultaneously. For $k \in \mathbb{N}$, denote by $q_k(x'|x)$ the conditional density of X_i given $X_{i-k} = x$. Let $\hat{q}_k(x'|x)$ and $\hat{q}_k(x'|x; \mathcal{Q}_\theta)$ be the nonparametric estimate and parametric estimate under H_0 of $q_k(x'|x)$, respectively. For appropriate distance measure $d(\cdot, \cdot)$ and weights $\omega_k > 0$, we may construct a test statistic of the form

$$\begin{aligned} T_n &= \sum_{k=1}^p \omega_k d(\hat{q}_k(\cdot|\cdot), \hat{q}_k(\cdot|\cdot; \mathcal{Q}_{\hat{\theta}})) \quad \text{or} \\ T_n &= \max_{1 \leq k \leq p} d(\hat{q}_k(\cdot|\cdot), \hat{q}_k(\cdot|\cdot; \mathcal{Q}_{\hat{\theta}})). \end{aligned} \quad (40)$$

It is beyond the scope of the present work to explore this approach and further research will be conducted in the future.

4. Finite sample performance

4.1. Kernel function, bandwidth selection, and $\mathfrak{X}_n$

In our data analysis we use the 4-th order kernel $K(u) = 2\varphi(u) - \varphi(u/\sqrt{2})/\sqrt{2}$, where $\varphi(u)$ is the standard Gaussian kernel. For nonparametric problems, the choice of bandwidth is usually more important than that of kernel function. For bandwidth b_n , we adopt the likelihood cross-validation bandwidth selection method. Denote by $\hat{q}_{-i}(x'|x; b_n)$ the estimate in (11) based all samples but the leave-one-out pair (X_i, X_{i+1}) and bandwidth b_n . The likelihood cross-validation bandwidth selection method then selects the optimal b_n^{opt}

$$b_n^{\text{opt}} = \underset{b_n}{\operatorname{argmax}} \sum_{i=0}^{n-1} \log \hat{q}_{-i}(X_{i+1}|X_i; b_n). \quad (41)$$

See Section 5.2.2 in Li and Racine (2007) for more discussion.

In Theorem 2, we need to select a set $\mathfrak{X}_n$ of grid points. For a realization $\{X_i\}$, let $\ell_{0.15}$ and $\ell_{0.85}$ be the 15 and 85 percentiles, respectively. Let s be the sample standard deviation of differences $\{X_{i+1} - X_i\}$. Then the region $\mathfrak{X} = \{(x, x') : x \in [\ell_{0.15}, \ell_{0.85}], |x' - x| \leq s\}$ contains a large proportion of points (X_i, X_{i+1}) , $i = 0, \dots, n$. Partition $[\ell_{0.15}, \ell_{0.85}]$ into 10 intervals of equal length, and denote the grid points by $x_j = \ell_{0.15} + j(\ell_{0.85} - \ell_{0.15})/10$, $j = 0, \dots, 10$. For each x_j , divide $[x_j - s, x_j + s]$ into five intervals of equal length with grid points $x_j \pm s$, $x_j \pm 0.5s$, x_j . We then take $\mathfrak{X}_n = \{(x_j, x_j \pm \tau s), \tau = 1.0, 0.5, 0, j = 0, \dots, 10\}$.

4.2. Accuracies of the asymptotic null distribution

Compared to marginal density estimation, transition density estimation requires larger sample sizes. With the development of modern technology, data sets with sizes of the order of tens of thousands have become available. We simulate $n = 10,000$ daily observations with $\Delta = 1/252$ (one year has approximately 252 trading days) from the Ornstein–Uhlenbeck process (Vasicek, 1977):

$$dX_t = \beta(v - X_t)dt + \sigma dW_t, \quad v > 0, \beta > 0, \sigma > 0. \quad (42)$$

We set $\theta = (\beta, v, \sigma) = (0.2, 0.06, 0.013)$. For simplicity write $X_i = X_{i\Delta}$. By Euler's discretization scheme (29), the true transition density of (42) is

$$q_X(x'|x; \theta) = (\sigma \sqrt{\Delta})^{-1} \varphi\{[x' - x - \beta(v - x)\Delta]/\sigma \sqrt{\Delta}\}.$$

We simulate 1000 realizations of size n from the null model (42). For each realization, we compute the test statistic

$$T_n = \sup_{(x, x') \in \mathfrak{X}_n} \frac{b_n \sqrt{n}}{\varphi_K} \frac{|\hat{q}_X(x'|x) - q_X(x'|x; \hat{\theta})|}{\sqrt{\hat{q}_X(x'|x)/\hat{\pi}_X(x)}}, \quad (43)$$

where $\hat{\theta}$ is the linear regression estimate of θ based on the approximation (29). Using the cross-validation method (41), we find that most realizations from (42) give the optimal bandwidth $b_n \approx 0.0005$. Since (41) is computationally expensive, instead of applying (41) for each realization, we set $b_n = 0.0005$ for all realizations to limit computation; the same technique is also used in Sections 4.3–4.6. We compare the empirical quantiles of these 1000 realized T_n with the asymptotic quantiles derived from Theorem 2. In particular, the asymptotic $(1 - \alpha)$ -quantile is $B_{m_n}(z_\alpha)$, where $B_k(z)$ and z_α are defined as in (18) and (20), respectively. Table 1 presents the empirical and asymptotic quantiles of T_n for different values of $1 - \alpha$. We see that the asymptotic quantiles approximate the empirical quantiles reasonably well.

4.3. Power study for Markov diffusion process

Let $\lambda \in [0, 1]$. Consider the true underlying process

$$dX_t = \beta(\nu - X_t)dt + [(1 - \lambda)\sigma + \lambda \cdot 0.07|X_t|^{0.7}]dW_t, \quad t > 0. \quad (44)$$

If $\lambda = 0$, then (44) reduces to the Vasicek (1977) model (42); if $\lambda = 1$, then it becomes a special example of the CKLS model (Chan et al., 1992):

$$dX_t = \beta(\nu - X_t) + \sigma X_t^\gamma dW_t. \quad (45)$$

For $\lambda \in (0, 1)$, the volatility term $(1 - \lambda)\sigma + \lambda \cdot 0.07|X_t|^{0.7}$ is a weighted version of the volatilities in the two models (42) and (45). The latter two models are among the most widely used interest rates models. In (44), we use the same setting for $\nu, \beta, \sigma, n, \Delta$ as in Section 4.2. We use (44) as our true data generating process and (42) as our null hypothesis H_0 . To study the power of testing H_0 , we use the asymptotic quantile from Table 1 with significance level $\alpha = 0.05$. For a fixed $\lambda \in [0, 1]$, the power is the proportion of realizations among 1000 realizations from (44) that reject H_0 . The size and power for (42) against (44) are 0.034, 0.068, 0.165, 0.428, 0.709, 0.885, corresponding to $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$.

4.4. Power study for jump–diffusion models

We now consider the power when testing (42) against the jump–diffusion model

$$dX_t = \beta(\nu - X_{t-})dt + \sigma(X_{t-})dW_t + J_t \cdot dN_{t-}, \quad (46)$$

where N_t is a Poisson process with intensity $\lambda(X_{t-})$, and $J_t \sim N(0, \eta^2)$ is the independent jump size. As in Aït-Sahalia et al. (2009), consider the specification $\lambda(x) = \lambda$, $\sigma(x) = \xi$, such that

$$\xi^2 + \lambda\eta^2 = \sigma^2, \quad \frac{\lambda\eta^2}{\xi^2 + \lambda\eta^2} = \frac{\tau}{1.1}, \quad \eta^2 = 2\xi^2,$$

where $\sigma = 0.013$ as in (42) and $\tau \in [0, 1]$, with $\tau = 0$ being the null model (42). For all cases, we use $b_n = 0.0005$ to limit computations. The size and power for (42) against (46) are 0.036, 0.032, 0.096, 0.424, 0.896, 0.994, corresponding to $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$.

4.5. Power study for models with market microstructure noise

Consider the following model with market microstructure noise

$$\begin{aligned} X_i &= Y_i + \varepsilon_i, \\ Y_i &= \theta_1 [(1 - \lambda)Y_{i-1} + \lambda|Y_{i-1}|] + \eta_i, |\theta_1| < 1, \lambda \in [0, 1] \end{aligned} \quad (47)$$

where $\varepsilon_i \sim N(0, \theta_3^2)$ and $\eta_i \sim N(0, \theta_2^2)$ are independent noises. Based on the contaminated observations $\{X_i\}$, we wish to test the null hypothesis $H_0: Y_i = \theta_1 Y_{i-1} + \eta_i$ about the unobservable process $\{Y_i\}$. In (47), the parameter λ measures the deviation from the null model, with $\lambda = 0$ being the null model and $\lambda = 1$ being the TAR model $Y_i = \theta_1 |Y_{i-1}| + \eta_i$.

Under H_0 , elementary calculations show that

$$\begin{aligned} \mathbb{E}(X_i^2) &= \frac{\theta_2^2}{1 - \theta_1^2} + \theta_3^2, \quad \text{Cov}(X_{i-1}, X_i) = \frac{\theta_1 \theta_2^2}{1 - \theta_1^2}, \\ \text{Cov}(X_{i-2}, X_i) &= \frac{\theta_1^2 \theta_2^2}{1 - \theta_1^2}. \end{aligned}$$

So, we can use moments based estimation method by replacing the theoretical expectations with their corresponding empirical versions. Using sample size $n = 2000$ and true parameters $(\theta_1, \theta_2, \theta_3) = (0.6, 0.2, 0.2)$, we find that most realizations give optimal bandwidth $b_n \approx 0.12$ for $\lambda \in [0, 1]$. As in Section 4.2, we use $b_n = 0.12$ for all realizations to limit computation. The size and power are 0.027, 0.212, 0.675, 0.943, 0.991, 1.000, corresponding to $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$.

4.6. Power study for stochastic volatility models

Let ε_i be i.i.d. standard normal random variables. Consider stochastic volatility model

$$X_i = \sigma_i \varepsilon_i,$$

with $\{\sigma_i\}$ being a stochastic process given by

$$\begin{aligned} \sigma_i^2 &= 0.01 \exp(v_i) \quad \text{and} \\ v_i &= \theta_1 + \theta_2 v_{i-1} + (1 - \lambda + \lambda \sqrt{0.1 + v_{i-1}^2}) \eta_i, \end{aligned} \quad (48)$$

where η_i are i.i.d. centered normal random variables with variance $\theta_3^2 > 0$. In (48), for $\lambda \neq 0$, $\{v_i\}$ satisfy a autoregressive conditional heteroscedastic model. As in Section 3.5, $\{X_i\}$ form a HMM with respect to the unobservable volatilities $\{\sigma_i\}$. We test the simple linear autoregressive null hypothesis $H_0: v_i = \theta_1 + \theta_2 v_{i-1} + \eta_i$.

Under H_0 , it is easy to check that

$$\begin{aligned}\log[\mathbb{E}(X_i^2)] &= 2\log(0.1) + \frac{\theta_1}{1-\theta_2} + \frac{\theta_3^2}{2(1-\theta_2^2)}, \\ \log[\mathbb{E}(X_i^4)] &= 4\log(0.1) + \log(3) + \frac{2\theta_1}{1-\theta_2} + \frac{2\theta_3^2}{1-\theta_2^2}, \\ \log[\mathbb{E}(X_{i-1}^2 X_i^2)] &= 4\log(0.1) + \frac{2\theta_1}{1-\theta_2} + \frac{\theta_3^3}{1-\theta_2}.\end{aligned}$$

So, we use a moments based method to estimate $(\theta_1, \theta_2, \theta_3)$ by replacing the theoretical expectations with their corresponding empirical versions. Using sample size $n = 20,000$ and true parameters $(\theta_1, \theta_2, \theta_3) = (0.3, 0.6, 0.4)$, we find that most realizations give optimal bandwidth $b_n \approx 0.02$ for $\lambda \in [0, 1]$. As in Section 4.2, we use $b_n = 0.02$ for all realizations to limit computation. The size and power are 0.074, 0.080, 0.124, 0.350, 0.778, 0.968, corresponding to $\lambda = 0.0, 0.2, 0.4, 0.6, 0.8, 1.0$.

5. Proofs

Recall that $\{X_i\}$ is a HMM with respect to the Markov chain $\{Y_i\}$. Also recall that, $\pi_X(x)$, $p_X(x, x')$ and $q_X(x'|x)$ are the marginal density of X_i , joint density of (X_{i-1}, X_i) and transition density or conditional density of X_i given $X_{i-1} = x$, respectively, and $q_{X|Y}(x|y)$ is the conditional density of X_i given $Y_i = y$. Throughout the proofs we let $c_1, c_2, \dots$, be constants that may vary from place to place.

5.1. A martingale decomposition argument

The key idea of our proofs is based on a martingale decomposition argument. First, we illustrate the basic idea. Let $\mathcal{F}_i = \sigma(X_j, Y_{j+1}, j \leq i)$ be the sigma field generated by $X_j, Y_{j+1}, j \leq i$. For $i \in \mathbb{Z}$, define the projection operator $\mathcal{P}_i Z = \mathbb{E}(Z | \mathcal{F}_i) - \mathbb{E}(Z | \mathcal{F}_{i-1})$ for $Z \in \mathcal{L}^1$. By the property of conditional expectation, it can be easily checked that

$$\begin{aligned}\mathbb{E}(\mathcal{P}_i Z | \mathcal{F}_{i-1}) &= \mathbb{E}[\mathbb{E}(Z | \mathcal{F}_i) - \mathbb{E}(Z | \mathcal{F}_{i-1}) | \mathcal{F}_{i-1}] \\ &= \mathbb{E}[\mathbb{E}(Z | \mathcal{F}_i) | \mathcal{F}_{i-1}] - \mathbb{E}(Z | \mathcal{F}_{i-1}) = 0.\end{aligned}$$

Thus, $\{\mathcal{P}_i\}_{i \in \mathbb{Z}}$ form a sequence of martingale difference operators with respect to the filtration $\{\mathcal{F}_i\}_{i \in \mathbb{Z}}$, in the sense that they can transform a sequence of random variables into martingale differences through projection. See Wu (2005) for more discussions.

Let g be any function satisfying $g(X_0, X_1) \in \mathcal{L}^1$ and define

$$S_n = \sum_{i=1}^n [\omega_i - \mathbb{E}(\omega_i)], \quad \text{where } \omega_i = g(X_{i-1}, X_i).$$

Suppose that we want to study the asymptotic behavior of S_n . By definition, it is easily seen that $\omega_i - \mathbb{E}(\omega_i) = \mathcal{P}_i \omega_i + \mathcal{P}_{i-1} \omega_i + [\mathbb{E}(\omega_i | \mathcal{F}_{i-2}) - \mathbb{E}(\omega_i)]$. Thus, we can write S_n as

$$\begin{aligned}S_n &= \sum_{i=1}^n \mathcal{P}_i \omega_i + \sum_{i=1}^n \mathcal{P}_{i-1} \omega_i + \sum_{i=1}^n [\mathbb{E}(\omega_i | \mathcal{F}_{i-2}) - \mathbb{E}(\omega_i)] \\ &:= M_n + N_n + R_n.\end{aligned}$$

In the above expression, since $\{\mathcal{P}_i\}_{i \in \mathbb{Z}}$ are martingale difference operators with respect to the filtration $\{\mathcal{F}_i\}_{i \in \mathbb{Z}}$, M_n is a martingale with respect to $\mathcal{F}_n$ and N_n is a martingale with respect to $\mathcal{F}_{n-1}$, and standard tools for martingales are applicable. The relative order of magnitude of M_n and N_n depends on the dimensionality of the nonparametric inference problem involved. Generally speaking, M_n and N_n are of the same order of magnitude for univariate nonparametric problems, but M_n dominates N_n for bivariate or multivariate nonparametric problems. An intuitive explanation is that, due to conditional expectation, $\mathcal{P}_{i-1} \omega_i = \mathbb{E}(\omega_i | \mathcal{F}_{i-1}) - \mathcal{P}_i(\omega_i | \mathcal{F}_{i-2})$ is smoother than $\mathcal{P}_i \omega_i = \omega_i - \mathbb{E}(\omega_i | \mathcal{F}_{i-1})$ and hence has smaller variance. Thanks to the highest amount of smoothing, R_n is often negligible under mild dependence conditions. We generally call (49) the martingale decomposition.

In Lemma 1 we establish a representation for R_n in (49). To this end we define

$$W_n(x, x') = \sum_{i=1}^n \{ \mathbb{E}[q_{X|Y}(x|Y_{i-1})q_{X|Y}(x'|Y_i)|Y_{i-1}] - \mathbb{E}[q_{X|Y}(x|Y_{i-1})q_{X|Y}(x'|Y_i)] \}. \quad (50)$$

It turns out that R_n is closely related to $W_n(x, x')$. In our subsequent proofs, we shall frequently use the following properties of conditional expectation.

- i. for a random variable $X \in \mathcal{L}^1$ and any sigma-field $\mathcal{F}$, we have $\mathbb{E}(X) = \mathbb{E}[\mathbb{E}(X | \mathcal{F})]$.
- ii. for a random variable $X \in \mathcal{L}^1$ and any two sigma-fields $\mathcal{F} \subset \mathcal{G}$, we have $\mathbb{E}(X | \mathcal{F}) = \mathbb{E}[\mathbb{E}(X | \mathcal{G}) | \mathcal{F}]$.

Lemma 1—Let $g(\cdot, \cdot)$ be a bivariate measurable function such that $g(X_0, X_1) \in \mathcal{L}^1$. Define

$$R_n = \sum_{i=1}^n \{ \mathbb{E}[g(X_{i-1}, X_i) | \mathcal{F}_{i-2}] - \mathbb{E}[g(X_{i-1}, X_i)] \}. \quad (51)$$

Assume for simplicity that $g(\cdot, \cdot)$ is a bounded function so that expectation and integral can be exchanged. Let $W_n(x, x')$ be as in (50). Then

$$R_n = \iint g(x, x') W_n(x, x') dx dx'. \quad (52)$$

Proof—Since $\{X_i\}$ is a HMM with respect to $\{Y_i\}$, given $\mathcal{F}_{i-1}$, the conditional distribution of X_i depends only on Y_i . We have

$$\mathbb{E}[g(X_{i-1}, X_i) | \mathcal{F}_{i-1}] = \int g(X_{i-1}, x') q_{X|Y}(x' | Y_i) dx'. \quad (53)$$

So, by the property of conditional expectation,

$$\begin{aligned}\mathbb{E}[g(X_{i-1}, X_i)|\mathcal{F}_{i-2}] &= \mathbb{E}\{\mathbb{E}[g(X_{i-1}, X_i)|\mathcal{F}_{i-1}|\mathcal{F}_{i-2}]\} \\ &= \int \mathbb{E}[g(X_{i-1}, x')q_{X|Y}(x'|Y_i)|\mathcal{F}_{i-2}]dx' .\end{aligned}$$

Notice that, by the HMM property,

$$\begin{aligned}\mathbb{E}[g(X_{i-1}, x')q_{X|Y}(x'|Y_i)|\mathcal{F}_{i-2}, Y_i] \\ = q_{X|Y}(x'|Y_i)\mathbb{E}[g(X_{i-1}, x')|Y_{i-1}].\end{aligned}$$

Thus, by the property of conditional expectation and the Markovian property of $\{Y_i\}$,

$$\begin{aligned}\mathbb{E}[g(X_{i-1}, x')q_{X|Y}(x'|Y_i)|\mathcal{F}_{i-2}] \\ &= \mathbb{E}\{\mathbb{E}[g(X_{i-1}, x')q_{X|Y}(x'|Y_i)|\mathcal{F}_{i-2}, Y_i]|\mathcal{F}_{i-2}\} \\ &= \mathbb{E}[g(X_{i-1}, x')|Y_{i-1}] \times \mathbb{E}[q_{X|Y}(x'|Y_i)|Y_{i-1}] \\ &= \int g(x, x')q_{X|Y}(x|Y_{i-1})dx \times \mathbb{E}[q_{X|Y}(x'|Y_i)|Y_{i-1}] \\ &= \int g(x, x')\mathbb{E}[q_{X|Y}(x|Y_{i-1})q_{X|Y}(x'|Y_i)|Y_{i-1}]dx.\end{aligned}\tag{54}$$

Also, notice that,

$$\begin{aligned}\mathbb{E}[g(X_{i-1}, X_i)] &= \mathbb{E}\{\mathbb{E}[g(X_{i-1}, X_i)|\mathcal{F}_{i-2}]\} \\ &= \iint g(x, x')\mathbb{E}[q_{X|Y}(x|Y_{i-1})q_{X|Y}(x'|Y_i)]dx dx',\end{aligned}\tag{55}$$

completing the proof.

Lemma 2 is needed to establish a uniform bound for $W_n(x, x')$ in Lemma 3.

Lemma 2—Let $g(\cdot, \cdot)$ be a differentiable bivariate function. For fixed $\delta_1, \delta_2 > 0$ and $a, b \in \mathbb{R}$, denote by $I_{a,b}$ the set $\{(u, v) \in \mathbb{R}^2: u \in [a, a + \delta_1], v \in [b, b + \delta_2]\}$. Then there exists a constant $C < \infty$, depending only on δ_1 and δ_2 , such that

$$\sup_{(u,v) \in I_{a,b}} g^2(u, v) \leq C \int_{I_{a,b}} \left\{ g^2(u, v) + \left[\frac{\partial g(u, v)}{\partial u} \right]^2 + \left[\frac{\partial g(u, v)}{\partial v} \right]^2 + \left[\frac{\partial^2 g(u, v)}{\partial u \partial v} \right]^2 \right\} du dv.$$

Proof—For $(u, v), (u_0, v_0) \in I_{a,b}$, let $\bar{g}(u, v, u_0, v_0) = g(u, v) - g(u_0, v) - g(u, v_0) + g(u_0, v_0)$. By the Cauchy–Schwarz inequality,

$$g^2(u, v) \leq 4[\bar{g}(u, v, u_0, v_0)^2 + g^2(u_0, v) + g^2(u, v_0) + g^2(u_0, v_0)]. \quad (56)$$

Notice that

$$\begin{aligned} |\bar{g}(u, v, u_0, v_0)| &= \left| \int_{u_0}^u \int_{v_0}^v \frac{\partial^2 g(u, v)}{\partial u \partial v} du dv \right| \\ &\leq \int_{I_{a,b}} \left| \frac{\partial^2 g(u, v)}{\partial u \partial v} \right| du dv, \end{aligned}$$

uniformly over $(u, v), (u_0, v_0) \in I_{a,b}$. Again, by the Cauchy–Schwarz inequality,

$$\sup_{(u,v) \in I_{a,b}} \bar{g}(u, v, u_0, v_0)^2 \leq \delta_1 \delta_2 \int_{I_{a,b}} \left[\frac{\partial^2 g(u, v)}{\partial u \partial v} \right]^2 du dv. \quad (57)$$

By Lemma 4 in Wu (2003), we have

$$\sup_{u \in [a, a+\delta_1]} g^2(u, v_0) \leq 2\delta_1^{-1} \int_a^{a+\delta_1} g^2(u, v_0) du + 2\delta_1 \int_a^{a+\delta_1} \left[\frac{\partial g(u, v_0)}{\partial u} \right]^2 du, \quad (58)$$

$$\sup_{v \in [b, b+\delta_2]} g^2(u_0, v) \leq 2\delta_2^{-1} \int_b^{b+\delta_2} g^2(u_0, v) dv + 2\delta_2 \int_b^{b+\delta_2} \left[\frac{\partial g(u_0, v)}{\partial v} \right]^2 dv. \quad (59)$$

Inserting (57)–(59) into the right-hand side of (56) and then taking an integral on both sides of the resulting inequality over $(u_0, v_0) \in I_{a,b}$, we obtain the desired result.

Lemma 3—Recall $W_n(x, x')$ in (50). Assume that Conditions 1 and 3 hold. Then

$$\left\| \sup_{(x,x') \in \mathcal{X}} |W_n(x, x')| \right\|_2 = O(\sqrt{n}).$$

Proof—By Lemma 2, it suffices to prove

$$\begin{aligned} \sup_{(x,x') \in \mathcal{X}} \left\{ \|W_n(x, x')\|_2^2 + \left\| \frac{\partial W_n(x, x')}{\partial x} \right\|_2^2 \right. \\ \left. + \left\| \frac{\partial W_n(x, x')}{\partial x'} \right\|_2^2 + \left\| \frac{\partial^2 W_n(x, x')}{\partial x \partial x'} \right\|_2^2 \right\} = O(n). \end{aligned}$$

First, we consider

$$\sup_{(x,x') \in \mathcal{X}} \|W_n(x, x')\|_2^2$$

. Under Conditions 1 and 3, the summands in $W_n(x, x')$ are α -mixing and uniformly bounded.

By Theorem 2.20 in Fan and Yao (2003), $\|W_n(x, x')\|_2^2 = O(n)$, uniformly over $(x, x') \in \mathcal{X}$.

For other terms, by the boundedness of $|\partial q_{X|Y}(x|y)/\partial x|$, we can exchange the order of the differentiation and expectation in $W_n(x, x')$ in view of the Dominated Convergence Theorem. Thus, they can be treated using the same argument as above.

5.2. Proof of Theorems 1 and 2

Recall $\hat{q}_X(x, x')$, $\hat{p}_X(x, x')$ and $\hat{\pi}_X(x)$ in (11)–(13), respectively. Define

$$\xi_i(x, x') = K_{b_n}(x - X_{i-1})K_{b_n}(x' - X_i), \quad (60)$$

$$S_n(x, x') = \frac{1}{nb_n^2} \sum_{i=1}^n \{\xi_i(x, x') - \mathbb{E}[\xi_i(x, x')]\}, \quad (61)$$

$$U_n(x, x') = b_n^{-2} \mathbb{E}[\xi_1(x, x')] - p_X(x, x'), \quad (62)$$

$$V_n(x, x') = p_X(x, x') [1 - \hat{\pi}_X(x)/\pi_X(x)]. \quad (63)$$

Then we have

$$\hat{\pi}_X(x) [\hat{q}_X(x' | x) - q_X(x' | x)] = S_n(x, x') + U_n(x, x') + V_n(x, x'), \quad (64)$$

where $S_n(x, x')$ is the stochastic part determining the asymptotic distribution of $\hat{q}_X(x' | x)$, $U_n(x, x')$ is the bias part due to $\hat{p}_X(x, x')$, and $V_n(x, x')$ is the bias part due to $\hat{\pi}_X(x)$. We shall treat them separately.

For $U_n(x, x')$, by Conditions 2 and 3 and the Taylor expansion of $p_X(x - ub_n, x' - vb_n)$ around (x, x') , elementary calculations show that

$$\begin{aligned} b_n^{-2} \mathbb{E}[\xi_1(x, x')] &= b_n^{-2} \iint K\left(\frac{x-x_0}{b_n}\right) K\left(\frac{x'-x_1}{b_n}\right) p_X(x_0, x_1) dx_0 dx_1 \\ &= \iint K(u)K(v) p_X(x - ub_n, x' - vb_n) du dv \\ &= p_X(x, x') + O(b_n^4), \end{aligned}$$

uniformly over $(x, x') \in \mathcal{X}$. Thus, $U_n(x, x') = O(b_n^4)$ uniformly over $(x, x') \in \mathcal{X}$.

Now we consider $S_n(x, x')$. By the martingale decomposition technique in (49), we have

$$S_n(x, x') = M_n(x, x') + N_n(x, x') + R_n(x, x'), \quad (65)$$

where

$$M_n(x, x') = \frac{1}{nb_n^2} \sum_{i=1}^n \mathcal{P}_i \xi_i(x, x'), \quad (66)$$

$$N_n(x, x') = \frac{1}{nb_n^2} \sum_{i=1}^n \mathcal{P}_{i-1} \xi_i(x, x'), \quad (67)$$

$$R_n(x, x') = \frac{1}{nb_n^2} \sum_{i=1}^n \{\mathbb{E}[\xi_i(x, x') | \mathcal{F}_{i-2}] - \mathbb{E}[\xi_i(x, x')]\}. \quad (68)$$

To establish a uniform bound for $N_n(x, x')$, we need the following Lemma 4.

Lemma 4—Let Conditions 2 and 3 hold. Define

$$H_n(x, x') = \sum_{i=1}^n \mathcal{P}_{i-1} \zeta_i(x, x'),$$

where $\zeta_i(x, x') = K_{b_n}(x - X_{i-1})q_{X|Y}(x' | Y_i).$ (69)

where $\zeta_i(x, x') = K_{b_n}(x - X_{i-1})q_{X|Y}(x' | Y_i).$

Assume $b_n \rightarrow 0$ and $\sup_n \log n / (nb_n) < \infty$. Then

$$\sup_{(x, x') \in \mathcal{X}} |H_n(x, x')| = O_{\mathbb{P}}(\sqrt{nb_n \log n}). \quad (70)$$

Proof—We shall use a chain argument. The basic idea is to approximate $H_n(x, x')$ by discrete versions over finer grid points. For simplicity write the compact set $\mathcal{X} = [0, T] \times [0, T]$ for some $0 < T < \infty$. Let $N = n^2$, $\omega_n = T/N$, $x_j = j\omega_n$, $0 \leq j \leq N$. Then $x_0, \dots, x_N$ partition $[0, T]$ into uniformly spaced intervals with equal length ω_n . For any $x, x' \in [0, T]$, there exist j and k such that $x \in [x_j, x_{j+1})$ and $x' \in [x_k, x_{k+1})$. Then we have

$$|K_{b_n}(x - X_{i-1}) - K_{b_n}(x_j - X_{i-1})| \leq c_1(x - x_j)/b_n \leq c_1\omega_n/b_n, \quad (71)$$

where $c_1 = \sup |K'(u)|$. Similarly,

$$|q_{X|Y}(x'|Y_i) - q_{X|Y}(x_k|Y_i)| \leq c_2(x' - x_k) \leq c_2\omega_n, \quad (72)$$

where $c_2 = \sup_{x,y} |\partial q_{X|Y}(x|y)/\partial x|$. Thus, by the boundedness of $K(u)$ and $q_{X|Y}(x|y)$, when $b_n \rightarrow 0$, there exists constant c_3 such that

$$\begin{aligned} |\zeta_i(x, x') - d_{ijk}| &\leq (c_1\omega_n/b_n + c_2\omega_n) \times \left[\sup_u |K(u)| + \sup_{x,y} q_{X|Y}(x|y) \right] \\ &\leq c_3\omega_n/b_n, \end{aligned} \quad (73)$$

where d_{ijk} is defined on discrete grid point (x_j, x_k) as

$$d_{ijk} = \zeta_i(x_j, x_k) = K_{b_n}(x_j - X_{i-1})q_{X|Y}(x_k|Y_i).$$

We then obtain

$$\begin{aligned} |\mathcal{P}_{i-1}\xi_i(x, x') - \mathcal{P}_{i-1}d_{ijk}| &\leq |\mathbb{E}\{[\xi_i(x, x') - d_{ijk}]|\mathcal{F}_{i-1}\}| \\ &\quad + |\mathbb{E}\{[\xi_i(x, x') - d_{ijk}]|\mathcal{F}_{i-2}\}| \\ &\leq 2c_3\omega_n/b_n. \end{aligned}$$

Notice that c_3 does not depend on the choices of i, j, k . So, $H_n(x, x')$ can be uniformly bounded as

$$\sup_{(x, x') \in \mathcal{X}} |H_n(x, x')| \leq \max_{0 \leq j, k \leq N} |D_n(j, k)| + 2c_3n\omega_n/b_n, \quad (74)$$

where

$$D_n(j, k) = \sum_{i=1}^n \mathcal{P}_{i-1}d_{ijk}.$$

Clearly, in (74), $n\omega_n/b_n = O[1/(nb_n)] \rightarrow 0$ with $\omega_n = O(1/n^2)$.

Now we consider $\max_{0 \leq j, k \leq N} |D_n(j, k)|$. Notice that for fixed j and k , $\{\mathcal{P}_{i-1}d_{ijk}\}_{i \in \mathbb{Z}}$ form martingale differences with respect to the filtration $\{\mathcal{F}_{i-1}\}_{i \in \mathbb{Z}}$. Elementary calculations show that there exists some constant $c_4 < \infty$ such that

$$\mathbb{E}[(\mathcal{P}_{i-1}d_{ijk})^2|\mathcal{F}_{i-2}] \leq \mathbb{E}(d_{ijk}^2|\mathcal{F}_{i-2}) \leq c_4b_n,$$

uniformly over i, j, k . So,

$$\sum_{i=1}^n \mathbb{E}[(\mathcal{P}_{i-1}d_{ijk})^2|\mathcal{F}_{i-2}] \leq c_4nb_n$$

. Assume without loss of generality that $\sup_{u,x,y} |K(u)q_{X|Y}(x|y)| \leq 1$. By Freedman's exponential inequality (Freedman, 1975) for bounded martingale differences, for any $c > 0$,

$$\begin{aligned} & \mathbb{P}(|D_n(j, k)| \geq c \sqrt{nb_n \log n}) \\ & \leq 2 \exp \left[-\frac{c^2 nb_n \log n}{2c \sqrt{nb_n \log n} + 2c_4 nb_n} \right] \\ & = 2 \exp \left[-\frac{c^2 \log n}{2c \sqrt{\log n / (nb_n)} + 2c_4} \right] = O(n^{-\lambda_c}), \end{aligned}$$

uniformly over j, k , where $\lambda_c = c^2/(2cc_5 + 2c_4)$ and $c_5 = \sup_n \sqrt{\log n / (nb_n)} < \infty$. Thus,

$$\begin{aligned} \mathbb{P} \left(\max_{0 \leq j, k \leq N} |D_n(j, k)| \geq c \sqrt{nb_n \log n} \right) & \leq \sum_{0 \leq j, k \leq N} \mathbb{P}(|D_n(j, k)| \geq c \sqrt{nb_n \log n}) \\ & = O(N^2 n^{-\lambda_c}) = O[n^{-(\lambda_c - 4)}]. \end{aligned}$$

Thus, by (74), the proof is completed by choosing a sufficiently large c so that $\lambda_c > 4$.

Lemmas 5 and 6 give uniform bounds for $N_n(x, x')$ and $R_n(x, x')$, respectively.

Lemma 5—Recall $N_n(x, x')$ in (67). Assume that the conditions in Lemma 4 hold. Then

$$\sup_{(x, x') \in \mathcal{X}} |N_n(x, x')| = O_P[\sqrt{\log n / (nb_n)}].$$

(75)

Proof—Applying the identity (53), we obtain

$$\begin{aligned} & \mathbb{E}[\xi_i(x, x')|\mathcal{F}_{i-1}] \\ & = K_{b_n}(x - X_{i-1}) \int K_{b_n}(x' - z)q_{X|Y}(z|Y_i)dv \\ & = b_n \int K(u)K_{b_n}(x - X_{i-1})q_{X|Y}(x' - ub_n|Y_i)du \\ & = b_n \int K(u)\xi_i(x, x' - ub_n)du, \end{aligned}$$

(76)

where $\xi_i(x, x')$ is defined as in (69). Thus, we have the identity

$$\begin{aligned}\mathbb{E}[\xi_i(x, x')|\mathcal{F}_{i-2}] &= \mathbb{E}\{\mathbb{E}[\xi_i(x, x')|\mathcal{F}_{i-1}]\mathcal{F}_{i-2}\} \\ &= b_n \int K(u)\mathbb{E}[\zeta_i(x, x' - ub_n)|\mathcal{F}_{i-2}]du.\end{aligned}$$

Combining the above two identities, we have $\mathcal{P}_{i-1}\zeta_i(x, x') = b_n \int K(u) \mathcal{P}_{i-1}\zeta_i(x, x' - ub_n)du$. Recall $H_n(x, x')$ in (69). Now we have

$$\begin{aligned}N_n(x, x') &= \frac{b_n}{nb_n^2} \int K(u)H_n(x, x' - ub_n)du \\ &= O_p[\sqrt{\log n/(nb_n)}]\end{aligned}$$

in view of Lemma 4.

Lemma 6—Recall $R_n(x, x')$ in (68). Assume that Conditions 1–3 hold. Then

$$\left\| \sup_{(x, x') \in \mathcal{X}} |R_n(x, x')| \right\|_2 = O(n^{-1/2}). \quad (77)$$

Proof—Recall $W_n(x, y)$ in (50). Applying Lemma 1 with $g(z, z') = K_{b_n}(x - z)K_{b_n}(x' - z')$, we obtain

$$\begin{aligned}R_n(x, x') &= (nb_n^2)^{-1} \iint K\left(\frac{x-z}{b_n}\right) K\left(\frac{x'-z'}{b_n}\right) W_n(z, z') dz dz' \\ &= n^{-1} \iint K(u)K(v)W_n(x - ub_n, x' - vb_n) du dv,\end{aligned} \quad (78)$$

in view of the change-of-variable $u = (x - z)/b_n$, $v = (x' - z')/b_n$. Thus, by Lemma 3, the proof is completed.

Lemma 7—Recall $\hat{\pi}_X(x)$ in (13). Assume that Conditions 1–3 hold. Then

$$\sup_{x \in [-T, T]} |\hat{\pi}_X(x) - \pi_X(x)| = O[b_n^4 + \sqrt{\log n/(nb_n)}]. \quad (79)$$

Proof—Applying the martingale decomposition technique in Section 5.1, we write

$$\hat{\pi}_X(x) - \pi_X(x) = Z_{n,1}(x) + Z_{n,2}(x) + Z_{n,3}(x),$$

where

$$\begin{aligned}
Z_{n,1}(x) &= \frac{1}{nb_n} \sum_{i=1}^n \mathcal{P}_{i-1} K_{b_n}(x - X_{i-1}), \\
Z_{n,2}(x) &= \frac{1}{nb_n} \sum_{i=1}^n \{\mathbb{E}[K_{b_n}(x - X_{i-1}) | \mathcal{F}_{i-2}] - \mathbb{E}K_{b_n}(x - X_{i-1})\}, \\
Z_{n,3}(x) &= b_n^{-1} \mathbb{E}K_{b_n}(x - X_{i-1}) - \pi_X(x).
\end{aligned}$$

We treat $Z_{n,1}(x)$, $Z_{n,2}(x)$ and $Z_{n,3}(x)$ separately. By the same argument in Lemma 4, we can show that $\sup_{x \in [-T, T]} |Z_{n,1}(x)| = O_{\mathbb{P}}[\sqrt{\log n / (nb_n)}]$. For $Z_{n,2}(x)$, notice that

$$\begin{aligned}
\mathbb{E}[K_{b_n}(x - X_{i-1}) | \mathcal{F}_{i-2}] &= \int K_{b_n}(x - z) q_{X|Y}(z | Y_{i-1}) dz \\
&= b_n \int K(u) q_{X|Y}(x - ub_n | Y_{i-1}) du.
\end{aligned}$$

Define

$$I_n(x) = \sum_{i=1}^n \{q_{X|Y}(x | Y_{i-1}) - \mathbb{E}[q_{X|Y}(x - ub_n | Y_{i-1})]\}.$$

By a similar argument in Lemma 3, it can be shown that $\sup_{x \in [-T, T]} |I_n(x)| = O_{\mathbb{P}}(\sqrt{n})$. Thus

$$Z_{n,2}(x) = \frac{1}{n} \int K(u) I_n(x - ub_n) du = O_{\mathbb{P}}(n^{-1/2}),$$

uniformly over $x \in [-T, T]$. By Taylor's expansion and Condition 2, it is easily seen that $Z_{n,3}(x) = O(b_n^4)$ uniformly over $x \in [-T, T]$. Thus, the desired result follows.

The following Lemma 8 is needed to study the conditional variance in Proposition 3 and quadratic characteristic of multidimensional martingale in Theorem 3.

Lemma 8—Let $\zeta_i(x, x')$ be as in (69). Define

$$I_n(x, x', z, z') = \sum_{i=1}^n \zeta_i(x, x') \zeta_i(z, z'). \quad (80)$$

Assume $nb_n \rightarrow \infty$. Then

$$\sup_{(x, x'), (z, z') \in \mathcal{X}} \|I_n(x, x', z, z')\|_2 = O(nb_n).$$

Proof—Let $d_i(x, x', z, z') = \zeta_i(x, x')\zeta_i(z, z') - \mathbb{E}[\zeta_i(x, x')\zeta_i(z, z') | \mathcal{F}_{i-2}]$. Then $\{d_i(x, x', z, z')\}_{i \in \mathbb{Z}}$ form martingale differences with respect to $\{\mathcal{F}_{i-1}\}_{i \in \mathbb{Z}}$. By the orthogonality of martingale differences, it is easy to see that

$$\begin{aligned} \left\| \sum_{i=1}^n d_i(x, x', z, z') \right\|_2^2 &= \sum_{i=1}^n \left\| d_i(x, x', z, z') \right\|_2^2 \\ &\leq \sum_{i=1}^n \left\| \zeta_i(x, x')\zeta_i(z, z') \right\|_2^2 = O(nb_n), \end{aligned}$$

uniformly over x, x', z, z' . Since $\mathbb{E}[\zeta_i(x, x')\zeta_i(z, z') | \mathcal{F}_{i-2}] = O(b_n)$ uniformly over i, x, x', z, z' . By the triangle inequality, the desired result then follows from

$$\begin{aligned} \|I_n(x, x', z, z')\|_2 &\leq \left\| \sum_{i=1}^n d_i(x, x', z, z') \right\|_2 \\ &\quad + \left\| \sum_{i=1}^n \mathbb{E}[\zeta_i(x, x')\zeta_i(z, z') | \mathcal{F}_{i-2}] \right\|_2. \quad \square \end{aligned}$$

Lemma 9—Let Conditions 2 and 3 hold. Define

$$J_n(x, x') = \sum_{i=1}^n \{\beta_i(x, x') - \mathbb{E}[\beta_i(x, x')]\},$$

where $\beta_i(x, x') = K_{b_n}^2(x - X_{i-1})q_{X|Y}(x' | Y_i)$.

Assume $b_n \rightarrow 0$ and $nb_n \rightarrow \infty$. Then

$$\sup_{(x, x') \in \mathcal{X}} \|J_n(x, x')\|_2 = O(\sqrt{nb_n}).$$

Proof—Apply the martingale decomposition technique in Section 5.1 and write $J_n(x, x')$ as

$$J_n(x, x') = \bar{J}_n(x, x') + \underline{J}_n(x, x'),$$

where

$$\begin{aligned} \bar{J}_n(x, x') &= \sum_{i=1}^n \mathcal{P}_{i-1} \beta_i(x, x'), \\ \underline{J}_n(x, x') &= \sum_{i=1}^n \{\mathbb{E}[\beta_i(x, x') | \mathcal{F}_{i-2}] - \mathbb{E}[\beta_i(x, x')]\}. \end{aligned}$$

Since $\{\beta_i(x, x')\}_{i \in \mathbb{Z}}$ form martingale differences with respect to $\{\mathcal{F}_{i-1}\}$, by the orthogonality of martingale differences, we have

$$\|\bar{J}_n(x, x')\|_2^2 = \sum_{i=1}^n \|\beta_{i-1}(x, x')\|_2^2 \leq \sum_{i=1}^n \|\beta_i(x, x')\|_2^2 = O(nb_n),$$

uniformly over x, x' . By the same argument in (54) and (55), we can write

$$\begin{aligned} \underline{J}_n(x, x') &= \int K_{b_n}^2(x - z) W_n(z, x') dz \\ &= b_n \int K^2(u) W_n(x - ub_n, x') du, \end{aligned}$$

where $W_n(x, x')$ is defined as in (50). By Lemma 3, $\|\underline{J}_n(x, x')\|_2 = O(b_n \sqrt{n})$ uniformly over x, x' . The desired result then follows from the triangle inequality

$$\|J_n(x, x')\|_2 \leq \|\bar{J}_n(x, x')\|_2 + \|\underline{J}_n(x, x')\|_2 = O(\sqrt{nb_n}) + O(b_n \sqrt{n}).$$

Proposition 3 and Theorem 3 present CLT and maximal deviation results for $M_n(x, x')$ in (66).

Proposition 3—Recall $M_n(x, x')$ in (66). Assume that Conditions 1–3 hold. Further assume that $b_n \rightarrow 0$ and $nb_n^2 \rightarrow \infty$. Let $(x, x') \in \mathcal{X}$. Then as $n \rightarrow \infty$,

$$\tilde{M}_n(x, x') := \frac{b_n \sqrt{n}}{\phi_K} \frac{M_n(x, x')}{\sqrt{p_X(x, x')}} \Rightarrow N(0, 1). \quad (81)$$

Moreover, for distinct points $(x_j, x'_j), j = 1, \dots, k$, in $\mathcal{X}$, $\tilde{M}_n(x_j, x'_j), j = 1, \dots, k$, are asymptotically independent.

Proof—We shall only show the convergence (81) for any fixed point $(x, x') \in \mathcal{X}$ since the asymptotic independence can be similarly treated by considering linear combinations of $\tilde{M}_n(x_j, x'_j), j = 1, \dots, k$ via the Cramér-Wold device. By Section 5.1, $\tilde{M}_n(x, x')$ is a martingale with respect to $\mathcal{F}_n$, so it suffices to verify the convergence of the conditional variance and the Lindeberg condition in order to prove (81). The convergence of the conditional variance is verified in the proof of Theorem 3; see q_{rs} in (85) with $r = s$. It remains to verify the Lindeberg condition.

Recall $\xi_i(x, x')$ in (60). Let $\gamma_i(x, x')$ be defined as in (83). Then $\tilde{M}_n(x, x') = \sum_{i=1}^n \gamma_i(x, x')$. Since (x, x') is fixed, we suppress the dependence on (x, x') and write $\xi_i = \xi_i(x, x')$ and $\gamma_i = \gamma_i(x, x')$. By the boundedness of $q_{XY}(x|y)$, it is easy to see that $|\mathbb{E}(\xi_i | \mathcal{F}_{i-1})| \leq c_1 b_n$ for some constant c_1 . For any $c_2 > 0$, let $c_3 = c_2 [\phi_K^2 p_X(x, x')]^{1/2}$. Because $b_n \rightarrow 0$ and $nb_n^2 \rightarrow \infty$, $c_3 \sqrt{nb_n^2} - c_1 b_n \geq c_3 \sqrt{nb_n^2}/2$ for large enough n . Therefore,

$$\begin{aligned}
& \sum_{i=1}^n \mathbb{E}(\gamma_i^2 \mathbf{1}_{|\gamma_i| \geq c_2}) \\
&= \frac{1}{\varphi_K^2 p_X(x, x') b_n^2} \mathbb{E}\{[\xi_1 - \mathbb{E}(\xi_1 | \mathcal{F}_0)]^2 \mathbf{1}_{|\xi_1 - \mathbb{E}(\xi_1 | \mathcal{F}_0)| \geq c_3 \sqrt{nb_n^2}}\} \\
&\leq \frac{1}{\varphi_K^2 p_X(x, x') b_n^2} \mathbb{E}\{(\xi_1^2 + c_1^2 b_n^2) \mathbf{1}_{|\xi_1| \geq c_3 \sqrt{nb_n^2} - c_1 b_n}\} \\
&= O(b_n^{-2}) \mathbb{E}(\xi_1^2 \mathbf{1}_{|\xi_1| \geq c_3 \sqrt{nb_n^2}/2}) + O(1) \mathbb{P}(|\xi_1| \geq c_3 \sqrt{nb_n^2}/2).
\end{aligned}$$

Applying the inequality $\mathbb{E}(\xi_1^2 \mathbf{1}_{|\xi_1| \geq \lambda}) \leq \mathbb{E}(|\xi_1|^3)/\lambda = O(b_n^2/\lambda) = o(b_n^2)$ with $\lambda = c_3 \sqrt{nb_n^2}/2 \rightarrow \infty$, we obtain $O(b_n^{-2}) \mathbb{E}(\xi_1^2 \mathbf{1}_{|\xi_1| \geq c_3 \sqrt{nb_n^2}/2}) \rightarrow 0$. By the boundedness of ξ_1 , $\mathbb{P}(|\xi_1| \geq c_3 \sqrt{nb_n^2}/2) \rightarrow 0$. Thus, we conclude that $\sum_{i=1}^n \mathbb{E}(\gamma_i^2 \mathbf{1}_{|\gamma_i| \geq c_2}) \rightarrow 0$, completing the proof.

Theorem 3—Recall $M_n(x, x')$ in (66). Let m_n , $B_n(z)$ and $\mathcal{X}_n$ be as in Theorem 2. Assume that Conditions 1–3 hold. Further assume that

$$m_n^2 (\log m_n)^3 [b_n^3 + (nb_n^2)^{-1}] \rightarrow 0. \quad (82)$$

Then for every $z \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} \mathbb{P} \left\{ \sup_{(x, x') \in \mathcal{X}_n} \frac{b_n \sqrt{n}}{\varphi_K} \frac{|M_n(x, x')|}{\sqrt{p_X(x, x')}} \leq B_{m_n}(z) \right\} = e^{-2e^{-z}}.$$

Proof—We follow the argument in Zhao and Wu (2008). Recall $\zeta_i(x, x') = K_{b_n}(x - X_{i-1})K_{b_n}(x' - X_i)$ in (60). Let

$$\gamma_i(x, x') = [nb_n^2 \varphi_K^2 p_X(x, x')]^{-1/2} \{\xi_i(x, x') - \mathbb{E}[\xi_i(x, x') | \mathcal{F}_{i-1}]\}. \quad (83)$$

For fixed $k \in \mathbb{N}$ distinct integers $0 \leq j_1, j_2, \dots, j_k \leq m_n - 1$, define the k -dimensional vector $\zeta_i = [\gamma_i(x_{j_1}, x'_{j_1}), \dots, \gamma_i(x_{j_k}, x'_{j_k})]^T$ and

$$M_{n,k} = \sum_{i=1}^n \zeta_i = [\tilde{M}_n(x_{j_1}, x'_{j_1}), \dots, \tilde{M}_n(x_{j_k}, x'_{j_k})]^T.$$

Here T denotes the transpose and $\tilde{M}_n(x, x')$ is defined as in (81). Then $\{\zeta_i\}_{i \in \mathbb{Z}}$ are k -dimensional vectors of martingale differences with respect to $\{\mathcal{F}_i\}_{i \in \mathbb{Z}}$.

Denote by Q_n the $k \times k$ quadratic characteristic matrix of $M_{n,k}$. That is,

$$Q_n = \sum_{i=1}^n \mathbb{E}(\xi_i \xi_i^T | \mathcal{F}_{i-1}) := (q_{rs})_{1 \leq r, s \leq k}. \quad (84)$$

Let $\tau_{rs} = \phi_K^2 [p_X(x_{j_r}, x'_{j_r}) p_X(x_{j_s}, x'_{j_s})]^{1/2}$. Then we can write q_{rs} as

$$q_{rs} = \sum_{i=1}^n \mathbb{E}[\gamma_i(x_{j_r}, x'_{j_r}) \gamma_i(x_{j_s}, x'_{j_s}) | \mathcal{F}_{i-1}] = q_{rs}(1) - q_{rs}(2), \quad (85)$$

where

$$q_{rs}(1) = \frac{1}{nb_n^2 \tau_{rs}} \sum_{i=1}^n \mathbb{E}[\xi_i(x_{j_r}, x'_{j_r}) \xi_i(x_{j_s}, x'_{j_s}) | \mathcal{F}_{i-1}],$$

$$q_{rs}(2) = \frac{1}{nb_n^2 \tau_{rs}} \sum_{i=1}^n \mathbb{E}[\xi_i(x_{j_r}, x'_{j_r}) | \mathcal{F}_{i-1}] \mathbb{E}[\xi_i(x_{j_s}, x'_{j_s}) | \mathcal{F}_{i-1}].$$

By (76), for all r, s ,

$$\begin{aligned} \|q_{rs}(2)\|_2 &= \frac{1}{nb_n^2 \tau_{rs}} \left\| b_n^2 \iint K(u) K(v) \sum_{i=1}^n \xi_i(x_{j_r}, x'_{j_r} - ub_n) \times \xi_i(x_{j_s}, x'_{j_s} - vb_n) du dv \right\|_2 \\ &= O(b_n), \end{aligned}$$

in view of the Cauchy–Schwarz inequality and Lemma 8. For $q_{rs}(1)$, we consider two cases $r \neq s$ and $r = s$ separately. For $r \neq s$, since $\max\{|x_{j_r} - x_{j_s}|, |x'_{j_r} - x'_{j_s}|\} \geq 2\omega b_n$ and K has support $[-\omega, \omega]$, we have $\xi_i(x_{j_r}, x'_{j_r}) \xi_i(x_{j_s}, x'_{j_s}) = 0$ and $q_{rs}(1) = 0$. For $r = s$, notice that

$$\mathbb{E}[\xi_i^2(x_{j_r}, x'_{j_r}) | \mathcal{F}_{i-1}] = b_n \int K(u) \beta_i(x, x' - ub_n) du,$$

where $\beta_i(x, x')$ is defined as in Lemma 9. Also,

$$\mathbb{E}[\xi_i^2(x_{j_r}, x'_{j_r})] = \mathbb{E}[\mathbb{E}[\xi_i^2(x_{j_r}, x'_{j_r}) | \mathcal{F}_{i-1}]] = b_n \int K(u) \mathbb{E}[\beta_i(x, x' - ub_n)] du$$

. Recall $J_n(x, x')$ in Lemma 9. Therefore, by Lemma 9,

$$\begin{aligned} \|q_{rr}(1) - \frac{1}{nb_n^2 \tau_{rr}} \sum_{i=1}^n \mathbb{E}[\xi_i^2(x_{j_r}, x'_{j_r})]\| &= \frac{1}{nb_n^2 \tau_{rr}} \left\| \int K(u) J_n(x, x' - ub_n) du \right\|_2 \\ &= O[(nb_n)^{-1/2}]. \end{aligned}$$

It is easily seen that $\mathbb{E}[\xi_i^2(x_{j_r}, x'_{j_r})] = b_n^2 \tau_{rr} [1 + O(b_n^2)]$. Thus, by the triangle inequality, $\|q_{rr}(1) - 1\|_2 = O[(nb_n)^{-1/2} + b_n^2]$. In summary, by (85), we have $\|q_{rs} - I_{rs}\|_{3/2} \leq \|q_{rs} - I_{rs}\|_2 = O[(nb_n)^{-1/2} + b_n]$ uniformly over $1 \leq r, s \leq k$. Here I_{rs} is the (r, s) -element of the $k \times k$ identity matrix. It is easily seen that $\sum_{i=1}^n \mathbb{E}|\gamma_i(x_{j_r}, x'_{j_r})|^3 = O[(nb_n^2)^{-1/2}]$ uniformly over $1 \leq r \leq k$. Then $\sum_{i=1}^n \mathbb{E}|\gamma_i(x_{j_r}, x'_{j_r})|^3 + \mathbb{E}(|q_{rs} - I_{rs}|^{3/2}) = O(\Omega_n)$ uniformly, where $\Omega_n = (nb_n^2)^{-1/2} + b_n^{3/2}$.

Under (82), it is easily shown that $[1 + B_{m_n}(z)]^4 \exp[B_{m_n}^2(z)/2] \Omega_n \rightarrow 0$ for every fixed z . For $j = 1, \dots, m_n$, define events

$$E_{m_n} = \left\{ \sup_{(x, x') \in \mathcal{X}_n} |\tilde{M}_n(x, x')| > B_{m_n}(z) \right\}.$$

Let $N_1, N_2, \dots$, be independent standard normals. By Theorem 1 (a multivariate version of it also holds) in Grama and Häusler (2006),

$$\begin{aligned} \mathbb{P} \left[\bigcap_{r=1}^k A_{j_r} \right] &= \mathbb{P} \left[\bigcap_{r=1}^k \{|N_r| > B_{m_n}(z)\} \right] [1 + o(1)] \\ &= \left(\frac{2e^{-z}}{m_n} \right)^k [1 + o(1)]. \end{aligned} \quad (86)$$

Here the first equality agrees with the intuition that $\tilde{M}_n(x_{j_1}, x'_{j_1}), \dots, \tilde{M}_n(x_{j_k}, x'_{j_k})$ are asymptotically independent standard normals since the limiting covariance matrix is the identity matrix; the second equality follows from $\mathbb{P}(N_1 > x) = [1 + o(1)]\varphi(x)/x$ as $x \rightarrow \infty$, where φ is the standard normal density.

Notice that $E_{m_n} = \bigcup_{j=1}^{m_n} A_j$. Therefore, by (86) and the inclusion–exclusion inequality, we have, for any fixed k and large enough n ,

$$\begin{aligned} \mathbb{P}(E_{m_n}) &\geq \sum_{j=1}^{m_n} \mathbb{P}[A_j] - \sum_{j_1 < j_2} \mathbb{P}[A_{j_1} \cap A_{j_2}] + \dots - \sum_{j_1 < \dots < j_{2k}} \mathbb{P} \left[\bigcap_{r=1}^{2k} A_{j_r} \right] \\ &= \sum_{r=1}^{2k} (-1)^{r-1} \binom{m_n}{r} \left(\frac{2e^{-z}}{m_n} \right)^r [1 + o(1)] \\ &= - \sum_{r=1}^{2k} \frac{(-2e^{-z})^r}{r!} [1 + o(1)]. \end{aligned}$$

Thus, $\liminf_{n \rightarrow \infty} \mathbb{P}(E_{m_n}) \geq - \sum_{r=1}^{2k} (-2e^{-z})^r / r!$. Similarly, applying the inclusion–exclusion inequality with an odd number of items in the expansion, we have

$$\limsup_{n \rightarrow \infty} \mathbb{P}(E_{m_n}) \leq - \sum_{r=1}^{2k-1} (-2e^{-z})^r / r!. \text{ The result then follows by letting } k \rightarrow \infty.$$

Proof of Theorems 1 and 2—Theorems 1 and 2 follows from Slutsky's theorem in view of (64), (65), Lemmas 5–7, Proposition 3, and Theorem 3. We omit the details.

Acknowledgments

I am grateful to Editor Professor Ronald Gallant, an Associate Editor, three anonymous referees, and Professor Wei Biao Wu for their constructive comments. This work was partially supported by a 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 NIH.

References

- Aït-Sahalia Y. Testing continuous-time models of the spot interest rate. *Review of Financial Studies*. 1996; 9:385–426.
- Aït-Sahalia Y, Fan J, Jiang J. Nonparametric tests of the Markov hypothesis in continuous-time models. *The Annals of Statistics*. 2010; 38:3129–3163.
- Aït-Sahalia Y, Fan J, Peng H. Nonparametric transition-based tests for jump–diffusions. *Journal of the American Statistical Association*. 2009; 104:1102–1116.
- Aït-Sahalia Y, Mykland PA, Zhang L. How often to sample a continuous-time process in the presence of market microstructure noise. *Review of Financial Studies*. 2005; 18:351–416.
- Andersen TG, Sørensen BE. GMM estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Business & Economic Statistics*. 1996; 14:328–352.
- Azzalini A, Bowman A. On the use of nonparametric regression for checking linear relationships. *Journal of the Royal Statistical Society Series B*. 1993; 55:549–557.
- Ball CA, Torous WN. The stochastic volatility of short-term interest rates: some international evidence. *The Journal of Finance*. 1999; 54:2339–2359.
- Bickel PJ, Ritov Y. Inference in hidden Markov models I: local asymptotic normality in the stationary case. *Bernoulli*. 1996; 2:199–228.
- Bickel PJ, Rosenblatt M. On some global measures of the deviations of density function estimates. *The Annals of Statistics*. 1973; 1:1071–1095.
- Bosq, D. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer; New York: 1998.
- Bradley, RC. Basic properties of strong mixing conditions. In: Eberlein, E.; Taqqu, MS., editors. *Dependence in Probability and Statistics*. Birkhauser; Boston: 1986. p. 165–192.
- Bradley RC. Basic properties of strong mixing conditions. A survey and some open questions. *Probability Surveys*. 2005; 2:107–144.
- Broto C, Ruiz E. Estimation methods for stochastic volatility models: a survey. *Journal of Economic Surveys*. 2004; 18:613–649.
- Chan KC, Karolyi AG, Longstaff FA, Sanders AB. An empirical comparison of alternative models of the short-term interest rate. *The Journal of Finance*. 1992; 47:1209–1227.
- Engle RF. Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation. *Econometrica*. 1982; 50:987–1008.
- Eubank RL, Speckman PL. Confidence bands in nonparametric regression. *Journal of the American Statistical Association*. 1993; 88:1287–1301.
- Fan Y, Li Q. Consistent model specification tests: omitted variables and semiparametric functional forms. *Econometrica*. 1996; 64:865–890.
- Fan, J.; Yao, Q. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer; New York: 2003.
- Fan J, Zhang W. Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*. 2000; 27:715–731.
- Fan J, Zhang C, Zhang J. Generalized likelihood ratio statistics and Wilks phenomenon. *The Annals of Statistics*. 2001; 29:153–193.
- Freedman DA. On tail probabilities for martingales. *The Annals of Probability*. 1975; 3:100–118.
- Gao J, King M. Adaptive testing in continuous-time diffusion models. *Econometric Theory*. 2004; 20:844–882.
- Genon-Catalot V, Jeantheau T, Larédo C. Stochastic volatility models as hidden Markov models and statistical applications. *Bernoulli*. 2000; 6:1051–1079.

- Grama IG, Haeusler E. An asymptotic expansion for probabilities of moderate deviations for multivariate martingales. *Journal of Theoretical Probability*. 2006; 19:1–44.
- Haggan V, Ozaki T. Modelling nonlinear random vibrations using an amplitude dependent autoregressive time series model. *Biometrika*. 1981; 68:189–196.
- Hansen LP, Scheinkman JA. Back to the future: generating moment implications for continuous time Markov processes. *Econometrica*. 1995; 63:767–804.
- Härdle W, Mammen E. Comparing nonparametric versus parametric regression fits. *The Annals of Statistics*. 1993; 21:1926–1947.
- Hong Y, Li H. Nonparametric specification testing for continuous-time models with applications to term structure of interest rates. *Review of Financial Studies*. 2005; 18:37–84.
- Hong Y, White H. Consistent specification testing via nonparametric series regression. *Econometrica*. 1995; 63:1133–1159.
- Hull J, White A. The pricing of options on assets with stochastic volatilities. *The Journal of Finance*. 1987; 42:281–300.
- Jacquier E, Polson NG, Rossi PE. Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business & Economic Statistics*. 1994; 12:371–417.
- Johnston GJ. Probabilities of maximal deviations for nonparametric regression function estimates. *Journal of Multivariate Analysis*. 1982; 12:402–414.
- Kim S, Shephard N, Chib S. Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies*. 1998; 65:361–393.
- Knafl G, Sacks J, Ylvisaker D. Confidence bands for regression functions. *Journal of the American Statistical Association*. 1985; 80:683–691.
- Li, Q.; Racine, J. *Nonparametric Econometrics*. Princeton University Press; Princeton, New Jersey: 2007.
- MacDonald, IL.; Zucchini, W. *Hidden Markov and Other Models for Discrete-Valued Time Series*. Chapman & Hall; London: 1997.
- Ruiz E. Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*. 1994; 63:289–306.
- Taylor SJ. Modeling stochastic volatility: a review and comparative study. *Mathematical Finance*. 1994; 4:183–204.
- Tong, H. *Nonlinear Time Series Analysis: A Dynamical System Approach*. Oxford University Press; Oxford: 1990.
- Vasicek OA. An equilibrium characterization of the term structure. *Journal of Financial Economics*. 1977; 5:177–188.
- Wu WB. Empirical processes of long-memory sequences. *Bernoulli*. 2003; 9:809–831.
- Wu WB. Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the USA*. 2005; 102:14150–14154. [PubMed: 16179388]
- Zhang L, Mykland PA, Ait-Sahalia Y. A tale of two time scales: determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*. 2005; 472:1394–1411.
- Zhao Z. Parametric and nonparametric models and methods in financial econometrics. *Statistics Surveys*. 2008; 2:1–42.
- Zhao Z. Density estimation for nonlinear parametric models with conditional heteroscedasticity. *Journal of Econometrics*. 2010; 155:71–82.
- Zhao Z, Wu WB. Confidence bands in nonparametric time series regression. *The Annals of Statistics*. 2008; 36:1854–1878.

Table 1

Comparison of empirical and asymptotic quantiles of T_n in (43).

$1 - \alpha$	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50
Empirical	1.86	1.98	2.06	2.13	2.21	2.26	2.31	2.36	2.41	2.46
Asymptotic	2.00	2.09	2.16	2.22	2.27	2.32	2.37	2.41	2.46	2.51
$1 - \alpha$	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	0.99
Empirical	2.52	2.57	2.63	2.71	2.78	2.87	2.98	3.11	3.33	3.73
Asymptotic	2.57	2.63	2.68	2.75	2.82	2.91	3.03	3.18	3.43	4.01