PAYMENTS CANADA

Testing for omitted variables in the diffusion matrix of a multivariate diffusion process with applications to term structure of interest rates

payments.ca

Testing for Omitted Variables in the Diffusion Matrix of a Multivariate Diffusion Process with Applications to Term Structure of Interest Rates *

Fuchun Li

Payments Canada

Abstract

A critical issue in specifying a diffusion process is whether a subset of state variables of the diffusion process can be omitted, because a simplified diffusion process derived from omitting a subset of the state variables could be analytically tractable and easy to implement. But such simplification should be quantitatively justified.

In this paper, we develop a consistent nonparametric test for omitted variables in each component of the diffusion matrix in a multivariate diffusion process. We show that each of these test statistics follows an asymptotic standard normal distribution under the null hypothesis that a subset of state variables can be omitted from the specification of the component of the diffusion matrix, while diverging to positive infinity against fixed alternatives and having nontrivial power against a class of local alternatives. Monte Carlo simulations show that our tests perform well. Applying our test statistics to the specification analysis of the volatilities in the term structure of interest rates, we obtain new empirical findings.

Keywords: Diffusion process; Omitted variables; Asymptotic convergence; Consistent tests; Kernel estimation; Term structure of interest rates.
JEL classification: C12, C14. G12, G20.

*We are grateful to seminar participants at 5th Dongbei Econometrics Workshop, Bank of Canada, Dongbei University of Financial and Economics, Shandong University, Federico Bandi, Segun Bewaji, Zengjin Chen, Cyrielle Chiron, Yangjin Fan, and Aman Ullah for their comments and suggestions. The views in this paper are exclusively those of the author and should not be attributed to the Payments Canada.

1 Introduction

Since the work by Merton (1973) and Black and Scholes (1973), continuous-time Markov models have been widely used to study the dynamics of economic variables, such as interest rates, consumption, saving, investment, stock prices, etc. The typical modeling approach in this literature is to assume that the data generating process describing the evolution of state variables follows a diffusion process, which is written as a stochastic differential equation of the state variables.

Testing the specification of a diffusion process has been an active research area in recent years. This includes tests for a parametric specification of a diffusion process versus a nonparametric alternative or a parametric specification of a diffusion process versus a semiparametric alternative. For examples, Aït-Sahalia (1996), Hong and Li (2005), Chen, Gao, and Tang (2008), and Aït-Sahalia, Fan and Peng (2009) develop consistent tests for the parametric specification of the transition density function of a diffusion process; Chen and Hong (2010) develop a test for the parametric specification of the characteristic function of a diffusion process; Kristensen (2011) develops a test for the parametric specification of a diffusion process versus a semiparametric diffusion process; Li (2019) develops tests for the parametric specification of the diffusion matrix in a diffusion process.

However, some important issues still have remained unexplored in the literature of testing the specification of a diffusion process. For example, a diffusion process is often assumed to follow a special case that simplifies the model specification by the dimension reduction of state variables for deriving closed-form solutions of various derivative securities, such as the case of multifactor term structure models (Dai and Singleton, 2000), or the case of the analysis of consumption, asset pricing, and investment (Merton 1973, and Abel, 1985).[1] A potentially serious problem associ-

[1]For example, in specifying a term structure model of interest rates, we are confronted with trade-offs between the richness of econometric representations of the state variables and the computational burdens of estimation and

ated with such a simplified diffusion process is model misspecification caused by omitted state variables, which could lead to misleading results in inference and hypothesis testing. Thus, it is important to develop a reliable specification test for the significance of a subset of state variables in a diffusion process, namely whether certain state variables can be omitted from the specification of a diffusion process. Although theoretical results have been provided for testing omitted variables in a traditional setup of the regression model (e.g., Aït-Sahalia, et al. (2001), Fan and Li (1996), and Lavergne and Vuong (2000)), to the best of our knowledge, no theoretical result has been developed to testing for omitted variables in a diffusion process.

In this paper, we focus on testing for omitted variables in the diffusion matrix of a multivariate diffusion process in the nonparametric context. The diffusion matrix of a multivariate diffusion process, as the second moment and the measurement of the volatility dynamics of the state variables, is a crucial factor in modeling the movements of state variables as well as the comovements among state variables. To test whether certain state variables can be omitted from the diffusion matrix of a multivariate diffusion process, we propose a consistent test for omitted variables in each component in the diffusion matrix, with nonparametrically specifying the functional form of each component in the diffusion matrix. Using theories of degenerate U-statistics of absolutely regular processes, each of all these test statistics is shown to follow an asymptotic standard normal distribution under the null hypothesis that a subset of state variables can be omitted from the specification of the component of the diffusion matrix, while diverging to positive infinity under fixed alternatives and having nontrivial power against a class of local alternatives. Monte Carlo simulations show that our tests have reasonable size and good power against a variety of alternatives.

---

pricing. A simplified term structure model yields essentially closed-form expressions for zero-coupon-bond prices, which greatly facilitates pricing and econometric implementation. As such, the need for model simplification is paramount because one may be very hard to obtain the results of the economic analysis without simplification. But such simplification should be quantitatively justified, i.e., hypothesis tests are needed.

To illustrate the empirical applications of our testing procedure, our tests are applied to the situation of dimension reduction in the specification of the diffusion matrix in a multivariate term structure model of interest rates. We particularly focus on a two-factor term structure model of interest rates. Using the daily 3-month Canadian Treasury Bills and 10-year Canadian bond yield as the proxies of the short term interest rate and long term interest rate, respectively, we examine whether the volatility of each interest rate process depends on the other interest rate as well as on its own value. Our empirical finding indicates that both the short term interest rate and the long term interest rate cannot be omitted for the specification of the volatilities of either short term interest rate process or long term interest rate process, suggesting that the two-factor CIR-type model used by Hull and White (1990) and the two-factor Brennan-Schwartz-type model used by Hsin (1995) are rejected. This empirical result is consistent with the expectations hypothesis that the instantaneous variance of the changes in the long term interest rate contains information about the future values of the short term interest rate, while the long term interest rate directly affects the instantaneous variance of the changes in the short term interest rates.

The remainder of the paper is organized as follows. In Section 2, we introduce the model, hypotheses of interest, and our test statistic for each component in the diffusion matrix. In Section 3, we derives the asymptotic null distribution of each of these test statistics, and discuss its asymptotic power property. In Section 4, we use Monte Carlo simulation approach to evaluate the finite sample performances of our test statistics. In Section 5, our test statistics are applied to examine dimension reduction in the specification of diffusion matrix in a two-factor term structure model of interest rates. Section 6 concludes. All proofs of the main results are given in Appendix A and Appendix B. Appendix C contains some technical lemmas that are used in the proofs of Appendix A and Appendix B.

2 The Model, Hypotheses, and Test Statistics

The model specified as the underlying process of a $d \times 1$ vector of state variables $\{x_t, t \geq t_0\}$ is a time-homogeneous Itô diffusion process represented by the following stochastic differential equation,

$$dx_t = \mu(x_t)dt + \sigma(x_t)dB_t, \tag{1}$$

with a given initial condition x_{t_0}, where $\mu(x_t) \equiv (\mu_1(x_t), ..., \mu_d(x_t))'$ is a $d \times 1$ drift vector, $\sigma(x_t) \equiv \{\sigma_{ij}(x_t)\}_{1 \leq i,j \leq d}$ is a $d \times d$ matrix, and $B_t \equiv (B_t^1, ..., B_t^d)'$ is a d-dimensional standard Brownian motion. Assume that x_{t_0} is independent of B_t and x_t takes values in $I \subset R^d$.

For $1 \leq i \leq d$, each coordinate x_t^i of the diffusion process x_t can be written as,

$$dx_t^i = \mu_i(x_t)dt + \sum_{j=1}^{d} \sigma_{ij}(x_s)dB_s^j. \tag{2}$$

Now, we assume that the process x_t is observed at equispaced times $\{t = t_1, t_2, ..., t_n\}$ in the time interval $[t_0, T]$, where T is a strict positive number ($T > t_0$). The observations can be expressed as $\{x_t = x_{t_0+\triangle_n}, x_{t_0+2\triangle_n}, ..., x_{t_0+n\triangle_n}\}$ at $\{t_1 = t_0 + \triangle_n, t_2 = t_0 + 2\triangle_n, ..., t_n = t_0 + n\triangle_n\}$, where $\triangle_n = (T - t_0)/n$ is the sampling interval. We use the notation $x_{n,t}$ to express the observation on the process x_t at $t_0 + t\triangle_n$. Thus, the data are given by a triangular array of random variables $\{x_{n,t}, 1 \leq t \leq n\}$.

Let $\mathcal{F}_{s,t}^n$ express the σ-algebra generated by $\{x_{n,t'} | t' = s, s+1, ..., t; 1 \leq s \leq t \leq n\}$. Following this literature (e.g., Harel and Puri, 1990), we define the β-mixing coefficient of the triangular array of $\{x_{n,t}, 1 \leq t \leq n\}$ by,

$$\beta_{n,\tau_n} \equiv sup_{s \leq n-\tau_n} E[sup_{A \in \mathcal{F}_{s+\tau_n,n}^n} \{P(A|\mathcal{F}_{1,s}^n) - P(A)\}], \tag{3}$$

which measures the degree of time-series dependence of $\{x_{n,t}, 1 \leq t \leq n\}$. Let $\{x_{n,t}, 1 \leq t \leq n\}$ be a strictly stationary triangular array process. $\{x_{n,t}, 1 \leq t \leq n\}$ is called absolutely regular, if $\beta_{n,\tau_n} \to 0$, as $\tau_n \to \infty$, where $\tau_n \varphi(\triangle_n) \to \infty$ as $n \to \infty$, and $\varphi(\cdot)$ is a continuous, positive, decreasing function on $[0, \delta_0)$ with δ_0 being a positive constant and $lim_{n\to\infty}\varphi(\triangle_n) \to 0$ as $n \to \infty$.

We assume that the sampling interval shrinks to zero as the sample size n tends to infinity (infill assumption), and the time span $T \to \infty$ as $n \to \infty$ (the long-span assumption). The infill assumption is crucial for the nonparametric estimation of the diffusion matrix, because the identification of the diffusion matrix depends on the local dynamics of the diffusion process. Under the long-span assumption, we can use the asymptotic independence implied by the absolutely regular property of $\{x_{n,t}, 1 \leq t \leq n\}$, which ensures that under certainty conditions, the β-mixing coefficient goes to zeros at a fast enough pace for deriving the asymptotic distributions under the null hypotheses.

Given the diffusion process x_t in (1), the $d \times d$ symmetric and non-negative matrix $a(x) \equiv \sigma(x)\sigma(x)'$, with general element $a_{ij}(x) = \sum_{l=1}^{d} \sigma_{il}(x)\sigma_{jl}(x)$, is its diffusion matrix for any $x \equiv (x^1, ..., x^d) \in I, 1 \leq i, j \leq d$. The dynamic properties of the diffusion process x_t are characterized by its transition density function, which depends on $\mu(\cdot)$ and $a(\cdot)$. In fact, if there exist a continuum of solutions in $\sigma(\cdot)$ to equation $a(x) \equiv \sigma(x)\sigma(x)'$, then the transition probability function is identical for each of these $\sigma(x)$ (Stroock and Varadhan, 1979). This indicates that the dynamic properties of the diffusion process x_t are determined entirely by the drift vector $\mu(\cdot)$ and diffusion matrix $a(\cdot)$, i.e., the pairs $(\mu(\cdot), a(\cdot))$.

Given that x_t is a diffusion process, $a(x_t)$ can be expressed as,

$$a(x_t) = lim_{\triangle_n \downarrow 0} \frac{1}{\triangle_n} E[(x_{t+\triangle_n} - x_t)(x_{t+\triangle_n} - x_t)' | x_t], \tag{4}$$

or equivalently, for $1 \leq i, j \leq d$, we have,

$$a_{ij}(x_t) = lim_{\triangle_n \downarrow 0} \frac{1}{\triangle_n} E[(x^i_{t+\triangle_n} - x^i_t)(x^j_{t+\triangle_n} - x^j_t)' | x_t]. \tag{5}$$

Let $x_t = (w_t', v_t')'$, where w_t is a $q \times 1$ vector $(1 \leq q \leq d-1)$, and v_t is a $(d-q) \times 1$ vector. The objective of this paper is to test whether the subset of state variables v_t is insignificant for the explanation of the instantaneous covariance $a_{ij}(x_t)$ between x_t^i and x_t^j conditional on x_t for $1 \leq i, j \leq d$. To achieve this objective, we define a diffusion matrix $b(w_t) = \{b_{ij}(w_t), 1 \leq i, j \leq d\}$ as follows,

$$b(w_t) = lim_{\triangle_n \downarrow 0} \frac{1}{\triangle_n} E[(x_{t+\triangle_n} - x_t)(x_{t+\triangle_n} - x_t)'|w_t], \tag{6}$$

or equivalently, we define,

$$b_{ij}(w_t) = lim_{\triangle_n \downarrow 0} \frac{1}{\triangle_n} E[(x_{t+\triangle_n}^i - x_t^i)(x_{t+\triangle_n}^j - x_t^j)|w_t], \;\; 1 \leq i, j \leq d. \tag{7}$$

The null hypothesis that we are interested can be characterized by the following form,

$$H_0^{ij} : Pr[a_{ij}(x_t) = b_{ij}(w_t)] = 1. \tag{8}$$

The alternative hypothesis is,

$$H_a^{ij} : Pr[a_{ij}(x_t) = b_{ij}(w_t)] < 1. \tag{9}$$

Let $u_{n,t} \equiv \frac{(x_{n,t+1}^i - x_{n,t}^i)(x_{n,t+1}^j - x_{n,t}^j)}{\triangle_n} - b_{ij}(w_{n,t})$, we define the distance measure between $a_{ij}(x_t)$ and $b_{ij}(w_t)$ as $D_{nij} \equiv E\{u_{n,t} E[u_{n,t}|x_{n,t}]\} = E[(E[u_{n,t}|x_{n,t}])^2]$. Under the Assumptions 1-4 in Section 3, using Itô's lemma to $\frac{(x_{n,s}^i - x_{n,t}^i)(x_{n,s}^j - x_{n,t}^j)}{\triangle_n}$, where $t \leq s < t+1$, we have,

$$\begin{aligned} lim_{n\to\infty} D_{nij} &= lim_{n\to\infty} E\Big\{\Big[E\Big[\frac{(x_{n,t+1}^i - x_{n,t}^i)(x_{n,t+1}^j - x_{n,t}^j)}{\triangle_n}\Big| x_{n,t}\Big] - b_{ij}(w_{n,t})\Big]\Big\}^2 \\ &= lim_{n\to\infty} E[a_{ij}(x_{n,t}) - b_{ij}(w_{n,t})]^2 \\ &= D_{ij} \geq 0, \end{aligned} \tag{10}$$

and $D_{ij} = 0$ in (10) holds if and only if H_0^{ij} is true. Hence D_{nij} can be used for the construction of a consistent test for H_0^{ij} against H_a^{ij}. To construct a feasible test statistic, we need to estimate $u_{n,t}$ and $E[u_{n,t}|x_{n,t}]$ by using the nonparametric kernel method. In order to avoid the random denominator problem associated with the kernel estimation method, instead of D_{nij}, we choose its density weighted version: $E_{nij} \equiv E\{u_{n,t} f_w(w_{n,t}) E[u_{n,t} f_w(w_{n,t})|x_{n,t}] f(x_{n,t})\} = E\{(E[u_{n,t} f_w(w_{n,t})|x_{n,t}])^2 f(x_{n,t})\}$, where $f_w(w_t)$ and $f(x_t)$ are the density function of w_t and the density function of x_t, respectively.

The sample analogue of E_{nij} is,

$$I_{nij} = n^{-1} \sum_t u_{n,t} f_w(w_{n,t}) E[f_w(w_{n,t}) u_{n,t} | x_{n,t}] f(x_{n,t}). \tag{11}$$

We estimate $u_{n,t}$ by $\hat{u}_{n,t} = \frac{[x^i_{n,t+1} - x^i_{n,t}][x^j_{n,t+1} - x^j_{n,t}]}{\triangle_n} - \hat{b}_{ij}(w_{n,t})$, where $\hat{b}_{ij}(w_{n,t})$ is a kernel estimator of $b_{ij}(w_{n,t})$, which is proposed by Bandi and Moloche (2018) and is defined as,

$$\hat{b}_{ij}(w_{n,t}) = \frac{\frac{1}{(n-1)a_n^q \triangle_n} \sum_{s \neq t} L(\frac{w_{n,s} - w_{n,t}}{a_n})[x^i_{n,s+1} - x^i_{n,s}][x^j_{n,s+1} - x^j_{n,s}]}{\hat{f}_{w_t}}, \tag{12}$$

where $L(\cdot)$ is a kernel function and a_n is a smoothing parameter, and $\hat{f}_{w_t}$ is the kernel estimator of $f_{w_t} \equiv f_w(w_{n,t})$ given by,

$$\hat{f}_{w_t} = \frac{1}{(n-1)a_n^q} \sum_{s \neq t} L(\frac{w_{n,s} - w_{n,t}}{a_n}). \tag{13}$$

We estimate $E[u_{n,t} f_w(w_{n,t})|x_{n,t}] f(x_{n,t})$ by,

$$[(n-1)h_n^d]^{-1} \sum_{s \neq t} \hat{u}_{n,s} \hat{f}_w(w_{n,s}) K(\frac{x_{n,s} - x_{n,t}}{h_n}), \tag{14}$$

where $K(\frac{x_{n,s} - x_{n,t}}{h_n}) = K(\frac{w_{n,s} - w_{n,t}}{h_n}, \frac{v_{n,s} - v_{n,t}}{h_n})$, $K(\cdot)$ is a kernel function, and h_n is a smoothing parameter. Since we allow that T goes to infinity as $n \to \infty$, the smoothing parameter can be denoted by $h_n(a_n)$ instead of $h_{n,T}(a_{n,T})$.

Inserting (12), (13), and (14) into (11) yields the estimator of I_{nij} by $\hat{I}_{nij}$,

$$\hat{I}_{nij} = \frac{1}{n(n-1)h_n^d} \sum_t \sum_{s \neq t} [\hat{u}_{n,t} \hat{f}_w(w_{n,t})][\hat{u}_{n,s} \hat{f}_w(w_{n,s})] K_{ts}, \tag{15}$$

where $K_{ts} = K(\frac{x_{n,t} - x_{n,s}}{h_n})$. Based on $\hat{I}_{nij}$, our test statistic is defined as,

$$Q_{nij} \equiv n h_n^{d/2} \hat{I}_{nij} / v_{nij}, \tag{16}$$

where $v^2_{nij} = \frac{2}{n(n-1)h_n^d} \sum_t \sum_{s \neq t} [\hat{u}^2_{n,t} \hat{f}^2_w(w_{n,t})][\hat{u}^2_{n,s} \hat{f}^2_w(w_{n,s})] K^2_{ts}$.

3 Asymmetric Properties of the Test Statistics

We make the following assumptions to derive the asymptotic properties of the test statistics, Q_{nij}, $1 \leq i, j \leq d$.

Assumption 1. *Let $D = \prod_{i=1}^{d}(l_i, r_i)$ be a product of intervals, (l_i, r_i), for $i = 1, \ldots, d$, where $-\infty \leq l_i < r_i \leq \infty$. On D, the functions $\mu(x), \sigma(x)$, and $b(w)$ satisfy the following conditions:*

(i) *$\mu(x)$, $\sigma(x)$, and $b(w_t)$ are continuously differentiable.*

(ii) *Let $|\cdot|$ denote the Max norm for a matrix. There exists a positive constant C_D such that for every $x \in D$,*

$$|\mu(x)|^2 + |\sigma(x)|^2 \leq C_D(1 + |x|^2). \tag{17}$$

$$|b(w)| \leq C_D(1 + |w|^2). \tag{18}$$

(iii) *There exists a nonnegative function $\varsigma(\cdot, \cdot)$ such that $E[\varsigma^4(x_{t_1}, x_{t_2})] \leq C_E$ for any $t_1, t_2 \in [t_0, T]$, where C_E s a positive constant and,*

$$|\mu(x) - \mu(y)| \leq \varsigma(x, y)|x - y|. \tag{19}$$

Assumption 2. *Let $E|x_{t_0}|^8 < \infty$. $\{x_{n,t}, 1 \leq t \leq n, n \geq 1\}$ is strictly stationary and absolutely regular with mixing coefficient $\beta_{n,\tau_n} = O(\lambda^{\tau_n \varphi(\triangle_n)})$ for some $0 < \lambda < 1$, where $\varphi(\triangle_n) = O(\frac{1}{log^{\lambda_0}(\triangle_n^{-1})})$, and λ_0 is a positive constant.*

Assumption 3. *Assume that both $K(\cdot)$ and $L(\cdot)$ are product kernel functions with order $r (r \geq 2)$. Let $k(\cdot)$ and $l(\cdot)$ be their corresponding univariate kernel functions, which satisfy the Lipschitz condition and are even, bounded functions with $k(s) = O(1 + |s|^{1+r+\kappa})^{-1}$ and $l(s) = O(1 + |s|^{1+r+\kappa})^{-1}$ for some $\kappa > 0$.*

Assumption 4. *Let $f(\cdot), f_w(\cdot) \in \mathcal{G}_r^{\infty}$ and $a_{ij}(\cdot), b_{ij}(\cdot) \in \mathcal{G}_r^5$, where $\mathcal{G}_r^{\alpha}$ is defined as the class of functions, $\{g : D \to R\}$, satisfying that g is r-times partially differentiable, and for some $\rho > 0$, $sup_{y \in \phi_\rho^x} |g(y) - g(x) - Q_{ij}(y,x)| \leq \bar{L}_0(x)|y-x|^r$, for all x, where $\phi_\rho^x = \{y : |y-x| < \rho\}$; $Q_{ij}(y,x)$ is a $(r-1)$ th degree homogeneous polynominal in $(y-x)$ with the coefficients being the partial derivatives of g at x of orders 1 through $(r-1)$, and $\bar{L}_0(x)$ is a continuous function and has finite α th moment. Also, $f(\cdot)$ is bounded.*

We briefly comment on the above assumptions. Assumption 1 (i) ensures that the coefficients of the stochastic differential equation (1) are locally Lipschitz under which a solution to (1) will be unique if it exists (Theorem 5.2.5 in Karatzas and Shreve, 1991). Assumption 1 (ii) ensures the existence of a solution to the stochastic differential equation (1) by preventing explosion of the process in finite expected time. Assumption 1 (iii) is made to derive some important moment inequalities for asymptotic analysis of our test statistics. In Assumption 2, we require that the triangular array of $\{x_{n,t}, 1 \leq t \leq n, n \geq 1\}$ be absolutely regular with geometric decay rate, under which the central limit theorem for degenerate U-statistics of absolutely regular processes can be used to derive the asymptotic distributions of our test statistics. The absolute regularity with geometric decay rate of a multivariate diffusion process can be shown by using Theorem 6.1 in Meyn and Tweedie(1993). Following this approach provided by Meyn and Tweedie (1993), it can be shown that the popular multivariate affine term structure models in Dai and Singleton (2000) satisfy the property of absolutely regular with a geometric decay rate, under certain restrictions of the coefficients. Assumption 3 characterizes the conditions on the rth order kernel functions used in the nonparametric kernel estimations of the diffusion matrixes of $a(x)$ and $b(w)$, and marginal density functions of x_t and w_t. Assumption 4 imposes smoothness and moments conditions on functions $a_{ij}(\cdot)$, $b_{ij}(\cdot)$, $f(\cdot)$, and $f_w(\cdot)$. For instance, $a_{ij}(\cdot) \in \mathcal{G}_r^5$ means that $a_{ij}(y)$ is differentiable up to order r and has Taylor expansion given by $(a_{ij}(x)+Q_{ij}(y,x))$ with the remainder satisfying a local Lipschitz condition. In addition, $a_{ij}(y)$ satisfies the finite $5th$ comment condition and all the partial derivatives (up to order r) of $a_{ij}(y)$ satisfy this moment condition as well.

With the above assumptions, the asymptotic null distribution and consistency of Q_{nij} are provided in the following theorem.

Theorem 1. *Suppose that* $h_n \to 0, nh_n^d \to \infty, na_n^q \to \infty, na_n^{2r}h_n^{d/2} \to 0, Th_n^{d/2} \to 0$ *and* $h_n^d/a_n^{2q} \to 0$*, as* $n \to \infty$*. Then under Assumptions 1-4, we have,*

(i) *under* H_0^{ij}*,* $Q_{nij} = nh_n^{d/2}\hat{I}_{nij}/v_{nij} \to N(0,1)$ *in distribution as* $n \to \infty$*, and* v_{nij}^2 *is a consistent estimator of* v_{ij}^2*, where* $v_{ij}^2 = 2E[(a_{ii}(x_t)a_{jj}(x_t)+a_{ij}^2(x_t))^2 f_w^4(w_t)f(x_t)]\int K^2(u)du$*;*

(ii) Q_{nij} *is consistent against* H_a^{ij}*.*

The proof of Theorem 1 is given in the Appendix A.

To test the null hypothesis H_0^{ij} against H_a^{ij} at the level α, we need to compare Q_{nij} to the critical value z_α from the $N(0,1)$ distribution, i.e., $z_{0.01} = 2.33, z_{0.05} = 1.64$, and $z_{0.01} = 1.28$ because the test Q_{nij} is one-sided. We reject the null hypothesis when $Q_{nij} > z_\alpha$.

In the rest of this section, we will analyze the local power properties of the test statistics for a sequence of local alternatives. We consider the sequence of local alternative hypothesis (Pitman alternative hypothesis) represented by,

$$LH : a_{ij}(x) = b_{ij}(w) + \gamma_n \triangle(x), \tag{20}$$

where $\int \triangle(x)dx = 0$, $\int |\triangle(x)|dx < \infty$, $\int \triangle^2(x)dx < \infty$, and $\gamma_n \to o$ as $n \to \infty$.

Theorem 2. *Under the assumptions of Theorem 1, if LH holds and* $\gamma_n = n^{-1/2}h_n^{-d/4}$*, then we have,* $Pr[Q_{nij} > z_\alpha] \to 1 - \Phi(z_\alpha - \frac{1}{v_{ij}}\int \triangle^2(x)f_w^2(w)f(x)dx)$*, where* $\Phi(\cdot)$ *is the distribution function of the standard normal random variable z.*

The proof of Theorem 2 is given in the Appendix A.

Theorem 2 indicates that Q_{nij} can detect local alternatives in LH that differ from the null by $O(n^{-1/2}h_n^{-d/4})$, and the slower h_n converges to zero, the larger is the asymptotic power of Q_{nij} in the sense that it can detect local alternatives in LH that are closer to the null.

4 Monte Carlo Results

We now conduct Monte Carlo simulations to assess the finite sample performance of $Q_{nij}, 1 \leq i, j \leq d$. We focus on the affine specification of the diffusion matrix in a three-dimensional diffusion process (see, e.g., Hong and Li (2005), and Dai and Singleton (2000)) given the importance of an affine model in the existing asset pricing literature. Since the diffusion matrix is a symmetric matrix, we need only evaluate the finite sample performance for our tests $Q_{n11}, Q_{n12}, Q_{n13}, Q_{n22}, Q_{n23}$, and Q_{n33}.

4.1 Size performance

To examine the size performance of Q_{nij}, $1 \leq i, j \leq d$, we simulate data from the following five data-generating processes (DGPs):

DGP1. Three-factor model with time-varying variances is specified,

$$d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt + \begin{pmatrix} \sqrt{x_{1t}} & 0 & 0 \\ 0 & \sqrt{x_{2t}} & 0 \\ 0 & 0 & \sqrt{x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix} \quad (21)$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\leq i,j\leq 3} \equiv \begin{pmatrix} x_{1t} & 0 & 0 \\ 0 & x_{2t} & 0 \\ 0 & 0 & x_{3t} \end{pmatrix}$, where $x_t = (x_{1t}, x_{2t}, x_{3t})$ hereafter. The parameters are taken from Aït-Sahalia and Kimmel (2010). The null hypotheses are, $Pr[a_{ii}(x_t) = b_{ii}(x_{it})] = 1$, for $i = 1, 2, 3$, where both $a_{ii}(\cdot)$ and $b_{ii}(\cdot)$ are unknown functions satisfied with the Assumption 1 and Assumption 4. The alternative hypotheses are: $Pr[a_{ii}(x_t) = b_{ii}(x_{it})] < 1$, for $i = 1, 2, 3$. DGP1 is used to investigate the size performance of Q_{n11}, Q_{n22}, and Q_{n33}.

DGP2. Three-factor model with nonlinear drift is specified,

$$d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} \alpha_{-1}x_{1t}^{-1}+\alpha_0+\alpha_1 x_{1t}+\alpha_2 x_{1t}^2 \\ \alpha_{-1}x_{2t}^{-1}+\alpha_0+\alpha_1 x_{2t}+\alpha_2 x_{2t}^2 \\ \alpha_{-1}x_{3t}^{-1}+\alpha_0+\alpha_1 x_{3t}+\alpha_2 x_{3t}^2 \end{pmatrix} + \begin{pmatrix} \sqrt{x_{1t}} & 0 & 0 \\ 0 & \sqrt{x_{2t}} & 0 \\ 0 & 0 & \sqrt{x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (22)$$

and its diffusion matrix is the same as in DGP1. The parameter values are set as $(\alpha_{-1},\alpha_0,\alpha_1,\alpha_2) = (0.00107,-0.0517,0.877,-4.604)$, which come from Hong and Li (2005).

The null hypotheses are, $Pr[a_{ii}(x_t)=b_{ii}(x)]=1$, for $i=1,2,3$, where both $a_{ii}(\cdot)$ and $b_{ii}(\cdot)$ are unknown functions satisfied with the Assumption 1 and Assumption 4, and the alternative hypotheses are, $Pr[a_{ii}(x_t)=b_{ii}(x)]<1$, for $i=1,2,3$. DGP2 is used to examine the size performance of Q_{nii} $(i=1,2,3)$ when the drift term is nonlinearly specified.

DGP3. The three-factor diffusion process with time-varying conditional variances and conditional covariance $(a_{12}(x_t)\neq 0)$ is specified as,

$$d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt + \begin{pmatrix} \sqrt{x_{1t}} & \sqrt{x_{2t}} & 0 \\ 0 & \sqrt{x_{2t}} & 0 \\ 0 & 0 & \sqrt{x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (23)$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\le i,j,\le 3} = \begin{pmatrix} x_{1t}+x_{2t} & x_{2t} & 0 \\ x_{2t} & x_{2t} & 0 \\ 0 & 0 & x_{3t} \end{pmatrix}$.

The null hypotheses are, $Pr[a_{11}(x_t)=b_{11}(x_{1t},x_{2t})]=1$, $Pr[a_{22}(x_t)=b_{22}(x_{2t})]=1$, $Pr[a_{33}(x_t)=b_{33}(x_{3t})]=1$, and $Pr[a_{12}(x_t)=b_{12}(x_{2t})]=1$, where $a_{11}(\cdot),a_{22}(\cdot),a_{33}(\cdot),b_{11}(\cdot,\cdot),b_{22}(\cdot),b_{33}(\cdot)$, and $b_{12}(\cdot)$ are unknown functions satisfied with the Assumption 1 and Assumption 4. DGP3 is used to examine the size performance of Q_{n12}, as well as the size performance of $Q_{nii}(i=1,2,3)$.

DGP4. The three-factor diffusion process with time-varying conditional variances and conditional covariance $(a_{13}(x_t)\neq 0)$ is specified as,

$$d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt + \begin{pmatrix} \sqrt{x_{1t}} & 0 & \sqrt{x_{1t}} \\ 0 & \sqrt{x_{2t}} & 0 \\ 0 & 0 & \sqrt{x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (24)$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\le i,j,\le 3} = \begin{pmatrix} 2x_{1t} & 0 & \sqrt{x_{1t}x_{3t}} \\ 0 & x_{2t} & 0 \\ \sqrt{x_{1t}x_{3t}} & 0 & x_{3t} \end{pmatrix}$. The null hypotheses are, $Pr[a_{ii}(x_t)=b_{ii}(x_{it})]=1$, for $i=1,2,3$, and $Pr[a_{13}(x_t)=b_{13}(x_{1t},x_{3t})]=1$, where $a_{ii}(\cdot),b_{ii}(\cdot),a_{13}(\cdot)$, and $b_{13}(\cdot,\cdot)$ are unknown functions satisfied with the Assumption 1 and Assumption 4, $i=1,2,3$. DGP4 is used to examine the size performance of Q_{n13}, as well as the size performance of $Q_{nii}(i=1,2,3)$.

DGP5. The three-factor diffusion process with time-varying conditional variance and conditional covariance ($a_{23}(x_t)\neq 0$) is specified as,

$$d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} = \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt + \begin{pmatrix} \sqrt{x_{1t}} & 0 & 0 \\ 0 & \sqrt{x_{2t}} & \sqrt{x_{2t}} \\ 0 & 0 & \sqrt{x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (25)$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\leq i,j,\leq 3} = \begin{pmatrix} x_{1t} & 0 & 0 \\ 0 & 2x_{2t} & \sqrt{x_{2t}x_{3t}} \\ & \sqrt{x_{2t}x_{3t}} & x_{3t} \end{pmatrix}$.

The null hypotheses are, $Pr[a_{ii}(x_t)=b_{ii}(x_{it})]=1$, for $i=1,2,3$, and $Pr[a_{23}(x_t)=b_{23}(x_{2t},x_{3t})]=1$, where $a_{ii}(\cdot),b_{ii}(\cdot)$, $i=1,2,3$, and $b_{23}(\cdot,\cdot)$ are unknown functions satisfied with the Assumption 1 and Assumption 4. DGP5 is used to examine the size performance of Q_{n23}, as well as the size performance of $Q_{nii}(i=1,2,3)$.

In this simulation studies, we simulate 1500 realizations of a random sample $\{x_t\}_{t=1}^n$ at daily frequency for $n=250,500,1000$, and 2500, respectively. We discard the first 500 observations to eliminate any start-up effects. These sample sizes correspond to $1,2,4$, and 10 years of daily data, respectively. T is set to $1,5$, and 10 to consider the impact of the time span on the performance of these tests. We use standard normal kernel functions for both $K(\cdot)$ and $L(\cdot)$ with smoothing parameters chosen by $h_{x_i}=x_{isd}n^{-1/4}$ and $a_{w_j}=w_{jsd}n^{-1/6}$, where x_{isd} and w_{jsd} are the standard deviations of $\{x_{it}\}_{t=1}^n$ ($i=1,2,3$) and $\{w_{jt}\}_{t=1}^n$ ($j=1$ or $j=1,2$), respectively. The above choices of h and a satisfy the conditions of Theorem 1. For DGP1, we simulate the data from the model transition density, which has a closed-form. For DGP2, DGP3, DGP4, and DGP5, their transition densities have no closed forms, we use Milstein's scheme (Kloeden and Platen, 1992) to simulate the data sets of the random sample. For $T=1$, Table 1 reports the rejection rates of Q_{n11},Q_{n22}, and

Q_{n33} under DGP1 and DGP2, and Table2 reports the rejection rates of $Q_{n11}, Q_{n22}, Q_{n33}, Q_{n12}, Q_{n13}$, and Q_{n23} under DGP3, DGP4, and DGP5, using the asymptotic critical values at 5% and 10%.[2] Table1 shows that the estimated sizes converge to their nominal sizes as n increases. Note that the impact of the nonlinear drift on the estimated sizes is minimal, suggesting that Q_{nii} $(i = 1,2,3)$ can achieve robustness to the drift specification. This result can be explained by the fact that the test statistics are independent of the specification of the drift matrix. Overall, the Monte Carol simulation results suggest that our test statistics Q_{nii} $(i = 1,2,3)$ has reasonable size performance for sample sizes as small as n = 250. Table2 indicates that under null hypothesis of either a conditional variance or a conditional covariance in the diffusion matrix, the corresponding test statistic Q_{nij} $(1 \leq i \leq j \leq d)$ shows that the estimated sizes converge to the nominal sizes as n increases.

4.2 Power performance

To investigate the power performance of Q_{nij} $(1 \leq i \leq j \leq d)$, we simulate the data from four other diffusion processes below.

DGP6. The three-factor diffusion process with time-varying conditional variance is specified,

$$\begin{aligned} d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} &= \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt \\ &+ \begin{pmatrix} \sqrt{x_{1t}+x_{2t}+x_{3t}} & 0 & 0 \\ 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} & 0 \\ 0 & 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (26) \end{aligned}$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\leq i,j\leq 3} \equiv \begin{pmatrix} x_{1t}+x_{2t}+x_{3t} & 0 & 0 \\ 0 & x_{1t}+x_{2t}+x_{3t} & 0 \\ 0 & 0 & x_{1t}+x_{2t}+x_{3t} \end{pmatrix}$.

We use DGP6 to examine the power performance of Q_{nii} $(i = 1,2,3)$. DGP1 is the null model.

[2] Simulation results for $T = 5,10$ are not presented but are available from the author. They are qualitatively similar to those for $T = 1$.

DGP7. The three-factor diffusion process with both time-varying conditional variance and conditional covariance is specified as,

$$\begin{aligned} d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} &= \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt \\ &+ \begin{pmatrix} \sqrt{x_{1t}+x_{2t}+x_{3t}} & \sqrt{x_{1t}+x_{2t}+x_{3t}} & 0 \\ 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} & 0 \\ 0 & 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (27) \end{aligned}$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\le i,j\le 3} \equiv \begin{pmatrix} 2(x_{1t}+x_{2t}+x_{3t}) & x_{1t}+x_{2t}+x_{3t} & 0 \\ x_{1t}+x_{2t}+x_{3t} & x_{1t}+x_{2t}+x_{3t} & 0 \\ 0 & 0 & x_{1t}+x_{2t}+x_{3t} \end{pmatrix}$.

We use DGP7 to examine the power performance of Q_{nii} $(i=1,2,3)$ and Q_{n12}. DGP3 is the null model.

DGP8. The three-factor diffusion process with time-varying conditional variance and conditional covariance is specified as,

$$\begin{aligned} d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} &= \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt \\ &+ \begin{pmatrix} \sqrt{x_{1t}+x_{2t}+x_{3t}} & 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} \\ 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} & 0 \\ 0 & 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (28) \end{aligned}$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\le i,j\le 3} \equiv \begin{pmatrix} 2(x_{1t}+x_{2t}+x_{3t}) & 0 & x_{1t}+x_{2t}+x_{3t} \\ 0 & x_{1t}+x_{2t}+x_{3t} & 0 \\ x_{1t}+x_{2t}+x_{3t} & 0 & x_{1t}+x_{2t}+x_{3t} \end{pmatrix}$.

We use DGP8 to examine the power performance of Q_{nii} $(i=1,2,3)$ and Q_{n13}. DGP4 is the null model.

DGP9. The three-factor diffusion process with time-varying conditional variance and conditional covariance is specified as,

$$\begin{aligned} d\begin{pmatrix} x_{1t} \\ x_{2t} \\ x_{3t} \end{pmatrix} &= \begin{pmatrix} 1-0.5x_{1t} & 0 & 0 \\ 0 & 2-2x_{2t} & 0 \\ 0 & 0 & 1-x_{3t} \end{pmatrix} dt \\ &+ \begin{pmatrix} \sqrt{x_{1t}+x_{2t}+x_{3t}} & 0 & 0 \\ 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} & \sqrt{x_{1t}+x_{2t}+x_{3t}} \\ 0 & 0 & \sqrt{x_{1t}+x_{2t}+x_{3t}} \end{pmatrix} d\begin{pmatrix} B_t^1 \\ B_t^2 \\ B_t^3 \end{pmatrix}, \quad (29) \end{aligned}$$

and its diffusion matrix is: $\{a_{ij}(x_t)\}_{1\leq i,j\leq 3} \equiv \begin{pmatrix} (x_{1t}+x_{2t}+x_{3t}) & 0 & 0 \\ 0 & 2(x_{1t}+x_{2t}+x_{3t}) & x_{1t}+x_{2t}+x_{3t} \\ 0 & x_{1t}+x_{2t}+x_{3t} & x_{1t}+x_{2t}+x_{3t} \end{pmatrix}$.

We use DGP9 to examine the power performance of Q_{nii} $(i = 1,2,3)$ and Q_{n23}. DGP5 is the null model.

Since the closed form transition densities are not available for these alternative models (DGP6, DGP7, DGP8, and DGP9), we use Milstein's scheme to simulate the data sets of the random sample. Both Table3 and Table 4 report the simulation results from DGP6, DGP7, DGP8, and DGP9, which indicate that the estimated powers of Q_{n11}, Q_{n22}, and Q_{n33} increase rapidly with respect to n, suggesting that Q_{n11}, Q_{n22}, and Q_{n33} have good power in detecting the misspecification caused by omitted variables in the conditional variances in DGP1,DGP3, DGP4, and DGP5 against their corresponding alternatives, even if there exists misspecification in other conditional variances or conditional covariances. Additionally, the simulation results show that Q_{n12}, Q_{n13} and Q_{n23} are quite powerful against the misspecification from omitted variables from the specifications of $a_{12}(x_t), a_{13}(x_t)$, and $a_{23}(x_t)$, respectively. The simulation results show that our tests perform rather well in detecting model misspecification coming from omitted variables in the conditional variances or the conditional covariances.

Table 1: **Percentage Rejection of True H_0**

n	5%	10%	5%	10%	5%	10%
	DGP1: Three-factor model with time-varying variances					
	$\text{H}_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t})] = 1$		$\text{H}_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$\text{H}_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.021	0.029	0.021	0.046	0.033	0.039
500	0.043	0.063	0.037	0.048	0.036	0.041
1000	0.045	0.065	0.039	0.063	0.036	0.052
2500	0.047	0.072	0.043	0.085	0.068	0.073
	DGP2: Three-factor model with nonlinear drift					
	$\text{H}_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t})] = 1$		$\text{H}_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$\text{H}_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.031	0.022	0.033	0.043	0.027	0.043
500	0.042	0.041	0.039	0.048	0.029	0.052
1000	0.045	0.063	0.041	0.055	0.033	0.066
2500	0.053	0.076	0.043	0.069	0.042	0.070

Table 1 reports the estimated sizes of the test statistics, Q_{n11}, Q_{n22}, and Q_{n33}. For DGP1, we simulate the data from the model transition density, which has a closed-form. For DGP2, its transition density has no closed form, we use Milstein's scheme (Kloeden and Platen, 1992) to simulate the data sets of the random sample. The data are simulated at a daily frequency.

Table 2: **Percentage Rejection of True H_0**

n	5%	10%	5%	10%	5%	10%
	DGP3: Three-factor model with time-varying variance and covaraince ($a_{12}(x_t) \neq 0$)					
	$\mathrm{H}_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t}, x_{2t})] = 1$		$\mathrm{H}_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$\mathrm{H}_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.009	0.042	0.023	0.048	0.019	0.041
500	0.028	0.044	0.024	0.052	0.022	0.045
1000	0.030	0.068	0.030	0.055	0.033	0.050
2500	0.036	0.077	0.042	0.063	0.036	0.066
	$\mathrm{H}_0^{12}: Pr[a_{12}(x_t) = b_{12}(x_{2t})] = 1$					
250	0.015	0.037				
500	0.027	0.059				
1000	0.036	0.078				
2500	0.042	0.090				
	DGP4: Three-factor model with time-varying variance and covaraince($a_{13}(x_t) \neq 0$)					
	$\mathrm{H}_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t})] = 1$		$\mathrm{H}_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$\mathrm{H}_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.033	0.041	0.018	0.039	0.013	0.030
500	0.036	0.045	0.024	0.062	0.023	0.034
1000	0.040	0.050	0.033	0.071	0.029	0.047
2500	0.043	0.065	0.040	0.083	0.032	0.065
	$\mathrm{H}_0^{13}: Pr[a_{13}(x_t) = b_{13}(x_{1t}, x_{3t})] = 1$					
250	0.019	0.025				
500	0.022	0.031				
1000	0.030	0.047				
2500	0.041	0.079				
	DGP5: Three-factor model with time-varying variance and covaraince($a_{23}(x_t) \neq 0$)					
	$\mathrm{H}_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t})] = 1$		$\mathrm{H}_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$\mathrm{H}_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.008	0.021	0.016	0.040	0.018	0.033
500	0.024	0.043	0.026	0.053	0.023	0.054
1000	0.029	0.048	0.031	0.061	0.031	0.085
2500	0.048	0.085	0.089	0.089	0.046	0.092
	$\mathrm{H}_0^{23}: Pr[a_{23}(x_t) = b_{23}(x_{2t}, x_{3t})] = 1$					
250	0.025	0.040				
500	0.029	0.058				
1000	0.042	0.079				
2500	0.044	0.092				

Table 2 reports the estimated sizes of the test statistics, $Q_{n11}, Q_{n22}, Q_{n33}, Q_{n12}, Q_{n13}$, and Q_{n23}. Their transition densities of these GDPs have no closed forms, we use Milstein's scheme (Kloeden and Platen, 1992) to simulate the data sets of the random sample at a daily frequency.

Table 3: **Percentage Rejection of True H_0**

n	5%	10%	5%	10%	5%	10%
	DGP6: Three-factor model with time-varying variance and covariance					
	$\mathrm{H}_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t})] = 1$		$\mathrm{H}_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$\mathrm{H}_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.240	0.283	0.223	0.251	0.290	0.318
500	0.352	0.379	0.415	0.423	0.400	0.436
1000	0.810	0.830	0.892	0.900	0.903	0.912
2500	0.960	0.980	1.000	1.000	1.000	1.000
	DGP7: Three-factor model with time-varying variance and covariance					
	$\mathrm{H}_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t}, x_{2t})] = 1$		$\mathrm{H}_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$\mathrm{H}_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.250	0.286	0.233	0.305	0.320	0.338
500	0.431	0.472	0.453	0.542	0.411	0.459
1000	0.845	0.867	0.877	0.911	0.852	0.891
2500	0.965	0.988	1.000	1.000	1.000	1.000
	$\mathrm{H}_0^{23}: Pr[a_{12}(x_t) = b_{12}(x_{2t})] = 1$					
250	0.351	0.370				
500	0.410	0.423				
1000	0.753	0.751				
2500	0.915	0.963				

Table 3 reports the estimated sizes of the test statistics, $Q_{n11}, Q_{n22}, Q_{n33}$, and Q_{n12}. Their transition densities of these GDPs have no closed forms, we use Milstein's scheme (Kloeden and Platen, 1992) to simulate the data sets of the random sample at a daily frequency.

Table 4: **Percentage Rejection of True H_0**

n	5%	10%	5%	10%	5%	10%
	DGP8: Three-factor model with time-varying variance and covariance					
	$H_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t})] = 1$		$H_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$H_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.272	0.293	0.303	0.332	0.238	0.251
500	0.304	0.391	0.412	0.502	0.305	0.458
1000	0.641	0.708	0.720	0.822	0.640	0.693
2500	0.891	0.923	0.956	1.000	0.963	1.000
	$H_0^{13}: Pr[a_{13}(x_t) = b_{13}(x_{1t}, x_{3t})] = 1$					
250	0.221	0.331				
500	0.481	0.509				
1000	0.545	0.781				
2500	0.932	1.000				
	DGP9: Three-factor model with time-varying variance and covariance					
	$H_0^{11}: Pr[a_{11}(x_t) = b_{11}(x_{1t})] = 1$		$H_0^{22}: Pr[a_{22}(x_t) = b_{22}(x_{2t})] = 1$		$H_0^{33}: Pr[a_{33}(x_t) = b_{33}(x_{3t})] = 1$	
250	0.162	0.211	0.218	0.279	0.231	0.256
500	0.392	0.421	0.345	0.412	0.351	0.418
1000	0.527	0.619	0.540	0.601	0.641	0.782
2500	0.934	0.976	0.891	0.942	0.935	0.987
	$H_0^{23}: Pr[a_{23}(x_t) = b_{23}(x_{2t}, x_{3t})] = 1$					
250	0.239	0.288				
500	0.367	0.430				
1000	0.564	0.673				
2500	0.947	0.965				

Table 4 reports the estimated sizes of the test statistics, $Q_{n11}, Q_{n22}, Q_{n33}, Q_{n12}, Q_{n13}$, and Q_{n23}. Their transition densities of these GDPs have no closed forms, we use Milstein's scheme (Kloeden and Platen, 1992) to simulate the data sets of the random sample at a daily frequency.

5 Applications to Term Structure of Interest Rates

In order to provide insights of the empirical applications of our testing procedure, our tests are applied to investigate the model specification of the term structures of interest rates. We particularly focus on the specification analysis of the diffusion matrix in a two-factor term structure model of interest rates, although our tests are applicable to many multivariate term structure models of interest rates in this literature, for example, the three-factor term structure models of interest rates in Dai and Singleton (2000) and Hong and Li (2005)).

Let r_t^1 and r_t^2 denote the short term rate of interest and the long term rate of interest, respectively, which can be expressed as follows,

$$dr_t^1 = \mu_1(r_t^1, r_t^2)dt + \sigma_1(r_t^1, r_t^2)dB_t^1, \tag{30}$$

$$dr_t^2 = \mu_2(r_t^1, r_t^2)dt + \sigma_2(r_t^1, r_t^2)dB_t^2, \tag{31}$$

where $\mu_1(r_t^1, r_t^2), \mu_2(r_t^1, r_t^2)$, $\sigma_1(r_t^1, r_t^2)$, and $\sigma_2(r_t^1, r_t^2)$ are unknown functions and satisfy assumptions 1-4. $(B_t^1, B_t^2)'$ is a 2-dimensional standard Brownian motion.

In the literature of modeling term structure of interest rates, two popular two-factor models are the Brennan-Schwartz model (1979), where $\sigma_1(r_t^1, r_t^2) = \sigma_1 r_t^1$ and $\sigma_2(r_t^1, r_t^2) = \sigma_2 r_t^2$, and the two-factor CIR model (Canabarro, 1995), where $\sigma_1(r_t^1, r_t^2) = \sigma_1\sqrt{r_t^1}$, and $\sigma_2(r_t^1, r_t^2) = \sigma_2\sqrt{r_t^2}$. In both the Brennan-Schwartz model and the two-factor CIR model, the volatility of each interest rate only depends on its current level.

We are interested in whether either r_t^1 or r_t^2 can be omitted from the specification of the volatilities of both r_t^1 and r_t^2. Namely, for the process r_t^1, we are interested in testing the following hypotheses,

$$H_0^{11} : Pr[\sigma_1(r_t^1, r_t^2) = \sigma_{11}^0(r_t^1)] = 1, \text{against } H_1^{11} : Pr[\sigma_1(r_t^1, r_t^2) = \sigma_{11}^0(r_t^1)] < 1 \tag{32}$$

and

$$H_0^{12}: Pr[\sigma_1(r_t^1, r_t^2) = \sigma_{12}^0(r_t^2)] = 1, \text{against } H_1^{12}: Pr[\sigma_1(r_t^1, r_t^2) = \sigma_{12}^0(r_t^2)] < 1, \tag{33}$$

where both $\sigma_{11}^0(r_t^1)$ and $\sigma_{12}^0(r_t^1)$ are unknown functions.

Similarly, for the process r_t^2, we are interested in testing the following hypotheses,

$$H_0^{21}: Pr[\sigma_2(r_t^1, r_t^2) = \sigma_{21}^0(r_t^1)] = 1, \text{against } H_1^{21}: Pr[\sigma_2(r_t^1, r_t^2) = \sigma_{21}^0(r_t^1)] < 1 \tag{34}$$

and

$$H_0^{22}: Pr[\sigma_2(r_t^1, r_t^2) = \sigma_{22}^0(r_t^2)] = 1, \text{against } H_1^{22}: Pr[\sigma_2(r_t^1, r_t^2) = \sigma_{22}^0(r_t^2)] < 1, \tag{35}$$

where both $\sigma_{21}^0(r_t^1)$ and $\sigma_{22}^0(r_t^1)$ are unknown functions.

We use daily 3-month Canadian Treasury bills and 10-year bond yield as the proxies of the short term interest rate and long term interest rate, respectively. The data covers the period from October 1, 2007 to January 31, 2021. We use the standard normal kernel functions $K(\cdot)$ and $L(\cdot)$ with smoothing parameters chosen by $h_r = \sigma_r n^{-1/4}$ and $a_w = \sigma_w n^{-1/5}$, where σ_r and σ_w are the sample standard deviations of r_t and w_t, respectively. Table 5 reports the description statistics of the two time series. Dickey-Fuller nonstationarity tests are conducted, and the presentence of a unit root is rejected for the two time series. Since the test is known to have low power, even slight rejection means that the existence of a unit root is unlikely. Both time series are plotted in Figure1. Due to the monetary policy shift, after 2007 year, the interest rates are characterized by substantially lower levels. We can make two observations from Figure1. First, both the short term and long term interest rates move up and down somewhat together (the correlation for the period above is 75%). Therefore, parallel shifts are common. Second, although long rates directionally follow short rates, they tend to lag in magnitude.

The test results are reported in Table 6. The large test statistics are providing overwhelming evidence that the omitted variable specifications of both $\sigma_1(r_t^1, r_t^2)$ and $\sigma_2(r_t^1, r_t^2)$ are rejected. To display possible reasons for the rejections of the omitted variable specifications of both $\sigma_1(r_1; r_2)$ and $\sigma_2(r_1, r_2)$, the nonparametric estimators of $\sigma_1(r_1, r_2)$ and $\sigma_2(r_1, r_2)$ are reported in Figure 2 and Figure 3, respectively. Given the nonparametric shapes of $\sigma_1(r_1, r_2)$ and $\sigma_2(r_1, r_2)$, the rejections are not surprising. The volatilities of both the short term and long term interest rates depend not only on the current value of the short term interest rate but also on the long term interest rate. This is consistent with the expectations theory that the current long term interest rate carries information about the future values of the short term interest rate, suggesting that the long term interest rate directly affects the volatility of the short term interest rates. Thus, it is reasonable that the specification of the volatility of the short term interest rate depends on the short term interest rate as well as the long term interest rate. Further, the volatility of the long term interest rate not only depends on the current long term interest rate but also the current short term interest rate because the long term interest rate is related to the current short term interest rate.

Table 5: **Summary statistics of the data and stationarity test**

	n	Mean	SD	ρ_1	ρ_3	ρ_5	ρ_7	ρ_9	ρ_{11}	ρ_{13}
r_t^1	3083	0.010	0.007	0.997	0.976	0.967	0.959	0.952	0.985	0.946
r_t^2	3083	0.023	0.008	0.996	0.987	0.980	0.973	0.965	0.959	0.952

Augmented Dickey-Fuller stationarity test

H_0	Test statistic	Critical value	Result
r_t^1 is not stationary	-3.91	-2.57	Rejected
r_t^2 is not stationary	-2.67	-2.57	Rejected

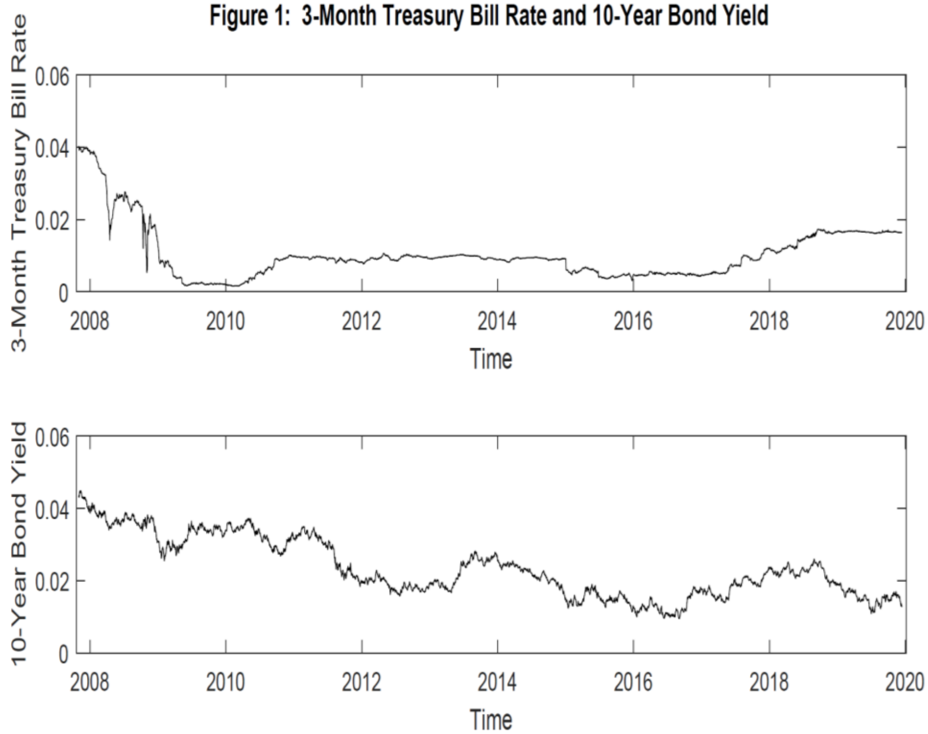
In Table 5, r_t^1 and r_t^2 denote the daily 3-month treasury bill rate and daily 10-year bond yield, respectively. The sample covers period from October 1 2007 to January 31 2020. ρ_l denotes the autocorrelation coefficient of order l. The augmented Dickey-Fuller test statistic is computed as $\hat{\tau} = \hat{\alpha}_l / ase(\hat{\alpha}_l)$ in the model: $\Delta r_t^i = \alpha_0 + \alpha_1 r_t^i + \sum_{j=1}^{p} \phi_j \Delta r_{t-j}^i + \varepsilon_t, i = 1, 2$. The value of ρ is set as the highest significant lag order from either the autocorrelation function or the partial autocorrelation function of the first differenced series (up to a maximum lag order of $\sqrt{n}$)

Table 6: **Testing for omitted variable in a two-factor term structure model**

Volatility	Null hypothesis	Test statistic	Result
$\sigma_1(r_1,r_2)$	$\Pr[\sigma_1(r_1,r_2)=\sigma^0_{11}(r_1)]=1$	$Q_{n11}=202.45$	Reject
	$\Pr[\sigma_1(r_1,r_2)=\sigma^0_{12}(r_2)]=1$	$Q_{n11}=311.21$	Reject
$\sigma_2(r_1,r_2)$	$\Pr[\sigma_2(r_1,r_2)=\sigma^0_{21}(r_1)]=1$	$Q_{n22}=158.56$	Reject
	$\Pr[\sigma_2(r_1,r_2)=\sigma^0_{22}(r_2)]=1$	$Q_{n22}=196.03$	Reject

Table 6 reports the test statistic values of Q_{n11} and Q_{n22} using the daily 3-month treasury bill rate and daily 10-year bond yield. The sample covers period from October 1 2007 to January 31 2020. The two-factor model is specified as: $dr_t^1=\mu_1(r_t^1,r_t^2)dt+\sigma_1(r_t^1,r_t^2)dB_t^1$ and $dr_t^2=\mu_2(r_t^1,r_t^2)dt+\sigma_2(r_t^1,r_t^2)dB_t^2$.

Figure 1: 3-Month Treasury Bill Rate and 10-Year Bond Yield



Figures 1 plots the daily 3-month Canadian Treasury bills and 10-year-bond yield. The data covers the period from October 1, 2007 to January 31, 2021.

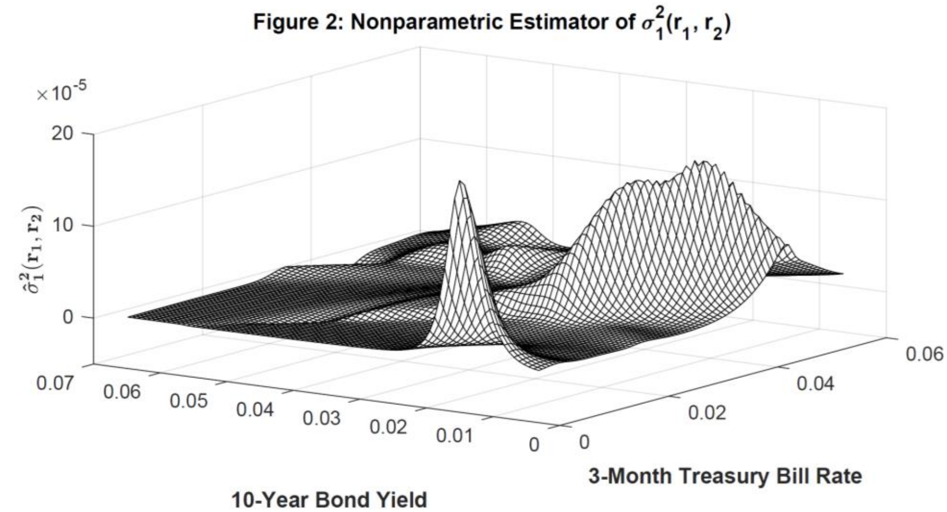


Figure2 plots the nonparametric estimator of $\sigma_1^2(r_t^1, r_t^2)$, which is the diffusion function of r_t^1 in (30), using the daily 3-month Canadian Treasury bills and 10-year-above bond yield. The data covers the period from October 1, 2007, to January 31, 2021.

Figure 3: Nonparametric Estimator of $\sigma_2^2(r_1, r_2)$

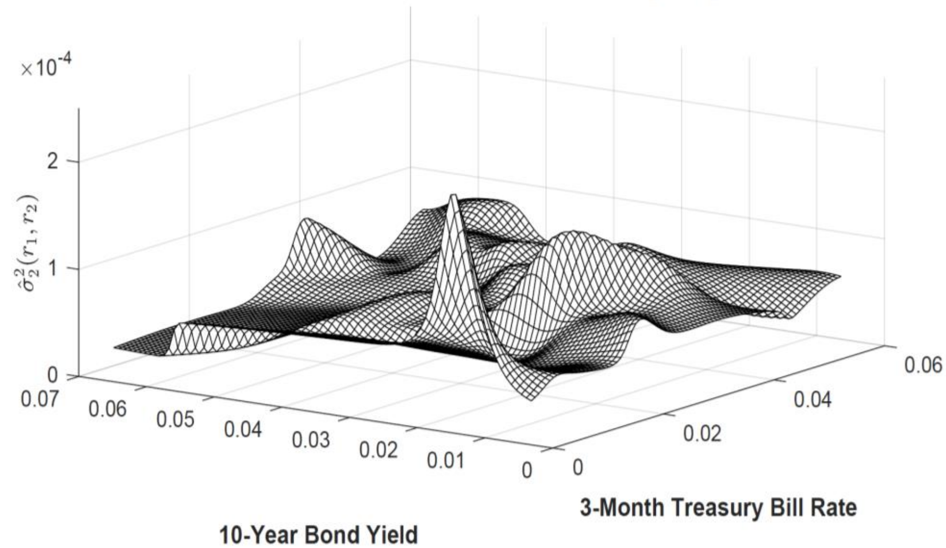


Figure3 plots the nonparametric estimator of $\sigma_2^2(r_t^1, r_t^2)$, which is the diffusion function of r_t^1 in (31), using the daily 3-month Canadian Treasury bills and 10-year-above bond yield. The data covers the period from October 1, 2007, to January 31, 2021.

6 Conclusion

In this paper, we have developed a nonparametric test for omitted variables in each component of the diffusion matrix in a multivariate diffusion process. Each of all these test statistics follows an asymptotic standard normal distribution under null hypothesis that a subset of state variables can be omitted from the specification of the componenet, while diverging to infinity if omitted variables do not exist in the component. Monte Carlo simulations show that our tests have reasonable size and good power against a variety of alternatives.

To illustrate the empirical relevance of our tests, we apply our tests to the situation of dimension reduction in the specification analysis of the diffusion matrix in a two-factor term structure model of interest rates. Using the daily 3-month Canadian Treasury Bills and 10-year Canadian bond yield as the proxies of the short term interest rate and long term interest rate, respectively. The empirical result indicates that the volatilities of both the short term and long term interest rates depend not only on the current value of the short term interest rate but also on the long term interest rate. Since interest rate volatility is of fundamental importance in valuing contingent claims and hedging interest rate risk, our results suggest that a volatility-simplified term structural model could lead to model misspecification, although it greatly facilitates pricing and econometric implementation.

In our future work, the results obtained in this paper are expected to be extended to test for the omitted variables of the diffusion matrix in a multivariate Jump-diffusion process, which is a valuable tool in asset pricing and hedging of derivative securities.

References

[1] Abel, A.B. (1985): "A Stochastic Model of Investment, Marginal q, and the Market Value of the Firm," *International Economic Review 26, 305-322.*

[2] Aït-Sahalia, Y. (1996): " Testing continuous-time models of the spot interest rate," *Review of Financial Studies 2, 385-426.*

[3] Aït-Sahalia, Y., P.J., Bickel, and T.M.Stoker (2001): "Goodness-of-Fit Tests for Kernel Regression with an Application to Option Implied Volatilities," *Journal of Econometrics, 105, 363-412.*

[4] Aï-Sahalia, Y. (2008): "Closed-Form Likelihood Expansions for Multivariate Diffusions," *The Annals of Statistics, 36, No 2, 906-937.*

[5] Aït-Sahalia, Y., J.Fan, and H.Peng (2009): "Nonparametric Transition-Based Tests for Jump Diffusions," *Journal of the American Statistical Association, 104, 1102-1116.*

[6] Aï-Sahalia, Y. and R.Kimmel (2010): "Estimating Affine Multifactor Term Structure Models Using Closed-Form Likelihood Expansions," *Journal of Financial Economics, 98, 113-144.*

[7] Bandi, F.M. and G.Moloche (2018): "On the Functional Estimation of Multivariate Diffusion Processes," *Econometric Theory, V.34, Issue 4, 896-946.*

[8] Black, F. and M.Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy, 81, 637-654.*

[9] Brennan, M. and E.Schwartz (1979): "A Continuous Time Approach to the Pricing of Fonds," *Journal of Banking and Finance, 3, 133-155.*

[10] Canabarro, E.(1995): "Where do one-factor interest rate models fail?" *The Journal of Fixed Income, 5(2), 31-52.*

[11] Chen, S. X., J.Gao, and C.Y.Tang (2008): "A Test for Model Specification of Diffusion Processes," *Annals of Statistics, 36, 167-198.*

[12] Chen, B. and Y. Hong (2010): "Characteristic Function-Based Testing for Multifactor Continuous-Time Markov Models Via Nonparametric Regression," *Econometric Theory, 26, 1115-1179.*

[13] Dai, Q., and K.Singleton (2000): "Specification Analysis of Affine Term Structure Models," *Journal of Finance, 55, 1943-1978.*

[14] Dvoretzky, A. (1972): "Asymptotic normality for sums of dependent random variables," *Berkeley Symposium on Mathematical Statistics and Probability, 1972: 513-535 (1972)*

[15] Fan, Y. and Q.Li (1996): "Consistent Model Specification Tests: Omitted Variables and Semiparametric Fucntions," *Econometrica, 64, 865-890.*

[16] Fan, Y. and Q. Li (1999): "Central Limit Theorem for Degenerate U-Statistics of Absolutely Regular Processes with Applications to Model Specification Testing," *Nonparametric Statistics, 10, 245-271.*

[17] Friedman, A. (1975): "Stochastic Differential Equations and Applicatins," *Vol. 1. Academic Press.*

[18] Harel, M. and M. Puri (1990): "Weak Convergence of the Serial Linear Rank Statistic with Unbounded Scores and Regression Constants under Mixing Conditions," *Journal of Statistical Planning and Inference, V.25, Issue 2, 163-186.*

[19] Hong,Y. and H.Li (2005): "Nonparametric Specification Testing for Continuous-Time Models with Applications to Term Structure of Interest Rates," *Review of Financial Studies, Vol.18, 37-84.*

[20] Hsin, C-W.(1995): "An Empirical Investigation of the Two-Factor Brennan-Schwartz Term Structure Model," *Review of Quantitative Finance and Accounting, 5, 71-92.*

[21] Hull, J. and A. White (1990): "Pricing Interest Rate Derivative Securities," *Review of Financial Studies, 3, 573-592.*

[22] Karatzas, I. and S.E.Shreve (1991): "Brownian Motion and Stochastic Calculus," *Springer, New York. MR1121940.*

[23] Kloeden, P.E. and E.Platen (1992): "Numerical Solution of Stochastic Differential Equations," *Springer-Verlag Berlin Heidelberg.*

[24] Kristensen (2011): "Semi-Nonparametric Estimation and Misspecification Testing of Diffusion Models," *Journal of Econometrics 164, 382-403.*

[25] Lavergne, P., and Q.H. Vuong (2000): "Nonparametric Significance Testing," *Econometric theory, 16, 576-601.*

[26] Li, F. (2019): "Testing for the Diffusion Matrix in a Continuous-Time Markov Process Model with Applications to the Term Structure of Interest Rates," *Journal of Financial Econometrics, Furthercoming.*

[27] Merton, R.C.(1973): "The Theory of Rational Option Pricing," *Bell Journal of Economics and Management Science 4, 141-183.*

[28] Meyn, S., and R.L.Tweedie (1993): "Stability of Markovian Processes III: Forster-Lyapunov Criteria for Continuous-Time Processses," *Adv. Appl.Prob, 25, 518-548.*

[29] Robinson, P.M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica, 56, 931-954.*

[30] Stroock, D.W. and S.R.S. Varadhan (1979): "Multidimensional Diffusion Processes," *Springer, New York. MR0532498.*

[31] Yoshihara, K. (1976): "Limiting Behavior of Generalized Quadratic Forms Generated by Absolutely Regular Sequences II," *Z. Wahrscheinlichkeitstheorie verw. Gebiete., 35, 237-252.*

Appendix A

Throughout this appendix, h_n and a_n are expressed by h and a, respectively. $\int_{t_0+t\triangle_n}^{t_0+(t+1)\triangle_n} G(u)du$ is denoted by $\int_{\triangle_n} G(u)du$, where $G(u)$ is any integral function. The symbol C denotes a generic big enough positive constant.

Proof of Theorem 1. Since the proof of (ii) in Theorem 1 is similar to and in fact much simpler than the proof of (i) in Theorem 1, we will only provide the proof of (i) in Theorem 1.

Recall that $\hat{u}_{n,t} = [x^i_{n,t+1} - x^i_{n,t}][x^i_{n,t+1} - x^j_{n,t}]/\triangle_n - \hat{b}_{ij}(w_{n,t})$, where $\hat{b}_{ij}(w_{n,t})$ is a kernel estimator of $b_{ij}(w_{n,t})$ that is defined by (12). Under null hypothesis, using Itô's lemma to $g_{ij}(x_s) = [x^i_{n,s} - x^i_{n,t}][x^i_{n,s} - x^j_{n,t}]/\triangle_n$, we can rewrite $\hat{u}_{n,t}$ as,

$$\begin{aligned} \hat{u}_{n,t} &= b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}) \\ &\quad + \frac{1}{\triangle_n}\int_{\triangle_n}\{[x^j_u - x^j_{n,t}]\mu_i(x_u) + [x^i_u - x^i_{n,t}]\mu_j(x_u)\}du \\ &\quad + \int_{\triangle_n}\frac{\partial g_{ij}(x_u)}{\partial x_u}\sigma(x_u)dB_u \\ &\equiv b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}) + c_{n,t} + \varepsilon_{n,t}, \end{aligned} \tag{A.1}$$

where $\frac{\partial g_{ij}(x_u)}{\partial x_u} = (\frac{\partial g_{ij}(x_u)}{\partial x^1_u}, \ldots, \frac{\partial g_{ij}(x_u)}{\partial x^d_u})$. Inserting $(A.1)$ into $\hat{I}_{nij}$ in (15) yields the following result,

$$\begin{aligned} \hat{I}_{nij} &= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{(b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})(b_{ij}(w_{n,s}) - \hat{b}_{ij}(w_{n,s}))\hat{f}_w(w_{n,s})\Big\}K_{ts} \\ &\quad + \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{\varepsilon_{n,t}\varepsilon_{n,s}\hat{f}_w(w_{n,t})\hat{f}_w(w_{n,s})\Big\}K_{ts} \\ &\quad + 2\frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{(b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})\hat{f}_w(w_{n,s})c_{n,s}\Big\}K_{ts} \\ &\quad + 2\frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{(b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})\hat{f}_w(w_{n,s})\varepsilon_{n,s}\Big\}K_{ts} \\ &\quad + 2\frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{\varepsilon_{n,t}\hat{f}_w(w_{n,t})c_{n,s}\hat{f}_w(w_{n,s})\Big\}K_{ts} \\ &\quad + \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{c_{n,t}\hat{f}_w(w_{n,t})c_{n,s}\hat{f}_w(w_{n,s})\Big\}K_{ts} \\ &\equiv J^{ij}_{n1} + J^{ij}_{n2} + 2J^{ij}_{n3} + 2J^{ij}_{n4} + 2J^{ij}_{n5} + J^{ij}_{n6}. \end{aligned} \tag{A.2}$$

We will prove Theorem 1 (i) by showing that $J_{nl}^{ij} = o_p((nh^{d/2})^{-1})$ for $l = 1,3,4,5,6$ and $nh^{d/2}J_{n2}^{ij}/v_{nij} \to N(0,1)$ in distribution. These results will be proven in Lemma A1-Lemma A6 below.

Lemma A1. Under the assumptions in Theorem 1, we have $J_{n1}^{ij} = o_p((nh^{d/2})^{-1})$.

Proof: using $\hat{f}_t = \frac{1}{(n-1)h^d}\sum_{s\neq t} K_{ts}$, $ab \leq \frac{1}{2}(a^2+b^2)$ (a and b are nonnegative real numbers), and $f_t \equiv f(x_{n,t})$, we have,

$$
\begin{aligned}
E|J_{n1}^{ij}| &\leq \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t} E|(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})(b_{ij}(w_{n,s})-\hat{b}_{ij}(w_{n,s}))\hat{f}_w(w_{n,s})K_{ts}| \\
&\leq \frac{1}{2n(n-1)h^d}\sum_t\sum_{s\neq t} E[(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))^2\hat{f}_w^2(w_{n,t})K_{ts}] \\
&\quad + \frac{1}{2n(n-1)h^d}\sum_t\sum_{s\neq t} E[(b_{ij}(w_{n,s})-\hat{b}_{ij}(w_{n,s}))^2\hat{f}_w^2(w_{n,s})K_{ts}] \\
&= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t} E[(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))^2\hat{f}_w^2(w_{n,t})K_{ts}] \\
&= \frac{1}{n}\sum_t E[(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))^2\hat{f}_w^2(w_{n,t})\hat{f}_t] \\
&\leq \frac{1}{n}\sum_t E[(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))^2\hat{f}_w^2(w_{n,t})f_t] \\
&\quad + \frac{1}{n}\sum_t E[(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))^2\hat{f}_w^2(w_{n,t})(\hat{f}_t - f_t)] \\
&= o((nh^{d/2})^{-1}), \qquad \text{(A.3)}
\end{aligned}
$$

by (i) and (ii) of Lemma B.

Lemma A2. Under the assumptions in Theorem 1, we have $nh^{d/2}J_{n2}^{ij} \to N(0, v_{ij}^2)$ in distribution, where $v_{ij}^2 = 2E[(a_{ii}(x_t)a_{jj}(x_t)+a_{ij}^2(x_t))^2 f_w^4(w_t)f(x_t)]\int K^2(u)du$.

Proof: J_{n2}^{ij} can be rewritten in the following way,

$$
\begin{aligned}
J_{n2}^{ij} &= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\varepsilon_{n,t}\varepsilon_{n,s}f_{w_t}f_{w_s}K_{ts} + \frac{2}{n(n-1)h^d}\sum_t\sum_{s\neq t}\varepsilon_{n,t}\varepsilon_{n,s}(\hat{f}_{w_t}-f_{w_t})f_{w_s}K_{ts} \\
&\quad + \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\varepsilon_{n,t}\varepsilon_{n,s}(\hat{f}_{w_t}-f_{w_t})(\hat{f}_{w_s}-f_{w_s})K_{ts} \\
&\equiv J_{n21}^{ij} + 2J_{n22}^{ij} + J_{n23}^{ij}. \qquad \text{(A.4)}
\end{aligned}
$$

We will prove $nh^{d/2}J_{n2}^{ij} \to N(0, v_{ij}^2)$ by showing that $nh^{d/2}J_{n21}^{ij} \to N(0, v_{ij}^2)$ in distribution, and $J_{n2l}^{ij} = o_p((nh^{d/2})^{-1})$ for $l = 2$ and $l = 3$. To apply Lemma C.1 to show $nh^{d/2}J_{n21}^{ij} \to N(0, v_{ij}^2)$ in distribution, we let $U_n \equiv \sum_{1\leq s<t\leq n} \varepsilon_{n,t}\varepsilon_{n,s} f_{w_t} f_{w_s} K_{st} = \sum\sum_{1\leq s\leq t\leq n} H(z_{n,t}, z_{n,s})$, where $z_{n,t} = (x_{n,t}, \varepsilon_{n,t}, f_{w_t})'$. We now show that conditions (A1), (A2), and (A3) in Lemma C.1 are satisfied. First we check (A1). Let $\tilde{z}_{n,t}{}_{t=1}^{n}$ be an i.i.d. sequence having the same marginal distribution as $\{z_{n,t}\}_{t=1}^{n}$. We have $\sigma_{nij}^2 \equiv E[H^2(\tilde{z}_{n,1}, \tilde{z}_{n,2})] = E\{E[\varepsilon_{n,t}^2|x_{n,t}]E[\varepsilon_{n,s}^2|x_{n,s}] f_{w_t}^2 f_{w_s}^2 K_{ts}^2\}$. Using Theorem 2.8 in Friedman(1975), we have,

$$
\begin{aligned}
E[\varepsilon_{n,t}^2|x_{n,t}] &= \frac{1}{\triangle_n^2}\int_{\triangle_n} E\Big\{[\frac{\partial g_{ij}(x_u)}{\partial x}\sigma(x_u)][\frac{\partial g_{ij}(x_u)}{\partial x}\sigma(x_u)]'|x_{n,t}\Big\}du \\
&= \frac{1}{\triangle_n^2}\int_{\triangle_n} \Big\{E[(x_u^j - x_{n,t}^j)^2 a_{ii}(x_{n,t})|x_{n,t}] + 2E[(x_u^i - x_{n,t}^i)(x_u^j - x_{n,t}^j)a_{ij}(x_{n,t})|x_{n,t}] \\
&\quad + E[(x_u^i - x_{n,t}^i)^2 a_{jj}(x_{n,t})|x_{n,t}]\Big\}du \\
&\quad + \frac{1}{\triangle_n^2}\int_{\triangle_n} \Big\{E(x_u^i - x_{n,t}^j)^2(a_{ii}(x_u) - a_{ii}(x_{n,t}))|x_{n,t}] \\
&+ 2E[(x_u^i - x_{n,t}^i)(x_u^j - x_{n,t}^j)(a_{ij}(x_u) - a_{ij}(x_{n,t}))|x_{n,t}] \\
&+ E[(x_u^i - x_{n,t}^i)^2(a_{ij}(x_u) - a_{ij}(x_{n,t}))|x_{n,t}]\Big\}du \\
&\equiv R_{n,t}^1 + R_{n,t}^2. \qquad (A.5)
\end{aligned}
$$

Under Assumptions1-4, applying Itô's lemma to $(x_u^j - x_{n,t}^j)^2$, $(x_u^i - x_{n,t}^i)^2$, and $x_u^j - x_{n,t}^j)(x_u^i - x_{n,t}^i)$, respectively, and using Lemma C.3 in Appendix C, we have,

$$
\begin{aligned}
\sigma_{nij}^2 &= E\{E[\varepsilon_{n,t}^2|x_{n,t}]E[\varepsilon_{n,s}^2|x_{n,s}] f_{w_t}^2 f_{w_s}^2 K_{ts}^2\} \\
&= E\{R_{n,t}^1 R_{n,s}^1 + R_{n,t}^2 R_{n,s}^2 + 2R_{n,t}^1 R_{n,s}^2] f_{w_t}^2 f_{w_s}^2 K_{ts}^2\} \\
&= E\{[a_{ii}(x_{n,t})a_{jj}(x_{n,t}) + a_{ij}^2(x_{n,t})][a_{ii}(x_{n,s})a_{jj}(x_{n,s}) + a_{ij}^2(x_{n,s})] f_{w_t}^2 f_{w_s}^2 K_{ts}^2\} + o(h^d) \\
&= \int [a_{ii}(x)a_{jj}(x) + a_{ij}^2(x)][a_{ii}(y)a_{jj}(y) + a_{ij}^2(y)] f_w^2(w_x) f_w^2(w_y) f(x) f(y) K^2(\frac{x-y}{h}) dxdy + o(h^d) \\
&= h^d\Big\{\int [a_{ii}(x)a_{jj}(x) + a_{ij}^2(x)]^2 f_w^4(w_s) f^2(x)dx \int K^2(u)du + o(1)\Big\}. \qquad (A.6)
\end{aligned}
$$

Similarly, we have $\mu_{n4} \equiv E[H^4(\tilde{z}_{n,1}, \tilde{z}_{n,2})] = E\{E[\varepsilon_{n,t}^4|x_{n,t}][\varepsilon_{n,s}^4|x_{n,s}]f_{w_t}^4 f_{w_s}^4 K_{ts}^4\} = O(h^d)$.

Denote $K((x - x_{n,t_l})/h) = K_{xt_l}$, $l = 1,2,3,4$, we have,

$$
\begin{aligned}
\tilde{\gamma}_{n14} &\equiv \int \{E[H(z, z_{n,t})H(z, z_{n,s})]\}^2 dF(z) \\
&= \int E[\varepsilon_{n,t}^4 f_{w_t}^4 | x_{n,t} = x] E[\varepsilon_{n,t_1} f_{w_{t_1}} \varepsilon_{n,t_2} f_{w_{t_2}} K_{xt_1} K_{xt_2}] E[\varepsilon_{n,t_3} f_{w_{t_3}} \varepsilon_{n,t_4} f_{w_{t_4}} K_{xt_3} K_{xt_4}] f(x) dx \\
&\leq \int E[\varepsilon_{n,t}^4 f_{w_t}^4 | x_{n,t} = x] \{E[\varepsilon_{n,t_1}^2 f_{w_{t_1}}^2 \varepsilon_{n,t_2}^2 f_{w_{t_2}}^2] E[K_{xt_1}^2 K_{xt_2}^2] E[\varepsilon_{n,t_3}^2 f_{w_{t_3}}^2 \varepsilon_{n,t_4}^2 f_{w_{t_4}}^2] E[K_{xt_3}^2 K_{xt_4}^2]\}^{1/2} f(x) dx \\
&\leq C \int E[\varepsilon_{n,t}^4 f_{w_t}^4 | x_{n,t} = x] f(x) dx h^{2d} = O(h^{2d}).
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\gamma}_{n22} &\equiv E[H^2(\tilde{z}_{n,1}, \tilde{z}_{n,2}) H^2(\tilde{z}_{n,1}, \tilde{z}_{n,3})] = \int\int\int E[\varepsilon_{n,t}^4 | x_{n,t} = x] E[\varepsilon_{n,t}^2 | x_{n,t} = y] E[\varepsilon_{n,t}^2 | x_{n,t} = z] \\
&\quad \times K^2((x-y)/h) K^2((x-z)/h) f(x) f(y) f(z) dx dy dz = O(h^{2d}).
\end{aligned}
$$

$$
\begin{aligned}
\gamma_{n11} &\equiv max_{t \neq s, t' \neq s'} E[H(z_{n,t}, z_{n,s})(z_{n,t'}, z_{n,s'})] \leq E[|\varepsilon_{n,t} f_{w_t} \varepsilon_{n,s} f_{w_s} \varepsilon_{n,t'} f_{w_{t'}} \varepsilon_{n,s'} f_{w_{s'}} K_{ts} K_{t's'}|] \\
&\leq [E|\varepsilon_{n,t} f_{w_t} \varepsilon_{n,s} f_{w_s} \varepsilon_{n,t'} f_{w_{t'}} \varepsilon_{n,s'} f_{w_{s'}}|^{\xi}]^{1/\xi} [E|K_{ts} K_{t's'}|]^{1/\eta} = O(h^{2d/\eta}),
\end{aligned}
$$

where $\eta = (1 - \xi^{-1})^{-1}$ and ξ is slightly larger than 2. Hence $\gamma_n \equiv max\{\gamma_{n11}, \tilde{\gamma}_{n22}, \tilde{\gamma}_{n14}\} = O(h^{2d/\eta})$, where $(1 < \eta < 2)$.

Similarly, we can prove that $\gamma_{n13} \equiv max_{t \neq s, t' \neq s'}[H(z_{n,t}, z_{n,s}) H^3(z_{n,t}, z_{n,s})] = O(h^{2d/\eta'})$ and $\gamma_{n22} \equiv max_{t \neq s, t' \neq s'}[H^2(z_{n,t}, z_{n,s}) H^3(z_{n,t'}, z_{n,s'})] = O(h^{2d/\eta'})$, where $1 < \eta' < 2$. Hence, $\bar{v}_n \equiv max\{\gamma_{n22}, \gamma_{n13}\} = O(h^{2d/\eta'})$. Summarizing the above results, we have shown that $\sigma_{nij}^2 = O(h^d)$, $\mu_{n4} = O(h^d)$, $\gamma_n = O(h^{2d/\eta})$, $\bar{v}_n = O(h^{2d/\eta'})$, where $1 < \eta < 2$, and $1 < \eta' < 2$. These results imply (i), (ii), and (iii) in (A1) of Lemma C.1.

Next, for $G(z_{n,t}, z_{n,s}) \equiv E[H(z, z_{n,t})H(z, z_{n,s})]$, we have,

$$
\begin{aligned}
G(z_{n,t}, z_{n,s}) &= \varepsilon_{n,t} f_{w_t} \varepsilon_{n,s} f_{w_s} f_{w_s} \int E[\varepsilon_{n,t'}^2 f_{w_t'} | x_{n,t'} = x] K(\frac{x_{n,t} - x}{h}) K(\frac{x_{n,s} - x}{h}) f(x) dx \\
&= \varepsilon_{n,t} f_{w_t} \varepsilon_{n,s} f_{w_s} f_{w_s} \int \sigma_{n,t'}^2(x) K(\frac{x_{n,t} - x}{h}) K(\frac{x_{n,s} - x}{h}) f(x) dx \\
&= h^d \varepsilon_{n,t} f_{w_t} \varepsilon_{n,s} f_{w_s} f_{w_s} \int \sigma_{n,t'}^2(x_{n,t} + hu) K(u) K(\frac{x_{n,t} - x_{n,s}}{h} + u) f(x_{n,t} + hu) du.
\end{aligned}
$$

Hence, it is straightforward to show that $v_G^2 \equiv E[G^2(z_{n,t},z_{n,t})] = O(h^{2d}), \mu_{nG2} \sim E[G^2(z_{n,t},z_{n,s})] = O(h^{2d}h^{d/\eta'})$ and $\gamma_{nG11} = O(h^{2d}h^{d/\eta'})$, where $\eta' > 1$. Finally, it is easy to show that M_n is bounded.

From now on, we let $m = \tau_n = [Clog^{1+\lambda_0}(n)]$. Using the Assumption 2, we have, $\beta_{n,\tau_n} = O(\lambda^{Clog^{1+\lambda_0}(n)\frac{1}{log^{\lambda_0}(\triangle_n^{-1})}}) = O(\lambda^{Clog(n)}) = O(\lambda^{-C\gamma log_\lambda(n)}) = O(n^{-C\gamma})$, where $\gamma = -log(\lambda) > 0$. Hence we have: for any given $i > 0$, $n^i \beta_{n,m}^{\delta/(1+\delta)} = o(1)$ as long as we choose C sufficiently large. Hence the conditions (i), (ii), and (iii) of (A2), and conditions (i) and (ii) of (A3) in Lemma C.1 are satisfied.

Using Lemma C.1 and the fact that $E(U_n) = 0$, we have $nh^{d/2}J_{n21}^{ij} = 2U_n/((n-1)h^{d/2}) = [\sqrt{2}n\sigma_{nij}/(n-1)h^{d/2}][\sqrt{2}U_n/(n\sigma_{nij})] \to N(0, v_{ij}^2)$ in distribution.

To prove $J_{n22}^{ij} = o_p((nh^{d/2})^{-1})$, we rewrite J_{n22}^{ij} as $\frac{1}{n^2(n-1)h^d a^q}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_2} \varepsilon_{n,t_1}\varepsilon_{n,t_2}(L_{t_1t_3} - a^q f_{t_1})f_{t_2}K_{t_1t_2}$ and we consider the second moment of J_{n22}^{ij}:

$$\begin{aligned} E_1 \;\; \equiv \;\; & E(J_{n22}^{ij})^2 = \frac{1}{[n^2(n-1)h^d a^q]^2}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_2}\sum\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4} E\Big\{\varepsilon_{n,t_1}\varepsilon_{n,t_2}(L_{t_1t_3} - a^q f_{t_1}) \\ & \times f_{t_2}K_{t_1t_2}\varepsilon_{n,t_4}\varepsilon_{n,t_5}(L_{t_4t_6} - a^q f_{t_4})f_{t_5}K_{t_4t_5}\Big\}. \end{aligned}$$

We consier three different cases for E_1 : (a) for at least three different $i \neq j, |t_i - t_j| > m$ for all $j \neq i$; (b) for exactly two different $j \neq i, |t_i - t_j| > m$; (c) all the remaining cases. We use E_{1a}, E_{1b}, and E_{1c} to denote these three cases. For case (a), since at least there exists one $i \in \{1,2,4,5\}$, using Lemma C.2 we have $E_{1a} \leq 0 + Cn^6\beta_{n,m}^{\delta/(1+\delta)} = o((nh^{d/2})^{-1})$. For case (b), we only need to consier $|t_3 - t_4| > m$ for all $i \neq 3$ and $|t_6 - t_i| > m$ for all $j \neq 6$ because otherwise $E_i(b)$ will be bounded by $Cn^6\beta_{n,m}^{\delta/(1+\delta)} = o((nh^{d/2})^{-1})$. Case (b) has n^4m^2 terms and they correspond to either (i) $|t_1 - t_4| \leq m$ or $|t_1 - t_2| \leq m$ and $|t_4 - t_5| \leq m$. We use $E_{1b}(1)$ and $E_{1b}(2)$ to denote these two subcases. Using Lemma C.2 four times, and $E_{t_1,t_4}[\varepsilon_{n,t_2}\varepsilon_{n,t_5}K_{t_1t_2}K_{t_4t_5}] \leq \{E_{t_1,t_4}[\varepsilon_{n,t_2}^2\varepsilon_{n,t_5}^2]E_{t_1,t_4}[K_{t_1t_2}^2K_{t_4t_5}^2]\}^{1/2}$, we have,

$$\begin{aligned} E_{1b}(1) \;\; \leq \;\; & \frac{1}{[n^2(n-1)h^d a^q]^2}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_2}\sum\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4} E\Big\{\varepsilon_{n,t_1}\varepsilon_{n,t_4}E_{t_1,t_4}[\varepsilon_{n,t_2}\varepsilon_{n,t_5}f_{t_5}f_{t_2}K_{t_1t_2}K_{t_4t_5}] \\ & \times [E_{t_1}(L_{t_1t_3} - a^q f_{t_1})][E_{t_4}(L_{t_4t_6} - a^q f_{t_4})]\Big\} \end{aligned}$$

$$\leq \ Cn^4m^2(n^3h^da^q)^{-2}O(a^{2(q+r)}h^d) = O(m^2a^{2r}(n^2h^d)^{-1}) = o((nh^{d/2})^{-1}). \tag{A.7}$$

By Lemma C.2, Lemma 4 in Robinson (1988), and the fact that $E[\varepsilon_{n,t_1}\varepsilon_{n,t_2}K_{t_1,t_2}] \leq \{E[\varepsilon_{n,t_1}^2\varepsilon_{n,t_2}^2]E[K_{t_1t_2}]\}^{1/2}$, we have,

$$\begin{aligned}
E_{1b}(2) &\leq \frac{1}{[n^2(n-1)h^da^q]^2}\sum\sum\sum_{t_1,t_2\neq t_1,t_3\neq t_1}\sum\sum\sum_{t_4,t_5\neq 4,t_6\neq t_4} \\
&\quad \times E\{K_{t_1t_2}\varepsilon_{n,t_1}\varepsilon_{n,t_2}[E_{t_1}(L_{t_1t_3}-a^qf_{t_1})]\}E\{K_{t_4t_5}\varepsilon_{n,t_4}\varepsilon_{n,t_5}[E_{t_4}(L_{t_4t_6}-a^qf_{t_4})]\} \\
&\leq Cn^2(n^3h^da^q)^{-2}\sum\sum_{t_1,t_2\neq t_1}\sum\sum_{t_4,t_5\neq t_4}E[|\varepsilon_{n,t_1}\varepsilon_{n,t_2}|K_{t_1t_2}D_f(w_{t_1})]E[|\varepsilon_{n,t_1}\varepsilon_{n,t_2}|K_{t_4t_5}D_f(w_{t_4})]a^{2(q+r)} \\
&= Cn^4m^2(n^3h^da^q)^{-2}O(a^{2(q+r)}h^q) = (n^2h^d)^{-1}O(m^2a^{2r}) = o((n^2h^d)^{-1}). \qquad \text{(A.8)}
\end{aligned}$$

Hence, we have $E_{1b} = o(nh^{d/2})$ by (A.7) and (A.8). For case (c), it has at most n^3m^3 terms and using Lemma C.2, we have $E_{1c} \leq Cn^3m^3(n^3h^da^q)^{-2}O(a^{q+r}h^d + h^{2d}) = o((nh^{d/2})^{-1})$.

Finally, we consider J_{n23}^{ij}.

$$\begin{aligned}
E|J_{n23}^{ij}| &\leq \frac{1}{2n(n-1)h^d}\sum_t\sum_{s\neq t}E\{\varepsilon_{n,t}^2(\hat{f}_{w_t}-f_{w_t})^2K_{ts}+\varepsilon_{n,s}^2(\hat{f}_{w_s}-f_{w_s})^2K_{ts}\} \\
&= n^{-1}\sum_t E[\varepsilon_{n,t}^2(\hat{f}_{w_s}-f_{w_s})^2\hat{f}_t] \\
&\leq n^{-1}\sum_t E[\varepsilon_{n,t}^2(\hat{f}_{w_s}-f_{w_s})^2f_t] + n^{-1}\sum_t E[\varepsilon_{n,t}^2(\hat{f}_{w_s}-f_{w_s})^2(\hat{f}_t-f_t)] \\
&= o((nh^{d/2})^{-1}), \qquad \text{(A.9)}
\end{aligned}$$

by (iii) and (iv) of Lemma B.

Lemma A3. Under the assumptions in Theorem 1, we have $J_{n3}^{ij} = o_p((nh^{d/2})^{-1})$.

Proof:

$$\begin{aligned}
|J_{n3}^{ij}| &\leq \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}|(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})\hat{f}_w(w_{n,s})c_{n,s}|K_{ts} \\
&\leq \frac{1}{2n}\sum_t(b_{ij}(w_{n,t})-\hat{b}_{ij}(w_{n,t}))^2\hat{f}_w^2(w_{n,t})f_t \\
&\quad +\frac{1}{2n}\sum_t(b_{ij}(w_{n,t})-\hat{b}_{ij}^2(w_{n,t}))^2\hat{f}_w^2(w_{n,t})(\hat{f}_t-f_t)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{2n}\sum_t c_{n,t}^2 \hat{f}_w^2(w_{n,t}) f_t \\
&+\frac{1}{2n}\sum_t c_{n,t}^2 \hat{f}_w^2(w_{n,t})(\hat{f}_t - f_t) \\
\equiv\ & J_{n31}^{ij} + J_{n32}^{ij} + J_{n33}^{ij} + J_{n34}^{ij}. \qquad \text{(A.10)}
\end{aligned}
$$

Using (ii) and (iii) of Lemma B, we have $J_{n31}^{ij} = o_p((nh^{d/2})^{-1})$ and $J_{n32}^{ij} = o_p((nh^{d/2})^{-1})$. For J_{n33}^{ij}, we have,

$$
\begin{aligned}
J_{n33}^{ij} &= \frac{1}{2n^3 a^{2q}} \sum_t \sum_{s \neq t} L^2\left(\frac{w_{n,s} - w_{n,t}}{a}\right) c_{n,t}^2 + \frac{1}{2n^3 a^{2q}} \sum\sum \sum_{s \neq s' \neq t} L\left(\frac{w_{n,s} - w_{n,t}}{a}\right) L\left(\frac{w_{n,s'} - w_{n,t}}{a}\right) \\
&\equiv J_{n331}^{ij} + J_{n332}^{ij}. \qquad \text{(A.11)}
\end{aligned}
$$

For J_{n331}^{ij}, using Hölder inequality for each integral in $c_{n,t}^2$, we have,

$$
\begin{aligned}
EJ_{n331}^{ij} &\leq \frac{C}{n^3 a^{2q} \triangle_n} \sum_t \sum_{s \neq t} E\Big\{ \int_{\triangle_n} [x_u^i - x_{n,t}^j]^2 \mu_i^2(x_u) du + \int_{\triangle_n} [x_u^i - x_{n,t}^i]^2 \mu_j^2(x_u) du \\
&\quad + \int_{\triangle_n} [b_{ij}(x_u) - b_{ij}(w_{n,t})]^2 du \Big\} \\
&= O((na^{2q})^{-1} \triangle_n) = o((nh^{d/2})^{-1}). \qquad \text{(A.12)}
\end{aligned}
$$

For J_{n332}^{ij}, we consider two different cases: (a) $min\{|s-t|, |s-s'| > m\}$ and (b) $min\{|s-t|, |s-t'|\} \leq m$. We use $J_{n332(a)}^{ij}$ and $J_{n332(b)}^{ij}$ to denote these two cases. For case (a), we have,

$$
\begin{aligned}
J_{n332(a)}^{ij} &\leq \frac{C}{n^3 a^{2q} \triangle_n} \sum_t \sum_{s \neq s' \neq t} E\Big\{ \Big[\int_{\triangle_n} (x_u^j - x_{n,t}^j)^2 \mu_i^2(x_u) du + \int_{\triangle_n} (x_u^i - x_{n,t}^i)^2 \mu_j^2(x_u) du \\
&\quad + \int_{\triangle_n} (b_{ij}(w_u) - b_{ij}(w_{n,t}))^2 \mu_i^2(x_u) du \Big] L\left(\frac{w_{n,s'} - w_{n,t}}{a}\right) \int L\left(\frac{w - w_{n,t}}{a}\right) dF_w(w) \\
&\quad + Cn^3 \beta_{n,m}^{\delta/(1+\delta)} \Big\} \\
&\leq \frac{C}{n^2 a^q \triangle_n} \sum_t \sum_{s' \neq t} E\Big\{ \Big[\int_{\triangle_n} (x_u^j - x_{n,t}^j)^2 \mu_i^2(x_u) du + \int_{\triangle_n} (x_u^i - x_{n,t}^i)^2 \mu_j^2(x_u) du \\
&\quad + \int_{\triangle_n} (b_{ij}(w_u) - b_{ij}(w_{n,t}))^2 \mu_i^2(x_u) du \Big] L\left(\frac{w_{n,s'} - w_{n,t}}{a}\right) + Cn^3 \beta_{n,m}^{\delta/(1+\delta)} \Big\} \\
&= O(a^{-q/2} \triangle_n) = o((nh^{d/2})^{-1}). \qquad \text{(A.13)}
\end{aligned}
$$

For case (b), we have at most mn^2 terms. Hence we have,

$$
\begin{aligned}
J^{ij}_{n332(b)} &\leq \frac{Cm}{na^{2q}\triangle_n}E\Big\{\Big[\int_{\triangle_n}(x^j_u - x^j_{n,t})^2\mu^2_i(x_u)du + \int_{\triangle_n}(x^i_u - x^i_{n,t})^2\mu^2_j(x_u)du \\
&\quad + \int_{\triangle_n}(b_{ij}(w_u) - b_{ij}(w_{n,t}))^2\mu^2_i(x_u)du\Big]L(\frac{w_s - w_{n,t}}{a})L(\frac{w_{s'} - w_{n,t}}{a})\Big\} \\
&= O(m(na^{2q})^{-1}\triangle_n) = o((nh^{d/2})^{-1}). \qquad \text{(A.14)}
\end{aligned}
$$

From (A.12) to (A.14), we have $J^{ij}_{n33} = o_p((nh^{d/2})^{-1})$. Given $J^{ij}_{n33} = o_p((nh^{d/2})^{-1})$ we can prove $J^{ij}_{n34} = o_p((nh^{d/2})^{-1})$ just as we prove $B_{n2} = o_p((nh^{d/2})^{-1})$ given that $B_{n1} = o_p((nh^{d/2})^{-1})$ in Lemma B.

Lemma A4: Under the assumptions in Theorem1, we have $J^{ij}_{n4} = o((nh^{d/2})^{-1})$.

Proof: using (v) and (vi) in Lemma B, we have,

$$
\begin{aligned}
J^{ij}_{n4} &= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}(b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})\hat{f}_w(w_{n,s})\varepsilon_{n,s}K_{ts} \\
&= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}(b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})f_w(w_{n,s})\varepsilon_{n,s}K_{ts} \\
&\quad + \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}(b_{ij}(w_{n,t}) - \hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})(\hat{f}_w(w_{n,s}) - f_w(w_{n,s}))\varepsilon_{n,s}K_{ts} \\
&= o_p((nh^{d/2})^{-1}). \qquad \text{(A.15)}
\end{aligned}
$$

Lemma A5: Under the assumptions in Theorem 1, we have $J^{ij}_{n5} = o((nh^{d/2}))$.

Proof: using (vii) and (viii) in Lemma (b), we have,

$$
\begin{aligned}
J^{ij}_{n5} &= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\varepsilon_{n,t}\hat{f}_w(w_{n,t})c_{n,s}\hat{f}_w(w_{n,s})K_{ts} \\
&= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\varepsilon_{n,t}f_w(w_{n,t})c_{n,s}\varepsilon_{n,t}\hat{f}_w(w_{n,s})K_{ts} \\
&+ \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\varepsilon_{n,t}(\hat{f}_w(w_{n,t}) - f_w(w_{n,t}))c_{n,s}\hat{f}_w(w_{n,s})K_{ts} \\
&= o_p((nh^{d/2})^{-1}) \qquad \text{(A.16)}
\end{aligned}
$$

Lemma A6: Under the assumptions in Theorem 1, we have $J^{ij}_{n6} = o((nh^{d/2})^{-1})$.

Proof: using $\hat{f}_t = \frac{1}{(n-1)h^d}\sum_{s\neq t} K_{ts}$ and the way as we have proven $J_{n33} = o_p((nh^{d/2})^{-1})$, we have,

$$
\begin{aligned}
|J_{n6}^{ij}| &\leq \frac{1}{2n(n-1)h^d}\sum\sum_{s\neq t} c_{n,t}^2 \hat{f}_w^2(w_{n,t})K_{ts} + \frac{1}{2n(n-1)h^d}\sum\sum_{t,s\neq t} c_{n,s}^2 \hat{f}_w^2(w_{n,s})K_{ts} \\
&= \frac{1}{n}\sum_t c_{n,t}^2 \hat{f}_w^2(w_{n,t}) f_t + \frac{1}{n}\sum_t c_{n,t}^2 \hat{f}_w^2(w_{n,t})(\hat{f}_t - f_t) = o_p((nh^{d/2})^{-1}). \qquad \text{(A.17)}
\end{aligned}
$$

Proof of Theorem 2:

Under LH, we have $\hat{u}_{n,t} = a_{ij}(x_{n,t}) - \hat{b}_{ij}(w_{n,t}) + c_{n,t} + \varepsilon_{n,t}$. Similar to the proof of Theorem 1, we can show that $J_{n3}^{ij} = o_p((nh^{d/2})^{-1}), J_{n4}^{ij} = o_p((nh^{d/2})^{-1}), J_{n5}^{ij} = o_p((nh^{d/2})^{-1})$, and $J_{n6}^{ij} = o_p((nh^{d/2})^{-1})$. Also, using the similar way as in the proof of Theorem 1, we can prove that under LH, $nh^{d/2}J_{n2}^{ij}$ converges to $N(0, v_{ij}^2)$. Consequently, $nh^{d/2}(\hat{I}_{nij} - J_{n1}^{ij})$ converges to $N(0, v_{ij}^2)$. We consider J_{n1}^{ij}.

$$
\begin{aligned}
J_{n1}^{ij} &= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{(a_{ij}(x_{n,t}) - \hat{b}_{ij}(w_{n,t}))\hat{f}_w(w_{n,t})(a_{ij}(x_{n,s}) - \hat{b}_{ij}(w_{n,s}))\hat{f}_w(w_{n,s})\Big\}K_{ts} \\
&= \frac{1}{n(n-1)h^d}\sum_t\sum_{s\neq t}\Big\{(a_{ij}(x_{n,t}) - b_{ij}(w_{n,t}))f_w(w_{n,t})(a_{ij}(x_{n,s}) - b_{ij}(w_{n,s}))f_w(w_{n,s})\Big\}K_{ts} \\
&\quad + o_p((nh^{d/2})^{-1}) \\
&= \frac{1}{n}\sum_{t=1}^{n}(a_{ij}(x_{n,t}) - b_{ij}(w_{n,t}))f_w(w_{n,t})\hat{E}[(a_{ij}(x_{n,t}) - b_{ij}(w_{n,t}))f_w(w_{n,t})|x_{n,t}]\hat{f}_t + o_p((nh^{d/2})^{-1}) \\
&= \gamma_n^2 E[\triangle^2(x_{n,t})f_w^2(w_{n,t})f(x_{n,t})] + o_p((nh^{d/2})^{-1}) \\
&= \gamma_n^2 \int \triangle^2(x) f_w^2(w) f^2(x)dx + o_p((nh^{d/2})^{-1}),
\end{aligned}
$$

which indicates that $Q_{nij} = nh^{d/2}\hat{I}_{nij}$ converges to $N(\int \triangle^2(x)f_w^2(w)f^2(x)dx, v_{ij}^2)$, i.e., $Pr[Q_{nij} > z_\alpha] \to 1 - \Phi(z_\alpha - \frac{1}{v_{ij}}\int \triangle^2(x)f_w^2(w)f^2(x)dx)$.

Appendix B. Lemma B used to prove the main results

Lemma B. Under the assumptions in Theorem 1, we have,

(i)$B_{n1} \equiv n^{-1}\sum_t E[(\hat{b}_{ij}(w_{n,t}) - b_{ij}(w_{n,t}))^2 \hat{f}^2_{w_t}] = o((nh^{d/2})^{-1})$.

(ii)$B_{n2} \equiv n^{-1}\sum_t E[(\hat{b}_{ij}(w_{n,t}) - b_{ij}(w_{n,t}))^2 \hat{f}^2_{w_t}(\hat{f}_t - f_t)] = o((nh^{d/2})^{-1})$.

(iii) $B_{n3} \equiv n^{-1}\sum_t E[\varepsilon^2_{n,t}(\hat{f}_{w_t} - f_{w_t})^2] = o((nh^{d/2})^{-1})$.

(iv)$B_{n4} \equiv n^{-1}\sum_t E[\varepsilon^2_{n,t}(\hat{f}_{w_t} - f_{w_t})^2(\hat{f}_t - f_t)] = o((nh^{d/2})^{-1})$.

(v) $B_{5n} \equiv (n(n-1)h^d)^{-1}\sum_t\sum_{s\neq t}[\hat{b}_{ij}(w_{n,t}) - b_{ij}(w_{n,t})]\hat{f}_{w_t}\varepsilon_{n,s}f_{w_s}K_{ts} = o_p((nh^{d/2})^{-1})$.

(vi) $B_{6n} \equiv (n(n-1)h^d)^{-1}\sum_t\sum_{s\neq t}[\hat{b}_{ij}(w_{n,t}) - b_{ij}(w_{n,t})]\hat{f}_{w_t}\varepsilon_{n,s}(\hat{f}_{w_s} - f_{w_s})K_{ts} = o_p((nh^{d/2})^{-1})$.

(vii) $B_{7n} \equiv (n(n-1)h^d)^{-1}\sum_t\sum_{s\neq t}\varepsilon_{n,t}f_{w_t}c_{n,s}\hat{f}_{w_s}K_{ts} = o_p((nh^{d/2})^{-1})$.

(viii)$B_{8n} \equiv (n(n-1)h^d)^{-1}\sum_t\sum_{s\neq t}\varepsilon_{n,t}(\hat{f}_{w_t} - f_{w_t})c_{n,s}\hat{f}_{w_s}K_{ts} = o_p((nh^{d/2})^{-1})$.

Proof of (i): denoting $\frac{[x^i_{n,s+1}-x^i_{n,s}][x^j_{n,s+1}-x^j_{n,s}]}{\triangle_n}$ by $X^{ij}_{ns}(\triangle_n)$, we have,

$$
\begin{aligned}
B_{n1} &= \frac{1}{n(n-1)^2a^{2q}}\Big\{\sum\sum_{s\neq t}E[(X^{ij}_{ns}(\triangle_n) - b_{ij}(w_{n,t}))^2L^2_{st}] \\
&\quad + \sum\sum_{s\neq t\neq l}E[(X^{ij}_{ns}(\triangle_n) - b_{ij}(w_{n,t}))L_{st}(X^{ij}_{nl}(\triangle_n) - b_{ij}(w_{n,t}))L_{lt}]\Big\} \\
&\equiv \frac{1}{n(n-1)^2a^{2q}}\Big\{B_{n11} + B_{n12}\Big\}
\end{aligned}
$$

Using Lemma C.2, Lemma C.3, and Lemma 5 in Robinson (1988), we have,

$$
\begin{aligned}
B_{n11} &\leq 2\sum\sum_{s\neq t}E[(b_{ij}(w_{n,s}) - b_{ij}(w_{n,t}))^2L^2_{st}] \\
&\quad + 2\sum\sum_{s\neq t}E[(X^{ij}_{ns}(\triangle_n) - b_{ij}(w_{n,s}))^2L^2_{st}] \\
&\leq 2\sum\sum_{s\neq t}\Big\{\int(b_{ij}(x) - b_{ij}(y))^2L^2(\frac{x-y}{a})dF_w(x)dF_w(y) \\
&\quad + E\Big\{[X^{ij}_{ns}(\triangle_n) - b_{ij}(w_{n,s})]^2\int L^2(\frac{x-w_{n,s}}{a})dF_w(x)\Big\} + C\beta^{\delta/(1+\delta)}_{n,s-t}\Big\}. \\
&\leq \sum\sum_{s\neq t}\Big\{Ca^{q+2} + Ca^qE[X^{ij}_{ns}(\triangle_n) - b_{ij}(w_{n,s})]^2\Big\} \\
&\quad + C\Big\{n\sum_{t=1}^{n-1}\beta^{\delta/(1+\delta)}_{n,\tau} - \sum_{t=1}^{n-1}\tau\beta^{\delta/(1+\delta)}_{n,\tau}\Big\} = O(n^2a^q) + O(n),
\end{aligned}
$$

because of $limi_{n\to\infty} n^{-1}\sum_{\tau=1}^{n-1}\tau\beta_{n,\tau}^{\delta/(1+\delta)} \leq \sum_{\tau=1}^{\infty}\beta_{n,\tau}^{\delta/(1+\delta)} < \infty$.

For B_{n12}, we first define $\Psi(z_r, z_s, z_t) \equiv [X_{ns}^{ij}(\triangle_n) - b_{ij}(w_{n,t})]L_{st}[X_{nr}^{ij}(\triangle_n) - b_{ij}(w_{n,t})]L_{rt}$. Since $\Psi(z_r, z_s, z_t)$ is not a symmetric function in its variables, we modify it to be a symmetric function in order to apply Lemma B.2 in Fan and Li (1999). We define,

$$\begin{aligned}
\bar{\Psi}(z_r, z_s, z_t) \;\equiv\;& [X_{ns}^{ij}(\triangle_n) - b_{ij}(w_{n,t})]L_{st}[X_{nr}^{ij}(\triangle_n) - b_{ij}(w_{n,t})]L_{rt} \\
&+[X_{nt}^{ij}(\triangle_n) - b_{ij}(w_{n,s})]L_{st}[X_{nr}^{ij}(\triangle_n) - b_{ij}(w_{n,s})]L_{rs} \\
&+[X_{ns}^{ij}(\triangle_n) - b_{ij}(w_{n,r})]L_{rs}[X_{nt}^{ij}(\triangle_n) - b_{ij}(w_{n,r})]L_{rt}
\end{aligned}$$

Note that since $\bar{\Psi}(z_r, z_s, z_t)$ has the same order as $\Psi(z_r, z_s, z_t)$, to simplify the proofs we will apply Lemma B.2 in Fan and Li (1999) directly to Ψ. Below we use $A \sim B$ to denote $A = O(B)$. Let M_{n12} and M_{n3} be defined as in the Lemma B.2 (Fan and Li, 1999), and $\eta = (1-\zeta^{-1})^{-1}$, where $\zeta > 2, 1 < \eta < 2$, then it is easy to see that,

$$\begin{aligned}
M_{n12} \;\sim\;& E\Big\{\int\Big\{|[X_{ns}^{ij}(\triangle_n) - b_{ij}(w_{n,t})]L_{st} \\
&\times[X_{nr}^{ij}(\triangle_n) - b_{ij}(w_{n,t})]L_{rt}\Big\}^{1+\delta} dF_w(w_r)\Big\} \\
\leq\;& \triangle_n^2 E\Big\{|X_{ns}^{ij}(\triangle_n)\triangle_n L_{st}|^{1+\delta} \\
&\times\int|X_{nr}^{ij}(\triangle_n)\triangle_n L_{rt}|^{1+\delta} dF_w(w_r)\Big\} \\
&+(\triangle_n)^{-1}E\Big|X_{ns}^{ij}(\triangle_n)\triangle_n b_{ij}(w_{n,t})L_{st}|^{1+\delta}\int L_{rt}^{1+\delta} dF_w(w_r)\Big\} \\
&+\triangle_n^{-1}E\Big\{[b_{ij}(w_{n,t})L_{st}]^{1+\delta}\int[X_{nr}^{ij}(\triangle_n)L_{rt}]^{1+\delta} dF_w(w_r)\Big\} \\
&+\triangle_n^{-1}E\Big\{[b_{ij}^2(w_{n,t})L_{st}]^{1+\delta}\int K_{rt}^{1+\delta} dF_w(w_r)\Big\} \\
\leq\;& \triangle_n^2 a^{q/\eta}E\Big\{|X_{ns}^{ij}(\triangle_n)\triangle_n L_{st}|[E[|X_{nr}^{ij}(\triangle_n)\triangle_n|]^{(1+\delta)\zeta}]^{1/\zeta} \\
&\times\int L(u)^{(1+\delta)\eta} f_w(w_{n,r}+au)du]^{1/\eta}\Big\} \\
&+\triangle_n^{-1}a^q E\Big\{|X_{ns}^{ij}(\triangle_n)b_{ij}(w_{n,t})L_{st}|^{1+\delta}\int[L(u)]^{1+\delta} f_w(w_{n,t}+au)du\Big\}
\end{aligned}$$

$$
\begin{aligned}
&+\triangle_n^{-1}a^{q/\eta}E\Big\{[b_{ij}^2(w_{n,t})L_{st}]^{1+\delta}[E|X_{nr}^{ij}(\triangle_n)\triangle_n|^{(1+\delta)\zeta}]]^{1/\zeta} \\
&\times[\int[L(u)]^{(1+\delta)\eta}f_w(w_{n,t}+au)du]^{1/\rho}\Big\} \\
&+\triangle_n^{-1}a^qE\Big\{[b_{ij}^2(w_{n,t})L_{st}]^{1+\delta}\int[L(u)]^{1+\delta}f_w(w_{n,t}+au)du\Big\} \\
\leq\ &\triangle_n^{-2}a^{2q/\eta}[E[|X_{ns}^{ij}\triangle_n|]^{(1+\delta)\zeta}]^{1/\zeta}[E[X_{nr}^{ij}\triangle_n|]^{(1+\delta)\zeta}]^{1/\zeta} \\
&+\triangle_n^{-1}a^{q(1+\eta)/\eta}[E[|X_{ns}^{ij}\triangle_n b_{ij}(w_{n,t})|]^{(1+\delta)\zeta}]^{1/\zeta} \\
&+\triangle_n^{-1}a^{2q/\eta}[E[b_{ij}(w_{n,t})]^{(1+\delta)\zeta}]^{1/\zeta}[E[|X_{ns}^{ij}(\triangle_n)\triangle_n|^{(1+\delta)\zeta}]]^{1/\zeta} \\
=\ &O(a^{2q/\eta}).
\end{aligned}
$$

Also it is easy to check that $M_{n3}=O(1)$. Let $\{\tilde{z}_t\}$ denotes an independent process that has the same marginal distribution as the dependent process $\{z_t\}$.

$$
\begin{aligned}
|E[\Psi(\tilde{z}_r,\tilde{z}_s,\tilde{z}_t)]| &= \Big|E\Big\{L_{st}E\Big\{[X_{ns}^{ij}(\triangle_n)-b_{ij}(w_{n,t})]|x_{n,s}\Big\} \\
&\quad\times L_{rt}E\Big\{[X_{nr}^{ij}(\triangle_n)-b_{ij}(w_{n,t})]|x_{n,t}\Big\}\Big\}\Big| \\
&\leq E\Big\{L_{rt}L_{st}[C\sqrt{\triangle_n}(1+\sqrt{1+x_{n,s}^2})+\Big|(b_{ij}(w_{n,s})-b_{ij}(w_{n,t})\Big|\Big\} \\
&\quad\times[C\sqrt{\triangle_n}(1+\sqrt{1+x_{n,r}^2})+\Big|(b_{ij}(w_{n,r})-b_{ij}(w_{n,t}))\Big|]\Big\} \\
&= o(a^{2q}\triangle_n)+O(a^{2q+v})\sqrt{\triangle_n}+O(a^{2(q+v)}).
\end{aligned}
$$

Thus, we have $B_{n1}=(n(n-1)^2a^{2q})^{-1}\{B_{n11}+B_{n12}\}=(n(n-1)^2a^{2q})^{-1}[O(n^2a^q)+O(n^3a^{2q}\triangle_n)+O(n^3a^{2q+v}\sqrt{\triangle_n})+O(n^3a^{2(q+v)})+O(n^2a^{2q/\eta})+O(n))]=O(a^2(na^q)^{-1})+O(\triangle_n)+O(a^v\sqrt{\triangle_n})+O(a^{2v})+O((na^{2q})^{-1})+O(n^{-2}a^{-2q})=(nh^{d/2})^{-1}[O(a^2h^{d/2}a^{-q})+O(Th^{d/2})+O(\sqrt{n}a^vh^{d/2}\sqrt{T})+O(na^{2v}h^{d/2})+O(h^{d/2}a^{2q/\eta-2q})+O((na^q)^{-1}(h^{d/2}a^{-q}))]=o((nh^{d/2})^{-1})$.

Proof of (ii): we have,

$$
\begin{aligned}
B_{n2} &= (n^3a^{2q})^{-1}\sum_t\sum_{s\neq t}\sum_{s'\neq t}E\Big\{[b_{ij}(w_{n,t})-X_{ns}^{ij}(\triangle_n)] \\
&\quad\times[b_{ij}(w_{n,t})-X_{ns'}^{ij}(\triangle_n)]L_{st}L_{s't}[n^{-1}\sum_{t'\neq t}(h^{-d}K_{t't}-f_t)]\Big\}.
\end{aligned}
$$

We consider two different cases for B_{n2}: (a) $min\{|t'-t|,|t'-s|,|t'-s'|\} > m$ and (b) $min\{|t'-t|,|t'-s|,|t'--s'|\} \leq m$. $B_{n2(a)}$ and $B_{n2(b)}$ are used to denote these two cases. Using Lemma C.2, we have,

$$\begin{aligned}
B_{n2(a)} &\leq (n^3a^{2q})^{-1}\sum_t\sum_{s\neq t}\sum_{s'\neq t'}\Big|E\Big\{[b_{ij}(w_{n,t})-X^{ij}_{ns}(\triangle_n)]L_{st} \\
&\quad \times[b_{ij}(w_{n,t})-X^{ij}_{ns'}(\triangle_n)]L_{s't}\int[h^{-d}K(\frac{x-x_{n,t}}{h})-f(x)]dF(x)\Big\}\Big| \\
&\quad +Cn^3\beta^{\delta/(1+\delta)}_{n,m} \\
&\leq Cn^{-1}\sum E[(\hat{b}_{ij}(w_{n,t})-b_{ij}(w_{n,t}))^2\hat{f}_w]+Cn^3\beta^{\delta/(1+\delta)}_{n,m}=o((nh^{d/2})^{-1}).
\end{aligned}$$

For case (b), assuming that $|t'-t|\leq m$, we have $|n^{-1}\sum_{t'\neq t}(h^{-d}L_{tt'}-f_t)|\leq C(nh^d)^{-1}m$ uniformly for any t. Thus we have,

$$\begin{aligned}
B_{n2(b)} &\leq (n^3a^{2q})^{-1}\sum_t\sum_{s\neq t}\sum_{s'\neq t'}E\Big\{\Big|[b_{ij}(w_{n,t})-X^{ij}_{ns}(\triangle_n)]L_{st} \\
&\quad \times[b_{ij}(w_{n,t})-X^{ij}_{ns}(\triangle_n)]L_{s't}n^{-1}\sum_{t'\neq t}[n^{-1}K(\frac{x-x_{n,t'}}{h})-f_{w_t}]\Big|\Big\} \\
&\leq Cm(nh^d)^{-1}n^{-1}\sum_t E[(\hat{b}_{ij}(w_{n,t})-b_{ij}(w_{n,t}))^2\hat{f}^2_{w_t}]=o((nh^{d/2})^{-1}).
\end{aligned}$$

Thus, $B_{n2}=B_{n2(a)}+B_{n2(b)}=o((nh^{d/2})-1)$.

Proof of (iii): we have,

$$\begin{aligned}
B_{n3} &= (n^3h^{2q})^{-1}\sum_t E[\varepsilon^2_{n,t}\sum^2_{s\neq t}\sum_{s\neq t}(L_{st}-a^qf_{w_t})\sum_{t'\neq t}(L_{t't}-a^qf_{w_t})] \\
&= \Big\{\sum\sum_{s\neq t}E[\varepsilon^2_{n,t}(L_{st}-a^qf_{w_t})^2]+\sum\sum\sum_{t\neq s\neq t'}E[\varepsilon^2_{n,t}(L_{st}-a^qf_{w_t})(L_{tt'}-a^qf_{w_t})]\Big\} \\
&\equiv (n^3a^{2q})^{-1}\{C_{n1}+C_{n2}\}.
\end{aligned}$$

By Lemma C.2, we have,

$$C_{n1} \leq \sum\sum_{t\neq s}\Big\{E\Big\{\varepsilon^2_{n,t}\int[L(\frac{w_{n,s}-w_{n,t}}{a}-a^qf_{w_t})]^2dF_w(w_{n,s})\Big\}+C\beta^{\delta/(1+\delta)}_{n,|s-t|}\Big\}$$

$$
\begin{aligned}
&\le \sum\sum_{t\neq s} Ca^q E\varepsilon_{n,t}^2 + C\Big\{\sum_{\tau=1}^{n-1}\beta_{n,\tau}^{\delta/(1+\delta)} - \sum_{\tau=1}^{n-1}\tau\beta_{n,\tau}^{\delta/(1+\delta)}\Big\} \\
&= O(n^2a^q)+O(n).
\end{aligned}
$$

By similar argument as Lemma A.1 in Fan and Li (1999), we have, $C_{n2} = O(n^3a^{2(q+v)}) + O(n^2a^{3q/(2+2\delta)}) + O(n)$, where $\delta \le 1/2$. Hence, we have that $B_{n3} = (n^3a^{2q})^{-1}[O(n^2a^{q+v}) + O(n^3a^{2(q+v)}) + O(n^2a^{3q/(2+2\delta)}) + O(n)] = o((nh^{d/2})^{-1})$. Thus, $B_{n3} = o(nh^{d/2-1})$.

Proof of (iv): given the proof of (iii),(iv) can be proven just as we prove (ii) given the proof of (i).

Proof of (v): B_{n5} can be rewritten as,

$$
B_{n5} = \frac{1}{n(n-1)^2h^d}\sum_{t_1}\sum_{t_2\neq t_1}\sum_{t_3\neq t_1}[b_{ij}(w_{n,t_1}) - X_{nt_3}^{ij}(\triangle_n)]\varepsilon_{n,t_2}f_w(w_{n,t_2})L_{t_1,t_3}K_{t_1,t_2}.
$$

Consider the second moment of B_{n5}, we have,

$$
\begin{aligned}
B_{n5}^2 &= (n(n-1)^2a^qh^d)^{-2}\sum\sum\sum_{t_2\neq t_1, t_3\neq t_1}\sum\sum\sum_{t_5\neq t_4, t_6\neq t_4} E\Big\{[b_{ij}(w_{n,t_1}) - X_{nt_3}^{ij}(\triangle_n)] \\
&\quad \times [b_{ij}(w_{n,t_4}) - X_{nt_6}^{ij}(\triangle_n)]\varepsilon_{n,t_2}\varepsilon_{n,t_4}f_w(w_{n,t_2})f_w(w_{n,t_5})L_{t_1t_3}L_{t_4t_6}L_{t_1t_2}L_{t_4t_5}\Big\}.
\end{aligned}
$$

We consider four different cases: (a) for all i's, $|t_i - t_j > m|$ for all $j \neq i$; (b) for exactly four different i's, $|t_i - t_j| > m$ for all $j \neq i$; (c) for exactly three different $i's$, $t_i - t_j| > m$ for all $j \neq i$; (d) all the other remaining cases. We use $B_{n5(a)}, B_{n5(b)}, B_{n5(c)}$, and $B_{n5(d)}$ to denote the four cases.

By Lemma C.2, we have $EB_{n5(a)} \le 0 + Cn^6\beta_{n,m}^{\delta/(1+\delta)} = o((n^2h^d)^{-1})$. For case (b), we just need to consider the case $|t_2 - t_1| \le m$. Let t_{-i} denote all t_i's with $j \neq i$. Using Lemma C2 for $i = 1,3,4,6$, we have,

$$
\begin{aligned}
EB_{n5(b)} &\le (n(n-1)^2a^qh^d)^{-2}\sum\sum\sum\sum\sum_{|t_i - t_{-i}|>m, i=1,3,4,6, |t_2-t_5|\le m} E\Big\{\varepsilon_{n,t_2}\varepsilon_{n,t_5}f_{w_{t_2}}f_{w_{t_5}} \\
&\times \; E_{t_1}\Big[E_{t_3}\Big[(b_{ij}(w_{n,t_1}) - b_{ij}(w_{n,t_3}))L_{t_1t_3} - \Big[\triangle_n^{-2}X_{nt_3}^{ij}(\triangle_n) - b_{ij}(w_{n,t_3})\Big]L_{t_1t_3}\Big]K_{t_1t_2}\Big\}
\end{aligned}
$$

$$\times \quad E_{t_4}\Big\{E_{t_6}\Big[(b_{ij}(w_{n,t_4})-b_{ij}(w_{n,t_6}))L_{t_4t_6}-\Big[\triangle_n^{-2}X_{nt_6}^{ij}(\triangle_n)-b_{ij}(w_{n,t_6})\Big]L_{t_4t_6}\Big]K_{t_4t_5}\Big\}\Big\}$$
$$+Cn^6\beta_{n,m}^{\delta/(1+\delta)}.$$
$$\leq \quad Cn^2(n^3a^qh^d)^{-2}\sum\sum\sum\sum_{t_1\neq t_4,|t_2-t_5|\leq m} E\varepsilon_{n,t_2}\varepsilon_{n,t_3}E_{t_1}\Big\{\Big[a^{q+v}D_{ij}(w_{n,t_1})+\triangle_n^{1/2}a^{q/2}$$
$$\times[\int(1+|x_{n,t_3}|^2)(\mu_i^2+\mu_j^2(x_{n,t_3})+1)dF(x_{n,t_3})]^{1/2}\Big]K_{t_1t_2}\Big\}E_{t_4}\Big\{\Big[a^{q+v}D_{ij}(w_{n,t_4})$$
$$+\triangle_n^{1/2}a^{q/2}[\int(1+|x_{n,t_3}|^2)(\mu_i^2+\mu_j^2(x_{n,t_3})+1)dF(x_{n,t_3})]^{1/2}\Big]K_{t_4t_5}\Big\}+Cn^6\beta_{n,m}^{\delta/(1+\delta)}$$
$$\leq \quad Cn^4(n^3a^qh^d)^{-2}\sum\sum_{|t_2-t_3|\leq m} h^{2d}E\Big[\varepsilon_{n,t_2}(a^{q+v}D_{ij}(w_{n,t_2})+\triangle_n^{1/2}a^{q/2})$$
$$\times\varepsilon_{n,t_5}(a^{q+v}D_{ij}(w_{n,t_5})+\triangle_n^{1/2}a^{q/2})\Big]+Cn^6\beta_{n,m}^{\delta/(1+\delta)}$$
$$= \quad O(n^{-6}a^{-2q}h^{-2d}(n^5mh^{2d}(a^{2(q+v)}+\triangle_na^q+\triangle_n^{1/2}a^{3q/2+v})))=o((n^2h^d)^{-1})$$

For case (c), we only need to consider the case $|t_2-t_3|\leq m$, and $|t_2-t_3|\leq m$ for exactly one $i\in\{1,3,4,6\}$. By symmetry we only consider $i=1$ and $i=3$. For $i=1$, using Lemma C.2 three times, we have,

$$EB_{n5(c)} \quad = \quad (n^3h^da^q)^{-2}\sum\sum\sum\sum\sum\sum_{|t_2-t_5|\leq m,|t_1-t_2|\leq m,|t_i-t_{-i}|>m,i=3,4,6} E\Big\{\varepsilon_{n,t_2}\varepsilon_{n,t_3}f_w(w_{n,t_2})f_w(w_{n,t_5})$$
$$\times K_{t_1t_2}E_{t_3}\Big\{\Big[(b_{ij}(w_{n,t_1})-b_{ij}(w_{n,t_3}))L_{t_1t_3}-\Big[\triangle_n^{-2}X_{nt_3}^{ij}(\triangle_n)-b_{ij}(w_{n,t_3})\Big]L_{t_1t_3}\Big]\Big\}$$
$$\times E_{t_4}\Big\{E_{t_6}\Big\{\Big[(b_{ij}(w_{n,t_4})-b_{ij}(w_{n,t_6}))L_{t_4t_6}-\Big[\triangle_n^{-2}X_{nt_6}^{ij}(\triangle_n)-b_{ij}(w_{n,t_6})\Big]L_{t_4t_6}\Big]K_{t_4t_5}\Big\}\Big\}\Big\}$$
$$+Cn^6\beta_{n,m}^{\delta/(1+\delta)}$$
$$\leq \quad Cn^2(n^3h^da^q)^{-2}\sum\sum\sum\sum_{|t_1-t_2|\leq m,|t_2-t_5|\leq m,t_4\neq t_5} E\Big\{\varepsilon_{n,t_2}\varepsilon_{n,t_5}K_{t_1t_2}$$
$$\times\Big[a^{q+v}D_{ij}(w_{n,t_1})+a^{q/2}\triangle_n^{1/2}\Big]E_{t_4}\Big\{\Big[a^{q+v}D_{ij}(w_{n,t_4})+a^{q/2}\triangle_n^{1/2}\Big]K_{t_4t_5}\Big\}\Big\}+Cn^6\beta_{n,m}^{\delta/(1+\delta)}$$
$$\leq \quad Cn^3h^d(n^3h^da^q)^{-2}\sum\sum\sum_{|t_1-t_2|\leq m,|t_2-t_5|\leq m} E\Big\{\varepsilon_{n,t_2}\varepsilon_{n,t_5K_{t_1t_2}\left[a^{q+v}D_{ij}(w_{n,t_1})+a^{q/2}\triangle_n^{1/2}\right]}$$
$$\times\Big[a^{q+v}D_{ij}(w_{n,t_5})+a^{q/2}\triangle_n^{1/2}\Big]\Big\}+Cn^6\beta_{n,m}^{\delta/(1+\delta)}$$
$$= \quad O(m^2n^4h^d(n^3h^da^q)^{-2}(a^{2(q+v)}+a^q\triangle_n+a^{3q/2+v}\triangle_n^{1/2}))$$

$$= \quad o((n^2h^d)^{-1}).$$

For $i=3$, using Lemma C.2 three times, we have,

$$\begin{aligned}
EB_{n5(c)} &= (n^3h^da^q)^{-2}\sum\sum\sum\sum\sum\sum_{|t_2-t_3|\le m,|t_2-t_5|\le m,|t_i-t_{-i}|>m,i=1,4,6} E\Big\{\varepsilon_{n,t_2}\varepsilon_{n,t_5}f_w(w_{n,t_2})f_w(w_{n,t_5}) \\
&\quad \times E_{t_1}\Big\{\Big[(b_{ij}(w_{n,t_1})-b_{ij}(w_{n,t_3}))L_{t_1t_3}-\Big[\triangle_n^{-2}X^{ij}_{nt_3}(\triangle_n)-b_{ij}(w_{n,t_3})\Big]L_{t_1t_3}\Big]K_{t_1t_2}\Big\} \\
&\quad \times E_{t_4}\Big\{E_{t_6}\Big\{\Big[(b_{ij}(w_{n,t_4})-b_{ij}(w_{n,t_6}))L_{t_4t_6}-\Big[\triangle_n^{-2}X^{ij}_{nt_6}(\triangle_n)-b_{ij}(w_{n,t_6})\Big]L_{t_4t_6}\Big]K_{t_4t_5}\Big\}\Big\}\Big\} \\
&\quad +Cn^6\beta_{n,m}^{\delta/(1+\delta)} \\
&\le Cn^3h^{2d}(n^3h^da^q)^{-2}\sum\sum\sum_{|t_2-t_3|\le m,|t_2-t_5|\le m} E\Big\{\varepsilon_{n,t_2}\varepsilon_{n,t_5}f_w(w_{n,t_2})f_w(w_{n,t_5}) \\
&\quad \times\Big[b_{ij}(w_{n,t_2})+b_{ij}(w_{n,t_3})+\triangle_n^{-1}\Big|\triangle_nX^{ij}_{nt_6}(\triangle_n)-b_{ij}(w_{n,t_3})\Big|\Big] \\
&\quad \times\Big[a^{q+v}D_{ij}(w_{n,t_5})+a^{q/2}\triangle_n^{1/2}\Big]\Big\}+Cn^6\beta_{n,m}^{\delta/(1+\delta)} \\
&= O(n^4m^2h^{2d}(n^3h^da^q)^{-2})(a^{q+v}+a^{q/2}\triangle_n^{1/2}))+Cn^6\beta_{n,m}^{\delta/(1+\delta)}=o((n^2h^d)^{-1}).
\end{aligned}$$

For case (d), we have at most n^3m^3 terms, we have $EB_{n5(d)}=O(n^3m^3(n^3h^da^q)^{-2}(a^{2q}+h^{2d})=o((n^2h^d)^{-1})$. Finally, using Chebychev's inequality,we have, $B_{n5}=o_p((nh^{d/2})^{-1})$.

Proof of (vi): the proof of (vi) follows the same steps as the proof of B_{n2} above except that we need to cite B_{n5} instead of B_{n1}.

Proof of (vii): B_{n7} can be written as $B_{n7}=(a^qh^dn(n-1)^2)^{-1}\sum\sum\sum_{t_2\neq t_1,t_3\neq t_2}\varepsilon_{n,t_1}f_w(w_{n,t_1})c_{n,t_2}L_{t_2t_3}K_{t_1t_2}$. The second moment of B_{n7} is,

$$E(B_{n7}^2)=(a^qh^dn(n-1)^2)^{-2}\sum\sum\sum_{t_1,t_{21},t_3\neq t_1}\sum\sum\sum_{t_4,t_5\neq t_4,t_6\neq t_4}E\Big[\varepsilon_{n,t_1}f_{w_{n,t_1}}c_{n,t_2}L_{t_2t_3}K_{t_1t_2}\varepsilon_{n,t_4}c_{n,t_5}f_{w_{n,t_4}}L_{t_5t_6}K_{t_4t_5}\Big]$$

We consider four different cases: (a) for i's, $|t_i-t_j|>m$ for all $j\neq i$; (b) for exactly four differnt i's , $|t_i-t_j|>m$ for all $j\neq i$; (c) for exactly three different i's, $|t_i-t_j|>m$ for all $j\neq i$; (d) all the other remaining cases. We use $E(B^2_{n7(j)}), j=a,b,c,d$, to express these four cases.

For case (a), using Lemma C.2, we have $E(B^2_{n7(a)}) \leq 0 + Cn^6\beta_{n,m}^{\delta/(1+\delta)}$. For case ($b$), we only need to consider the case $|t_1 - t_4| \leq m$. Otherwise, it must be either t_1 or t_4 is more than m periods away from other indices, and $E(B_{n7(b)})$ is bounded by $O(n^6\beta_{n,m}^{\delta/(1+\delta)})$. Given $|t_1 - t_4| \leq m$, we know that t_i is at least m period away from any other indices, for $i = 2,3,5,6$. Using Lemma C.2 four times, we have,

$$
\begin{aligned}
EB_{n7(b)} &\leq (n^3h^da^q)^{-2}\sum\sum\sum\sum\sum\sum_{|t_i-t_j|>m,|t_1-t_4|\leq m,i=2,3,5,6,i\neq j}\Big|E\varepsilon_{n,t_1}\varepsilon_{n,t_4}f_{w_{n,t_1}}f_{w_{n,t_4}}E_{t_2}\Big\{c_{n,t_2} \\
&\quad \times E_{t_3}(L_{t_2t_3})K_{t_1t_2}\Big\}E_{t_5}\Big\{c_{n,t_5}E_{t_6}(L_{t_5t_6})K_{t_4t_5}\Big\}\Big| + Cn^6\beta_{n,m}^{\delta/(1+\delta)} \\
&\leq Cn^2a^{2q}(n^3h^da^q)^{-2}\sum\sum\sum\sum_{|t_i-t_j|>m,|t_1-t_4|\leq m,i=2,5} E\Big\{|\varepsilon_{n,t_1}\varepsilon_{n,t_4}|E_{t_2}\Big[|c_{n,t_2}|K_{t_1t_2}\Big] \\
&\quad \times E_{t_5}\Big[|c_{n,t_5}|K_{t_4t_5}\Big]\Big\} + Cn^6\beta_{n,m}^{\delta/(1+\delta)} \\
&\leq Cn^2a^{2q}(n^3h^da^q)^{-2}\sum\sum\sum\sum_{|t_i-t_j|>m,|t_1-t_4|\leq m,i=2,5} E\Big\{|\varepsilon_{n,t_1}\varepsilon_{n,t_4}|\Big[E_{t_2}|c_{n,t_2}|^\xi\Big]^{1/\xi}\Big[E_{t_2}K^\eta_{t_1t_2}\Big]^{1/\eta} \\
&\quad \times\Big[E_{t_5}|c_{n,t_5}|^\xi\Big]^{1/\xi}\Big[E_{t_5}K^\eta_{t_4t_5}\Big]^{1/\eta}\Big\} + Cn^6\beta_{n,m}^{\delta/(1+\delta)} \\
&\leq Cn^4a^{2q}h^{2d/\eta}\triangle_n(n^3h^da^q)^{-2}\sum\sum_{|t_2-t_4|\leq m} E|\varepsilon_{n,t_1}\varepsilon_{n,t_{1=4}}| \\
&= O(mTn^{-2}h^{2q/\eta-2d}) = o((n^2h^d)^{-1}),
\end{aligned}
$$

where ξ is slightly larger than 2 and $\eta = (1-\xi^{-1})^{-1}$ $(1 < \eta < 2)$. For case (c), we only need to consider $|t_1 - t_4| \leq m$ and either $|t_1 - t_i| \leq m$ or $|t_4 - t_i| \leq m$ for exactly one $i \in \{2,3,4,5\}$. By symmetry we only consider $i = 2$ and $i = 3$. For $i = 2$, we have,

$$
\begin{aligned}
EB_{n7(c)} &\leq (n^3h^da^q)^{-2}\sum\sum\sum\sum\sum\sum_{|t_1-t_4|\leq m,|t_1-t_2|\leq m,|t_i-t_j|\geq m,i\neq 3,5,6,i\neq j}\Big|E\Big\{\varepsilon_{n,t_1}\varepsilon_{n,t_4}f_{w_{n,t_1}}f_{w_{n,t_4}}K_{t_1t_2} \\
&\quad \times E_{t_3}\Big[c_{n,t_2}L_{t_2t_3}\Big]E_{t_3}\Big\{c_{n,t_5}E_{t_6}(L_{t_5t_6})K_{t_4t_5}\Big\}\Big\}\Big| + Cn^6\beta_{n,m}^{\delta/(1+\delta)} \\
&\leq Cn^2(n^3a^qh^d)^{-2}a^{2q}\sum\sum\sum\sum_{|t_1-t_4|\leq m,|t_1-t_2|\leq m,t_{45}} E\Big\{|\varepsilon_{n,t_1}\varepsilon_{n,t_4}c_{n,t_2}K_{t_1t_2}|E_{t_5}[|c_{n,t_5}K_{t_4t_5}|]\Big\} \\
&\quad + Cn^6\beta_{n,m}^{\delta/(1+\delta)}
\end{aligned}
$$

$$\leq \ Cn^2({}^3a^qh^d)^{-2}a^{2q}\triangle_n^{1/2}h^{1/2}\sum\sum\sum\sum_{|t_1-t_4|\leq m,|t_1-t_2|\leq m} E\left\{|\varepsilon_{n,t_1}\varepsilon_{n,t_4}c_{n,t_2}K_{t_1t_2}|\right\}+Cn^6\beta_{n,m}^{\delta/(1+\delta)}$$
$$= \ O(n^{-2}h^{-d}(m^2T/n))+Cn^6\beta_{n,m}^{\delta/(1+\delta)}=o((n^2h^d)^{-1}).$$

Similarly, we can get $EB_{n7(c)}=o((n^2h^d)^{-2})$ for $i=3$. For case (d), since $EB_{n7(d)}$ has at most n^3m^3 terms, we have $EB_{n7(d)}^2\leq Cm^3n^3(n^3h^da^q)^{-2}(h^{2d}+\triangle_n a^{2q})=o((n^2h^d)^{-1})$.

Proof of (viii): we can prove (viii) by following the similar steps as we prove (vi).

Appendix C. Technical Lemmas

Lemma C.1. *Let $\{X_{n,t},t\leq n\}$ be strictly stationary and satisfy the absolutely regular condition with mixing coefficient $\beta_{n,\tau}$ for each row of $\{X_{n,t}\}$. Define a degenerate U-statistic of the triangular arrays of $\{X_{n,t}\}$ as $U_n=\sum\sum_{1\leq s\leq t\leq n}H_n(X_{n,t},X_{n,s})$, where $H_n(\cdot,\cdot)$ depends on n and satisfies $\int H_n(x,y)dF_n(x)=0$ for all y and $F_n(\cdot)$ is the marginal distribution function of $X_{n,t}$. If Assumption (A1), Assumption (A2), and Assumption (A3) in Fan and Li (1999) are satisfied by each row in $X_{n,t}$, then we have $\sqrt{2}U_n/(n\sigma_n)\to N(0,1)$ in distribution as $n\to\infty$, where $\sigma_n^2=E[H_n^2(X_{n,1},X_{n,2})]$.*

Proof: The proof of Lemma C.1 is omitted because one can prove Lemma C.1 by straightforwardly following the way as the proof of Theorem 2.1 in Fan and Li (1999). The central limit theorem (Theorem 2.2 in Dvoretzky, 1972) for double arrays of random variables is used to prove Lemma C.1.

Lemma C.2. *Let $\xi_{n,1},\ldots,\xi_{n,n}$ be random vectors taking values in R^d satisfying an absolute regularity condition with the mixing coefficient $\beta_{n,\tau}$, where $\beta_{n,\tau}\equiv Sup_{s+\tau\leq n}E[Sup_{A\in M_{s+\tau,n}}\{Pr(A|M_{1,s}^n(n))-Pr(A)\}]$, where $M_{s,t}^n$ is the σ algebra generated by $(\xi_{n,s},\ldots,\xi_{n,t})$. Let $h(y_1,\ldots,y_k)$ be a Boreal measurable function such that , for some $\delta>0$,*

$\bar{M}_n=max\left\{E[|h(\xi_{n,i_1},\ldots,\xi_{n,i_k})|^{1+\delta}],E[E_{\xi_{n,i_1},\ldots,\xi_{n,i_j}}|h(\xi_{n,i_1},\ldots,\xi_{n,i_k})|^{1+\delta}]\right\}$ exists. Then,

$$|E[h(\xi_{n,i_1},...,\xi_{n,i_k})] - E[E_{\xi_{n,i_1},...,\xi_{n,i_j}} h(\xi_{n,i_1},...,\xi_{n,i_k})]| \leq 4\bar{M}_n^{1/(1+\delta)} \beta_{n,\tau}^{\delta/(1+\delta)},$$

where, $E_{\xi_{n,i_1},...,\xi_{n,i_j}} h(\xi_{n,i_1},...,\xi_{n,i_k}) = \int h(y_{i_1},...,y_{i_k}) = \int h(y_{i_1},...,y_{i_j},\xi_{n,i_1,...,\xi_{n,i_k}}) dF_n^{i_1,...,i_j}(y_{i_1},...,y_{i_j})$, $\tau = i_{j+1} - i_j$, and $F_n^{i_1,...,i_j}(y_{i_1},...,y_{i_j})$ is the distribution function of random vectors $\xi_{n,i_1},...,\xi_{n,i_j}$ and $i_1 < i_2 < ... < i_j$.

Proof: the proof of Lemma C.2 is omiited because it is the straightforward extension of Lemma 1 in Yoshihara (1976) to a triangular array case.

Lemma C.3. $E[|x_{t_0}|^{2l}] < \infty$ *for some positive integer. For* $t' \in (t_0 + j\triangle_n, T)$*, we have* $E^j[|x_{t'} - x_{n,j}|^{2l}] \leq D_n[1 + |x_{n,j}|^{2l}][t' - t_0 - j\triangle_n]^l$*, where* $D_n = 2^{2(2l-1)} C_D^{2l} e^{2l(2l+1)C_D^2(t'-t_0-j\triangle_n)}[(t'-t_0-j\triangle_n)^l + (l(2l-1))^l]$.

Proof: Lemma C.3 directly follows Theorem 2.2 of Friedman (1975) by replacing the unconditional expectation by conditional expectations.