Singapore Management University

Institutional Knowledge at Singapore Management University

Research Collection School Of Economics School of Economics

1-2012

Conditional Independence Specification Testing for Dependent Processes with Local Polynomial Quantile Regression

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Halbert L. WHITE
University of California - San Diego

Follow this and additional works at: https://ink.library.smu.edu.sg/soe_research

Part of the Econometrics Commons

Citation

Su, Liangjun and Halbert L. White. 2012. "Conditional Independence Specification Testing for Dependent Processes with Local Polynomial Quantile Regression." In *Essays in Honor of Jerry Hausman*, Advances in Econometrics, vol. 29, edited by Badi H. Baltagi, R. Carter Hill, Whitney K. Newey, and Halbert L. White, 355-434. Bingley: Emerald.

This Book Chapter is brought to you for free and open access by the School of Economics at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection School Of Economics by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email cherylds@smu.edu.sg.

Conditional Independence Specification Testing for Dependent Processes with Local Polynomial Quantile Regression*

Liangjun Su[a] and Halbert White[b]
[a] School of Economics, Singapore Management University, Singapore, 178903
[b] Department of Economics, UCSD, La Jolla, CA 92093-0508

November 12, 2011

Abstract

We provide straightforward new nonparametric methods for testing conditional independence using local polynomial quantile regression, allowing weakly dependent data. Inspired by Hausman's (1978) specification testing ideas, our methods essentially compare two collections of estimators that converge to the same limits under correct specification (conditional independence) and that diverge under the alternative. To establish the properties of our estimators, we generalize the existing nonparametric quantile literature not only by allowing for dependent heterogeneous data but also by establishing a weak consistency rate for the local Bahadur representation that is uniform in both the conditioning variables and the quantile index. We also show that, despite our nonparametric approach, our tests can detect local alternatives to conditional independence that decay to zero at the parametric rate. Our approach gives the first nonparametric tests for time-series conditional independence that can detect local alternatives at the parametric rate. Monte Carlo simulations suggest that our tests perform well in finite samples. Our tests have a variety of uses in applications, such as testing conditional exogeneity or Granger non-causality.

Key Words: Conditional independence, Empirical process, Granger causality, Local polynomial, Quantile regression, Specification test, Uniform local Bahadur representation.

1 Introduction

Hausman's (1978) seminal paper on specification testing opened the way to a broad array of methods for assessing the validity of econometric models and their resulting insights. The fundamental idea of comparing two estimators, both consistent under correct specification, but divergent under misspecification applies not only to detecting incorrect parametric functional form for conditional means, variances, or other aspects of the conditional distribution of a variable of interest, but also to detecting failures of exogeneity – the stochastic orthogonality condition between observable and unobservable drivers of

*Address correspondence to: Halbert White, Department of Economics, UCSD, La Jolla, CA 92093-0508, USA. Phone: +1 858 534-3502; e-mail: hwhite@weber.ucsd.edu. The first author gratefully acknowledges the financial support from a research grant (Grant number: C244/MSS8E004) from Singapore Management University.

the dependent variable of interest ensuring that structural features of interest can be recovered from observable data.

Although functional form misspecification can be considerably mitigated using nonparametric methods, exogeneity in one form or another remains a crucial assumption; its failure results in estimators being largely uninformative about structural objects of interest (see, e.g., White and Chalak, 2012). There remains a clear need to develop convenient nonparametric methods for exogeneity specification testing; Hausman's (1978) approach provides a valuable foundation on which to construct such tests.

Over the years since Hausman's paper appeared, various alternatives to strict exogeneity (observable causes (X) independent of unobservable causes (U)) useful for identifying economic structure of interest have emerged. One important alternative is that introduced by Barnow, Cain, and Goldberger (1980), termed "selection on observables," namely that observable causes are independent of unobservable causes, conditional on some further observables, say W. We write this conditional independence as $X \perp U \mid W$, following Dawid (1979). This condition plays a key role in the identification and estimation of treatment effects (White and Lu, 2011; White and Chalak, 2012). Similar conditions play a key role in recovering structural features in many other contexts, as catalogued by Chalak and White (2011).

Because U is unobservable and, in the general case, not estimable, indirect methods for testing $X \perp U \mid W$ have been developed, based on the fact that, under additional plausible assumptions, $X \perp U \mid W$ implies certain conditional independence relations among observables, e.g., $X \perp S \mid W$, where S is observable (White and Chalak, 2010). Conditional independence also plays a key role in other important contexts. For example, tests of Granger non-causality in distribution (Granger, 1969; Granger and Newbold, 1986) are tests of conditional independence among observables. And, as White and Lu (2011) show, such tests can be used to test structural non-causality under appropriate conditions.

Accordingly, our main goal and contribution here is to provide straightforward and powerful new nonparametric methods for testing conditional independence. Hausman's (1978) ideas provide the basic foundation for our approach; however, as is now often the case, our methods compare two *collections* of estimators that converge to the same limits under correct specification (conditional independence) and that diverge under the alternative.

We construct our tests using local polynomial quantile regression, allowing weakly dependent data. This yields specification testing methods suitable for either cross-section or time-series data. In pursuing our main goal in this way, we make a number of further related contributions. Specifically, we generalize the existing nonparametric quantile literature not only by allowing for dependent heterogeneous data but also by establishing a weak consistency rate for the local Bahadur representation that is uniform in *both* the conditioning variables and the quantile index. We also show that, despite our nonparametric approach, our tests can detect local alternatives to conditional independence that decay to zero at the *parametric* rate, in contrast to the tests of Huang (2010) and of Su and White (2007, 2008, 2011). Although other tests can also detect local alternatives at the parametric rate (Linton and Gonzalo, 1997; Delgado and González-Manteiga, 2001; Song, 2009; Huang and White, 2010), those tests are for independent identically distributed (IID) data and do not necessarily extend easily to the time-series case. Our tests are thus the first for time series conditional independence that can detect local alternatives at the parametric rate.

The rest of the paper is organized as follows. In Section 2 we describe quantile regression and its relation to conditional independence. Section 3 introduces the local polynomial quantile regression estimator and studies its uniform local Bahadur representation. We apply this representation result to testing conditional independence in Section 4, where we also conduct some Monte Carlo simulations to evaluate the finite sample performance of our tests. Section 5 provides a summary and conclusion. All proofs are provided the appendix.

2 Quantile Regression and Conditional Independence

Let $\{(Y_t, X_t)\}$ denote a time series of random vectors, with Y_t a scalar for simplicity. Let $m_t(\tau, x)$ define the τth conditional quantile function of Y_t given $X_t = x \in \mathbb{R}^d$, that is, the τth conditional regression quantile. Specifically,

$$m_t(\tau, x) \equiv \inf \{y : F_t(y|x) \geq \tau\},$$

where $F_t(\cdot|x)$ denotes the conditional cumulative distribution function (CDF) of Y_t given $X_t = x$. Let $\rho_\tau(z) = z(\tau - \mathbf{1}(z \leq 0))$ be the "check" function, with $\mathbf{1}(\cdot)$ being the usual indicator function. It is well known that the τth conditional quantile $m_t(\tau, x)$ solves the minimization problem

$$m_t(\tau, x) = \underset{q \in \mathcal{Q}}{\arg\min} E\left[\rho_\tau(Y_t - q(X_t)|X_t = x)\right], \tag{2.1}$$

where $\mathcal{Q}$ is a given space of measurable functions defined on $\mathbb{R}^d$. As is common, we assume that the solution to this minimization problem is unique.

Often, the distribution of $\{(X_t, Y_t)\}$ is assumed to be stationary, so the conditional quantile function $m_t(\tau, x)$ is not time-varying; in this case, we write $m_t(\tau, x) = m(\tau, x)$ for all $t \geq 1$. Here, we do not assume stationarity; however, under conditional stationarity of Y_t given X_t, we again have $m_t = m$.

Koenker and Bassett (1978) pioneered quantile regression, treating the linear parametric case where $\mathcal{Q} = \{q : q(x) = \beta^T x,\ \beta \in \mathbb{R}^d\}$ and ^{T} denotes the transpose operator. Subsequently, nonparametric quantile regression has been studied by Bhattacharya and Gangopadhyay (1990), Chaudhuri (1991), White (1992), Fan, Hu, and Truong (1994), He and Shao (1996), Welsh (1996), Yu and Jones (1998), Hong (2000), and Lu, Hui, and Zhao (2001), among others. Here, we apply local polynomial methods, as described in the next section.

Our main focus of interest is the conditional independence of Y_t and Z_t given X_t, $Y_t \perp Z_t \mid X_t$. Let $m(\tau, x)$ and $m(\tau, x, z)$ define the τth conditional quantile functions of Y_t given $X_t = x$ and $(X_t, Z_t) = (x, z)$, respectively. Then $Y_t \perp Z_t \mid X_t$ if and only if the following null hypothesis holds:

$$H_0 : \Pr\left[m(\tau, X_t, Z_t) = m(\tau, X_t)\right] = 1 \text{ for all } \tau \in (0, 1). \tag{2.2}$$

An important special case is that of Granger non-causality (Granger, 1969). Let $\mathcal{X}_{t-1} = (X_{t-1}, \cdots, X_{t-p_x})^T$ and $\mathcal{Y}_{t-1} = (Y_{t-1}, \cdots, Y_{t-p_y})^T$, and let $m(\tau, \mathcal{Y}_{t-1})$ and $m(\tau, \mathcal{X}_{t-1}, \mathcal{Y}_{t-1})$ denote the τth conditional quantiles of Y_t given $\mathcal{Y}_{t-1}$ and $(\mathcal{X}_{t-1}, \mathcal{Y}_{t-1})$, respectively. Finite-order Granger non-causality in distribution is the condition that $Y_t \perp \mathcal{X}_{t-1} \mid \mathcal{Y}_{t-1}$ (for additional related concepts, see White and Lu, 2010). Then $Y_t \perp \mathcal{X}_{t-1} \mid \mathcal{Y}_{t-1}$ if and only if

$$H_0^G : \Pr\left[m(\tau, \mathcal{X}_{t-1}, \mathcal{Y}_{t-1}) = m(\tau, \mathcal{Y}_{t-1})\right] = 1 \text{ for all } \tau \in (0, 1).$$

Recently, Jeong and Härdle (2008) proposed a test of a version of this hypothesis with fixed τ by extending the work of Zheng (1998) from the IID case to the time-series case.

An important feature of H_0 is that it involves all quantiles $\tau \in (0,1)$ and all values in the joint support of the conditioning variables. This generally requires the convergence of the quantile estimators underlying the test to be uniform in both τ and the conditioning variables.

3 Local Polynomial Quantile Regression and Uniform Bahadur Representation

3.1 The local polynomial quantile regression estimator

When the distribution of Y_t given X_t is stationary, and if $m(\tau, x)$ is a sufficiently smooth function of x, for any $\widetilde{x}$ in a neighborhood of x, we have

$$\begin{aligned} m(\tau, \widetilde{x}) &= m(\tau, x) + \sum_{1 \le |\boldsymbol{j}| \le p} \frac{1}{\boldsymbol{j}!} D^{|\boldsymbol{j}|} m(\tau, x) (\widetilde{x} - x)^{\boldsymbol{j}} + o(\|\widetilde{x} - x\|^p) \\ &\equiv \sum_{0 \le |\boldsymbol{j}| \le p} \beta_{\boldsymbol{j}}(\tau, x; h) ((\widetilde{x} - x)/h)^{\boldsymbol{j}} + o(\|\widetilde{x} - x\|^p), \text{ say.} \end{aligned}$$

Here, we use the notation of Masry (1996): Letting $j_1, ..., j_d$ be non-negative integers, $\boldsymbol{j} \equiv (j_1, ..., j_d)$, $|\boldsymbol{j}| \equiv \sum_{i=1}^d j_i$, $x^{\boldsymbol{j}} \equiv \boldsymbol{\Pi}_{i=1}^d x_i^{j_i}$, $\sum_{0 \le |\boldsymbol{j}| \le p} \equiv \sum_{k=0}^p \sum_{\substack{j_1=0 \\ j_1+...+j_d=k}}^k \cdots \sum_{j_d=0}^k$, $D^{|\boldsymbol{j}|} m(\tau, x) \equiv \frac{\partial^{|\boldsymbol{j}|} m(\tau, x)}{\partial^{j_1} x_1 ... \partial^{j_d} x_d}$, $\beta_{\boldsymbol{j}}(\tau, x; h) \equiv \frac{h^{|\boldsymbol{j}|}}{\boldsymbol{j}!} D^{|\boldsymbol{j}|} m(\tau, x)$, where $\boldsymbol{j}! \equiv \boldsymbol{\Pi}_{i=1}^{d_1} j_i!$, and $h = h(n)$ is a bandwidth parameter that scales the distance between $\widetilde{x}$ and x.

With observations $\{(Y_t, X_t)\}_{t=1}^n$, we estimate the quantile regression function by local polynomial quantile regression. For this, let $\boldsymbol{\beta}$ denote the vector formed by stacking the vectors $\beta_{\boldsymbol{j}}$ in lexicographic order. Define

$$\widehat{\boldsymbol{\beta}}(\tau, x; h) \equiv \arg\min_{\boldsymbol{\beta}} Q_n(\tau, x; \boldsymbol{\beta}) \equiv n^{-1} \sum_{t=1}^n \rho_\tau \left(Y_t - \sum_{0 \le |\boldsymbol{j}| \le p} \beta_{\boldsymbol{j}} ((X_t - x)/h)^{\boldsymbol{j}} \right) K((x - X_t)/h), \quad (3.1)$$

where K is a nonnegative kernel function on $\mathbb{R}^d$. The conditional quantile $m(\tau, x)$ and its derivatives up to pth order are then estimated respectively by

$$\widehat{m}(\tau, x) = \widehat{\beta}_{\mathbf{0}}(\tau, x; h) \text{ and } \widehat{D}^{|\boldsymbol{j}|} m(\tau, x) = (\boldsymbol{j}!/h^{|\boldsymbol{j}|}) \widehat{\beta}_{\boldsymbol{j}}(\tau, x; h), \quad 0 \le |\boldsymbol{j}| \le p.$$

In particular, a local linear approach obtains when $p = 1$. See Fan, Hu, and Truong (1994) and Yu and Jones (1998), among many others.

To proceed, we introduce some notation. Let $N_l = (l + d - 1)!/(l!(d-1)!)$ be the number of distinct d-tuples $\boldsymbol{j}$ with $|\boldsymbol{j}| = l$. This denotes the number of distinct l-th order partial derivatives of $m(\tau, x)$ with respect to x. Arrange the N_l d-tuples as a sequence in a lexicographical order (with highest priority to last position), so that $\phi_l(1) \equiv (0, 0, ..., l)$ is the first element in the sequence and $\phi_l(N_l) \equiv (l, 0, ..., 0)$ is the last element, and let ϕ_l^{-1} denote the mapping inverse to ϕ_l. Let $N = \sum_{l=1}^p N_l$. For each $\boldsymbol{j}$ with $0 \le |\boldsymbol{j}| \le 2p$, let $\mu_{\boldsymbol{j}} = \int_{\mathbb{R}^d} x^{\boldsymbol{j}} K(x) dx$, and define the $N \times N$ dimensional matrix $\mathbb{H}$ and $N \times 1$ matrix $\mathbb{B}$

by

$$\mathbb{H}=\begin{bmatrix}\mathbb{H}_{0,0} & \mathbb{H}_{0,1} & \cdots & \mathbb{H}_{0,p}\\ \mathbb{H}_{1,0} & \mathbb{H}_{1,1} & \cdots & \mathbb{H}_{1,p}\\ \vdots & \vdots & \ddots & \vdots\\ \mathbb{H}_{p,0} & \mathbb{H}_{p,1} & \cdots & \mathbb{H}_{p,p}\end{bmatrix},\ \mathbb{B}=\begin{bmatrix}\mathbb{H}_{0,p+1}\\ \mathbb{H}_{1,p+1}\\ \vdots\\ \mathbb{H}_{p,p+1}\end{bmatrix}, \tag{3.2}$$

where $\mathbb{H}_{i,j}$ are $N_i \times N_j$ dimensional matrices whose (l,s) elements are $\mu_{\phi_i(l)+\phi_j(s)}$.

3.2 Assumptions

A Bahadur representation is an approximation of the sample quantiles by the empirical distribution function (Bahadur, 1966). Local Bahadur representations of conditional quantiles have been previously considered in a number of papers, including He and Shao (1996), Honda (2000), Lu, Hui, and Zhao (2001). In particular, Honda (2000) establishes a Bahadur representation that is uniform in the conditioning variables. More recently, Kong, Linton, and Xia (2010) establish a strong uniform consistency rate for the Bahadur representation of local polynomial M-regression estimates; there, too, the uniform rate is obtained only in the conditioning variables. In this section, we provide conditions sufficient to obtain a local Bahadur representation for $\widehat{\boldsymbol{\beta}}(\tau, x; h)$, uniform in both τ and the conditioning variables.

For given n, let $\left\{(Y_{nt}, X_{nt}) \in \mathbb{R}\times\mathbb{R}^d\right\}_{t=1}^n$ be a sequence of time-series random vectors. The triangular-array notation $\{(Y_{nt}, X_{nt})\}_{t=1}^n$ facilitates the study of asymptotic local power properties of many testing problems, including ours. Nevertheless, to avoid complicated notation we will suppress reference to the n subscript in what follows; in particular, we write $Y_t = Y_{nt}$, $X_t = X_{nt}$. For example, we will denote the conditional CDF of Y_{nt} given X_{nt} as $F_t(\cdot|X_t)$, instead of $F_{nt}(\cdot|X_{nt})$.

Next, let $\mathcal{T} \subset (0,1)$ and for $(\tau, u) \in \mathcal{T}\times\mathbb{R}$, define

$$\psi_\tau(u) \equiv \tau - \mathbf{1}(u \le 0).$$

For simplicity, we let the supports of the X_t's and Y_t's be time-invariant. For simplicity, we also suppose that the conditional support of Y_t given X_t coincides with Y_t's unconditional support. These restrictions can be straightforwardly relaxed, but with a considerable proliferation of notation. We thus let $\mathcal{X}$ denote the common support of the X_t's and $\mathcal{Y}$ denote the common support of the Y_t's. We let $\|\cdot\|$ denote the Euclidean norm. Although n is implicit for Y_t and X_t in what follows, the stated conditions hold for $n = 1, 2, \cdots$, and the referenced bounding constants or functions do not depend on n.

Assumption A1. $\{(Y_t, X_t)\}$ is a strong mixing process with mixing coefficients $\alpha(s)$ such that $\sum_{s=0}^{\infty} s^3\alpha(s)^{\eta/(4+\eta)} \le C < \infty$ for some $\eta > 0$ with $\eta/(4+\eta) \le 1/2$.

Assumption A2. (i) X_t is continuously distributed, with probability density function (PDF) $f_t(\cdot)$ bounded with bounded first order derivatives on $\mathcal{X}$ for each $t = 1, 2, \cdots$. (ii) The conditional CDF $F_t(\cdot|X_t)$ of Y_t given X_t has Lebesgue density $f_t(\cdot|X_t)$ such that $\sup_{y:\, F_t(y|X_t)\in\mathcal{T}} f_t(y|X_t) \le C_1 < \infty$ for all t, and for all $y_1, y_2 \in \mathcal{Y}$, $|f_t(y_1|X_t) - f_t(y_2|X_t)| \le C_2(X_t)\,|y_1 - y_2|$ a.s. for all t, where $C_2(\cdot)$ is a continuous function. (iii) The joint PDF $f_{ts}(\cdot,\cdot)$ of (X_t, X_s) is bounded for all $t, s = 1, 2, \cdots$.

Assumption A3. For all $(\tau, x) \in \mathcal{T}\times\mathcal{X}$, the conditional quantile function defined by $m_t(\tau, x) \equiv \inf\{y \in \mathcal{Y} : F_t(y|x) \ge \tau\}$ satisfies: (i) $m_t(\tau, x) = m(\tau, x) + n^{-1/2}c(\tau, x, t/n)$ where $c(\tau, x, t/n)$ is uniformly bounded for all $(\tau, x) \in \mathcal{T}\times\mathcal{X}$ and $t \le n$; (ii) $m(\tau, x)$ is bounded uniformly in $(\tau, x) \in \mathcal{T}\times\mathcal{X}$.

It is Lipschitz continuous in (τ, x) and for each $\tau \in \mathcal{T}$ has all partial derivatives with respect to x up to order $p+1$; (iii) The $(p+1)$th order partial derivatives with respect to x, i.e., $D^{\boldsymbol{k}} m(\tau, x)$ with $|\boldsymbol{k}| = p+1$, are uniformly bounded in $(\tau, x) \in \mathcal{T} \times \mathcal{X}$ and are Hölder continuous in (τ, x) with exponent $\gamma_0 > 0 : |D^{\boldsymbol{k}} m(\tau, x) - D^{\boldsymbol{k}} m(\widetilde{\tau}, \widetilde{x})| \leq C_3 (|\tau - \widetilde{\tau}|^{\gamma_0} + ||x - \widetilde{x}||^{\gamma_0})$ for some constant $C_3 < \infty$, and for all τ, $\widetilde{\tau} \in \mathcal{T}$ and x, $\widetilde{x} \in \mathcal{X}$ and all $\boldsymbol{k}$ such that $|\boldsymbol{k}| = p+1$.

Assumption A4. $\underline{\lim}_{n\to\infty} n^{-1} \sum_{t=1}^{n} f_t(m(\tau, x)|x)\ f_t(x) > 0$ uniformly in $(\tau, x) \in \mathcal{T} \times \mathcal{X}$.

Assumption A5. The kernel function $K(\cdot)$ is a product kernel of $k(\cdot)$, which is a symmetric density function with compact support $\mathcal{A} \equiv [-c_k, c_k]$. $\sup_{a\in\mathcal{A}} |k(a)| \leq \bar{c}_1 < \infty$, and $|k(a) - k(\widetilde{a})| \leq \bar{c}_2 |a - \widetilde{a}|$ for all a, $\widetilde{a} \in \mathbb{R}$ and some $\bar{c}_2 < \infty$. The functions $H_{\boldsymbol{j}}(x) = x^{\boldsymbol{j}} K(x)$ for all $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p+1$ are Lipschitz continuous. $\mathbb{H}$ defined in (3.2) is positive definite.

Assumption A6. As $n \to \infty$, (i) $h \to 0$, $nh^{2d}/(\log n)^2 \to \infty$, $nh^{d+2(p+1)} \to c \in [0, \infty)$, $h^{d/2}/(\kappa_n \log\log n) \to 0$, and $h^{(p+1)}/\kappa_n \to 0$ for some non-increasing positive sequence κ_n. (ii) Let $h = n^{-1/\varsigma_1}$ and $\kappa_n = n^{-\varsigma_\kappa}$ where $\varsigma_1 > 0$, $\varsigma_\kappa \geq 0$ and $4\varsigma_\kappa + d/\varsigma_1 - 1 < 0$. There exists $\upsilon \in (\varsigma_\kappa, (1 - d/\varsigma_1)/2)$ such that $16/\eta > [5/2 + d + 2\upsilon + 3d/(2\varsigma_1) + (2N+1)\varsigma_\kappa]/(1 - 2\upsilon - d/\varsigma_1) - 1$.

Assumption A1 restricts the process $\{(Y_t, X_t)\}$ to be strong mixing with mixing rates decaying sufficiently fast. It does not require stationarity. Assumption A2 imposes smoothness conditions on the functions $f_t(\cdot)$, $f_t(\cdot|X_t)$, and $f_{ts}(\cdot, \cdot)$. Assumptions A3-A4 enable us to establish the uniform local Bahadur representation for our local polynomial estimates. In particular, A3 allows us to establish a uniform Bahadur representation for asymptotically stationary process. Assumptions A5 and A6 specify typical conditions on the kernel and bandwidth used in local polynomial regression. In particular, Assumption A6 implies that $nh^d \kappa_n^2 \to \infty$ as $n \to \infty$.

3.3 Uniform local Bahadur representation

We now show that with the above assumptions, the local polynomial quantile estimator $\widehat{\boldsymbol{\beta}}(\tau, x; h)$ has a Bahadur representation uniform in both τ and x. For this, we introduce some additional notation. Let $\mu((X_t - x)/h)$ be an $N \times 1$ vector that contains the regressors $((X_t - x)/h)^{\boldsymbol{j}}$ in the local polynomial quantile regression (see (3.1)) in the lexicographical order. For example, if $p = 1$, then $\mu((X_t - x)/h) = (1, (X_t - x)^T/h)^T$. Let $\mu_{tx} \equiv \mu((X_t - x)/h)$. Define

$$H_n(\tau, x) \equiv \frac{1}{nh^d} \sum_{t=1}^{n} K((x - X_t)/h)\ f_t(m(\tau, X_t)|X_t)\ \mu_{tx}\mu_{tx}^T, \text{ and}$$

$$J_n(\tau, x) \equiv \frac{1}{\sqrt{nh^d}} \sum_{t=1}^{n} K((x - X_t)/h)\ \mu_{tx}\, \psi_\tau\left(Y_t - \boldsymbol{\beta}_0(\tau, x; h)^T \mu_{tx}\right).$$

Theorem 3.1 *Suppose Assumptions A1-A6 hold. Then*

$$\sqrt{nh^d}\left(\widehat{\boldsymbol{\beta}}(\tau, x; h) - \boldsymbol{\beta}(\tau, x; h)\right) = H_n(\tau, x)^{-1} J_n(\tau, x) + o_P(\kappa_n) \text{ uniformly in } (\tau, x) \in \mathcal{T} \times \mathcal{X}.$$

In particular,

$$\sqrt{nh^d}\,(\widehat{m}(\tau, x) - m(\tau, x)) = e_1^T H_n(\tau, x)^{-1} J_n(\tau, x) + o_P(\kappa_n) \text{ uniformly in } (\tau, x) \in \mathcal{T} \times \mathcal{X},$$

where $e_1 = (1, 0, ..., 0)^T$ *is an* N*-vector.*

Theorem 3.1 generalizes the local Bahadur representation results in the literature. This uniform result is useful for many statistical applications, where one usually requires κ_n to be 1 or $h^{d/2}$. If $\kappa_n = 1$, we can choose $\varsigma_\kappa = 0$ in Assumption A6(ii), and the last two conditions in Assumption A6(i) are automatically satisfied. To construct a conditional quantile-based test that can detect deviations from the null at the parametric rate, one typically needs $\kappa_n = h^{d/2}$. In this case, $\varsigma_\kappa = d/(2\varsigma_1)$, and the first condition in Assumption A6(ii) implies that $nh^{3d} \to \infty$. The following corollary is handy for deriving the asymptotic properties of our test statistic.

Corollary 3.2 *Suppose Assumptions A1-A6 hold with κ_n and ς_κ in Assumption A6 replaced by $h^{d/2}$ and $d/(2\varsigma_1)$, respectively. Suppose $nh^{2(p+1)} = o(1)$. Then, uniformly in $(\tau, x) \in \mathcal{T} \times \mathcal{X}$,*

$$\sqrt{nh^d}\left(\widehat{m}(\tau,x) - m(\tau,x)\right) = e_1^T H(\tau,x)^{-1} \overline{J}_n(\tau,x)\left[1 + o_P(1)\right] + o_P(h^{d/2}),$$

where $\overline{J}_n(\tau,x) \equiv n^{-1/2}h^{-d/2}\sum_{t=1}^n K\left((x - X_t)/h\right)\mu_{tx}\,\psi_\tau\left(Y_t - m(\tau, X_t)\right)$, $H(\tau,x) \equiv \overline{f}(\tau,x)\,\mathbb{H}$, $\overline{f}(\tau,x) \equiv \lim_{n\to\infty}\overline{f}_n(\tau,x)$, *and* $\overline{f}_n(\tau,x) \equiv n^{-1}\sum_{t=1}^n f_t\left(m(\tau,x)\,|x\right)\, f_t(x)$.

If we assume the process $\{(X_t, Y_t)\}$ is stationary, then the conditional and marginal PDF's $f_t\left(m(\tau,x)\,|x\right)$ and $f_t(x)$ can be written $f\left(m(\tau,x)\,|x\right)$ and $f(x)$, so that $H(\tau,x) = f\left(m(\tau,x)\,|x\right)\, f(x)\,\mathbb{H}$. This result is frequently used in the next section.

4 Testing Conditional Independence

4.1 Motivation and the test statistic

As discussed in Section 2, we wish to test

$$H_0 : \Pr\left[m(\tau, X_t, Z_t) = m(\tau, X_t)\right] = 1 \text{ for all } \tau \in (0,1),$$

where X_t and Z_t are random vectors of dimension d_X and d_Z, respectively. An obvious way to test this hypothesis would be to compare estimators of $m(\tau, X_t, Z_t)$ and $m(\tau, X_t)$ for all τ and all admissible $W_t \equiv \left(X_t^T, Z_t^T\right)^T$. This approach clearly would give a form of Hausman test. As White (1994, ch.9) shows, however, Hausman tests can also be formulated as $m-$tests, that is, tests of specific moment restrictions that hold under correct specification and fail otherwise. Such $m-$tests are often especially convenient, both for analysis and computation.

To formulate a corresponding $m-$test for H_0, let $u_{t\tau} \equiv Y_t - m(\tau, X_t)$ and $\varepsilon_{t\tau} \equiv Y_t - m(\tau, X_t, Z_t)$, and recall that $\psi_\tau(u) \equiv \tau - \mathbf{1}(u < 0)$. Then $u_{t\tau} = \varepsilon_{t\tau}$ under H_0 and H_0 holds if and only if

$$H_0^* : E\left[\psi_\tau(u_{t\tau})\,|W_t\right] = 0 \text{ a.s. for all } \tau \in (0,1).$$

This hypothesis has the form of a conditional moment restriction, involving the generalized residuals $\psi_\tau(u_{t\tau})$. Two challenges are apparent here. First, for each τ, there is an infinite number of unconditional moment restrictions implied by $E\left[\psi_\tau(u_{t\tau})\,|W_t\right] = 0$. Second, we must accommodate the fact that τ can take a continuum of values.

Results of Stinchcombe and White (1998, SW) allow us to convert conditional to unconditional moment restrictions in a convenient way. Specifically, SW provide conditions under which

$$\begin{aligned} E\left[\psi_\tau\left(u_{t\tau}\right)|W_t\right] &= 0 \text{ a.s. for all } \tau \in (0,1) \\ &\quad \text{if and only if} \\ E\left[\psi_\tau\left(u_{t\tau}\right)\varphi(W_t,\gamma)\right] &= 0 \text{ almost everywhere (a.e.) in } (\tau,\gamma) \in (0,1)\times\Gamma \end{aligned} \tag{4.1}$$

where $\Gamma \subset \mathbb{R}^{d_\Gamma}$ is a properly chosen set with typical choices $d_\Gamma = d_X + d_Z$ or $d_X + d_Z + 1$, and φ is a *generically comprehensively revealing* (GCR) or *comprehensively revealing* (CR) function. Examples of GCR functions include

(1) $\varphi(W_t,\gamma) = \exp\left(i\gamma^T W_t\right)$,
(2) $\varphi(W_t,\gamma) = \sin\left(\gamma^T W_t\right)$,
(3) $\varphi(W_t,\gamma) = \exp\left(\gamma^T W_t\right)$,

where $i = \sqrt{-1}$. The following CR functions are frequently used in the literature:

(4) $\varphi(W_t,\gamma) = \mathbf{1}\left(W_t \leq \gamma\right)$,
(5) $\varphi(W_t,\gamma) = \mathbf{1}\left(\beta^T W_t \leq \alpha\right)$ with $\gamma = \left(\alpha, \beta^T\right)^T$,

where $\mathbf{1}\left(W_t \leq \gamma\right) = \Pi_{i=1}^{d_X+d_Z}\mathbf{1}\left(W_{ti} \leq \gamma_i\right)$, and W_{ti} and γ_i are the ith elements of W_t and γ respectively. See SW for primitive conditions for GCR or CR functions.

A remarkable property of GCR functions is that if φ is GCR, then deviations from the null hypothesis can be detected by essentially any choice of $\gamma \in \Gamma$, where Γ can be chosen as any small compact set with non-empty interior. In contrast, for CR functions the set Γ may have to be $\mathbb{R}^{d_\Gamma}$ in order to ensure consistency of the associated test. Also, different choices of φ result in different local power properties. There is no general way to choose the "optimal" φ to conduct a test because such a function φ will depend on the underlying data generating process or the true alternative. For this reason, it is desirable to establish a general theory that covers a large class of (G)CR functions φ.

Given a sample $\{(Y_t, W_t)\}_{t=1}^n$, define the empirical process

$$S_n^{(1)}\left(\tau,\gamma\right) \equiv n^{-1/2}\sum_{t=1}^n \psi_\tau\left(u_{t\tau}\right)\varphi(W_t,\gamma).$$

Since $u_{t\tau}$ is not observable, in practice we replace it with $\widehat{u}_{t\tau}$, where $\widehat{u}_{t\tau} \equiv Y_t - \widehat{m}(\tau, X_t)$, and $\widehat{m}(\tau, x)$ denotes the p-th order local polynomial quantile regression estimate of $m(\tau,x)$. A practical test of CI can be based on the process

$$S_n^{(2)}\left(\tau,\gamma\right) \equiv n^{-1/2}\sum_{t=1}^n \psi_\tau\left(\widehat{u}_{t\tau}\right)\varphi(W_t,\gamma).$$

The limiting distribution of $S_n^{(2)}\left(\tau,\gamma\right)$ is different from that of $S_n^{(1)}\left(\tau,\gamma\right)$, a consequence of the "parameter estimation error" problem. As we explain shortly, this causes great difficulty in proposing a bootstrap test statistic whose limiting distribution coincides with the limiting null distribution of the test statistic. In addition, the indicator function in $\psi_\tau\left(\cdot\right)$ is not a smooth function, which makes the asymptotic analysis of $S_n^{(2)}\left(\tau,\gamma\right)$ intractable: even if we assume that the conditional quantile function $m\left(\tau,x\right)$ belongs to a certain smooth class of functions (e.g., Van der Vaart and Wellner (1996, p.154)) so that $S_n^{(1)}\left(\tau,\gamma\right)$ obeys a version of the Donsker theorem, it is hard (if even possible) to ensure that the local polynomial quantile estimate $\widehat{m}(\tau, x)$ also belongs to the same class. We therefore can not apply empirical process theory to study $S_n^{(2)}(\tau, \gamma)$ directly. Instead, we propose to approximate the indicator function by a smooth function $G(\cdot)$ and consider the stochastic process

$$S_n(\tau, \gamma) \equiv n^{-1/2} \sum_{t=1}^{n} [\tau - G(-\widehat{u}_{t\tau}/\lambda_n)] \varphi(W_t, \gamma),$$

where $G(\cdot)$ is a function that behaves like a CDF with uniformly bounded derivatives up to third order and $\lambda_n \to 0$ is a smoothing parameter.

The process $S_n(\tau, \gamma)$ will be the main ingredient of our test statistic. Under some regularity conditions, it converges to a mean-zero Gaussian process under the null and diverges for some value(s) of (τ, γ) under the alternative. Consequently, we accommodate the continuum of values for τ and γ using the Cramér-von Mises test statistics

$$CM_n \equiv \int_{\Gamma} \int_{\mathcal{T}} |S_n(\tau, \gamma)|^2 \Psi_1(d\tau) \Psi_2(d\gamma), \tag{4.2}$$

where $\mathcal{T} = [\underline{\tau}, \overline{\tau}]$ is a subset of $(0, 1)$, and $\Psi_1(\cdot)$ and $\Psi_2(\cdot)$ are weighting functions satisfying some mild conditions.

One can also consider the Kolmogorov-Smirnoff test statistic

$$KS_n \equiv \sup_{\tau \in \mathcal{T}} \sup_{\gamma \in \Gamma} |S_n(\tau, \gamma)|.$$

But KS_n is much more computationally demanding than CM_n, so we focus on the CM_n statistic.

As we show next, despite the nonparametric quantile regression, our tests can detect local alternatives that decay to zero at the parametric rates, in sharp contrast with the tests of Su and White (2007, 2008, 2011). More importantly, since our tests only involve d_X-dimensional smoothing, it is less severely subject to the "curse of dimensionality" problem than some of the earlier tests. In addition, our tests allow for weakly dependent data, and they are asymptotically pivotal under the null hypothesis for independent or martingale difference sequence (m.d.s.) data.

4.2 Asymptotic null distribution

We add the following assumptions.

Assumption B1. For all $\tau \in (0, 1)$, $E[\psi_\tau(Y_t - m(\tau, W_t))|W_t] = 0$ a.s.

Assumption B2. $\{(Y_t, W_t)\}$ is a strictly stationary strong mixing process with mixing coefficients $\alpha(s)$ such that (i) $\sum_{s=0}^{\infty} s^5 \alpha(s)^{\eta/(6+\eta)} \leq C < \infty$ for some $\eta > 0$ with $\eta/(6 + \eta) \leq 1/2$; (ii) $\frac{\eta}{2+\eta} + \frac{1+\nu}{6\overline{\nu}} < 1$ where $\overline{\nu} = 1 \wedge (r\nu)$, and r and ν are specified in Assumption B4 below.

Assumption B3. (i) The conditional CDF of Y_t given W_t, $F_{Y|W}(\cdot|W_t)$, and its Lebesgue density function $f_{Y|W}(\cdot|W_t)$ have continuous derivatives up to qth order denoted respectively by $F_{Y|X}^{(s)}(\cdot|X_t)$ and $f_{Y|W}^{(s)}(\cdot|W_t)$, $s = 1, \cdots, q$. $f_{Y|W}(\cdot|W_t)$ is Lipschitz continuous a.s., and $F_{Y|W}^{(q)}(\cdot|W_t)$ and $f_{Y|W}^{(q)}(\cdot|W_t)$ are bounded and uniformly continuous on $\mathbb{R}$ a.s. (ii) Let $V_t \equiv (Y_t, W_t^T)^T$. The joint PDF $f_{t_1,\cdots,t_{12}}(\cdot)$ of $(V_{t_1}, \cdots, V_{t_{12}})$ exists and is bounded for all $t_1, \cdots, t_{12} \in \{1, 2, \cdots, n\}$.

Assumption B4. $\varphi(\cdot,\cdot)$ is uniformly bounded by c_φ on the support of W_t and Γ. For some constants $r \geq 2$ and $\nu \in (0,1]$, either one of the following conditions holds: (i) $\varphi(\cdot,\gamma)$ is Hölder continuous with respect to γ in the sense that for some measurable function $C_\varphi(\cdot)$ with $E(|C_\varphi(W)|^r) < \infty$,

$$|\varphi(W,\gamma) - \varphi(W,\gamma')| \leq C_\varphi(W) \|\gamma - \gamma'\|^\nu$$

for all $\gamma, \gamma' \in \Gamma$; (ii) $\varphi(\cdot,\gamma)$ is locally uniformly L_r-continuous with respect to γ in the sense that for some constant $C_\varphi > 0$,

$$\left\{ E\left[\sup_{\gamma' : \|\gamma - \gamma'\| \leq \delta} |\varphi(W_t,\gamma) - \varphi(W_t,\gamma')|^r \right] \right\}^{1/r} \leq C_\varphi \delta^\nu$$

for all $\gamma \in \Gamma$ and all small positive $\delta = o(1)$.

Assumption B5. (i) $\sup_{u\in\mathbb{R}} |G(u)| \leq \bar{c}_G$ for some $\bar{c}_G < \infty$, $\lim_{u\to-\infty} G(u) = 0$, and $\lim_{u\to\infty} G(u) = 1$. (ii) $G(\cdot)$ is three times differentiable with derivatives denoted by $G^{(s)}(\cdot)$ for $s = 1,2,3$; $G(\cdot)$ and its first derivative $G^{(1)}(\cdot)$ are uniformly bounded, and the integrals $\int_{-\infty}^{\infty} \left|G^{(s)}(u)\right| du$, $s = 1,2,3$, are finite. (iii) $g(\cdot) \equiv G^{(1)}(\cdot)$ is symmetric over its support. There exists an integer $q \geq 2$ such that $\int_{-\infty}^{\infty} u^s g(u)\, du = \delta_{s0}$ for $s = 0, 1, \cdots, q-1$ and $\int_{-\infty}^{\infty} u^q g(u)\, du < \infty$, where δ_{s0} is Kronecker's delta. (iv) For some $c_G < \infty$ and $A_G < \infty$, either $G^{(3)}(u) = 0$ for $|u| > A_G$ and for $u, u' \in \mathbb{R}$, $\left|G^{(3)}(u) - G^{(3)}(u')\right| \leq c_G |u - u'|$, or $G^{(3)}(u)$ is differentiable with $|G^{(4)}(u)| \leq c_G$ and for some $\gamma_0 > 1$ and $|G^{(4)}(u)| \leq c_G |u|^{-\gamma_0}$ for all $|u| > A_G$.

Assumption B6. As $n \to \infty$, $\lambda_n \to 0$, $n\lambda_n^{2q} \to 0$, $n^2 \lambda_n^3 h^{7d_X/2} / \log n \to \infty$, and $n^3 \lambda_n^6 h^{4d_X} / (\log n)^4 \to \infty$.

Assumption B1 says that for each τ, $m(\tau, W_t)$ is the τth conditional quantile function of Y_t given W_t. Assumption B2 strengthens the mixing conditions in Assumption A2. The first condition of B2 is used to determine the sixth moment of a second-order U-statistic, whereas the second condition is used together with Assumption B4 to prove the stochastic equicontinuity of a certain empirical process. Assumption B3 imposes some smoothness conditions on the conditional CDF $F_{Y|W}(\cdot|W_t)$ and PDF $f_{Y|W}(\cdot|W_t)$. The uniform boundedness of the joint PDF $f_{t_1,\cdots,t_{12}}(\cdot)$ facilitates the determination of the six moments of certain U-statistics. Assumptions B4(i) and (ii) parallel Conditions (3.1) and (3.2) in Chen, Linton, and Van Keilegom (2003). It is easy to verify that the five examples after (4.1) satisfy either condition (i) or (ii) in B4. In all but example (3), $\varphi(\cdot,\cdot)$ is uniformly bounded no matter whether we allow the support of W_t to be compact or not. In the case where W_t is compactly supported, $\varphi(\cdot,\cdot)$ is also uniformly bounded in example (3). Assumption B5(i) is required because we use G to approximate the indicator function. Nevertheless, G does not need to be bounded between 0 and 1, nor does it need to be monotone. Assumptions B5(ii)-(iv) specify smoothness conditions on G. In particular, Assumption B5(iii) requires that the first derivative function g behaves like a symmetric qth order kernel and Assumption B5(iv) is used in studying the remainder term of a third order Taylor expansion. If $q = 2$, the CDF for the standard normal distribution meets all the conditions on G; if $q = 4$, one can use the integral of the fourth order Gaussian or Epanechnikov kernel as G. Assumption B6 specifies conditions on the smoothing parameters λ_n and h. Note that the last requirement in the assumption implies that $n^{-1/2} h^{-d_X/2} \sqrt{\log n} = o(\lambda_n)$, i.e., $n\lambda_n^2 h^{d_X} / \log n \to \infty$. We can set $h = n^{-1/\varsigma_1}$

and $\lambda_n \propto n^{-1/\varsigma_2}$ so that Assumptions A6 and B6 are both satisfied. We then need

$$\max\left(\frac{6\varsigma_1}{4\varsigma_1 - 7d_X}, \frac{6\varsigma_1}{3\varsigma_1 - 4d_X}\right) < \varsigma_2 < 2q.$$

When the dimension d_X of the conditioning variable X_t is small and η is small enough in Assumption B2, $q = 2$ will suffice. For example, if $d = 1$, $p = 1$, $q = 2$, $h \propto n^{-1/3.5}$, then one can choose $\varsigma_2 \in (42/13, 4)$; if $d = 2$, $p = 3$, $q = 2$, $h \propto n^{-1/7}$, then one can choose $\varsigma_2 \in (42/13, 4)$.

Let $\eta_t(\gamma, \tau) = [\varphi(W_t, \gamma) - c_0 b(X_t, \gamma)]\psi_\tau(\varepsilon_{t\tau})$, where $c_0 = e_1^T \mathbb{HB}$ and $b(X_t, \gamma) \equiv E[\varphi(W_t, \gamma)|X_t]$. The following theorem shows that $S_n(\cdot, \cdot)$ converges weakly to a Kiefer process under the null hypothesis. We let $\Rightarrow$ denote weak convergence. Here and below, we use η^c to denote the complex conjugate of η.

Theorem 4.1 *Suppose the conditions of Corollary 3.2 hold. Suppose Assumptions B1-B6 hold. Then under* H_0

$$S_n(\cdot, \cdot) \Rightarrow S_\infty(\cdot, \cdot),$$

where $S_\infty(\cdot, \cdot)$ *is a mean-zero Gaussian process with covariance kernel* $\Upsilon(\tau, \tau'; \gamma, \gamma') \equiv E[S_\infty(\tau, \gamma) S_\infty(\tau', \gamma')] = E[\eta_1(\gamma, \tau)\eta_1^c(\gamma', \tau')] + \sum_{i=1}^\infty E\left[\eta_1(\gamma, \tau)\eta_{1+i}^c(\gamma', \tau')\right] + E\left[\eta_{1+i}(\gamma, \tau)\eta_1^c(\gamma', \tau')\right]$.

Remark 1. Theorem 4.1 indicates that the process $\{S_n(\tau, \gamma) : \tau \in \mathcal{T}, \gamma \in \Gamma, n \geq 1\}$ converges to a zero-mean Gaussian process under the null hypothesis of CI. By the continuous mapping theorem,

$$CM_n \Rightarrow \int_\Gamma \int_{\mathcal{T}} |S_\infty(\tau, \gamma)|^2 \Psi_1(d\tau)\Psi_2(dw),$$

provided Ψ_1 and Ψ_2 are well behaved. The covariance kernel of the limiting process $\{S_\infty(\tau, \gamma) : \tau \in \mathcal{T}, \gamma \in \Gamma\}$ depends on the (G)CR function $\varphi(\cdot, \cdot)$ and the dependence structure in the data. There is thus no way to tabulate the critical values for our test, so we will provide a method to obtain bootstrap p-values. Note that the term $c_0 b(X_t, \gamma)$ in the definition of $\eta_t(\gamma, \tau)$ reflects the cost paid for replacing $m(\tau, X_t)$ with its local polynomial estimate. (We can show that $c_0 = 1$ for the local linear quantile regression estimate and lies strictly between 0 and 1 for general local polynomial regression with $p \geq 2$.) This term has to be taken into account when one proposes a bootstrap procedure to obtain the p-values.

Remark 2. Let $\mathfrak{F}_{t-1} \equiv \sigma(W_t, Y_{t-1}, W_{t-1}, Y_{t-2}, W_{t-2} \cdots)$. If $\{\psi_\tau(\varepsilon_{t\tau}), \mathfrak{F}_t\}$ is an m.d.s. for each τ (e.g., when $\{(Y_t, W_t)\}$ is an independent sequence), then

$$\Upsilon(\tau, \tau'; \gamma, \gamma') = E[\eta_1(\gamma, \tau)\eta_1^c(\gamma', \tau')].$$

In this special case, the limiting process $\{S_\infty(\tau, \gamma) : \tau \in \mathcal{T}, \gamma \in \Gamma\}$ is asymptotically pivotal. But this still depends on the chosen (G)CR and kernel functions.

4.3 Consistency and asymptotic local power properties

Now we study the consistency and asymptotic local power properties of tests based on $S_n(\cdot, \cdot)$. First, we show that the tests are consistent.

Theorem 4.2 *Suppose the conditions of Corollary 3.2 hold. Suppose Assumptions B1-B6 hold. Then under* $H_1 : H_0$ *is false, for each* $(\tau, \gamma) \in \mathcal{T} \times \Gamma$

$$n^{-1/2} S_n(\tau, \gamma) \xrightarrow{P} E\{f(m(\tau, W_t)|W_t)\varphi(W_t, \gamma)[m(\tau, W_t) - m(\tau, X_t)]\}.$$

Consequently, the (G)CR nature of the function φ implies that $E\{f(m(\tau, W_t)|W_t)\varphi(W_t,\gamma)[m(\tau, W_t) - m(\tau, X_t)]\} \neq 0$ in a set with positive measure, so the CM_n test statistic will diverge to ∞ under the alternative.

To study the local power properties of the tests based upon $S_n(\cdot,\cdot)$, we consider the quantile regression model (3.1) with the following class of local alternatives:

$$H_{1n}: m(\tau, W_t) = m(\tau, X_t) + n^{-1/2}\delta(\tau, W_t), \tag{4.3}$$

where $\delta(\cdot,\cdot)$ is a non-constant measurable function. To facilitate our analysis, we add the following assumption.

Assumption B7. (i) $\delta(\tau, W)$ is uniformly bounded and uniformly continuous on $\mathcal{T}$ and the support of W_t. (ii) $n^{-1}\sum_{t=1}^{n} E\{f(m(\tau, W_t)|W_t)\delta(\tau, W_t)[\varphi(W_t,\gamma) - c_0 b(X_t,\gamma)]\} = \mathbf{\Delta}(\tau,\gamma) + o(1)$ uniformly in $(\tau,\gamma) \in \mathcal{T}\times\Gamma$.

The above assumption is not minimal. The uniform boundedness and continuity of $\delta(\tau, W)$ greatly simplify our proofs.

Theorem 4.3 *Suppose the conditions of Corollary 3.2 hold. Suppose Assumptions B1-B7 hold. Then under H_{1n},*

$$S_n(\cdot,\cdot) \Rightarrow S_\infty(\cdot,\cdot) + \mathbf{\Delta}(\cdot,\cdot)$$

Theorem 4.3 implies that the CM_n test has non-trivial power in detecting $n^{-1/2}$-local alternatives provided $\mathbf{\Delta}(\tau,\gamma) \neq 0$ for (τ,γ) in a set of positive measure on $\mathcal{T}\times\Gamma$.

4.4 A bootstrap version of the test

From the previous section, we see that the asymptotic null distributions of the CM_n test statistics are generally not asymptotically pivotal, so the critical values for these tests cannot be tabulated. In this section, we propose a bootstrap version of our test, which is in the spirit of the block bootstrap (e.g., Bühlmann, 1994) but differs from the latter in several ways.

Let $\widehat{b}(X_j,\gamma)$ denote a local linear estimate of $b(X_j,\gamma)$ with kernel $K(\cdot)$ and bandwidth h_b. Let

$$S_n^*(\tau,\gamma) \equiv n^{-1/2}\sum_{i=1}^{n-L+1}\zeta_i\sum_{j=i}^{i+L-1}[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)]\left[\varphi(W_j,\gamma) - c_0\widehat{b}(X_j,\gamma)\right],$$

where $L \equiv L(n)$ denotes the block length and $\{\zeta_i\}_{i=1}^{n-L+1}$ is a sequence of random variables. The requirements on L and ζ_i are stated in the next assumption.

Assumption B8. (i) $\{\zeta_t\}_{t=1}^{n-L+1}$ are IID and independent of the process $\{(Y_t, W_t)\}$. (ii) $E(\zeta_t) = 0$, $E(\zeta_t^2) = 1/L$, and $E(\zeta_t^4) = O(1/L^2)$. (iii) As $n\to\infty$, $L\to\infty$ and $L/n^{1/2}\to 0$.

Like Inoue (2001), we will generate ζ_t independently from $N(0,1/L)$. Using $S_n^*(\cdot,\cdot)$, we construct the bootstrap version CM_n^* of the test statistic CM_n. We repeat this procedure B times to obtain the sequence $\{CM_{n,j}^*\}_{j=1}^{B}$. We reject the null when, for example, $p^* = B^{-1}\sum_{j=1}^{B}1(CM_n \leq CM_{n,j}^*)$ is smaller than the desired significance level. Let $\overset{p}{\Rightarrow}$ denote weak convergence in probability, as defined by Giné and Zinn (1990).

Theorem 4.4 *Suppose the conditions of Corollary 3.2 hold. Suppose Assumptions B1-B8 hold. Then under either H_0 or H_{1n}*

$$S_n^*(\cdot,\cdot) \overset{p}{\Rightarrow} S_\infty(\cdot,\cdot).$$

Remark 3. First, if $\{\psi_\tau(\varepsilon_{t\tau}), \mathfrak{F}_t\}$ is an m.d.s., we do not need to mimic the dependence structure in the data so we can take $L=1$ and our bootstrap is essentially a wild bootstrap:

$$S_n^W(\tau,\gamma) \equiv n^{-1/2} \sum_{i=1}^{n} \zeta_i \left[\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right] \varphi(W_j,\gamma),$$

and there is also no need to account for the parameter estimation error by recentering $\varphi(W_j,\gamma)$ around $c_0\widehat{b}(X_j,\gamma)$. Second, if $\{\psi_\tau(\varepsilon_{t\tau}), \mathfrak{F}_t\}$ is not an m.d.s., the limiting Gaussian process under the null hypothesis has the long-run covariance kernel defined in Theorem 4.1 and the wild bootstrap does not work, because it ignores the dependence structure of the data. Third, the parameter estimation error generally cannot be ignored in the bootstrap procedure. To see why, consider the following "naive" bootstrap process

$$S_n^\dagger(\tau,\gamma) \equiv n^{-1/2} \sum_{i=1}^{n-L+1} \zeta_i \sum_{j=i}^{i+L-1} \left[\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right] \varphi(W_j,\gamma).$$

One can decompose $\frac{1}{\sqrt{L}} \sum_{j=i}^{i+L-1} \left[\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right] \varphi(W_j,\gamma)$ into

$$\begin{aligned} &\frac{1}{\sqrt{L}} \sum_{j=i}^{i+L-1} \psi_\tau(\varepsilon_{j\tau}) \varphi(W_j,\gamma) + \frac{1}{\sqrt{L}} \sum_{j=i}^{i+L-1} [\mathbf{1}(\varepsilon_{j\tau} \le 0) - G(-\varepsilon_{j\tau}/\lambda_n)] \varphi(W_j,\gamma) \\ &+ \frac{1}{\sqrt{L}} \sum_{j=i}^{i+L-1} [G(-\varepsilon_{j\tau}/\lambda_n) - G(-\widehat{u}_{j\tau}/\lambda_n)] \varphi(W_j,\gamma), \end{aligned}$$

where the first term is our main object of interest, the second term represents the error due to the approximation of the indicator function by the smooth function $G(\cdot)$, and the third term reflects the parameter estimation error due to the estimation of $m(\tau,X_j)$ by $\widehat{m}(\tau,X_j)$ (under the local alternative H_{1n}, the difference between $m(\tau,X_j)$ and $m(\tau,W_j)$ does not enter the asymptotics of $S_n^\dagger(\tau,\gamma)$). Under weak conditions, we can show that the second term is $o_P(1)$ uniformly in (τ,γ), and the third term is also $o_P(1)$ uniformly in (τ,γ) provided $L^{1/2}\upsilon_n = o(1)$ where $\upsilon_n \equiv n^{-1/2}h^{-d_X/2}\sqrt{\log n} + h^{p+1}$ is the uniform probability order of the estimation error, i.e., $\max_{1\le j\le n} \sup_{\tau\in\mathcal{T}} |\widehat{m}(\tau,X_j) - m(\tau,X_j)| = O_P(\upsilon_n)$. (In the decomposition of $S_n(\tau,\gamma)$, the above third term corresponds to $\frac{1}{\sqrt{n}} \sum_{j=1}^{n} [G(-\varepsilon_{j\tau}/\lambda_n) - G(-\widehat{u}_{j\tau}/\lambda_n)] \varphi(W_j,\gamma)$, which is $O_P(1)$ instead.) It follows that

$$\begin{aligned} & E\left[S_n^\dagger(\tau,\gamma) S_n^{\dagger c}(\tau',\gamma') | \mathfrak{D}_n\right] \\ = \; & n^{-1} \sum_{i=1}^{n-L+1} \frac{1}{L} \sum_{j=i}^{i+L-1} \left[\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right] \varphi(W_j,\gamma) \left[\tau' - G\left(-\widehat{u}_{j\tau'}/\lambda_n\right)\right] \varphi^c(W_j,\gamma') \\ = \; & n^{-1} \sum_{i=1}^{n-L+1} \frac{1}{L} \sum_{j=i}^{i+L-1} \psi_\tau(\varepsilon_{j\tau}) \varphi(W_j,\gamma) \psi_{\tau'}(\varepsilon_{j\tau'}) \varphi^c(W_j,\gamma') + o_P(1) \\ \nrightarrow \; & \Upsilon(\tau,\tau';\gamma,\gamma') \text{ in probability,} \end{aligned}$$

where $\mathfrak{D}_n \equiv \{(Y_t, W_t)\}_{t=1}^n$. That is, conditional on $\mathfrak{D}_n$, $S_n^\dagger(\cdot,\cdot)$ cannot converge to $S_\infty(\cdot,\cdot)$, as it does not have the correct covariance kernel $\Upsilon(\cdot,\cdot;\cdot,\cdot)$. Fourth, as an alternative one can replace our bootstrap procedure by the block bootstrap of Bühlmann (1994):

$$S_n^\ddagger(\tau,\gamma) = n^{-1/2} \sum_{i=1}^{n-L+1} \sum_{j=s_i}^{s_i+L-1} \left[\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right] \left[\varphi(W_j,\gamma) - c_0 \widehat{b}(X_j,\gamma)\right],$$

where the s_i's are IID Uniform($\{1, 2, \cdots, n-L+1\}$). We conjecture that such a block bootstrap procedure is asymptotically equivalent to ours but the proof strategy will be quite different.

Remark 4. Theorem 4.4 shows that each bootstrapped process $\{S_n^*(\cdot,\cdot)\}$ converges weakly to the relevant Gaussian process, thus providing a valid asymptotic basis for approximating the limiting null distribution of test statistics based on $\{S_n(\cdot,\cdot)\}$. But we are only able to prove the above theorem under the sequence of local alternatives converging to 0 at the $n^{-1/2}$-rate (see H_{1n}). This is a phenomenon associated with many bootstrap versions of tests that aim not to re-estimate the model under investigation and are thus computationally attractive.

4.5 Monte Carlo simulations

In this subsection we conduct some Monte Carlo experiments to evaluate the finite sample performance of our tests. We consider four data generating processes (DGPs):

DGP 1.

$$\begin{aligned} Y_t &= \beta Z_t + X_t + \varepsilon_{Yt}, \\ Z_t &= X_t + 0.25X_t^2 + \varepsilon_{Zt}, \end{aligned}$$

where X_t is IID $U(-1,1)$, ε_{Yt} is IID $N(0,1)$, ε_{Zt} is IID, computed as the sum of 48 independent random variables, each uniformly distributed on [-0.25,0.25], $\{X_t\}$, $\{\varepsilon_{Yt}\}$, and $\{\varepsilon_{Zt}\}$ are mutually independent, and $\beta = \rho/(2\sqrt{1-\rho^2})$. It is easy to verify that ρ characterizes the conditional correlation coefficient of Y_t and Z_t given X_t.

DGP 2.

$$\begin{pmatrix} Y_t \\ Z_t \end{pmatrix} | X_t \sim N\left(\mathbf{0}, \begin{pmatrix} 1 + X_t^2 & \rho\sqrt{(1+X_t^2)(0.5+2X_t^2)} \\ \rho\sqrt{(1+X_t^2)(0.5+2X_t^2)} & 0.5 + 2X_t^2 \end{pmatrix}\right),$$

where $X_t = 0.5 + 0.5X_{t-1} + \varepsilon_{Xt}$, and ε_{Xt} is generated in the same way that ε_{Zt} is generated in DGP 1. Note that ρ is also the conditional correlation coefficient of Y_t and Z_t given X_t.

DGP 3.

$$\begin{aligned} Y_t &= 0.1 + \phi(Y_{t-1}) Y_{t-1} + \beta Z_{t-1} + \varepsilon_{Yt}, \\ Z_t &= 0.5 Z_{t-1} + \sqrt{1-0.5^2}\varepsilon_{Zt}, \end{aligned}$$

where $\varepsilon_{Yt} = 0.5\varepsilon_{Yt-1} + \sqrt{1-0.5^2} e_{Y_t}$, e_{Y_t} and ε_{Zt} are independently generated in the same way that ε_{Zt} is generated in DGP 1, $\phi(\cdot)$ is the standard normal PDF, and $\beta = \rho/(2\sqrt{1-\rho^2})$ with ρ denoting the conditional correlation coefficient of Y_t and Z_{t-1} given Y_{t-1}.

DGP 4.

$$\begin{aligned}
Y_t &= \frac{2}{1+\exp\left(1-0.5X_t^2\right)}+\sqrt{\vartheta_{Yt}}\varepsilon_{Yt}, \\
\vartheta_{Yt} &= 0.05+0.9\vartheta_{Y,t-1}+0.05Y_{t-1}^2+0.1X_t^2, \\
Z_t &= 1+X_t+\sqrt{\vartheta_{Zt}}\varepsilon_{Zt}, \\
\vartheta_{Zt} &= 0.05+0.7\vartheta_{Z,t-1}+0.2Z_{t-1}^2+0.2X_t^2,
\end{aligned}$$

where $X_t = 0.5X_{t-1} + \sqrt{1-0.5^2}\varepsilon_{Xt}$, ε_{Yt} and ε_{Zt} are both $t(3)/\sqrt{3}$ and have correlation given by ρ, and ε_{Xt} is generated as X_t in DGP 1, independently of $(\varepsilon_{Yt}, \varepsilon_{Zt})$.

Clearly, DGP 1 generates IID data $\{(Y_t, Z_t, X_t)\}$ whereas the other DGPs generate time-series dependent observations. $\{\psi_\tau(\varepsilon_{t\tau}),\ \mathfrak{F}_t\}$ forms an m.d.s. in both DGPs 1 and 2, but not in DGPs 3-4. Note that our test is based on local polynomial quantile regressions, which typically require compactly supported conditioning variables. This motivates the otherwise awkward way we generate ε_{Zt} in DGP 1, ε_{Xt} in DGP 2, and e_{Y_t} and ε_{Zt} in DGP 3. According to the central limit theorem, we can treat these variables as being nearly standard normal random variables but with compact support [-12, 12]. In all DGPs except DGP 3, we are interested in testing whether Y_t and Z_t are conditionally independent given X_t. In DGP 3, we test whether Y_t and Z_{t-1} are independent conditional on Y_{t-1}; i.e., the null hypothesis in this case is that Z_t does not Granger-cause Y_t at the first-order distributional level. Obviously, in all DGPs, the null hypotheses are satisfied if and only if the parameter ρ takes the value 0. The larger the value of $|\rho|$, the stronger the conditional dependence between Y_t and Z_t (or Z_{t-1} in DGP 3).

To construct the test statistics, we estimate the conditional quantile function $m(\tau, x)$ using locally linear quantile regression $(p = 1)$. We choose the normalized Epanechnikov kernel (with variance 1): $K(u) = \frac{3}{4}(1-\frac{1}{5}u^2)\mathbf{1}\left(|u| \le \sqrt{5}\right)$. Since there is no data-driven procedure to choose the bandwidth for quantile regression, to estimate the τth conditional quantile of Y_t given X_t, we choose a preliminary bandwidth according to the rule of thumb recommended by Yu and Jones (1998):

$$h_{0\tau} = s_X n^{-1/5}\left\{\tau(1-\tau)[\phi\left(\Phi^{-1}(\tau)\right)]^{-2}\right\}^{1/5},$$

where s_X is the standard deviation of X_t, and ϕ and Φ are the standard normal PDF and CDF, respectively. Since undersmoothing is required for our test, we modify the above choice of bandwidth to

$$h_{0\tau} = s_X n^{-1/\theta}\left\{\tau(1-\tau)[\phi\left(\Phi^{-1}(\tau)\right)]^{-2}\right\}^{1/5},$$

where $3 < \theta < 4$. We study the behavior of our tests with different choices of λ_n in order to examine the sensitivity of our test to the bandwidth sequence. Robinson (1991, p.448) and Lee (2003, p.16) propose very similar devices. Note that these choices for $h_{0\tau}$ and the kernel function meet the requirements for our test. Through a preliminary simulation study, we find our bootstrap-based test is not sensitive to the choice of θ when we take $\theta \in (3, 4)$. So we fix $\theta = 3.5$ for our simulation results.

To construct the bootstrap tests, we need to estimate $b(X_t, \gamma)$. Again, we apply the local linear estimation method by regressing $\varphi(W_t, \gamma)$ on X_t to obtain the estimate $\widehat{b}(X_t, \gamma)$. We choose the bandwidth by the rule of thumb: $h_b = 2s_X n^{-1/5}$. To construct the CM_n test statistics, we consider five cases of the (G)CR functions φ listed after eq. (4.1). We also need to choose the integrating functions $\Psi_1(\cdot)$ and $\Psi_2(\cdot)$ (see eq. (4.2)). We treat all quantiles τ as equally important, so we choose $\Psi_1(d\tau) = 1/(\overline{\tau} - \underline{\tau})$ if $\tau \in [\underline{\tau}, \overline{\tau}]$ and 0 otherwise. Following common practice in the parametric quantile regression literature, we set $\underline{\tau} = 0.1$ and $\overline{\tau} = 0.9$. The choice of $\Psi_2(\cdot)$ depends on the nature of the (G)CR function φ and the ease of implementation. To obtain the CM_n test statistics, we need to compute the integral

$$I_{ts} \equiv \int \varphi(W_t, \gamma)\varphi^c(W_s, \gamma) d\Psi_2(\gamma).$$

Even though numerical integration is possible, it is computationally costly, especially when the dimension of γ is high. To save time in computation, we choose Ψ_2 to ensure that I_{ts} can be calculated analytically. Let $N(\mathbf{0}, \Sigma)$ denote a multivariate normal distribution with mean zero and variance-covariance matrix Σ. Corresponding to the five choices of the GCR (or CR) functions, we consider the following integrating functions $\Psi_2(\cdot)$:

(1) When $\varphi(W_t, \gamma) = \exp\left(iw^T W_t\right)$, choose Ψ_2 to be the multivariate standard normal CDF. Then $I_{ts} = \exp(-\sum_{i=1}^{d_X+d_Z}(W_{ti} - W_{si})^2/2)$, and we denote the resulting test statistic as CM_{1n}. Here I is an identity matrix and W_{ti} denotes the ith element of W_t.

(2) When $\varphi(W_t, \gamma) = \sin\left(\gamma^T W_t\right)$, choose Ψ_2 to be the multivariate standard normal CDF. Then $I_{ts} = [\exp(-\sum_{i=1}^{d_X+d_Z}(W_{ti} - W_{si})^2/2) - \exp(-\sum_{i=1}^{d_X+d_Z}(W_{ti} + W_{si})^2/2)]/2$, and we denote the resulting test statistic as CM_{2n}.

(3) When $\varphi(W_t, \gamma) = \exp\left(\gamma^T W_t\right)$, we need to ensure that the values of $\gamma^T W_t$ are not too large or small in absolute value. (Note that $\exp(u)$ is close to linear if $|u|$ is close to 0, close to 0 if u is too small, and explodes quickly when u is too large. In such cases, the test will not be well behaved. See Bierens (1990).) We thus follow Bierens's (1990) advice and transform W_t to make sure each element W_{ti} of W_t lies between 0 and 1: $\overline{W}_{ti} = 1/(1 + \exp(-W_{ti}))$, $i = 1, \cdots, d_X + d_Z$. Then we choose $\Psi_2 \sim N(\mathbf{0}, \frac{1}{d_X+d_Z} I)$. In this case, $I_{ts} = \exp((d_X + d_Z)\sum_{i=1}^{d_X+d_Z}\left(\overline{W}_{ti} + \overline{W}_{si}\right)^2/2)$. We denote the resulting test statistic as CM_{3n}.

(4) When $\varphi(W_t, \gamma) = \mathbf{1}(W_t \leq \gamma)$, we consider two choices for $\Psi_2(\cdot)$. First, we choose Ψ_2 to be the multivariate standard normal CDF. Then $I_{ts} = \Pi_{i=1}^{d_X+d_Z}[1 - \Phi(W_{ti} \vee W_{si})]$, and we denote the resulting test statistic as CM_{4n}. Second, we choose Ψ_2 to be the empirical distribution of W_t; then $I_{ts} \equiv \frac{1}{n}\sum_{l=1}^{n}\Pi_{i=1}^{d_X+d_Z}\mathbf{1}(W_{ti} \vee W_{si} \leq W_{li})$, and we denote the resulting test statistic as CM_{4bn}.

(5) When $\varphi(W_t, \gamma) = \mathbf{1}\left(\beta^T W_t \leq \alpha\right)$ with $\gamma = \left(\alpha, \beta^T\right)^T$, we follow Escanciano (2006) and set

$$\Psi_2(dw) = F_{n,\beta}(d\alpha)\, d\beta$$

where $F_{n,\beta}(\cdot)$ denotes the empirical distribution function of $\left\{\beta^T W_t\right\}_{t=1}^n$ and $d\beta$ denotes the uniform density on the unit sphere. Then I_{ts} can be calculated analytically, but the exact formula is cumbersome. See Appendix B in Escanciano (2006) for a simple algorithm to compute I_{ts}.

We first focus on the finite sample performance of our tests under the null. Tables 1-4 report the empirical rejection frequencies of the CM_n tests at the 5% nominal level for DGPs 1-4, respectively, where $\rho = 0$. We use 1000 replications for each sample size n and 500 bootstrap resamples in each replication. To examine the sensitivity of our tests to the choice of block size L and the smoothing parameter λ_n, we set $\lambda_n = 0$, 0.001, and 0.01, and choose $L = \lceil cn^{1/4} \rfloor$ for three choices of c: 1, 2, 4. When $\lambda_n = 0$, we effectively replace the approximating function $G(-\widehat{u}_{t\tau}/\lambda_n)$ by the indicator function $\mathbf{1}(\widehat{u}_{t\tau} \leq 0)$. This allows us to examine whether the use of indicator function can be justified in practice when one needs to estimate the conditional quantile function but is not sure whether the estimate belongs to the same class of smooth functions as the original quantile function.

Table 1: Finite sample rejection frequency under the null (DGP 1, nominal level: 0.05)

n	Tests	$\lambda_n = 0,\ L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.001, L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.01,\ L = \lceil cn^{1/4} \rfloor$		
		$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$
100	CM_1	0.128	0.099	0.063	0.062	0.045	0.019	0.061	0.045	0.020
	CM_2	0.054	0.050	0.028	0.059	0.052	0.028	0.064	0.055	0.031
	CM_3	0.290	0.296	0.298	0.100	0.095	0.088	0.096	0.092	0.086
	CM_4	0.273	0.243	0.210	0.080	0.077	0.038	0.086	0.072	0.048
	CM_{4b}	0.359	0.321	0.269	0.100	0.080	0.042	0.102	0.082	0.050
	CM_5	0.589	0.539	0.455	0.102	0.076	0.044	0.097	0.082	0.045
200	CM_1	0.094	0.083	0.058	0.056	0.048	0.024	0.055	0.046	0.029
	CM_2	0.061	0.053	0.036	0.057	0.058	0.034	0.055	0.051	0.040
	CM_3	0.177	0.191	0.183	0.090	0.084	0.081	0.079	0.084	0.078
	CM_4	0.173	0.150	0.132	0.067	0.058	0.035	0.066	0.057	0.040
	CM_{4b}	0.217	0.203	0.171	0.079	0.063	0.045	0.080	0.064	0.042
	CM_5	0.326	0.315	0.253	0.099	0.076	0.060	0.095	0.076	0.050

Table 2: Finite sample rejection frequency under the null (DGP 2, nominal level: 0.05)

n	Tests	$\lambda_n = 0,\ L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.001, L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.01,\ L = \lceil cn^{1/4} \rfloor$		
		$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$
100	CM_1	0.145	0.119	0.065	0.071	0.050	0.017	0.066	0.045	0.018
	CM_2	0.062	0.050	0.033	0.055	0.045	0.026	0.054	0.045	0.028
	CM_3	0.285	0.282	0.299	0.064	0.067	0.060	0.062	0.063	0.064
	CM_4	0.332	0.316	0.264	0.082	0.069	0.040	0.073	0.065	0.032
	CM_{4b}	0.487	0.443	0.366	0.089	0.067	0.032	0.078	0.070	0.030
	CM_5	0.704	0.678	0.613	0.093	0.074	0.045	0.088	0.076	0.047
200	CM_1	0.109	0.089	0.071	0.059	0.056	0.033	0.059	0.059	0.034
	CM_2	0.065	0.050	0.040	0.061	0.049	0.034	0.057	0.049	0.036
	CM_3	0.223	0.224	0.232	0.066	0.065	0.058	0.068	0.064	0.063
	CM_4	0.243	0.221	0.186	0.076	0.065	0.045	0.070	0.062	0.044
	CM_{4b}	0.344	0.322	0.267	0.079	0.056	0.034	0.068	0.060	0.033
	CM_5	0.502	0.484	0.451	0.085	0.076	0.061	0.084	0.076	0.055

Table 3: Finite sample rejection frequency under the null (DGP 3, nominal level: 0.05)

n	Tests	$\lambda_n = 0,\ L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.001, L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.01,\ L = \lceil cn^{1/4} \rfloor$		
		$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$
100	CM_1	0.156	0.125	0.068	0.061	0.046	0.016	0.063	0.044	0.019
	CM_2	0.075	0.066	0.038	0.067	0.060	0.034	0.071	0.057	0.031
	CM_3	0.457	0.469	0.466	0.094	0.081	0.078	0.092	0.081	0.080
	CM_4	0.288	0.270	0.228	0.062	0.047	0.024	0.063	0.049	0.019
	CM_{4b}	0.277	0.261	0.224	0.055	0.045	0.018	0.060	0.048	0.021
	CM_5	0.817	0.795	0.721	0.103	0.085	0.054	0.096	0.080	0.050
200	CM_1	0.089	0.084	0.064	0.052	0.048	0.028	0.053	0.047	0.031
	CM_2	0.053	0.049	0.037	0.050	0.054	0.031	0.045	0.048	0.035
	CM_3	0.332	0.338	0.350	0.070	0.072	0.062	0.079	0.072	0.065
	CM_4	0.212	0.199	0.170	0.062	0.052	0.035	0.058	0.053	0.040
	CM_{4b}	0.211	0.200	0.171	0.059	0.050	0.036	0.057	0.047	0.038
	CM_5	0.623	0.598	0.545	0.071	0.072	0.049	0.067	0.068	0.044

Table 4: Finite sample rejection frequency under the null (DGP 4, nominal level: 0.05)

n	Tests	$\lambda_n = 0$, $L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.001$, $L = \lceil cn^{1/4} \rfloor$			$\lambda_n = 0.01$, $L = \lceil cn^{1/4} \rfloor$		
		$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$	$c = 1$	$c = 2$	$c = 4$
100	CM_1	0.162	0.122	0.054	0.081	0.050	0.012	0.079	0.053	0.013
	CM_2	0.067	0.044	0.028	0.054	0.039	0.019	0.054	0.038	0.022
	CM_3	0.274	0.273	0.295	0.082	0.076	0.056	0.078	0.074	0.054
	CM_4	0.241	0.217	0.190	0.064	0.050	0.029	0.068	0.052	0.027
	CM_{4b}	0.258	0.243	0.207	0.062	0.056	0.026	0.062	0.051	0.028
	CM_5	0.784	0.755	0.682	0.090	0.078	0.030	0.093	0.072	0.038
200	CM_1	0.217	0.180	0.096	0.088	0.069	0.057	0.082	0.700	0.050
	CM_2	0.052	0.049	0.030	0.046	0.044	0.033	0.049	0.045	0.034
	CM_3	0.249	0.250	0.249	0.084	0.071	0.063	0.078	0.076	0.064
	CM_4	0.226	0.216	0.192	0.074	0.058	0.040	0.076	0.060	0.041
	CM_{4b}	0.218	0.214	0.190	0.073	0.058	0.042	0.072	0.056	0.036
	CM_5	0.622	0.594	0.550	0.081	0.070	0.057	0.084	0.073	0.055

We summarize some important findings from Tables 1-4. First, when $\lambda_n = 0$, the sizes of our tests are highly distorted, whereas for $\lambda_n = 0.001$ or 0.01, they are reasonably well behaved. This indicates that the use of indicator function is questionable and thus we only focus on the case where $\lambda_n = 0.001$ or 0.01. Second, our tests depend on the choice of block size L (or equivalently c in the table): when $\lambda_n = 0.001$ or 0.01, the tests tend to be oversized for smaller values of block size ($c = 1, 2$) and close to the nominal levels or a little bit undersized when $c = 4$. Third, there is some level variation due to different choices of φ, but this is not large: the levels of CM_{3n} and CM_{5n} tend to be inflated more often than for the other tests. Fourth, as the sample size doubles, the levels of all tests improve.

Figures 1-4 display the powers of our tests at the 5% level for the block size $L = \lceil cn^{1/4} \rfloor$ ($c = 2$) and smoothing parameter $\lambda_n = 0.01$. To compare the tests on an equal basis, we consider not only the power of bootstrap-based tests but also the size-corrected power obtained by using critical values simulated from 250 replications under the null hypothesis of conditional independence ($\rho = 0$) for the four DGPs introduced above. In either case, we use 250 replications for each value of $\rho \in [-0.9, 0.9]$; the bootstrap tests are based on 500 bootstrap resamples in each replication. We summarize some of the main findings: (a) As the degree of conditional dependence ($|\rho|$) increases, the powers of all tests increase. (b) With or without size correction, the CM_{5n} test dominates the other tests in terms of power for all DGPs examined here. (c) The CM_{4bn} and CM_{3n} tests tend to be dominated by other tests in terms of size-corrected power. (d) The performance of the other tests tends to be DGP-dependent. Overall, CM_{5n} with $L = \lceil 4n^{1/4} \rfloor$ ($c = 4$) provide reliable level and power performance.

5 Summary and Conclusion

We provide straightforward new nonparametric methods for testing conditional independence using local polynomial quantile regression, allowing weakly dependent data. Inspired by Hausman (1978), our methods essentially compare two collections of estimators that converge to the same limits under correct specification (conditional independence) and that diverge under the alternative.

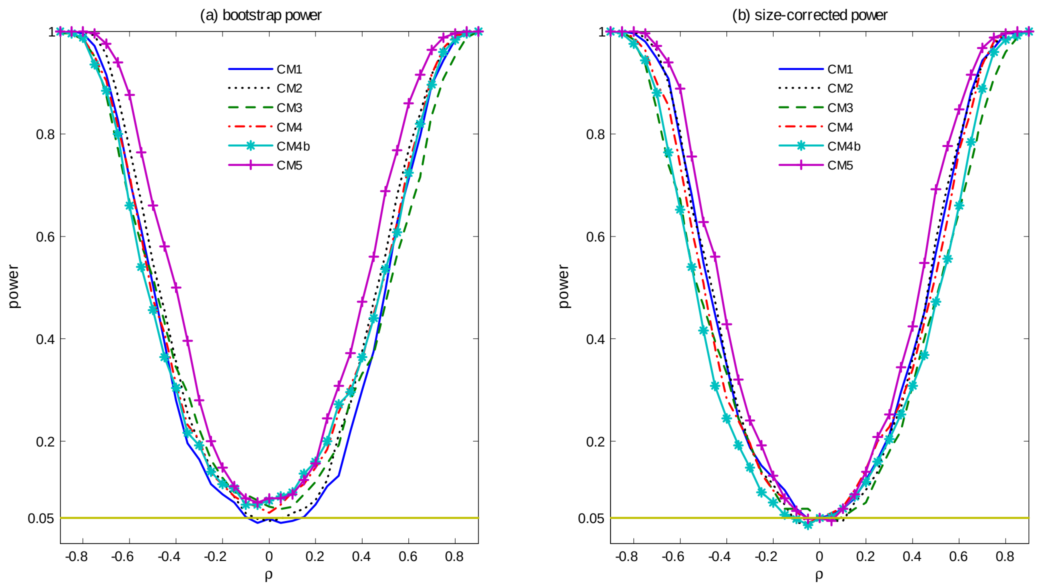

Figure 1: Power function for DGP 1 (n =100, $\lambda_n = 0.01$, nominal level: 0.05)

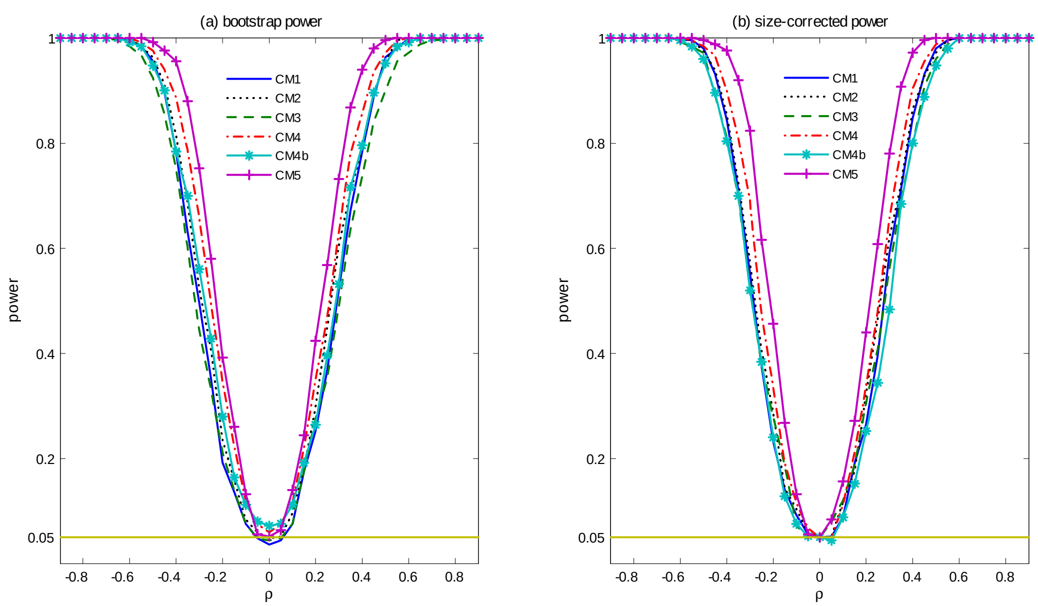

Figure 2: Power function for DGP 2 (n =100, $\lambda_n = 0.01$, nominal level: 0.05)

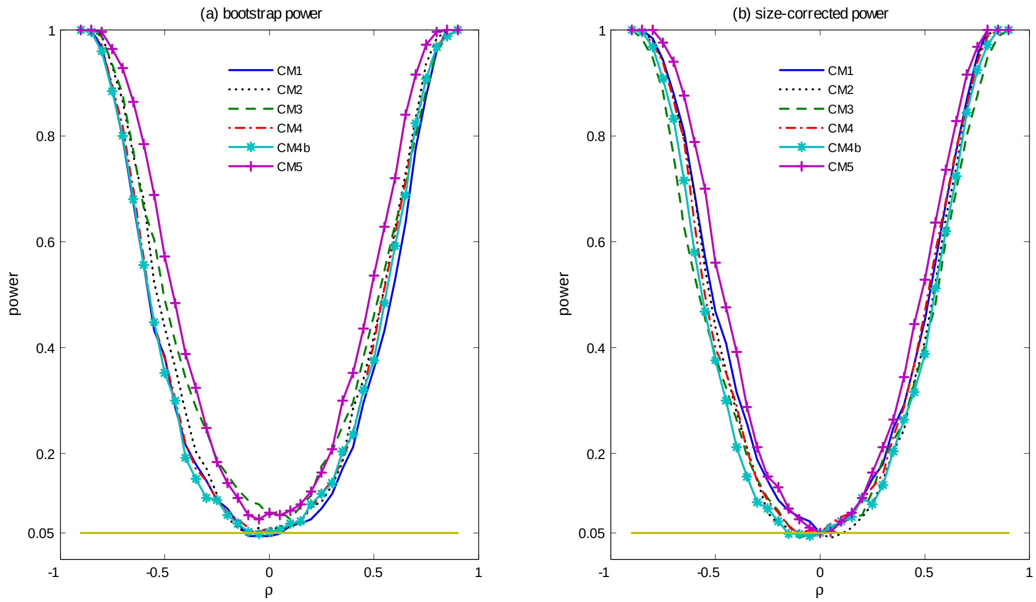

Figure 3: Power function for DGP 3 (n =100, $\lambda_n = 0.01$, nominal level: 0.05)

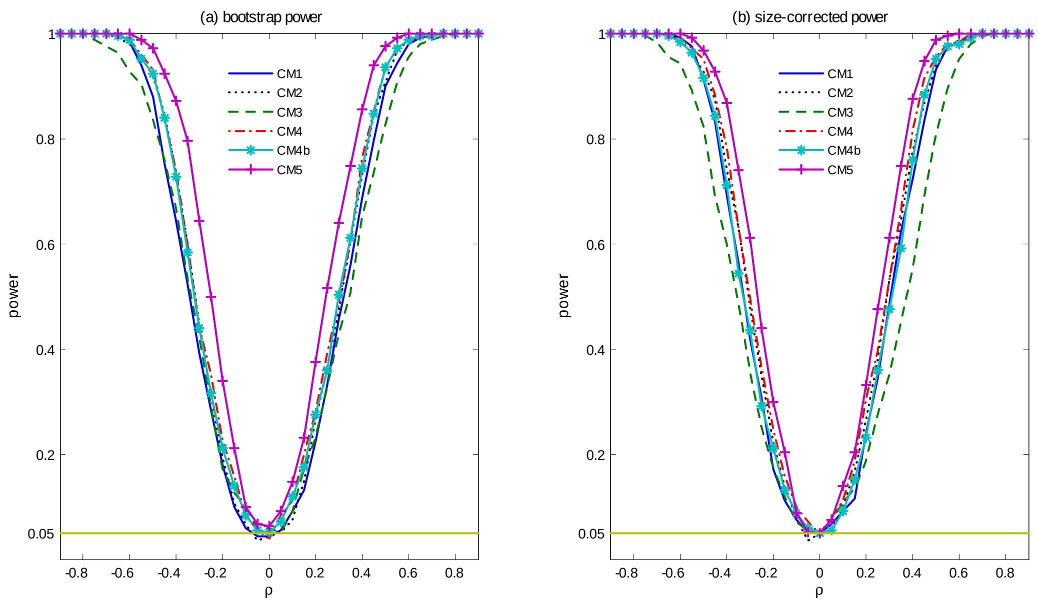

Figure 4: Power function for DGP 4 (n =100, $\lambda_n = 0.01$, nominal level: 0.05)

In addition, we generalize the existing nonparametric quantile literature not only by allowing for dependent heterogeneous data but also by establishing a weak consistency rate for the local Bahadur representation that is uniform in both the conditioning variables and the quantile index. We also show that, despite our nonparametric approach, our tests can detect local alternatives to conditional independence that decay to zero at the parametric rate. Our tests are the first for time-series conditional independence that can detect local alternatives at the parametric rate. Monte Carlo simulations suggest that our tests perform well in finite samples. Our tests have a variety of uses in applications, such as testing for failure of conditional exogeneity or for Granger non-causality.

REFERENCES

Andrews, D. W. K. & Pollard, D. (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review,* 62, 119-132.

Bahadur, R. R. (1966), A note on quantiles in large samples. *Annals of Mathematical Statistics*, 37, 577–580.

Barnow, B., Cain, G. & Goldberger A. (1980). Issues in the analysis of selectivity bias, in E. W. Stromsdorfer and George Farkas (eds.), *Evaluation Studies Review Annual*, Vol. 5, Beverly Hills, CA.: Sage Publications, pp. 43-59.

Bhattacharya, P. K. & Gangopadhyay, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Annals of Statistics,* 18, 1400-1415.

Billingsley, P. (1999). *Convergence of Probability Measures.* 2nd edition. Wiley: New York.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica,* 58, 1443-1458.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction.* Springer: New York.

Bühlmann, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics,* 22, 995-1012.

Chalak, K. & White, H. (2011). An extended class of instrumental variables for the estimation of causal effects. *Canadian Journal of Economics*, 44, 1-51.

Chaudhuri, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics,* 19, 760-777.

Chen, X., Linton, O. & van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica,* 71, 1591-1608.

Dawid, A. D. (1979). Conditional independence in statistical theory. *Journal of the Royal Statistical Society, Series B* 41, 1-31.

Delgado, M. A. & González-Manteiga, W. (2001). Significance testing in nonparametric regression based on the bootstrap. *Annals of Statistics,* 29, 1469-1507.

Escanciano, J. C. (2006). A consistent diagnostic test for regression models using projections. *Econometric Theory,* 22, 1030-1051.

Fan, J., Hu, T-C., & Truong, Y. K. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics,* 21, 433-446.

Fitzenberger, B. (1997). The moving block bootstrap and robust inference for linear least squares and quantile regressions. *Journal of Econometrics,* 82, 235-287.

Giné, E. & Zinn, J., (1990). Bootstrapping general empirical measures. *Annals of Probability,* 18, 851-869.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica,* 37, 424–438.

Granger, C. W. J. & Newbold, P. (1986). *Forecasting Economic Time Series.* 2nd edition. Orlando: Academic Press

Hall, P. & Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press: San Diego.

Hansen, B. E., (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory,* 24, 726-748.

Hausman, J. (1978). Specification tests in econometrics. *Econometrica*, 78, 1251-1271.

He, X. & Shao, Q-M. (1996). A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic design. *Annals of Statistics,* 24, 2608-2630.

Honda, T. (2000). Nonparametric estimation of a conditional quantile for α-mixing processes. *Annals of Institute of Statistical Mathematics,* 52, 459-470.

Huang, M. & White, H. (2010). A flexible test for conditional independence. Working Paper, Dept. of Economics, UCSD.

Huang, T.-M. (2010) Testing conditional independence using maximal nonlinear conditional independence. *Annals of Statistics,* 38, 2047-2091.

Inoue, A. (2001). Testing for distributional change in time series. *Econometric Theory, 17*, 156-187.

Jeong, K. & Härdle, W. K. (2008). A consistent nonparametric test for causality in quantile. Working paper, Center for Applied Statistics and Economics, Humboldt-Universität zu Berlin.

Koenker, R. (2005). *Regression Quantiles*. New York: Cambridge University Press.

Koenker, R. & Bassett, G. (1978). Regression quantiles. *Econometrica,* 46, 33-50.

Koenker, R. & Zhao, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory,* 12, 793-813.

Kong, E., Linton, O., & Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory,* 26, 1529-1564.

Koul, H. L., (1991). A weak convergence result useful in robust autoregression. *Journal of Statistical Planning and Inference,* 29, 1291-1308.

Lee, S. (2003). Efficient semiparametric estimation of partially linear quantile regression model. *Econometric Theory,* 19, 1-31.

Linton, O. & Gozalo, P. (1997). Conditional independence restrictions: testing and estimation. *Discussion Paper*, Cowles Foundation for Research in Economics, Yale University.

Lu, Z., Hui, Y. V., & Zhao, Q. (2001). Local linear quantile regression under dependence: Bahadur representation and applications. Mimeo, City University of Hong Kong.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency rates. *Journal of Time Series Analysis,* 17, 571-599.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 2.

Robinson, P. M. (1991). Consistent nonparametric entropy-based testing. *Review of Economic Studies,* 58, 437-453.

Ruppert, D. & Carroll, R. J. (1980). Trimmed least squares estimation in the linear model. *Journal of the American Statistical Association,* 75, 828-838.

Shen, J. & Huang, Y-M. (1998). Nonparametric Estimation in a model with a trend. *Statistical Inference for Stochastic Processes,* 1, 43-60.

Song, K. (2009). Testing conditional independence via Rosenblatt Transform. *Annals of Statistics,* 37, 4011-4045.

Stinchcombe, M. B. & White, H. (1998). Consistent Specification testing with nuisance parameters present only under the alternative. *Econometric Theory,* 14, 295-325.

Su, L. & White, H. (2007). A consistent characteristic-function-based test for conditional independence. *Journal of Econometrics,* 141, 807-834.

Su, L. & White, H. (2008). A nonparametric Hellinger metric test for conditional independence. *Econometric Theory* 24, 829-864.

Su, L. & White, H. (2011). Testing conditional independence via empirical likelihood. Discussion paper, Dept. of Economics, UCSD.

Sun, S. & Chiang, C-Y. (1997). Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. *Journal of Applied Mathematics and Stochastic Analysis,* 10, 3-20.

Van der Vaart, A. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics.* New York: Springer.

Welsh, A. H. (1996). Robust estimation of smooth regression and spread functions and their derivatives. *Statistica Sinica,* 6, 347-366.

White, H. (1992). Nonparametric estimation of conditional quantiles using neural networks. In *Proceedings of the Symposium on the Interface*. New York: Springer, pp. 190-199.

White, H. (1994). *Estimation, Inference and Specification Analysis*. New York: Cambridge University Press.

White, H. & Chalak, K. (2010). Testing a conditional form of exogeneity. *Economic Letters,* 109, 88-90.

White, H. & Chalak, K. (2012). Identification and identification failure for treatment effects using structural systems. *Econometric Reviews,* forthcoming.

White, H. & Lu, X. (2010). Granger causality and dynamic structural systems. *Journal of Financial Econometrics*, 8, 193-243.

White, H. & Lu, X. (2011). Causal diagrams for treatment effect estimation with application to efficient covariate selection. *Review of Economics and Statistics,* forthcoming.

Yu, K. & Jones, M. C. (1998). Local linear quantile regression. *Journal of American Statistical Association,* 93*, 228-237.*

Zheng, J. X. (1998). A consistent nonparametric test of parametric regression models under conditional quantile restrictions. *Econometric Theory,* 14, 123-138.

A Proof of the Main Results

Let $\mu_{ix} \equiv \mu(X_i - x)/h$ and $K_{ix} \equiv K\left((x - X_i)/h\right)$. Let E_i denote expectation conditional on X_i. We use C to signify a generic constant whose exact value may vary from case to case and a^T to denote the transpose of a unless otherwise stated. We write $A_n \simeq B_n$ to signify that $A_n = B_n[1 + o_P(1)]$ as $n \to \infty$.

First we state a lemma that is used in the proof of Theorem 3.1.

Lemma A.1 *Let $\mathcal{V}_n(\tau, x; \Delta)$ be a vector function that satisfies (i) $-\Delta^T \mathcal{V}_n(\tau, x; \lambda\Delta) \geq -\Delta^T \mathcal{V}_n(\tau, x; \Delta)$ for all $\lambda \geq 1$ and $(\tau, x) \in \mathcal{T} \times \mathcal{X}$, (ii) $sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}} sup_{\|\Delta\|\leq M} \|\mathcal{V}_n(\tau, x; \Delta) + \mathcal{H}_n(\tau, x)\Delta - \mathcal{A}_n(\tau, x)\| = o_P(\kappa_n)$, where Δ may depend on (τ, x), $0 < M < \infty$, $\inf_{(\tau,x)\in\mathcal{T}\times\mathcal{X}} \lambda_{\min}(\mathcal{H}_n(\tau, x)) > 0$ a.s. as $n \to \infty$, and $\|\mathcal{A}_n(\tau, x)\| = O_P(1)\ \forall (\tau, x) \in \mathcal{T} \times \mathcal{X}$. Suppose that $\Delta_{n\tau x}$ satisfies $sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}} ||\mathcal{V}_n(\tau, x; \Delta_{n\tau x})|| = o_P(\kappa_n)$. Then*

(a) $sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}} \|\Delta_{n\tau x}\| = O_P(1)$;

(b) $\Delta_{n\tau x} = \mathcal{H}_n(\tau, x)^{-1} \mathcal{A}_n(\tau, x) + o_P(\kappa_n)$ uniformly in $(\tau, x) \in \mathcal{T} \times \mathcal{X}$.

This extends the pointwise result of Koenker and Zhao (1996, p.809) to a uniform result.

To prove Theorem 3.1, we need some additional notation. Let $\boldsymbol{\beta}_{0\tau x} = \boldsymbol{\beta}(\tau, x; h)$ denote the vector that contains the true value of $m(\tau, x)$ and its scaled partial derivatives with respect to x. Let $\boldsymbol{\beta}_{\tau x}$ denote the stack of the quantile regression coefficients $\beta_{\boldsymbol{j}}$, $0 \leq |\boldsymbol{j}| \leq p$, in the lexicographical order (see (3.1)), where the dependence of $\beta_{\boldsymbol{j}}$ on (τ, x) is made explicitly, but we suppress the dependence of both $\boldsymbol{\beta}_{0\tau x}$ and $\boldsymbol{\beta}_{\tau x}$ on h. Define

$$a_n \equiv \sqrt{nh^d},\ \widehat{\Delta}_{\tau x} \equiv a_n(\widehat{\boldsymbol{\beta}}(\tau, x; h) - \boldsymbol{\beta}_{0\tau x}),\ \text{and } \Delta_{\tau x} \equiv a_n(\boldsymbol{\beta}(\tau, x; h) - \boldsymbol{\beta}_{0\tau x}).$$

It follows that

$$\widehat{\Delta}_{\tau x} = \arg \min_{\Delta_{\tau x}\in\mathbb{R}^N} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \left(\boldsymbol{\beta}_{0\tau x} + a_n^{-1}\Delta_{\tau x}\right)^T \mu_{ix} \right) K_{ix}. \tag{A.1}$$

Let $V_n(\tau, x; \Delta) \equiv a_n^{-1} \sum_{i=1}^n \psi_\tau(Y_i - \left(\boldsymbol{\beta}_{0\tau x} + a_n^{-1}\Delta\right)^T \mu_{ix})\mu_{ix}K_{ix}$, and $\overline{V}_n(\tau, x; \Delta) \equiv a_n^{-1} \sum_{i=1}^n E_i[\psi_\tau((Y_i - (\boldsymbol{\beta}_{0\tau x} + a_n^{-1}\Delta)^T\mu_{ix})]\mu_{ix}K_{ix}$. The following lemmas constitute the main steps in the proof of Theorem 3.1.

Lemma A.2 *Suppose Assumptions A1-A6 hold. Then $||V_n(\tau, x; 0)|| = O_P(1)$ for each $(\tau, x) \in \mathcal{T} \times \mathcal{X}$.*

Lemma A.3 *Suppose Assumptions A1-A6 hold. Then*
$\sup_{\tau\in\mathcal{T}} \sup_{x\in\mathcal{X}} \sup_{\|\Delta\|\leq M} \left\| V_n(\tau, x; \Delta) - V_n(\tau, x; 0) - \left[\overline{V}_n(\tau, x; \Delta) - \overline{V}_n(\tau, x; 0)\right] \right\| = o_P(\kappa_n)$.

Lemma A.4 *Suppose Assumptions A1-A6 hold. Then*
$\sup_{\tau\in\mathcal{T}} \sup_{x\in\mathcal{X}} \sup_{\|\Delta\|\leq M} \left\| \overline{V}_n(\tau, x; \Delta) - \overline{V}_n(\tau, x; 0) + H_n(\tau, x)\Delta \right\| = o_P(\kappa_n)$.

Lemma A.5 *Suppose Assumptions A1-A6 hold. Then* $\sup_{\tau\in\mathcal{T}} \sup_{x\in\mathcal{X}} ||V_n(\tau, x; \widehat{\Delta}_{\tau x})|| = o_P(\kappa_n)$.

Proof of Theorem 3.1 We prove the theorem by checking that the conditions of Lemma A.1 hold with $\mathcal{A}_n(\tau, x) = V_n(\tau, x; 0)$, $\mathcal{V}_n(\tau, x; \Delta) = V_n(\tau, x; \Delta)$, $\mathcal{H}_n(\tau, x) = H_n(\tau, x)$, and $\Delta_{n\tau x} = \widehat{\Delta}_{\tau x}$. By Assumption A4, $\mathcal{H}_n(\tau, x)$ is positive definite a.s. as $n \to \infty$ for each $(\tau, x) \in (\mathcal{T}, \mathcal{X})$. By Lemma A.2, $||\mathcal{A}_n(\tau, x)|| = O_P(1)\ \forall (\tau, x) \in \mathcal{T} \times \mathcal{X}$. By Lemma A.5, $\sup_{\tau\in\mathcal{T}} \sup_{x\in\mathcal{X}} ||\mathcal{V}_n(\tau, x, \widehat{\Delta}_{\tau x})|| = o_P(\kappa_n)$. By Lemmas A.3-A.4,

$$\sup_{\tau\in\mathcal{T}} \sup_{x\in\mathcal{X}} \sup_{\|\Delta\|\leq M} |V_n(\tau, x; \Delta) - V_n(\tau, x; 0) + H_n(\tau, x)\Delta| = o_P(\kappa_n),$$

so that condition (ii) in Lemma A.1 is satisfied. Noting that $\psi_\tau(y)$ is a non-decreasing function of y, the function

$$-\Delta^T V_n(\tau, x; \lambda\Delta) = a_n^{-1} \sum_{i=1}^{n} \psi_\tau(Y_{ni} - \boldsymbol{\beta}_{0\tau x}^T \mu_{ix} - \lambda a_n^{-1}\Delta^T\mu_{ix}) \left(-\Delta^T\mu_{ix}\right) K_{ix}$$

is also non-decreasing in λ. This implies that condition (i) in Lemma A.1 holds. Consequently, we have $\sqrt{nh^d}(\widehat{\boldsymbol{\beta}}(\tau, x) - \boldsymbol{\beta}_{0\tau x}) = H_n(\tau, x)^{-1} V_n(\tau, x; 0) + o_P(\kappa_n)$ uniformly in $(\tau, x) \in \mathcal{T} \times \mathcal{X}$. ■

Proof of Corollary 3.2 We prove the corollary by showing (i) $\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}\|H_n(\tau,x)-H(\tau,x)\| = O_P(n^{-1/2}h^{-d/2}\sqrt{\log n}+h^{\gamma_0}) = o_P(1)$, and (ii) $J_n(\tau,x)=\overline{J}_n(\tau,x)+o_P\left(h^{d/2}\right)$. Then $\sqrt{nh^d}(\widehat{m}(\tau,x)-m(\tau,x)) = e_1^T H(\tau,x)^{-1}\overline{J}_n(\tau,x)[1+o_P(1)]+o_P\left(h^{d/2}\right)$. The proof of (i) is similar to but simpler than that of Corollary 2 in Masry (1996) because we only need convergence in probability, whereas Masry proved almost sure convergence. For (ii), let

$$R_n(\tau,x) = \frac{1}{\sqrt{nh^d}}\sum_{t=1}^{n}\left\{\mathbf{1}\left(Y_t\le m(\tau,X_t)\right)-\mathbf{1}(Y_t\le\boldsymbol{\beta}_0(\tau,x;h)^T\mu_{ix})\right\}\mu_{ix}K_{ix}.$$

Then $J_n(\tau,x)=\overline{J}_n(\tau,x)+R_n(\tau,x)$. We can write $R_n(\tau,x)$ as $\{R_n(\tau,x)-E[R_n(\tau,x)]\}+E[R_n(\tau,x)]$. The last term is $O(\sqrt{nh^d}h^{p+1})$ uniformly in (τ,x). Following the proof for W_{n1} in Lemma A.3, we can show the first term is $o_P(h^{d/2})$. Then $J_n(\tau,x)=\overline{J}_n(\tau,x)+o_P(h^{d/2})$ because $O(\sqrt{nh^d}h^{p+1})=o_P\left(h^{d/2}\right)$ by assumption. ■

Proof of Lemma A.1 To save space, let $\mathcal{A}_{n\tau x}\equiv\mathcal{A}_n(\tau,x)$ and $\sup_{\tau,x}\equiv\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}$. Fix $\epsilon>0$, $\sigma>0$.

$$\begin{aligned}
& P\left(\sup_{\tau,x}\inf_{\|\Delta\|=M}\left[-\Delta^T\mathcal{V}_n(\tau,x;\Delta)\right]<\sigma M\right) \\
\le\ & P\left(\sup_{\tau,x}\inf_{\|\Delta\|=M}-\Delta^T\mathcal{V}_n(\tau,x;\Delta)<\sigma M,\ \sup_{\tau,x}\inf_{\|\Delta\|=M}-\Delta^T\left[-\mathcal{H}_n(\tau,x)\Delta+\mathcal{A}_{n\tau x}\right]\ge 2\sigma M\right) \\
& +P\left(\sup_{\tau,x}\inf_{\|\Delta\|=M}-\Delta^T\left[\mathcal{A}_{n\tau x}-\mathcal{H}_n(\tau,x)\Delta\right]<2\sigma M\right) \\
\equiv\ & A_{1n}+A_{2n},\ \text{say.}
\end{aligned}\tag{A.2}$$

Noting that $\sup_{\tau,x}\inf_{\|\Delta\|=M}-\Delta^T\mathcal{V}_n(\tau,x;\Delta)<\sigma M$ and $\sup_{\tau,x}\inf_{\|\Delta\|=M}-\Delta^T\left[-\mathcal{H}_n(\tau,x)\Delta+\mathcal{A}_{n\tau x}\right]$ $\ge 2\sigma M$ implies that

$$\begin{aligned}
& \sup_{\tau,x}\left[\sup_{\|\Delta\|=M}\left\{-\Delta^T\left[-\mathcal{H}_n(\tau,x)\Delta+\mathcal{A}_{n\tau x}\right]-\left(-\Delta^T\mathcal{V}_n(\tau,x;\Delta)\right)\right\}\right] \\
=\ & \sup_{\tau,x}\sup_{\|\Delta\|=M}\left(-\Delta^T\left[-\mathcal{H}_n(\tau,x)\Delta+\mathcal{A}_{n\tau x}\right]\right)-\inf_{\tau,x}\inf_{\|\Delta\|=M}\left(-\Delta^T\mathcal{V}_n(\tau,x;\Delta)\right) \\
\ge\ & 2\sigma M-\sigma M=\sigma M,
\end{aligned}$$

we have

$$\begin{aligned}
A_{1n} &\le P\left(\sup_{\tau,x}\sup_{\|\Delta\|=M}\Delta^T\left[\mathcal{V}_n(\tau,x;\Delta)+\mathcal{H}_n(\tau,x)\Delta-\mathcal{A}_{n\tau x}\right]\ge\sigma M\right) \\
&\le P\left(\sup_{\tau,x}\sup_{\|\Delta\|=M}\left\|\mathcal{V}_n(\tau,x;\Delta)+\mathcal{H}_n(\tau,x)\Delta-\mathcal{A}_{n\tau x}\right\|\ge\sigma\right),
\end{aligned}\tag{A.3}$$

where the last line follows from the fact $\left\|\Delta^T B\right\|\le\|\Delta\|\,\|B\|$. For A_{2n}, noting that

$$\begin{aligned}
& -\Delta^T\left[\mathcal{A}_{n\tau x}-\mathcal{H}_n(\tau,x)\Delta\right] \\
=\ & -\Delta^T\mathcal{A}_{n\tau x}+\Delta^T\mathcal{H}_n(\tau,x)\Delta\ge-\|\Delta\|\,\|\mathcal{A}_{n\tau x}\|+c_1\|\Delta\|^2\ \ \forall\ (\tau,x)\in\mathcal{T}\times\mathcal{X},
\end{aligned}$$

where $c_1\equiv\inf_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}\lambda_{\min}\left(\mathcal{H}_n(\tau,x)\right)>0$ a.s. as $n\to\infty$, we have

$$\sup_{\tau,x}\inf_{\|\Delta\|=M}\left\{-\Delta^T\left[\mathcal{A}_{n\tau x}-\mathcal{H}_n(\tau,x)\Delta\right]\right\}\ge-M\|\mathcal{A}_{n\tau x}\|+c_1M^2.$$

It follows that

$$\begin{aligned} A_{2n} &= P\left(\sup_{\tau,x}\inf_{\|\Delta\|=M}\left\{-\Delta^T\left[\mathcal{A}_{n\tau x}-\mathcal{H}_n(\tau,x)\Delta\right]\right\}<2\sigma M\right) \\ &\leq P\left(-\|\mathcal{A}_{n\tau x}\|+c_1M<2\sigma\right)=P\left(\|\mathcal{A}_{n\tau x}\|>c_1M-2\sigma\right). \end{aligned} \tag{A.4}$$

This, together with (A.2) and (A.3), implies that

$$\begin{aligned} & P\left(\sup_{\tau,x}\inf_{\|\Delta\|=M}\left[-\Delta^T\mathcal{V}_n(\tau,x;\Delta)\right]<\sigma M\right) \\ \leq\ & P\left(\sup_{\tau,x}\sup_{\|\Delta\|=M}\|\mathcal{V}_n(\tau,x;\Delta)+\mathcal{H}_n(\tau,x)\Delta-\mathcal{A}_{n\tau x}\|\geq\sigma\right)+P\left(\|\mathcal{A}_{n\tau x}\|\geq c_1M-2\sigma\right). \end{aligned}$$

Given condition (ii) and the fact that $\|\mathcal{A}_{n\tau x}\|=O_P(1)$, one can choose $M>0$ and $n_0>0$ such that for $n\geq n_0$,

$$P\left(\sup_{\tau,x}\inf_{\|\Delta\|=M}\left[-\Delta^T\mathcal{V}_n(\tau,x;\Delta)\right]<\sigma M\right)<\epsilon. \tag{A.5}$$

Next, consider the case $\|\Delta\|\geq M$. Let $\lambda^*\equiv\|\Delta\|/M$ and $\Delta^*\equiv\Delta/\lambda^*$. Then $||\Delta^*||=M$. By condition (i), we have

$$-\Delta^{*T}\mathcal{V}_n(\tau,x;\Delta)=-\Delta^{*T}\mathcal{V}_n(\tau,x;\lambda^*\Delta^*)\geq-\Delta^{*T}\mathcal{V}_n(\tau,x;\Delta^*).$$

It follows that $\|\mathcal{V}_n(\tau,x;\Delta)\|\geq-\Delta^{*T}\mathcal{V}_n(\tau,x;\Delta^*)/M$. This, together with (A.5), implies

$$P\left(\sup_{\tau,x}\inf_{\|\Delta\|\geq M}\|\mathcal{V}_n(\tau,x;\Delta)\|<\sigma\right)\leq P\left(\sup_{\tau,x}\inf_{\|\Delta^*\|=M}-\Delta^{*T}\mathcal{V}_n(\tau,x;\Delta^*)<\sigma M\right)<\epsilon.$$

Noting that $\sup_{\tau,x}\|\mathcal{V}_n(\tau,x;\Delta_{n\tau x})\|=o_P(\kappa_n)$, we have $P\left(\kappa_n^{-1}\sup_{\tau,x}\|\mathcal{V}_n(\tau,x;\Delta_{n\tau x})\|\geq\sigma\right)\leq\epsilon$ for large enough n, say $n\geq n_1$. It follows that when $n\geq n_0\vee n_1$, we have

$$\begin{aligned} & P\left(\sup_{\tau,x}\|\Delta_{n\tau x}\|\geq M\right) \\ \leq\ & P\left(\sup_{\tau,x}\|\Delta_{n\tau x}\|\geq M,\kappa_n^{-1}\sup_{\tau,x}\|\mathcal{V}_n(\tau,x;\Delta_{n\tau x})\|<\sigma\right)+P\left(\kappa_n^{-1}\sup_{\tau,x}\|\mathcal{V}_n(\tau,x;\Delta_{n\tau x})\|\geq\sigma\right) \\ \leq\ & P\left(\kappa_n^{-1}\sup_{\tau,x}\inf_{\|\Delta\|\geq M}\|\mathcal{V}_n(\tau,x;\Delta)\|<\sigma\right)+\epsilon\leq 2\epsilon. \end{aligned}$$

That is, $\sup_{\tau,x}\|\Delta_{n\tau x}\|=O_P(1)$. Then by condition (ii), we have

$$\mathcal{V}_n(\tau,x;\Delta_{n\tau x})+\mathcal{H}_n(\tau,x)\Delta_{n\tau x}-\mathcal{A}_n(\tau,x)=o_P(\kappa_n)\ \text{ uniformly in }\ (\tau,x)\in\mathcal{T}\times\mathcal{X}.$$

It follows that $\mathcal{H}_n(\tau,x)\Delta_{n\tau x}=\mathcal{A}_n(\tau,x)+o_P(\kappa_n)$ uniformly in (τ,x) as $\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{X}}\|\mathcal{V}_n(\tau,x;\Delta_{n\tau x})\|$ $=o_P(\kappa_n)$. The result then follows. ■

Proof of Lemma A.2 Let $\mathbb{V}(\tau,x)\equiv-[V_n(\tau,x;0)-\overline{V}_n(\tau,x;0)]$. Then by the Minkowski inequality, $\|V_n(\tau,x;0)\|\leq\|\mathbb{V}(\tau,x)\|+\left\|\overline{V}_n(\tau,x;0)\right\|\equiv V_{1n}+V_{2n}$, where we suppress the dependence of V_{1n} and V_{2n} on (τ,x). Let $R_i(\tau,x)\equiv m(\tau,X_i)-\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}$. Then

$$R_i(\tau,x)=(p+1)\sum_{|\boldsymbol{k}|=p+1}\frac{1}{\boldsymbol{k}!}(X_i-x)^{\boldsymbol{k}}\int_0^1 D^{\boldsymbol{k}}m(\tau,x+s(X_i-x))(1-s)^p\,ds. \tag{A.6}$$

Following Masry (1996), we can show that $R_i(\tau,x) = O_P\left(h^{p+1}\right)$ uniformly in (τ,x) on the set $\{i : \|X_i - x\| \leq Ch\}$. By the mean value expansion, Assumptions A2(ii), A3(i), A5 and A6 we have

$$\begin{aligned} V_{2n} &= \left\| a_n^{-1}\sum_{i=1}^{n}\left[F_i(m_i(\tau,X_i)|X_i) - F_i(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i)\right]\mu_{ix}K_{ix}\right\| \\ &\leq Ca_n^{-1}\sum_{i=1}^{n}\left\|\left[n^{-1/2}c(\tau,X_i,i/n) + R_i(\tau,x)\right]\mu_{ix}K_{ix}\right\| \leq Ca_n^{-1}\left(n^{-1/2}+h^{p+1}\right)\sum_{i=1}^{n}\|\mu_{ix}K_{ix}\| \\ &= O_P\left(h^{d/2} + n^{1/2}h^{d/2+p+1}\right) = O_P(1). \end{aligned}$$

Now let $\xi_{ik} = \left[\mathbf{1}(Y_i \leq \boldsymbol{\beta}_{0\tau x}^T\mu_{ix}) - F_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i\right)\right]\mu_{ix,k}K_{ix}$, where $\mu_{ix,k}$ denotes the kth element of the N-vector μ_{ix}, $k = 1,2,\cdots,N$. Let $\mathbb{V}_k(\tau,x)$ denote the kth element of the N-vector $\mathbb{V}(\tau,x)$. Then $E_i[\xi_{ik}] = 0$, $\mathbb{V}_k(\tau,x) = a_n^{-1}\sum_{i=1}^n \xi_{ik}$, and $E[\mathbb{V}_k(\tau,x)] = 0$. By Assumption A1 and the Davydov inequality (e.g., Bosq, 1996, p.19), we have

$$\begin{aligned} \operatorname{Var}(\mathbb{V}_k(\tau,x)) &= a_n^{-2}\sum_{i=1}^{n}E\left(\xi_{ik}^2\right) + 2a_n^{-2}\sum_{1\leq i<j\leq n}^{n}\operatorname{Cov}\left(\xi_{ik},\xi_{jk}\right) \\ &\leq a_n^{-2}\sum_{i=1}^{n}E\left(\xi_{ik}^2\right) + 2c_2 n^{-1}\sum_{1\leq i<j\leq n}^{n}\alpha(j-i)^{\eta/(2+\eta)} \\ &\leq n^{-1}h^{-d}\sum_{i=1}^{n}E\left(\mu_{ix,k}^2K_{ix}^2\right) + 2c_2\sum_{s=1}^{\infty}\alpha(s)^{\eta/(2+\eta)} = O(1), \end{aligned}$$

where $\sup_{n\geq 1}\max_{1\leq i\leq n}E(h^{-d}\left\|\mu_{ix,k}K_{ix}\right\|_{2+\eta}^2) \leq c_2 < \infty$ by the compactness of $K(\cdot)$. Thus $\mathbb{V}_k(\tau,x) = O_P(1)$ by the Chebyshev inequality. It follows that $\mathbb{V}(\tau,x) = O_P(1)$. ■

Proof of Lemma A.3 Let $\mu_{ix,k}$ denote the kth element of the N-vector μ_{ix}, $k = 1,2,\cdots,N$. Let

$$\mathbb{S}_{nk}(\tau,x;\Delta) = a_n^{-1}\sum_{i=1}^{n}\left\{s_{ni,k}(\tau,x;\Delta) - E_i[s_{ni,k}(\tau,x;\Delta)]\right\},$$

where $s_{ni,k}(\tau,x;\Delta) = \left[\mathbf{1}(Y_i \leq (\boldsymbol{\beta}_{0\tau x} + a_n^{-1}\Delta)^T\mu_{ix}) - \mathbf{1}(Y_i \leq \boldsymbol{\beta}_{0\tau x}^T\mu_{ix})\right]\mu_{ix,k}K_{ix}$. Note that $\mathbb{S}_{nk}(\tau,x;\Delta)$ is the kth element of $-\{V_n(\tau,x;\Delta) - V_n(\tau,x;0) - \left[\overline{V}_n(\tau,x;\Delta) - \overline{V}_n(\tau,x;0)\right]\}$. It suffices to show that for each $k = 1,2,\cdots,N$,

$$\sup_{x\in\mathcal{X}}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}\left|\mathbb{S}_{nk}(\tau,x;\Delta)\right| = o_P(\kappa_n). \tag{A.7}$$

By the Minkowski inequality, (A.7) will hold if

$$\sup_{x\in\mathcal{X}}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}\left|\mathbb{S}_{nk}^{+}(\tau,x;\Delta)\right| = o_P(\kappa_n) \text{ and } \sup_{x\in\mathcal{X}}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}\left|\mathbb{S}_{nk}^{-}(\tau,x;\Delta)\right| = o_P(\kappa_n), \tag{A.8}$$

where $\mathbb{S}_{nk}^{+}$ and $\mathbb{S}_{nk}^{-}$ are analogous to $\mathbb{S}_{nk}$ but with $\mu_{ix,k}$ replaced by $\mu_{ix,k}^{+} \equiv \max\left(\mu_{ix,k},0\right)$ and $\mu_{ix,k}^{-} \equiv \max\left(-\mu_{ix,k},0\right)$, respectively. We only show the first part of (A.8), as the other case is similar.

Let $\mathbf{D} \equiv \{\Delta : \|\Delta \leq M\|\}$. By selecting $n_0 = O\left(d_{0n}^{-N}\right)$ grid points, $\Delta_1,\cdots,\Delta_{n_0}$, we can cover $\mathbf{D}$ by $\mathbf{D}_s = \{\Delta : \|\Delta - \Delta_s\| \leq d_{0n}\}$ where $d_{0n} = \kappa_n/\log\log n$. Let $d_{1n} = n^{-1/2}/\log\log n$, and $d_{2n} = n^{-1/2}/\log\log n$. By selecting $n_1 = O\left(d_{1n}^{-1}\right)$ grid points, $\tau_1,\tau_2,\cdots,\tau_{n_1}$ we cover the compact set $\mathcal{T}$ by $\mathcal{T}_j = \{\tau : |\tau - \tau_j| \leq d_{1n}\}$ for $j = 1,\cdots,n_1$. Similarly, we select $n_2 = O\left(h^{-d}d_{2n}^{-d}\right)$ grid points $x_1,...,x_{n_2}$ to cover the compact set $\mathcal{X}$ by $\mathcal{X}_l = \{x : \|x - x_l\| \leq d_{2n}h\}$, $l = 1,...,n_2$.

Let $\varphi_{ix}(\tau,\Delta)=(\boldsymbol{\beta}_{0\tau x}+a_n^{-1}\Delta)^T\mu_{ix}$. By the definition above (A.6), $\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}=m(\tau,x)+R_i(\tau,x)$. Then by Assumptions A3 (ii)-(iii), we can quantify several objects that are used subsequently:

$$\begin{aligned}\overline{d}_{1n} &\equiv \sup_{|\tau-\tau^*|\leq d_{1n},||\Delta-\Delta^*||\leq d_{0n}}\ \sup_{x\in\mathcal{X},\{K_{ix}>0\}}|\varphi_{ix}(\tau,\Delta)-\varphi_{ix}(\tau^*,\Delta^*)|\leq C(d_{1n}+h^{p+1}d_{1n}^{\gamma_0}+a_n^{-1}d_{0n}),\\ \overline{d}_{2n} &\equiv \sup_{|\tau-\tau^*|\leq d_{1n}}\ \sup_{x\in\mathcal{X},\{K_{ix}>0\}}\left|\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}-\boldsymbol{\beta}_{0\tau^* x}^T\mu_{ix}\right|\leq C(d_{1n}+h^{p+1}d_{1n}^{\gamma_0}),\\ \overline{d}_{3n} &\equiv \sup_{(\tau,\Delta)\in\mathcal{T}\times\mathbf{D}}\ \sup_{\|x-x^*\|\leq hd_{2n},\{K_{ix}>0\}}|\varphi_{ix}(\tau,\Delta)-\varphi_{ix^*}(\tau,\Delta)|\leq C\left(hd_{2n}+h^{p+1}(hd_{2n})^{\gamma_0}+a_n^{-1}d_{2n}\right).\end{aligned}$$

For brevity, let

$$\begin{aligned}\alpha_i^0(\tau,x,\Delta,\lambda_1) &\equiv \mathbf{1}(Y_i\leq\varphi_{ix}(\tau,\Delta)+\lambda_1\overline{d}_{1n})-F_i(\varphi_{ix}(\tau,\Delta)+\lambda_1\overline{d}_{1n}|X_i),\\ \alpha_i(\tau,x,\Delta) &\equiv \mathbf{1}(Y_i\leq\varphi_{ix}(\tau,\Delta))-F_i(\varphi_{ix}(\tau,\Delta)|X_i),\\ \beta_i^0(\tau,x,\lambda_2) &\equiv \mathbf{1}\left(Y_i\leq\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}+\lambda_2\overline{d}_{2n}\right)-F_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}+\lambda_2\overline{d}_{2n}|X_i\right),\\ \beta_i(\tau,x) &\equiv \mathbf{1}\left(Y_i\leq\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}\right)-F_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i\right).\end{aligned}$$

Clearly, $\alpha_i^0(\tau,x,\Delta,0)=\alpha_i(\tau,x,\Delta)$, and $\beta_i^0(\tau,x,0)=\beta_i(\tau,x)$. Let

$$\mathbb{W}(\tau,x,\Delta)\equiv a_n^{-1}\sum_{i=1}^n\alpha_i(\tau,x,\Delta)\mu_{ix,k}^+K_{ix},\ \text{and}\ \mathbb{W}^*(\tau,x)\equiv a_n^{-1}\sum_{i=1}^n\beta_i(\tau,x)\mu_{ix,k}^+K_{ix}.$$

Then $\mathbb{S}_{nk}^+(\tau,x;\Delta)=\mathbb{W}(\tau,x,\Delta)-\mathbb{W}^*(\tau,x)$. Fix $x_l\in\mathcal{X}_l$. Then

$$\begin{aligned}&\sup_{x\in\mathcal{X}}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}\mathbb{S}_{nk}^+(\tau,x;\Delta)\\ \leq\ &\max_{1\leq l\leq n_2}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}|\mathbb{W}(\tau,x_l,\Delta)-\mathbb{W}^*(\tau,x_l)|+\max_{1\leq l\leq n_2}\sup_{x\in\mathcal{X}_l}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}|\mathbb{D}(\tau,x,\Delta)-\mathbb{D}(\tau,x_l,\Delta)|\\ \equiv\ &W_{n1}+W_{n2},\end{aligned}$$

where $\mathbb{D}(\tau,x,\Delta)\equiv\mathbb{W}(\tau,x,\Delta)-\mathbb{W}^*(\tau,x)$. It suffices to show $W_{na}=o_P(\kappa_n)$ for $a=1,2$.

Step 1. We show $W_{n1}=o_P(1)$. Fix $(\tau_j,\Delta_s)\in\mathcal{T}_j\times\mathbf{D}_s$. Then

$$\begin{aligned}W_{n1} &= \max_{1\leq l\leq n_2}\max_{1\leq j\leq n_1}\sup_{\tau\in\mathcal{T}_j}\max_{1\leq s\leq n_0}\sup_{\Delta\in\mathbf{D}_s}|\mathbb{W}(\tau,x_l,\Delta)-\mathbb{W}^*(\tau,x_l)|\\ &\leq \max_{1\leq l\leq n_2}\max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0}|\mathbb{W}(\tau_j,x_l,\Delta_s)-\mathbb{W}^*(\tau_j,x_l)|\\ &\quad+\max_{1\leq l\leq n_2}\max_{1\leq j\leq n_1}\sup_{\tau\in\mathcal{T}_j}\max_{1\leq s\leq n_0}\sup_{\Delta\in\mathbf{D}_s}|\mathbb{D}_s(\tau,x_l,\Delta)-\mathbb{D}_s(\tau_j,x_l,\Delta_s)|\\ &\equiv W_{n11}+W_{n12}.\end{aligned}\tag{A.9}$$

It suffices to show that $W_{n1a}=o_P(1)$ for $a=1,2$.

Let $\epsilon>0$. Let $\xi_{ni}(\tau,x,\Delta)\equiv[\mathbf{1}(Y_i\leq\varphi_{ix}(\tau,\Delta))-\mathbf{1}\left(Y_i\leq\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}\right)]\mu_{ix,k}^+K_{ix}$, and $\overline{\xi}_{ni}(\tau,x,\Delta)\equiv\xi_{ni}(\tau,x,\Delta)-E_i[\xi_{ni}(\tau,x,\Delta)]$. By Assumptions A1, A2(ii) and A5, one can readily show that there exist some positive constants c_{ξ_1} and c_{ξ_2} such that

$$\left|\overline{\xi}_{ni}(\tau,x,\Delta)\right|\leq c_{\xi_1}\ \text{and}\ E\left[\sum_{i=a+1}^{a+p_n}\overline{\xi}_{ni}(\tau,x,\Delta)\right]^2\leq c_{\xi_2}a_n^{-1}p_nh^d$$

for any $a = 1, \cdots, p_n$ and $1 \leq p_n \leq n/2$. Let $\upsilon \in (\varsigma_\kappa, (1-d/\varsigma_1)/2)$. We can apply Lemma C.3 with $M_n = c_{\xi_1}$, $p_n = n^{1/2-\upsilon}h^{d/2}$, $\epsilon^* = n^{-1/2}h^{d/2}\kappa_n\epsilon$ and $\sigma^2(p_n) = c_{\xi_2}a_n^{-1}p_n h^d$ to obtain

$$\begin{aligned}
P\left(W_{n11} > \kappa_n\epsilon\right) &\leq n_0 n_1 n_2 \max_{1\leq l\leq n_2}\max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0} P\left(\left|\mathbb{W}\left(\tau_j, x_l, \Delta_s\right) - \mathbb{W}^*\left(\tau_j, x_l\right)\right| > \kappa_n\epsilon\right) \\
&= n_0 n_1 n_2 \max_{1\leq l\leq n_2}\max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0} P\left(\left|n^{-1}\sum_{i=1}^{n}\overline{\xi}_{ni}\left(\tau_j, x_l, \Delta_s\right)\right| > n^{-1/2}h^{d/2}\kappa_n\epsilon\right) \\
&\leq n_0 n_1 n_2 C_0 \exp\left(-\frac{\kappa_n^2 h^d\epsilon^2}{C_1 c_{\xi_2}a_n^{-1}h^d + 2C_2 c_{\xi_1}n^{-\upsilon}h^d\kappa_n\epsilon}\right) \\
&\quad + n_0 n_1 n_2 C_3\sqrt{\frac{c_{\xi_1}}{n^{-1/2}h^{d/2}\kappa_n\epsilon}}n^{1/2+\upsilon}h^{-d/2}\alpha\left(p_n+1\right) \\
&= n_0 n_1 n_2 C_0 \exp\left(-\frac{\epsilon^2}{C_1 c_{\xi_2}n^{-1/2}h^{-d/2}\kappa_n^{-2} + 2C_2 c_{\xi_1}n^{-\upsilon}\kappa_n^{-1}\epsilon}\right) \\
&\quad + n_0 n_1 n_2 O\left(n^{3/4+\upsilon}h^{-3d/4}\kappa_n^{-1/2}\right)\alpha(n^{1/2-\upsilon}h^{d/2}) \\
&= n_0 n_1 n_2 C_0 \exp\left(-\frac{\epsilon^2}{C_1 c_{\xi_2}n^{-1/2}h^{-d/2}\kappa_n^{-2} + 2C_2 c_{\xi_1}n^{-\upsilon}\kappa_n^{-1}\epsilon}\right) \\
&\quad + o\left(n^{3/4+\upsilon}\kappa_n^{-1/2}d_{0n}^{-N}d_{1n}^{-1}d_{2n}^{-d}h^{-3d/4}(n^{1/2-\upsilon}h^{d/2})^{-\beta_0}\right) \\
&\equiv p_{n1} + p_{n2},
\end{aligned}$$

where $\beta_0 = 1 + 16/\eta$. Clearly, by Assumption A6 the first term p_{n1} is $o(1)$ provided $\upsilon > \varsigma_\kappa$. Noting that $d_{0n} = \kappa_n/\log\log n$, $d_{1n} = d_{2n} = n^{-1/2}/\log\log n$, and $h = n^{-1/\varsigma_1}$, we have

$$\begin{aligned}
p_{n2} &= o\left(n^{3/4+\upsilon}\kappa_n^{-1/2}d_{0n}^{-N}\ d_{1n}^{-1}d_{2n}^{-d}h^{-3d/4}(n^{1/2-\upsilon}h^{d/2})^{-\beta_0}\right) \\
&= o\left(n^{3/4+\upsilon}\kappa_n^{-N-1/2}n^{(d+1)/2}h^{-3d/4}(n^{1/2-\upsilon}h^{d/2})^{-\beta_0}\left(\log\log n\right)^{N+d+1}\right) \\
&= o\left(\left(n^{3/4+\upsilon}n^{(d+1)/2}n^{-(1/2-\upsilon)\beta_0}h^{-(3+2\beta_0)d/4}\kappa_n^{-N-1/2}\right)\left(\log\log n\right)^{N+d+1}\right) \\
&= o\left(n^{5/4+d/2+\upsilon-(1/2-\upsilon)\beta_0}h^{-(3d+2\beta_0)/4}\kappa_n^{-N-1/2}\left(\log\log n\right)^{N+d+1}\right) \\
&= o\left(n^{5/4+d/2+\upsilon+3d/(4\varsigma_1)+(N+1/2)\varsigma_\kappa-(1/2-\upsilon-d/(2\varsigma_1))\beta_0}\left(\log\log n\right)^{N+d+1}\right).
\end{aligned}$$

Then $p_{n2} = o(1)$ because $\beta_0 > [5/2+d+2\upsilon+3d/(2\varsigma_1)+(2N+1)\varsigma_\kappa]/(1-2\upsilon-d/\varsigma_1)$ and $\upsilon < (1-d/\varsigma_1)/2$ by Assumption A6(iii). It follows that

$$W_{n11} = o_P(\kappa_n). \tag{A.10}$$

Now consider W_{n12}. By the monotonicity of the indicator function and the nonnegativity of $\mu_{ix,k}^{+}K_{ix}$, we have that for any $(\tau, \Delta) \in \mathcal{T}_j \times \mathbf{D}_s$,

$$\begin{aligned}
&\mathbb{D}\left(\tau, x, \Delta\right) - \mathbb{D}\left(\tau_j, x, \Delta_s\right) \\
= \ & a_n^{-1}\sum_{i=1}^{n}\left\{\left[\alpha_i\left(\tau, x, \Delta\right) - \alpha_i\left(\tau_j, x, \Delta_s\right)\right] - \left[\beta_i\left(\tau, x\right) - \beta_i\left(\tau_j, x\right)\right]\right\}\mu_{ix,k}^{+}K_{ix} \\
\geq \ & a_n^{-1}\sum_{i=1}^{n}\left\{\left[\alpha_i^0\left(\tau_j, x, \Delta_s, 1\right) - \alpha_i\left(\tau_j, x, \Delta_s\right)\right] - \left[\beta_i^0\left(\tau_j, x, -1\right) - \beta_i\left(\tau_j, x\right)\right]\right\}\mu_{ix,k}^{+}K_{ix} \\
& + a_n^{-1}\sum_{i=1}^{n}\left[F_i\left(\varphi_{ix}\left(\tau_j, \Delta_s\right) + \overline{d}_{1n}|X_i\right) - F_i(\varphi_{ix}\left(\tau, \Delta\right)|X_i)\right]\mu_{ix,k}^{+}K_{ix}
\end{aligned}$$

$$+a_n^{-1}\sum_{i=1}^{n}[F_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i\right)-F_i(\boldsymbol{\beta}_{0\tau_j x}^T\mu_{ix}-\overline{d}_{2n}|X_i)]\}\mu_{ix,k}^{+}K_{ix}.$$

Similarly,

$$\begin{aligned}
&\mathbb{D}(\tau,x,\Delta)-\mathbb{D}(\tau_j,x,\Delta_s)\\
\leq\ & a_n^{-1}\sum_{i=1}^{n}\left\{\left[\alpha_i^0(\tau_j,x,\Delta_s,-1)-\alpha_i(\tau_j,x,\Delta_s)\right]-\left[\beta_i^0(\tau_j,x,1)-\beta_i(\tau_j,x)\right]\right\}\mu_{ix,k}^{+}K_{ix}\\
&+a_n^{-1}\sum_{i=1}^{n}\left[F_i\left(\varphi_{ix}(\tau_j,\Delta_s)-\overline{d}_{01n}|X_i\right)-F_i(\varphi_{ix}(\tau,\Delta)|X_i)\right]\mu_{ix,k}^{+}K_{ix}\\
&+a_n^{-1}\sum_{i=1}^{n}[F_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i\right)-F_i(\boldsymbol{\beta}_{0\tau_j x}^T\mu_{ix}+\overline{d}_{1n}|X_i)]\mu_{ix,k}^{+}K_{ix}.
\end{aligned}$$

It follows that

$$\begin{aligned}
&\max_{1\leq j\leq n_1}\sup_{\tau\in\mathcal{T}_j}\max_{1\leq s\leq n_0}\sup_{\Delta\in\mathbf{D}_s}|\mathbb{D}(\tau,x,\Delta)-\mathbb{D}(\tau_j,x,\Delta_s)|\\
\leq\ & \max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0}\left|a_n^{-1}\sum_{i=1}^{n}\left[\alpha_i^0(\tau_j,x,\Delta_s,\pm1)-\alpha_i(\tau_j,x,\Delta_s)\right]\mu_{ix,k}^{+}K_{ix}\right|\\
&+\max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0}\left|a_n^{-1}\sum_{i=1}^{n}\left[\beta_i^0(\tau_j,x,\mp1)-\beta_i(\tau_j,x)\right]\mu_{ix,k}^{+}K_{ix}\right|\\
&+\max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0}\left|a_n^{-1}\sum_{i=1}^{n}\left[F_i\left(\varphi_{ix}(\tau_j,\Delta_s)+\overline{d}_{1n}|X_i\right)-F_i\left(\varphi_{ix}(\tau_j,\Delta_s)-\overline{d}_{1n}|X_i\right)\right]\mu_{ix,k}^{+}K_{ix}\right|\\
&+\max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0}\left|a_n^{-1}\sum_{i=1}^{n}[F_i(\boldsymbol{\beta}_{0\tau_j x}^T\mu_{ix}+\overline{d}_{2n}|X_i)-F_i(\boldsymbol{\beta}_{0\tau_j x}^T\mu_{ix}-\overline{d}_{2n}|X_i)]\mu_{ix,k}^{+}K_{ix}\right|\\
\equiv\ & D_{n1}(x)+D_{n2}(x)+D_{n3}(x)+D_{n4}(x).
\end{aligned}\tag{A.11}$$

Let $\alpha_{i,js}(x)=\left[\alpha_i^0(\tau_j,x,\Delta_s,\pm1)-\alpha_i(\tau_j,x,\Delta_s)\right]\mu_{ix,k}^{+}K_{ix}$. Then one can readily show that there exist some positive constants c_{α_1} and c_{α_2} such that $|\alpha_{i,js}(x)|\leq c_{\alpha_1}$ and $E[\sum_{i=a+1}^{a+p_n}\alpha_{i,js}(x)]^2\leq c_{\alpha_2}a_n^{-1}p_nh^d$ for any $a=1,\cdots,p_n$ and $1\leq p_n\leq n/2$. Following the proof for W_{n11}, we can readily show that

$$P\left(\max_{1\leq l\leq n_2}D_{n1}(x_l)\geq\kappa_n\epsilon\right)\leq n_0n_1n_2\max_{1\leq l\leq n_2}\max_{1\leq j\leq n_1}\max_{1\leq s\leq n_0}P\left(\left|n^{-1}\sum_{i=1}^{n}\alpha_{i,js}(x_l)\right|\geq n^{-1}a_n\kappa_n\epsilon\right)=o(1).$$

Similarly, one can show that $P(\max_{1\leq l\leq n_2}D_{n2}(x_l)\geq\kappa_n\epsilon)=o(1)$. Next, by the mean value expansion and Assumptions A2, A5, and A6,

$$\begin{aligned}
\max_{1\leq l\leq n_2}D_{n3}(x_l)\ &\leq\ C\overline{d}_{1n}\max_{1\leq l\leq n_2}\left|a_n^{-1}\sum_{i=1}^{n}\mu_{ix_l,k}^{+}K_{ix_l}\right|=O_P\left(n^{1/2}h^{d/2}\overline{d}_{1n}\right)\\
&=\ O_P\left(n^{1/2}h^{d/2}(d_{1n}+h^{p+1}d_{1n}^{\gamma_0}+a_n^{-1}d_{0n})\right)=o_P(\kappa_n).
\end{aligned}$$

Similarly $\max_{1\leq l\leq n_2}D_{n4}(x_l)=O_P\left(n^{1/2}h^{d/2}\overline{d}_{2n}\right)=o_P(\kappa_n)$. These results, together with (A.11), imply that

$$W_{n12}=o_P(\kappa_n).\tag{A.12}$$

Combining (A.9), (A.10) and (A.12) yields $W_{n1} = o_P(\kappa_n)$.

Step 2. We show $W_{n2} = o_P(1)$. Write

$$\begin{aligned}
\mathbb{D}(\tau, x, \Delta) - \mathbb{D}(\tau, x_l, \Delta) &= a_n^{-1} \sum_{i=1}^{n} \{\alpha_i(\tau, x, \Delta) - \beta_i(\tau, x)\} \left(\mu_{ix,k}^{+} K_{ix} - \mu_{ix_l,k}^{+} K_{ix_l}\right) \\
&\quad + a_n^{-1} \sum_{i=1}^{n} [\alpha_i(\tau, x, \Delta) - \alpha_i(\tau, x_l, \Delta)] \mu_{ix_l,k}^{+} K_{ix_l} \\
&\quad + a_n^{-1} \sum_{i=1}^{n} [\beta_i(\tau, x_l) - \beta_i(\tau, x)] \mu_{ix_l,k}^{+} K_{ix_l} \\
&\equiv D_{n1}(\tau, x, x_l, \Delta) + D_{n2}(\tau, x, x_l, \Delta) + D_{n3}(\tau, x, x_l), \text{ say.} \qquad \text{(A.13)}
\end{aligned}$$

It suffices to show that each of the three terms on the r.h.s. of (A.13) is $o_P(\kappa_n)$ uniformly.

First, we consider $D_{n1}(\tau, x, x_l, \Delta)$. Assumption A5 implies that for all $\|x_1 - x_2\| \leq \delta \leq c_k$,

$$|K(x_2) - K(x_1)| \leq \delta K^*(x_1), \qquad \text{(A.14)}$$

where $K^*(x) = \overline{C}\mathbf{1}(\|x\| \leq 2dc_k)$ for some constant $\overline{C}$ that depends on $\bar{c}_1$ and $\bar{c}_2$ in the assumption. For any $x \in \mathcal{X}_l$, $\|x - x_l\|/h \leq d_{2n}$. It follows from (A.14) that $|K_{ix} - K_{ix_l}| \leq d_{2n} K_{ix_l}^*$ with $K_{ix_l}^* \equiv K^*((x_l - X_i)/h)$, and

$$\begin{aligned}
\left| \left(\frac{X_i - x}{h}\right)^{\boldsymbol{k}} K_{ix} - \left(\frac{X_i - x_l}{h}\right)^{\boldsymbol{k}} K_{ix_l} \right| &\leq \left| \left(\frac{X_i - x}{h}\right)^{\boldsymbol{k}} \right| |K_{ix} - K_{ix_l}| + \left| \left(\frac{X_i - x}{h}\right)^{\boldsymbol{k}} - \left(\frac{X_i - x_l}{h}\right)^{\boldsymbol{k}} \right| K_{ix_l} \\
&\leq (2c_k)^{|\boldsymbol{k}|} d_{2n} K_{ix_l}^* + (2c_k)^{|\boldsymbol{k}|-1} d_{2n} K_{ix_l} \mathbf{1}(|\boldsymbol{k}| > 0) \\
&\leq C d_{2n}(K_{ix_l}^* + K_{ix_l}).
\end{aligned}$$

With this, we can show that for any $x \in \mathcal{X}_l$ such that $\|x - x_l\|/h \leq d_{2n}$, $||\mu_{ix,k}^{+} K_{ix} - \mu_{ix_l,k}^{+} K_{ix_l}|| \leq C d_{2n}(K_{ix_l}^* + K_{ix_l})$. It follows that

$$\begin{aligned}
& \max_{1 \leq l \leq n_2} \sup_{x \in \mathcal{X}_l} \sup_{\tau \in \mathcal{T}} \sup_{\|\Delta\| \leq M} |D_{n1}(\tau, x, x_l, \Delta)| \\
\leq\ & \max_{1 \leq l \leq n_2} \sup_{x \in \mathcal{X}_l} \sup_{\tau \in \mathcal{T}} \sup_{\|\Delta\| \leq M} a_n^{-1} \sum_{i=1}^{n} |\alpha_i(\tau, x, \Delta) - \beta_i(\tau, x)| \left| \mu_{ix,k}^{+} K_{ix} - \mu_{ix_l,k}^{+} K_{ix_l} \right| \\
\leq\ & \max_{1 \leq l \leq n_2} \sup_{x \in \mathcal{X}_l} a_n^{-1} \sum_{i=1}^{n} \left| \mu_{ixk}^{+} K_{ix} - \mu_{ix_l,k}^{+} K_{ix_l} \right| \\
\leq\ & C \max_{1 \leq l \leq n_2} a_n^{-1} d_{2n} \sum_{i=1}^{n} (K_{ix_l}^* + K_{ix_l}) = O_P\left(nh^d a_n^{-1} d_{2n}\right) = o_P(\kappa_n). \qquad \text{(A.15)}
\end{aligned}$$

Now we consider $D_{n2}(\tau, x, x_l, \Delta)$ defined in (A.13). Let $\lambda_3 \in \mathbb{R}$. Define

$$\mathcal{D}_{n2}(\tau, x_l, \Delta, \lambda_3) = a_n^{-1} \sum_{i=1}^{n} [1(Y_i \leq \varphi_{ix_l}(\tau, \Delta) + \lambda_3 \overline{d}_{3n}) - F_i\left(\varphi_{ix_l}(\tau, \Delta) + \lambda_3 \overline{d}_{3n} | X_i\right)] \mu_{ix_l,k}^{+} K_{ix_l}.$$

Note that $\max_{1 \leq l \leq n_2} \sup_{\tau \in \mathcal{T}} \sup_{\|\Delta\| \leq M} |\mathcal{D}_{n2}(\tau, x_l, \Delta, \lambda_3) - \mathcal{D}_{n2}(\tau, x_l, \Delta, 0)|$ is exactly like the object W_{n11} defined in (A.9). Following the proof of the probability order of W_{n11}, we can also show that

$$P\left(\max_{1 \leq l \leq n_2} \sup_{\tau \in \mathcal{T}} \sup_{\|\Delta\| \leq M} |\mathcal{D}_{n2}(\tau, x_l, \Delta, \lambda_3) - \mathcal{D}_{n2}(\tau, x_l, \Delta, 0)| \geq \kappa_n \epsilon \right) = o(1) \text{ for each } \lambda_3. \qquad \text{(A.16)}$$

Again, by the monotonicity of the indicator function and the CDF F_i, we have

$$
\begin{aligned}
& \max_{1\leq l\leq n_2}\sup_{x\in\mathcal{X}_l}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}|D_{n2}(\tau,x,x_l,\Delta)| \\
= & \max_{1\leq l\leq n_2}\sup_{x\in\mathcal{X}_l}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}\left|a_n^{-1}\sum_{i=1}^n[\alpha_i(\tau,x,\Delta)-\alpha_i(\tau,x_l,\Delta)]\,\mu_{ix_l,k}^+K_{ix_l}\right| \\
\leq & \max_{1\leq l\leq n_2}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}|\mathcal{D}_{n2}(\tau,x_l,\Delta,1)-\mathcal{D}_{n2}(\tau,x_l,\Delta,0)| \\
& +\max_{1\leq l\leq n_2}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}|\mathcal{D}_{n2}(\tau,x_l,\Delta,-1)-\mathcal{D}_{n2}(\tau,x_l,\Delta,0)| \\
& +\max_{1\leq l\leq n_2}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}\left|a_n^{-1}\sum_{i=1}^n[F_i(\varphi_{ix_l}(\tau,\Delta)+\overline{d}_{3n}|X_i)-F_i(\varphi_{ix_l}(\tau,\Delta)-\overline{d}_{3n}|X_i)]\mu_{ix_l,k}^+K_{ix_l}\right|.
\end{aligned}
$$

The first two terms are $o_P(1)$ by (A.16). For the last term, a mean value expansion implies that it is no bigger than

$$2C_1\overline{d}_{3n}a_n^{-1}\sup_{1\leq l\leq n_2}\sum_{i=1}^n\mu_{ix_l,k}^+K_{ix_l}=O_P\left(nh^d\overline{d}_{3n}a_n^{-1}\right)=o_P(\kappa_n).$$

It follows that

$$\max_{1\leq l\leq n_2}\sup_{x\in\mathcal{X}_l}\sup_{\tau\in\mathcal{T}}\sup_{\|\Delta\|\leq M}|D_{n2}(\tau,x,x_l,\Delta)|=o_P(\kappa_n). \tag{A.17}$$

Analogously, one can show that $\max_{1\leq l\leq n_2}\sup_{x\in\mathcal{X}_l}\sup_{\tau\in\mathcal{T}}|D_{n3}(\tau,x,x_l)|=o_P(1)$. This, together with (A.13), (A.16) and (A.17), implies that $W_{n2}=o_P(1)$. ■

Proof of Lemma A.4 Let $\overline{H}_n(\tau,x)\equiv n^{-1}\sum_{i=1}^n f_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i\right)\mu_{ix}\mu_{ix}^TK_{ix}$. Noting that $m(\tau,X_i)-\beta_{0\tau x}^T\mu_{ix}=O_P\left(h^{p+1}\right)$ uniformly in (τ,x) on the set $\{i:K_{ix}>0\}$, it is easy to show that uniformly in (τ,x), $\overline{H}_n(\tau,x)=n^{-1}h^{-d}\sum_{i=1}^n f_i(m(\tau,X_i)|X_i)\mu_{ix}\mu_{ix}^TK_{ix}+O_P\left(h^{p+1}\right)=H_n(\tau,x)+o_P(\kappa_n)$. Then by the Minkowski inequality, we have that

$$
\begin{aligned}
& \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}}\sup_{\|\Delta\|\leq M}\left\|\overline{V}_n(\tau,x;\Delta)-\overline{V}_n(\tau,x;0)+H_n(\tau,x)\Delta\right\| \\
\leq & \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}}\sup_{\|\Delta\|\leq M}\left\|\overline{V}_n(\tau,x;\Delta)-\overline{V}_n(\tau,x;0)+\overline{H}_n(\tau,x)\Delta\right\|+o_P(\kappa_n).
\end{aligned}
$$

By Assumptions A2, A5 and A6,

$$
\begin{aligned}
& \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}}\sup_{\|\Delta\|\leq M}\left\|\overline{V}_n(\tau,x;\Delta)-\overline{V}_n(\tau,x;0)+\overline{H}_n(\tau,x)\Delta\right\| \\
= & \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}}\sup_{\|\Delta\|\leq M}\left\|a_n^{-1}\sum_{i=1}^n\left[F_i\left(\left(\boldsymbol{\beta}_{0\tau x}+a_n^{-1}\Delta\right)^T\mu_{ix}|X_i\right)-F_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i\right)\right]\mu_{ix}K_{ix}-\overline{H}_n(\tau,x)\Delta\right\| \\
= & \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}}\sup_{\|\Delta\|\leq M}\left\|a_n^{-2}\sum_{i=1}^n\int_0^1\left[f_i\left(\left(\boldsymbol{\beta}_{0\tau x}+sa_n^{-1}\Delta\right)^T\mu_{ix}|X_i\right)-f_i\left(\boldsymbol{\beta}_{0\tau x}^T\mu_{ix}|X_i\right)\right]ds\ \mu_{ix}\mu_{ix}^TK_{ix}\Delta\right\| \\
\leq & \sup_{x\in\mathcal{X}}\sup_{\|\Delta\|\leq M}a_n^{-3}\sum_{i=1}^n\left\|C_2(X_i)\Delta^T\mu_{ix}\mu_{ix}\mu_{ix}^TK_{ix}\Delta\right\|\leq CM^2\sup_{x\in\mathcal{X}}a_n^{-3}\sum_{i=1}^n\|\mu_{ix}\|^3K_{ix} \\
= & O_P\left(nh^da_n^{-3}\right)=o_P(\kappa_n).\ ■
\end{aligned}
$$

Proof of Lemma A.5 By the proof of Lemma A2 in Ruppert and Carroll (1980) and Assumptions

A5-A6,

$$
\begin{aligned}
\sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}}||V_n(\tau,x;\widehat{\Delta}_{\tau x})|| &= \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}}\left\|n^{-1/2}\sum_{i=1}^n\psi_\tau(Y_i-\widehat{\boldsymbol{\beta}}_{\tau x}^T\mu_{ix})\mu_{ix}K_{ix}\right\| \\
&\leq \sup_{\tau\in\mathcal{T}}\sup_{x\in\mathcal{X}} n^{-1/2}\sum_{i=1}^n\mathbf{1}(Y_i-\widehat{\boldsymbol{\beta}}_{\tau x}^T\mu_{ix}=0)\left\|\mu_{ix}K_{ix}\right\| \\
&\leq 2Nn^{-1/2}\max_{1\leq i\leq n}\sup_{x\in\mathcal{X}}\|\mu_{ix}K_{ix}\| = o_P(\kappa_n). \blacksquare
\end{aligned}
$$

NOTATION. To prove the main results in Section 4, we apply some propositions in the next appendix. For notational simplicity, let $m_{i\tau}\equiv m(\tau,W_i)$, $\varepsilon_{i\tau}\equiv Y_i-m_{i\tau}$, $u_{i\tau}\equiv Y_i-m(\tau,X_i)$, and $\widehat{u}_{i\tau}\equiv Y_i-\widehat{m}(\tau,X_i)$. Then by Corollary 3.2, it is standard to show that $\max_{1\leq i\leq n}\sup_{\tau\in\mathcal{T}}|\widehat{u}_{i\tau}-u_{i\tau}|$ $=o_P(n^{-1/2}h_X^{-d_X/2}\sqrt{\log n})$. Clearly, $u_{i\tau}=\varepsilon_{i\tau}$ under H_0 and $u_{i\tau}=\varepsilon_{i\tau}+n^{-1/2}\delta(\tau,W_i)$ under H_{1n}. Let $\varphi_{i\gamma}=\varphi(W_i,\gamma)$. We use $F(\cdot|W_i)$, $F(\cdot|X_i)$, $f(\cdot|W_i)$, and $f(\cdot|X_i)$ to denote $F_{Y|W}(\cdot|W_i)$, $F_{Y|X}(\cdot|X_i)$, $f_{Y|W}(\cdot|W_i)$, and $f_{Y|X}(\cdot|X_i)$, respectively. Recall that S^c denotes the complex conjugate of S.

Proof of Theorem 4.1 The proof is a special case of that of Theorem 4.3, so we omit it. $\blacksquare$

Proof of Theorem 4.2 Following the proof of Theorem 4.3 below, we can show that $n^{-1/2}S_n(\tau,\gamma)=n^{-1/2}S_{n3}(\tau,\gamma)+o_P(1)$, where $S_{n3}(\tau,\gamma)\equiv n^{-1/2}\sum_{i=1}^n\{G(-\varepsilon_{i\tau}/\lambda_n)-G(-\widehat{u}_{i\tau}/\lambda_n)\}\varphi_{i\gamma}$. It follows that

$$
\begin{aligned}
n^{-1/2}S_n(\tau,\gamma) &= n^{-1}\sum_{i=1}^n\{G(-\varepsilon_{i\tau}/\lambda_n)-G(-\widehat{u}_{i\tau}/\lambda_n)\}\varphi_{i\gamma}+o_P(1) \\
&= \lambda_n^{-1}n^{-1}\sum_{i=1}^n\left\{G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau}-\varepsilon_{i\tau})\right\}\varphi_{i\gamma}+o_P(1) \\
&= \lambda_n^{-1}n^{-1}\sum_{i=1}^n G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)[m(\tau,W_i)-m(\tau,X_i)]\varphi_{i\gamma} \\
&\quad -\lambda_n^{-1}n^{-1}\sum_{i=1}^n G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)[m(\tau,X_i)-\widehat{m}(\tau,X_i)]\varphi_{i\gamma}+o_P(1) \\
&= \lambda_n^{-1}n^{-1}\sum_{i=1}^n G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)[m(\tau,W_i)-m(\tau,X_i)]\varphi_{i\gamma}+o_P(1) \\
&= n^{-1}\sum_{i=1}^n f(m_{i\tau}|W_i)[m(\tau,W_i)-m(\tau,X_i)]\varphi_{i\gamma}+o_P(1) \\
&= \boldsymbol{\Delta}_\delta(\tau,\gamma)+o_P(1),
\end{aligned}
$$

where $\boldsymbol{\Delta}_\delta(\tau,\gamma)\equiv E\{f(m_{i\tau}|W_i)[m(\tau,W_i)-m(\tau,X_i)]\varphi_{i\gamma}\}$. When φ is (G)CR, $\boldsymbol{\Delta}_\delta(\tau,\gamma)\neq 0$ in a set of positive Lebesgue measure. The test statistic thus diverges to ∞ under the alternative. $\blacksquare$

Proof of Theorem 4.3 Decompose $S_n(\tau,\gamma)=n^{-1/2}\sum_{i=1}^n[\tau-G(-\widehat{u}_{i\tau}/\lambda_n)]\varphi_{i\gamma}$ as follows:

$$
\begin{aligned}
S_n(\tau,\gamma) &= n^{-1/2}\sum_{i=1}^n[\tau-\mathbf{1}(\varepsilon_{i\tau}<0)]\varphi_{i\gamma}+n^{-1/2}\sum_{i=1}^n[\mathbf{1}(\varepsilon_{i\tau}<0)-G(-\varepsilon_{i\tau}/\lambda_n)]\varphi_{i\gamma} \\
&\quad +n^{-1/2}\sum_{i=1}^n[G(-\varepsilon_{i\tau}/\lambda_n)-G(-\widehat{u}_{i\tau}/\lambda_n)]\varphi_{i\gamma} \\
&\equiv S_{n1}(\tau,\gamma)+S_{n2}(\tau,\gamma)+S_{n3}(\tau,\gamma), \text{ say.}
\end{aligned}
$$

By Propositions B.4 and B.7, we have that, uniformly in (τ,γ), $S_{n2}(\tau,\gamma)=o_P(1)$, and that

$$S_{n3}(\tau,\gamma)=-c_0 n^{-1/2}\sum_{i=1}^{n} b(X_i,\gamma)\psi_\tau(\varepsilon_{i\tau})+\Delta(\tau,\gamma)+o_P(1),$$

where $c_0=e_1^T\mathbb{H}^{-1}\mathbb{B}$, $\mathbb{B}=\int\mu(v)K(v)dv$, and $b(X_j,\gamma)=E[\varphi(W_j,\gamma)|X_j]$. It follows that $S_n(\tau,\gamma)$ $\equiv\overline{S}_n(\tau,\gamma)+\boldsymbol{\Delta}(\tau,\gamma)+o_P(1)$, where

$$\overline{S}_n(\tau,\gamma)\equiv n^{-1/2}\sum_{i=1}^{n}\left[\varphi_{i\gamma}-c_0 b(X_i,\gamma)\right]\psi_\tau(\varepsilon_{i\tau}).$$

It suffices to show that $\overline{S}_n(\cdot,\cdot)\Rightarrow S_\infty(\cdot,\cdot)$, where $S_\infty(\cdot,\cdot)$ is defined in Theorem 4.1. Define the pseudometric ρ_d on $(\mathcal{T},\Gamma)$:

$$\rho_d((\tau,\gamma),(\tau',\gamma'))\equiv\left\{E\left|\eta_i(\gamma,\tau)-\eta_i(\gamma',\tau')\right|^r\right\}^{1/r},$$

where $r\geq 2$ and $\eta_i(\gamma,\tau)\equiv[\varphi(W_i,\gamma)-c_0 b(X_i,\gamma)]\psi_\tau(\varepsilon_{i\tau})$. By Theorem 10.2 of Pollard (1990), this follows if we have (i) total boundedness of a pseudometric space $((\mathcal{T},\Gamma),\rho_d)$, (ii) stochastic equicontinuity of $\left\{\overline{S}_n(\tau,\gamma):n\geq 1\right\}$, and (iii) finite dimensional (fidi) convergence.

Consider the class of functions

$$\mathcal{F}_1\equiv\left\{f_{(\tau,\gamma)}:(\tau,\gamma)\in\mathcal{T}\times\Gamma\right\},$$

where $f_{(\tau,\gamma)}:[0,1]\times W\to\mathbb{R}$ is defined by

$$f_{(\tau,\gamma)}(U_i,W_i)\equiv[\varphi(W_i,\gamma)-c_0 b(X_i,\gamma)][\tau-1(U_i\leq\tau)],$$

and $U_i\equiv F(Y_i|W_i)$. Let $\delta<1$ and (τ',γ') be generic element in $\mathcal{T}\times\Gamma$. Noting that

$$\begin{aligned}\eta_i(\tau,\gamma)-\eta_i(\tau',\gamma') &= [\varphi(W_i,\gamma)-c_0 b(X_i,\gamma)][\tau-1(U_i\leq\tau)-\tau'+1(U_i\leq\tau')]\\ &\quad+[\varphi(W_i,\gamma)-c_0 b(X_i,\gamma)-\varphi(W_i,\gamma')+c_0 b(X_i,\gamma')][\tau'-1(U_i\leq\tau')]\\ &\equiv \eta_{i1}(\tau,\tau',\gamma)+\eta_{i2}(\tau',\gamma,\gamma'),\text{ say,}\end{aligned}$$

by the repeated use of C_r-inequality, the uniform boundedness of $\varphi(\cdot,\cdot)$ and Assumption B4, we have

$$\begin{aligned}&E\sup_{|\tau-\tau'|\leq\delta_1,\|\gamma-\gamma'\|\leq\delta_2,\sqrt{\delta_1^2+\delta_2^2}\leq\delta}\left|\eta_i(\tau,\gamma)-\eta_i(\tau',\gamma')\right|^r\\ \leq\;& 2^{r-1}E\sup_{|\tau-\tau'|\leq\delta_1,\|\gamma-\gamma'\|\leq\delta_2,\sqrt{\delta_1^2+\delta_2^2}\leq\delta}\left|\eta_{i1}(\tau,\tau',\gamma)\right|^r\\ &+2^{r-1}E\sup_{|\tau-\tau'|\leq\delta_1,\|\gamma-\gamma'\|\leq\delta_2,\sqrt{\delta_1^2+\delta_2^2}\leq\delta}\left|\eta_{i2}(\tau',\gamma,\gamma')\right|^r\\ \leq\;& CE\sup_{|\tau-\tau'|\leq\delta_1}\left|\tau-\tau'\right|^r+CE\sup_{|\tau-\tau'|\leq\delta_1}\left|1(U_i\leq\tau)-1(U_i\leq\tau')\right|^r\\ &+CE\sup_{\|\gamma-\gamma'\|\leq\delta_2}\left|\varphi(W_i,\gamma)-\varphi(W_i,\gamma')\right|^r+CE\sup_{\|\gamma-\gamma'\|\leq\delta_2}\left|b(X_i,\gamma)-b(X_i,\gamma')\right|^r\\ \leq\;& C\delta_1^r+CP(|U_i-\tau|\leq\delta_1)+C\delta_2^{r\nu}+C\delta_2^{r\nu}\\ \leq\;& C\delta_1^r+C\delta_1+2C\delta_2^{r\nu}\leq 2C\delta^{1\wedge(r\nu)}=2C\delta^{\overline{\nu}},\end{aligned}$$

where $\overline{\nu}=\min(1,r\nu)$. That is, $\mathcal{F}_1$ is a class of uniformly bounded functions satisfying L_r-continuity. L_r-continuity implies that the bracketing number satisfies

$$N\left(\epsilon,\mathcal{F}_1,\|\cdot\|_{L_r(P)}\right)\leq C\left(\frac{1}{\epsilon}\right)^{(1+d_\Gamma)/\overline{\nu}},$$

which in conjunction with Assumption B2(ii) implies that

$$\int_0^1 \epsilon^{-\frac{\eta}{2+\eta}} N\left(\epsilon, \mathcal{F}_1, \|\cdot\|_{L_r(P)}\right)^{\frac{1}{6}} d\epsilon \leq C \int_0^1 \epsilon^{-\frac{\eta}{2+\eta}-\frac{1+d_\Gamma}{6\nu}} d\epsilon < \infty.$$

If follows that conditions (i)-(ii) are satisfied by Theorem 2.2 of Andrews and Pollard (1994). The fidi convergence holds by the Cramér-Wold device and a central limit theorem for bounded random variables under strong mixing conditions. See Corollary 5.1 in Hall and Heyde (1980, p. 132). We are left to demonstrate that the sample covariance kernel converges to that of the limiting Gaussian process $S_\infty(\cdot,\cdot)$. By the Davydov inequality,

$$\left|E\left[\overline{S}_n(\tau,\gamma)\overline{S}_n^c(\tau',\gamma')\right]\right| = \left|\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n E\left[\eta_i(\gamma,\tau)\eta_i^c(\gamma',\tau')\right]\right| \leq \frac{4c_\varphi^2}{n}\sum_{i=1}^n\sum_{j=1}^n \alpha(|i-j|) \leq 4c_\varphi^2\sum_{s=0}^\infty \alpha(s) < \infty.$$

It follows that $E\left[\overline{S}_n(\tau,\gamma)\overline{S}_n^c(\tau',\gamma')\right]$ is absolutely convergent, and $E\left[\overline{S}_n(\tau,\gamma)\overline{S}_n^c(\tau',\gamma')\right] \to E[\eta_1(\gamma,\tau)\eta_1^c(\gamma',\tau')] + \sum_{i=1}^\infty E\left[\eta_1(\gamma,\tau)\eta_{1+i}^c(\gamma',\tau')\right] + E\left[\eta_{1+i}(\gamma,\tau)\eta_1^c(\gamma',\tau')\right]$. This completes the proof of the theorem. ■

Proof of Theorem 4.4 Let P^* denote the probability conditional on the original sample $\mathfrak{D}_n \equiv \{(Y_t, W_t)\}_{t=1}^n$. Let E^* denote the expectation with respect to P^*. Rewrite $S_n^*(\tau,\gamma) = \sum_{i=1}^{n-L+1} s_{ni}(\zeta_i;\tau,\gamma)$, where

$$s_{ni}(\zeta_i;\tau,\gamma) \equiv n^{-1/2}\zeta_i \sum_{j=i}^{i+L-1}\left[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)\right]\left[\varphi_{jw} - c_0\widehat{b}(X_j,\gamma)\right] \tag{A.18}$$

and $\widehat{u}_{j\tau} \equiv Y_j - \widehat{m}(\tau, X_j)$. Define the envelope function of s_{ni} as

$$\overline{s}_{ni}(\zeta_i) \equiv |\zeta_i| n^{-1/2} \sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}\left|\sum_{j=i}^{i+L-1}\left[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)\right]\left[\varphi_{jw} - c_0\widehat{b}(X_j,\gamma)\right]\right|. \tag{A.19}$$

Conditional on $\mathfrak{D}_n$, the triangular array $\{s_{ni}(\zeta_i;\tau,\gamma)\}$ is independent within rows, so we can apply Theorem 10.6 of Pollard (1990) to show the weak convergence of $S_n^*(\cdot,\cdot)$ to $S_\infty(\cdot,\cdot)$. Recall that Pollard's theorem allows the function $s_{ni}(\cdot;\cdot,\cdot)$ to depend on both n and i.

Define the pseudo-metric

$$\rho_n(\tau,\tau';\gamma,\gamma') \equiv \left\{\sum_{i=1}^n E^*\left[\left|s_{ni}(\zeta_i;\tau',\gamma') - s_{ni}(\zeta_i;\tau,\gamma)\right|^2\right]\right\}^{1/2}. \tag{A.20}$$

By Theorem 10.6 of Pollard (1990), it suffices to verify the following five conditions:

(i) $\{s_{ni}\}$ is manageable in the sense of Definition 7.9 of Pollard (1990);

(ii) $E^*[S_n^*(\tau,\gamma)S_n^{*c}(\tau',\gamma)] \xrightarrow{P} \Upsilon(\tau,\tau';\gamma,\gamma')$ for every (τ,γ), (τ',γ') in $\mathcal{T}\times\Gamma$;

(iii) $\overline{\lim}_{n\to\infty}\sum_{i=1}^n E^*\left(\overline{s}_{ni}^2(z_i)\right)$ is stochastically bounded;

(iv) $\sum_{i=1}^n E^*\left(\overline{s}_{ni}^2(z_i)\ \mathbf{1}(\overline{s}_{ni}(z_i) > \epsilon)\right) \xrightarrow{P} 0$ for each $\epsilon > 0$;

(v) $\rho(\tau,\tau';\gamma,\gamma') \equiv \text{plim}_{n\to\infty}\rho_n(\tau,\tau';\gamma,\gamma')$ is well defined and, for all deterministic sequences $\{\tau_n',\gamma_n'\}$ and $\{\tau_n,\gamma_n\}$, if $\rho(\tau_n,\tau_n';\gamma_n,\gamma_n') \to 0$ then $\rho_n(\tau_n,\tau_n';\gamma_n,\gamma_n') \xrightarrow{P} 0$.

Step 1. We verify condition (i). In order for the triangular array of process $\{s_{ni}(\zeta_i;\tau,\gamma)\}$ to be manageable with respect to the envelope $\overline{s}_{ni}(\zeta_i)$, we need to find a deterministic function $\lambda(\epsilon_0)$ that bounds the the covering number of $\boldsymbol{\alpha}\odot\mathbf{S}_n \equiv \{\alpha_i s_{ni}(\zeta_i;\tau,\gamma) : (\tau,\gamma)\in\mathcal{T}\times\Gamma,\ \alpha_i$ are nonnegative finite constants for all $i = 1,\cdots n\}$ with $\sqrt{\log\lambda(\epsilon_0)}$ integrable. Here, the covering number refers to the smallest number of closed balls with radius $(\epsilon_0/2)\sqrt{\sum_{i=1}^n \alpha_i^2 \left|\overline{s}_{ni}(\zeta_i)\right|^2}$ whose unions cover $\boldsymbol{\alpha} \odot \mathbf{S}_n$. It follows that within each closed ball

$$\sum_{i=1}^n \alpha_i^2 E^* \left|s_{ni}(\zeta_i;\tau,\gamma) - s_{ni}(\zeta_i;\tau',\gamma')\right|^2 \leq \frac{\epsilon_0^2}{4}\sum_{i=1}^n \alpha_i^2 E^* \left|\overline{s}_{ni}(\zeta_i)\right|^2 \ \forall \epsilon_0 \in (0,1]. \tag{A.21}$$

First, we study the term on the left hand side (l.h.s.) of (A.21). Let $\widetilde{\varphi}_{j\gamma} = \varphi_{j\gamma} - c_0\widehat{b}(X_j,\gamma)$ and $\overline{\varphi}_{j\gamma} = \varphi_{j\gamma} - c_0 b(X_j,\gamma)$. By Propositions B.9-B.10 , we have that uniformly in (τ,γ),

$$L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)]\widetilde{\varphi}_{j\gamma} = L^{-1/2}\sum_{j=i}^{i+L-1}\psi_\tau(\varepsilon_{j\tau})\overline{\varphi}_{j\gamma} + o_P(1). \tag{A.22}$$

Note that the local alternative does not contribute to the above equation because $L^{-1/2}\lambda_n^{-1}\sum_{j=i}^{i+L-1} G^{(1)}(-\varepsilon_{j\tau}/\lambda_n)\, n^{-1/2}\delta(\tau,W_j)\,\widetilde{\varphi}_{j\gamma} = O_P(\sqrt{L/n}) = o_P(1)$ by Proposition B.3(i) and the boundedness of $\delta(\cdot,\cdot)$ and φ. It follows that

$$\begin{aligned}
&\sum_{i=1}^n \alpha_i^2 E^* \left|s_{ni}(\zeta_i;\tau,\gamma) - s_{ni}(\zeta_i;\tau',\gamma')\right|^2 \\
=\ &\frac{1}{n}\sum_{i=1}^{n-L+1}\alpha_i^2\frac{1}{L}\left|\sum_{j=i}^{i+L-1}[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)]\,\widetilde{\varphi}_{j\gamma} - \sum_{j=i}^{i+L-1}[\tau' + G(-\widehat{u}_{j\tau'}/\lambda_n)]\,\widetilde{\varphi}_{j\gamma'}\right|^2 \\
=\ &\frac{1}{n}\sum_{i=1}^{n-L+1}\alpha_i^2\frac{1}{L}\left|\sum_{j=i}^{i+L-1}\overline{\varphi}_{j\gamma}\psi_\tau(\varepsilon_{j\tau}) - \sum_{j=i}^{i+L-1}\overline{\varphi}_{j\gamma'}\psi_{\tau'}(\varepsilon_{j\tau'})\right|^2 + o_P(1) \\
&\xrightarrow{P}\sum_{i=1}^\infty \alpha_i^2\left[\Upsilon(\tau,\tau;\gamma,\gamma) - 2\Upsilon(\tau,\tau';\gamma,\gamma') + \Upsilon(\tau',\tau';\gamma',\gamma')\right] \\
\equiv\ &\sum_{i=1}^\infty \alpha_i^2\rho^2(\tau,\tau';\gamma,\gamma'),\ \text{say.}
\end{aligned} \tag{A.23}$$

Next, we study the term on the right hand side (r.h.s.) of (A.21). By Propositions B.9-B.10,

$$\begin{aligned}
\sum_{i=1}^n \alpha_i^2 E^* \left|\overline{s}_{ni}(\zeta_i)\right|^2 &= \frac{1}{n}\sum_{i=1}^n \alpha_i^2 \sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}\left|L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)]\,\widetilde{\varphi}_{j\gamma}\right|^2 \\
&\leq \frac{1}{n}\sum_{i=1}^n \alpha_i^2 \sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}\left|L^{-1/2}\sum_{j=i}^{i+L-1}\psi_\tau(\varepsilon_{j\tau})\,\overline{\varphi}_{j\gamma} + o_P(1)\right|^2 = O_P(1),
\end{aligned}$$

where the last equality follows because $\{L^{-1/2}\sum_{j=i}^{i+L-1}\psi_\tau(\varepsilon_{j\tau})\,\overline{\varphi}_{j\gamma}\}$ is an empirical process indexed by (τ,γ) by the proof of Theorem 4.3. It follows that $\sum_{i=1}^n \alpha_i^2 E^* \left|\overline{s}_{ni}(\zeta_i)\right|^2 = O_P(1)$. This, together with (A.21) and (A.23), implies that for any small $\epsilon_1 > 0$, there exists a large constant $M_1 \equiv M_1(\epsilon_1)$ such that the following holds

$$\sum_{i=1}^\infty \alpha_i^2\rho^2(\pi,\pi';\tau,\tau') \leq \frac{\epsilon_0^2}{4}M_1 \text{ for sufficiently large } n \tag{A.24}$$

on a set with probability $1-\epsilon_1$.

Now, partition the compact set $\mathcal{T}$ by finite points $\underline{\tau} = \tau_0 < \tau_1 < \cdots < \tau_{N1-1} < \tau_{N1} = \overline{\tau}$ such that $|\tau_j - \tau_{j-1}| = \delta_1$. By selecting grid points $\gamma_1, \cdots, \gamma_{N2}$, we can cover the compact set Γ by $\Gamma_k = \{\gamma : \|\gamma - \gamma_k\| \le \delta_2\}$. Let $(\tau, \gamma) \in [\tau_{j-1}, \tau_j] \times \Gamma_k$ $(j, k = 1, 2, \cdots)$. Note that $||\psi_\tau(\varepsilon_{i\tau})\overline{\varphi}_{i\gamma} - \psi_{\tau_j}(\varepsilon_{i\tau_j})\overline{\varphi}_{i\gamma}||_{2+\eta}^{2+\eta} \le C\delta_1$ and $||[\overline{\varphi}_{i\gamma} - \overline{\varphi}_{i\gamma_k}]\psi_\tau(\varepsilon_{i\tau_j})||_{2+\eta}^{2+\eta} \le C\delta_2$. Denote $\varphi_n(\tau, \gamma) = \sum_{i=1}^n \overline{\varphi}_{i\gamma}\psi_\tau(\varepsilon_{i\tau})$. Let $\delta \equiv \sqrt{\delta_1^2 + \delta_2^2} < 1$. Then by the Cauchy-Schwarz and Davydov inequalities, we have

$$\begin{aligned}
\Upsilon(\tau_j, \tau; \gamma_k, \gamma) &= \lim_{n\to\infty} \frac{1}{n} E\left|\varphi_n(\tau, \gamma) - \varphi_n(\tau_j, \gamma_k)\right|^2 \\
&\le 2\lim_{n\to\infty} \frac{1}{n} E\left|\sum_{i=1}^n \overline{\varphi}_{i\gamma}\left[\psi_\tau(\varepsilon_{i\tau}) - \psi_{\tau_j}(\varepsilon_{i\tau_j})\right]\right|^2 + 2\lim_{n\to\infty} \frac{1}{n} E\left|\sum_{i=1}^n [\overline{\varphi}_{i\gamma} - \overline{\varphi}_{i\gamma_k}]\psi_{\tau_j}(\varepsilon_{i\tau_j})\right|^2 \\
&\le C\left[O(\delta_1 + \delta_2) + \delta_1^{2/(2+\eta)} \sum_{s=1}^\infty \alpha(s)^{1/(2+\eta)} + \delta_2^{2/(2+\eta)} \sum_{s=1}^\infty \alpha(s)^{1/(2+\eta)}\right] \le C\delta^{2/(2+\eta)},
\end{aligned}$$

where the exact values of C vary across lines. This implies that

$$\rho^2(\tau_j, \tau; \gamma_k, \gamma) = \Upsilon(\tau, \tau; \gamma, \gamma) - 2\Upsilon(\tau_j, \tau; \gamma_k, \gamma) + \Upsilon(\tau_j, \tau_j; \gamma_k, \gamma_k) \le C_1\delta^{2/(2+\eta)}$$

for large enough C_1. Consequently, if we choose $\delta = \epsilon_0^{2+\eta}$, then $\sum_{i=1}^\infty \alpha_i^2\rho^2(\pi_j, \pi; \tau_k, \tau) \le C_1\epsilon_0^2 \sum_{i=1}^\infty \alpha_i^2$, so that (A.24) can be satisfied for sufficiently large n and M_1. It follows that the capacity bound is $O(\delta^{-2}) = O(\epsilon_0^{-2(2+\eta)})$ and the integrability condition is satisfied.

Step 2. We verify condition (ii). Recall $\widetilde{\varphi}_{j\gamma} = \varphi_{j\gamma} - c_0\widehat{b}(X_j, \gamma)$ and $\overline{\varphi}_{j\gamma} = \varphi_{j\gamma} - c_0 b(X_j, \gamma)$. By Propositions B.9-B.10, $L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - G(-\widehat{u}_{j\tau}/\lambda_n)]\widetilde{\varphi}_{j\gamma} = L^{-1/2}\sum_{j=i}^{i+L-1}\psi_\tau(\varepsilon_{j\tau})\overline{\varphi}_{j\gamma} + o_P(1)$ uniformly in (τ, γ). It follows that

$$\begin{aligned}
E^*[S_n^*(\tau, \gamma)S_n^{*c}(\tau', \gamma')] &= n^{-1}\sum_{i=1}^{n-L+1}\frac{1}{L}\sum_{j_1=i}^{i+L-1}\sum_{j_2=i}^{i+L-1}[\tau - G(-\widehat{u}_{j_1\tau}/\lambda_n)]\widetilde{\varphi}_{j_1\gamma}[\tau' - G(-\widehat{u}_{j_2\tau'}/\lambda_n)]\widetilde{\varphi}_{j_2\gamma'}^c \\
&= n^{-1}\sum_{i=1}^{n-L+1}\frac{1}{L}\sum_{j_1=i}^{i+L-1}\sum_{j_2=i}^{i+L-1}\psi_\tau(\varepsilon_{j_1\tau})\overline{\varphi}_{j_1\gamma}\psi_\tau(\varepsilon_{j_2\tau'})\overline{\varphi}_{j_2\gamma'}^c + o_P(1) \\
&\equiv \overline{S}_n^* + o_P(1),
\end{aligned}$$

where we suppress the dependence of $\overline{S}_n^* \equiv \overline{S}_n^*(\tau, \gamma, \tau', \gamma')$ on $(\tau, \gamma, \tau', \gamma')$. First,

$$E(\overline{S}_n^*) = \frac{1}{n}\sum_{i=1}^{n-L+1}\frac{1}{L}\sum_{j_1=i}^{i+L-1}\sum_{j_2=i}^{i+L-1}E[\psi_\tau(\varepsilon_{j_1\tau})\psi_\tau(\varepsilon_{j_2\tau'})\overline{\varphi}_{j_1\gamma}\overline{\varphi}_{j_2\gamma'}^c] \to \Upsilon(\tau, \tau'; \gamma, \gamma').$$

To show $\text{Var}(\overline{S}_n^*) = o(1)$, let $\xi_{ni}^* \equiv \xi_{ni}^*(\tau, \tau'; \gamma, \gamma') = L^{-2}\sum_{j_1=i}^{i+L-1}\sum_{j_2=i}^{i+L-1}\psi_\tau(\varepsilon_{j_1\tau})\psi_\tau(\varepsilon_{j_2\tau'})\overline{\varphi}_{j_1\gamma}\overline{\varphi}_{j_2\gamma'}^c$, and let $\xi_{ni}(\tau, \gamma) \equiv L^{-1}\sum_{j=i}^{i+L-1}\psi_\tau(\varepsilon_{j\tau})\overline{\varphi}_{j\gamma}$. Then by the Cauchy inequality,

$$\|\xi_{ni}^*\|_8 = \|\xi_{ni}(\tau, \gamma)\xi_{ni}(\tau', \gamma')\|_8 \le \|\xi_{ni}(\tau, \gamma)\|_{16}\|\xi_{ni}(\tau', \gamma')\|_{16}.$$

By Lemma 3.1 of Andrews and Pollard (1994) with Q =16, $\|\xi_{ni}(\tau, \gamma)\|_{16}^{16} = E|\frac{1}{L}\sum_{j=i}^{i+L-1}\psi_\tau(\varepsilon_{j\tau})\varphi_{j\gamma}|^{16}$ $= O(L^{-8})$. Consequently, $E|\xi_{ni}^*|^8 = O(L^{-8})$. Let $\kappa_{4n} = \sup_{i\le n}\sup_{\tau,\tau';\gamma,\gamma'}E|\xi_{ni}^*|^8 = O(L^{-8})$ and $\kappa_{2n} = \sup_{i\le n}\sup_{\tau,\tau';\gamma,\gamma'}E|\xi_{ni}^*|^4 = O(L^{-4})$. By Lemma A.1(b) of Inoue (2001) with $\delta = 2$ (see also Lemma 9 of Bühlmann (1994)),

$$E\left|\frac{L}{n}\sum_{i=1}^{n-L+1}\xi_{ni}^*\right|^4 = O\left(L^4 n^{-4} L^2\left(n^2\kappa_{4n}^{1/2} + n\kappa_{2n}\right)\right) = O(n^{-2}L^2) = o(1).$$

Hence $\overline{S}_n^* = \Upsilon(\tau,\tau';\gamma,\gamma') + o_P(1)$ by the Chebyshev inequality.

Step 3. We verify condition (iii). This follows from the proof in Step 1 by taking $\alpha_i = 1\ \forall i$.

Step 4. We verify condition (iv). By the conditional Chebyshev inequality and Propositions B.9-B.10,

$$\begin{aligned} P^*\left(\overline{s}_{ni}(\zeta_i) > \epsilon\right) &\leq \frac{L}{n\epsilon^2}\left\{\sup_{\gamma\in -}\sup_{\tau\in\mathcal{T}}\left|\frac{1}{\sqrt{L}}\sum_{j=i}^{i+L-1}\left(\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right)\widetilde{\varphi}_{j\gamma}\right|\right\}^2 \\ &= \frac{L}{n\epsilon^2}\left\{\sup_{\gamma\in -}\sup_{\tau\in\mathcal{T}}\left|\frac{1}{\sqrt{L}}\sum_{j=i}^{i+L-1}\psi_\tau\left(\varepsilon_{j\tau}\right)\overline{\varphi}_{j\gamma}\right|^2 + o_P(1)\right\} = O_P\left(\frac{L}{n}\right). \end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned} &\sum_{i=1}^n E^*\left(\left|\overline{s}_{ni}(\zeta_i)\right|^2 \mathbf{1}\left(\overline{s}_{ni}(\zeta_i) > \epsilon\right)\right) \\ = {} & \frac{1}{n}\sum_{i=1}^n E^*\left\{\zeta_i^2 \sup_{\gamma\in -}\sup_{\tau\in\mathcal{T}}\left|\sum_{j=i}^{i+L-1}\left[\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right]\widetilde{\varphi}_{j\gamma}\right|^2 \mathbf{1}\left(\overline{s}_{ni}(\zeta_i) > \epsilon\right)\right\} \\ \leq {} & \frac{1}{n}\sum_{i=1}^n\left\{\frac{1}{L^2}\sup_{\gamma\in -}\sup_{\tau\in\mathcal{T}}\left|\sum_{j=i}^{i+L-1}\left[\tau - G\left(-\widehat{u}_{j\tau}/\lambda_n\right)\right]\widetilde{\varphi}_{j\gamma}\right|^4 P^*\left(\overline{s}_{ni}(\zeta_i) > \epsilon\right)\right\}^{1/2} = O_P(\sqrt{L/n}) = o_P(1). \end{aligned}$$

The result follows.

Step 5. We verify condition (v). From the verification of condition (i), we know that $\rho^2(\tau,\tau';\gamma,\gamma') = \mathrm{plim}_{n\to\infty}\,\rho_n^2(\tau,\tau';\gamma,\gamma')$ is well defined. If $\rho(\tau_n,\tau_n';\gamma_n,\gamma_n') \to 0$, then $\rho_n(\tau_n,\tau_n';\gamma_n,\gamma_n') \leq |\rho_n(\tau_n,\tau_n';\gamma_n,\gamma_n') - \rho(\tau_n,\tau_n';\gamma_n,\gamma_n')| + \rho(\tau_n,\tau_n';\gamma_n,\gamma_n') \xrightarrow{P} 0$. ■

B Propositions

In this appendix, we prove some propositions used in the proof of Theorems 4.1-4.4 and that apply some technical lemmas in the next appendix. Recall $\varphi_{i\gamma} \equiv \varphi(W_i,\gamma)$, $m_{i\tau} \equiv m(\tau, W_i)$, $g(\cdot) \equiv G^{(1)}(\cdot)$, and $\mu_{ix} \equiv \mu((X_i - x)/h)$. Let $K_{ij} \equiv K((X_i - X_j)/h)$, $\mu_{i,j} \equiv \mu((X_i - X_j)/h)$, and $g^{(s-1)}(\cdot) \equiv G^{(s)}(\cdot)$ for $s = 2, 3$. Here we use $E_i(\cdot)$ to denote expectation conditional on $W_i \equiv \left(X_i^T, Z_i^T\right)^T$ instead of X_i.

Proposition B.1 *(i)* $E\left[G\left(-\varepsilon_{i\tau}/\lambda_n\right)\right] - \tau = O\left(\lambda_n^q\right)$;
(ii) $E\left[G\left(-\varepsilon_{i\tau}/\lambda_n\right) - \mathbf{1}\left(\varepsilon_{i\tau} \leq 0\right)\right]^2 = O\left(\lambda_n\right)$;
(iii) $\lambda_n^{-1}E\left[G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right] = E\left[f\left(m_{i\tau}|W_i\right)\right] + O\left(\lambda_n^q\right)$.

Proof. Under Assumptions B3(i) and B5(i)-(iii), using change of variables, integration by parts, and a qth order Taylor expansion yields

$$\begin{aligned} & E\left[G\left(-\varepsilon_{i\tau}/\lambda_n\right)|W_i\right] \\ = {} & E\left[G\left(-\left(Y_i - m_{i\tau}\right)/\lambda_n\right)|W_i\right] = \int_{-\infty}^{\infty} G\left(-\left(y - m_{i\tau}\right)/\lambda_n\right)dF\left(y|W_i\right) \\ = {} & -\int_{-\infty}^{\infty} F\left(y|W_i\right)dG\left(-\left(y - m_{i\tau}\right)/\lambda_n\right) = \int_{-\infty}^{\infty} F\left(m_{i\tau} + v\lambda_n|W_i\right)g\left(-v\right)dv \end{aligned}$$

$$= \tau + \int_{-\infty}^{\infty} \sum_{s=1}^{q} \frac{1}{s!} F^{(s)}\left(m_{i\tau}|W_i\right) v^s \lambda_n^s g\left(v\right) dv + r_{ni} = \tau + \frac{\lambda_n^q \kappa_q}{q!} F^{(q)}\left(m_{i\tau}|W_i\right) + r_{ni},$$

where $\kappa_q \equiv \int_{-\infty}^{\infty} v^q g\left(v\right) dv,\ r_{ni} \equiv (\lambda_n^q/(q-1)!) \int_{-\infty}^{\infty} \int_0^1 [F^{(q)}\left(m_{i\tau} + sv\lambda_n|W_i\right) - F_i^{(q)}\left(m_{i\tau}|W_i\right)] v^q g\left(v\right) \left(1-s\right)^{q-1} dsdv$. By Assumption B3(i), the dominated convergence theorem, and the law of iterated expectations, $E\left|r_{ni}\right| = o(\lambda_n^q)$, and $E\left[G\left(-\varepsilon_{i\tau}/\lambda_n\right)\right] = \tau + \frac{\lambda_n^q \kappa_q}{q!} E\left[F^{(q)}\left(m_{i\tau}|W_i\right)\right] + o\left(\lambda_n^q\right)$. This proves (i). For (ii), using Assumptions B1, B5(i) and (iii), we have $G\left(0\right) \equiv \int_{-\infty}^{0} g\left(v\right) dv = 0.5$, and

$$\begin{aligned}
& E\left[G\left(-\varepsilon_{i\tau}/\lambda_n\right) - \mathbf{1}\left(\varepsilon_{i\tau} \le 0\right)\right]^2 \\
=\ & E\left[G^2\left(-\varepsilon_{i\tau}/\lambda_n\right)\right] + E\left[\mathbf{1}\left(\varepsilon_{i\tau} \le 0\right)\right] - 2E\left[G\left(-\varepsilon_{i\tau}/\lambda_n\right) \mathbf{1}\left(\varepsilon_{i\tau} \le 0\right)\right] \\
=\ & E\left[\int_{-\infty}^{\infty} G^2\left(-\left(y - m_{i\tau}\right)/\lambda_n\right) dF\left(y|W_i\right)\right] + \tau - 2E\left[\int_{-\infty}^{m_{i\tau}} G\left(-\left(y - m_{i\tau}\right)/\lambda_n\right) dF\left(y|W_i\right)\right] \\
=\ & -E\left[\int_{-\infty}^{\infty} F\left(m_{i\tau} + \lambda_n v|W_i\right) dG^2\left(-v\right)\right] + \tau - 2\left\{G\left(0\right)\tau + E\left[\int_{-\infty}^{0} F\left(m_{i\tau} + \lambda_n v|W_i\right) g\left(-v\right) dv\right]\right\} \\
=\ & -\tau \int_{-\infty}^{\infty} dG^2\left(-v\right) + \tau - 2\tau + O\left(\lambda_n\right) = O\left(\lambda_n\right).
\end{aligned}$$

By Assumptions B3(i) and B5(ii)-(iii),

$$\begin{aligned}
\lambda_n^{-1} E\left[G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right] &= E\left[\lambda_n^{-1} \int_{-\infty}^{\infty} G^{(1)}\left(-\left(y - m_{i\tau}\right)/\lambda_n\right) f\left(y|W_i\right) dy\right] \\
&= E\left[\int_{-\infty}^{\infty} f\left(m_{i\tau} + \lambda_n v|W_i\right) g\left(v\right) dv\ \right] \\
&= E\left[f\left(m_{i\tau}|W_i\right) + \int_{-\infty}^{\infty} \sum_{s=1}^{q} \frac{1}{s!} f_i^{(s)}\left(m_{i\tau}|W_i\right) \lambda_n^s v^s g\left(v\right) dv\right] + o\left(\lambda_n^q\right) \\
&= E\left[f\left(m_{i\tau}|W_i\right)\right] + \frac{\lambda_n^q \kappa_q}{q!} E\left[f^{(q)}\left(m_{i\tau}|W_i\right)\right] + o\left(\lambda_n^q\right) = E\left[f\left(m_{i\tau}|W_i\right)\right] + O\left(\lambda_n^q\right).
\end{aligned}$$

■

Proposition B.2 *(i)* $\sup_{\tau \in \mathcal{T}} \frac{1}{n\lambda_n} \sum_{i=1}^{n} E\left|G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| = O\left(1\right);$
(ii) $\sup_{\tau \in \mathcal{T}} \frac{1}{n\lambda_n^2} \sum_{i=1}^{n} E\left|G^{(2)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| = O\left(1\right);$
(iii) $\sup_{\tau \in \mathcal{T}} \frac{1}{n\lambda_n^3} \sum_{i=1}^{n} E\left|G^{(3)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| = O\left(1\right).$

Proof. By Assumptions A3(i), B3(i), and B5(ii)-(iii), we have, uniformly in τ,

$$\begin{aligned}
\lambda_n^{-1} E\left|G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| &= E\left[\lambda_n^{-1} \int_{-\infty}^{\infty} \left|g\left(-\left(y - m_{i\tau}\right)/\lambda_n\right)\right| f\left(y|W_i\right) dy\right] \\
&= -E\left[\int_{-\infty}^{\infty} f\left(m_{i\tau} + \lambda_n v|W_i\right) \left|g\left(v\right)\right| dv\right] \\
&= E\left[\int_{-\infty}^{\infty} \left(f\left(m_{i\tau}|W_i\right) + \sum_{s=1}^{q} \frac{1}{s!} f^{(s)}\left(m_{i\tau}|W_i\right) \lambda_n^s v^s\right) \left|g\left(v\right)\right| dv\right] + o\left(\lambda_n^q\right) \\
&= E\left[f\left(m_{i\tau}|W_i\right)\right] \int_{-\infty}^{\infty} \left|g\left(v\right)\right| dv + O\left(\lambda_n\right) = O\left(1\right).
\end{aligned}$$

Similarly, uniformly in τ,

$$\begin{aligned}\lambda_n^{-2}E\left|G^{(2)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right| &= E\left[\lambda_n^{-2}\int_{-\infty}^{\infty}\left|g^{(1)}\left(-\left(y-m_{i\tau}\right)/\lambda_n\right)\right|f\left(y|W_i\right)dy\right]\\ &= -\lambda_n^{-1}E\left[\int_{-\infty}^{\infty}f\left(m_{i\tau}+\lambda_n v|W_i\right)\left|g^{(1)}\left(v\right)\right|dv\right]\\ &= E\left[\int_{-\infty}^{\infty}\left(f\left(m_{i\tau}|W_i\right)+\sum_{s=1}^{q}\frac{1}{s!}f^{(s)}\left(m_{i\tau}|W_i\right)\lambda_n^s v^s\right)\operatorname{sgn}\left(g^{(1)}\left(v\right)\right)dg\left(v\right)\right]+o(\lambda_n^{q-1})\\ &= E\left[f^{(1)}\left(m_{i\tau}|W_i\right)\right]\int_{-\infty}^{\infty}\left|g^{(1)}\left(v\right)\right|dv+O\left(\lambda_n\right)=O\left(1\right).\end{aligned}$$

The proof of (iii) is similar and thus omitted. ■

Proposition B.3 *(i)* $\sup_{\tau\in\mathcal{T}}\frac{1}{n\lambda_n}\sum_{i=1}^n\left|G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right|=O_P\left(1\right)$;
(ii) $\sup_{\tau\in\mathcal{T}}\frac{1}{n\lambda_n^2}\sum_{i=1}^n\left|G^{(2)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right|=O_P(1+n^{-1/2}\lambda_n^{-3/2}\sqrt{\log n})$;
(iii) $\sup_{\tau\in\mathcal{T}}\frac{1}{n\lambda_n^3}\sum_{i=1}^n\left|G^{(3)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right|=O_P(1+n^{-1/2}\lambda_n^{-5/2}\sqrt{\log n})$.

Proof. We only show (i) since the other cases are similar. By Proposition B.2(i), it suffices to show that $\sup_{\tau\in\mathcal{T}}\frac{1}{n}\sum_{i=1}^n b_{ni}\left(\tau\right)=O_P\left(1\right)$, where $b_{ni}\left(\tau\right)\equiv\lambda_n^{-1}[|G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)|-E|G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)|]$. Let $b_n\left(\tau\right)\equiv\frac{1}{n}\sum_{i=1}^n b_{ni}\left(\tau\right)$. Noting that $\left\|\lambda_n^{-1}g\left(-\varepsilon_{i\tau}/\lambda_n\right)\right\|_{2+\eta}=O(\lambda_n^{-(1+\eta)/(2+\eta)})$, we have by the Davydov inequality that

$$\begin{aligned}\operatorname{Var}\left(b_n\left(\tau\right)\right) &= \frac{1}{n^2}\sum_{i=1}^n\operatorname{Var}\left(b_{ni}\left(\tau\right)\right)+\frac{2}{n^2}\sum_{1\le i<j\le n}^{n}E\left[b_{ni}\left(\tau\right)b_{nj}\left(\tau\right)\right]\\ &\le O\left(n^{-1}\lambda_n^{-1}\right)+C\max_{1\le i\le n}\left\|b_{ni}\left(\tau\right)\right\|_{2+\eta}^2\sum_{i=1}^n\sum_{s=1}^{\infty}\alpha\left(s\right)^{\eta/(2+\eta)}\\ &= O\left(n^{-1}\lambda_n^{-1}+n^{-1}\lambda_n^{-2(1+\eta)/(2+\eta)}\right)=o\left(1\right).\end{aligned}$$

It follows that $b_n\left(\tau\right)=o_P\left(1\right)$ for each τ. Following exactly the same argument as used in the proof of the uniform consistency of kernel density estimators, we can show that $\sup_{\tau\in\mathcal{T}}\left|b_n\left(\tau\right)\right|=O_P(n^{-1/2}\lambda_n^{-1/2}\sqrt{\log n})$. It follows that $\sup_{\tau\in\mathcal{T}}n^{-1}\lambda_n^{-1}\sum_{i=1}^n\left|G^{(1)}\left(-\varepsilon_{i\tau}/\lambda_n\right)\right|=O\left(1\right)+O_P(n^{-1/2}\lambda_n^{-1/2}\sqrt{\log n})=O_P\left(1\right)$. ■

Proposition B.4 $\mathbb{V}_{n1}\left(\tau,\gamma\right)\equiv n^{-1/2}\sum_{i=1}^n[\mathbf{1}\left(\varepsilon_{i\tau}<0\right)-G\left(-\varepsilon_{i\tau}/\lambda_n\right)]\varphi_{i\gamma}=o_P\left(1\right)$ *uniformly in* $(\tau,\gamma)\in\mathcal{T}\times\Gamma$.

Proof. Let $\mathbb{V}_{n1,1}\left(\tau,\gamma\right)\equiv n^{-1/2}\sum_{i=1}^n\left\{\mathbf{1}\left(\varepsilon_{i\tau}<0\right)-G\left(-\varepsilon_{i\tau}/\lambda_n\right)-\tau+E_i[G\left(-\varepsilon_{i\tau}/\lambda_n\right)]\right\}\varphi_{i\gamma}$ and $\mathbb{V}_{n1,2}\left(\tau,\gamma\right)\equiv n^{-1/2}\sum_{i=1}^n\left\{\tau-E_i[G\left(-\varepsilon_{i\tau}/\lambda_n\right)]\right\}\varphi_{i\gamma}$. Then $\mathbb{V}_{n1}\left(\tau,\gamma\right)=\mathbb{V}_{n1,1}\left(\tau,\gamma\right)+\mathbb{V}_{n1,2}\left(\tau,\gamma\right)$. By Proposition B.1(i) and the uniform boundedness of φ, $\sup_{\gamma\in-}\sup_{\tau\in\mathcal{T}}\left|\mathbb{V}_{n1,2}\left(\tau,\gamma\right)\right|=O\left(n^{1/2}\lambda_n^q\right)=o_P\left(1\right)$. We partition the compact set $\mathcal{T}$ by $\overline{n}_1$ points $\underline{\tau}=\tau_0<\tau_1<\tau_2<\cdots<\tau_{\overline{n}_1}=\overline{\tau}$ such that $\left|\tau_j-\tau_{j-1}\right|=\delta_{1n}=1/\log\log n$. Let $\mathcal{T}_1=[\tau_0,\tau_1]$ and $\mathcal{T}_j=(\tau_{j-1},\tau_j]$ for $j=2,\cdots,\overline{n}_1$. Let $\tau\in(\tau_j,\tau_{j+1}]$. We cover the compact set Γ by $\Gamma_k=\{\gamma:\left\|\gamma-\gamma_k\right\|\le\delta_{2n}\}$ for $k=1,\cdots,\overline{n}_2$, where $\delta_{2n}=1/(\log\log n)^{d_\Gamma}$, d_Γ is the dimension of $\gamma\in\Gamma$, and $\overline{n}_2=O(\log\log n)$.

Note that

$$\begin{aligned}&\sup_{\gamma\in-}\sup_{\tau\in\mathcal{T}}\left|\mathbb{V}_{n1,1}\left(\tau,\gamma\right)\right|\\ \le\ &\max_{1\le l\le\overline{n}_2}\max_{1\le j\le\overline{n}_1}\left|\mathbb{V}_{n1,1}\left(\tau_j,\gamma_l\right)\right|+\max_{1\le l\le\overline{n}_2}\max_{1\le j\le\overline{n}_1}\sup_{\gamma\in-_l}\sup_{\tau\in\mathcal{T}_j}\left|\mathbb{V}_{n1,1}\left(\tau,\gamma\right)-\mathbb{V}_{n1,1}\left(\tau_j,\gamma_l\right)\right|\\ \equiv\ &V_{n11}+V_{n12}.\end{aligned}$$

Fix $\epsilon > 0$. Let $\xi_{i,jl} \equiv \{\mathbf{1}(\varepsilon_{i\tau_j} < 0) - G(-\varepsilon_{i\tau_j}/\lambda_n) - \tau_j + E_i[G(-\varepsilon_{i\tau_j}/\lambda_n)]\}\varphi_{i\gamma_l}$. Then $\left|\xi_{i,jl}\right| \leq 2\overline{c}_G c_\varphi$, $E\left(\xi_{i,jl}\right) = 0$ and $E\left(\xi_{i,jl}^2\right) \leq C\overline{\lambda}_n$ by Propositions B.1(i)-(ii) where $\overline{\lambda}_n \equiv \lambda_n + \lambda_n^q$. By the Davydov inequality, one can show that there exists some positive constant c_ξ such that $E[\sum_{i=s+1}^{s+p_n} \xi_{i,jl}]^2 \leq c_\xi p_n \overline{\lambda}_n$ for any $s = 1, \cdots, p_n$ and $1 \leq p_n \leq n/2$. We can apply Lemma C.3 with $M_n = 2c_\varphi$, $p_n = n^{1/2-\upsilon}$ for some $\upsilon > 0$, $\epsilon^* = n^{-1/2}\epsilon$, and $\sigma^2(p_n) = c_\xi p_n \overline{\lambda}_n$ to obtain

$$\begin{aligned}
P\left(V_{n11} > \epsilon\right) &\leq \overline{n}_1\overline{n}_2 \max_{1\leq l\leq \overline{n}_2} \max_{1\leq j\leq \overline{n}_1} P\left(\left|n^{-1}\sum_{i=1}^{n}\xi_{i,jl}\right| > n^{-1/2}\epsilon\right) \\
&\leq \overline{n}_1\overline{n}_2 C_0 \exp\left(-\frac{\epsilon^2}{C_1 c_\xi \overline{\lambda}_n + 4c_\varphi C_2 n^{-\upsilon}\epsilon}\right) + \overline{n}_1\overline{n}_2 C_3\sqrt{\frac{2}{n^{-1/2}c_\varphi\epsilon}}\, n^{1/2+\upsilon}\alpha(n^{1/2-\upsilon}+1) \\
&= \overline{n}_1\overline{n}_2 \exp\left(-\frac{\epsilon^2}{C_1 c_\xi \overline{\lambda}_n + 4c_\varphi C_2 n^{-\upsilon}\epsilon}\right) + o\left(n^{3/4+\upsilon}(n^{1/2-\upsilon})^{-\beta_0}\left(\log\log n\right)^2\right) \\
&\equiv p_{n1} + p_{n2}, \qquad \text{(B.1)}
\end{aligned}$$

where $\beta_0 = 1 + 16/\eta$. Clearly, the first term p_{n1} is $o(1)$ provided $\upsilon > 0$. $p_{n2} = o(n^{3/4+\upsilon-(1/2-\upsilon)\beta_0}$ $(\log\log n)^2) = o(1)$ provided $\beta_0 > (3/4+\upsilon)/(1/2-\upsilon)$ and $\upsilon \in (0, 1/2)$.

Now consider the class of functions

$$\mathcal{F}_2 \equiv \left\{b_{(\tau,\gamma)} : (\tau,\gamma) \in \mathcal{T} \times -\right\}$$

where $b_{(\tau,\gamma)} : \mathbb{R} \times \mathbb{R}^{d_X+d_Z} \to \mathbb{R}$ is defined by

$$b_{(\tau,\gamma)}\left(Y_i, W_i\right) \equiv \left\{G\left(\frac{-Y_i + m\left(\tau, W_i\right)}{\lambda_n}\right) - E_i\left[G\left(\frac{-Y_i + m\left(\tau, W_i\right)}{\lambda_n}\right)\right]\right\}\varphi\left(W_i,\gamma\right).$$

Let $\overline{G}\left(-\varepsilon_{i\tau}/\lambda_n\right) \equiv G\left(-\varepsilon_{i\tau}/\lambda_n\right) - E_i[G\left(-\varepsilon_{i\tau}/\lambda_n\right)$ and $\varsigma_i\left(\tau,\gamma\right) \equiv \overline{G}\left(-\varepsilon_{i\tau}/\lambda_n\right)\varphi_{i\gamma}$. Then

$$\begin{aligned}
\varsigma_i\left(\tau,\gamma\right) - \varsigma_i\left(\tau',\gamma'\right) &= \left[\overline{G}\left(-\varepsilon_{i\tau}/\lambda_n\right) - \overline{G}\left(-\varepsilon_{i\tau'}/\lambda_n\right)\right]\varphi_{i\gamma} + \overline{G}\left(-\varepsilon_{i\tau'}/\lambda_n\right)\left[\varphi_{i\gamma} - \varphi_{i\gamma'}\right] \\
&\equiv \varsigma_{i1}\left(\tau,\tau',\gamma\right) + \varsigma_{i2}\left(\tau',\gamma,\gamma'\right), \text{ say.}
\end{aligned}$$

Following the arguments in the proof of Theorem 4.3 and Proposition B.7 below, by the C_r inequality and Assumptions A4, B4, and B5 we have

$$\begin{aligned}
& E \sup_{|\tau-\tau'|\leq\delta_1, \|\gamma-\gamma'\|\leq\delta_2, \sqrt{\delta_1^2+\delta_2^2}\leq\delta} \left|\varsigma_i\left(\tau,\gamma\right) - \varsigma_i\left(\tau',\gamma'\right)\right|^r \\
\leq\ & 2^{r-1}E \sup_{|\tau-\tau'|\leq\delta_1, \|\gamma-\gamma'\|\leq\delta_2, \sqrt{\delta_1^2+\delta_2^2}\leq\delta} \left|\varsigma_{i1}\left(\tau,\tau',\gamma\right)\right|^r \\
& +2^{r-1}E \sup_{|\tau-\tau'|\leq\delta_1, \|\gamma-\gamma'\|\leq\delta_2, \sqrt{\delta_1^2+\delta_2^2}\leq\delta} \left|\varsigma_{i2}\left(\tau',\gamma,\gamma'\right)\right|^r \\
\leq\ & CE \sup_{|\tau-\tau'|\leq\delta_1} \left|G\left(-\varepsilon_{i\tau}/\lambda_n\right) - G\left(-\varepsilon_{i\tau'}/\lambda_n\right)\right| + CE \sup_{\|\gamma-\gamma'\|\leq\delta_2} \left|\varphi\left(W_i,\gamma\right) - \varphi\left(W_i,\gamma'\right)\right|^r \\
\leq\ & CE\left|\sum_{j=1}^{3}\lambda_n^{-j}G^{(j)}\left(\frac{-Y_i + m\left(\tau,W_i\right)}{\lambda_n}\right)\right| \sup_{|\tau-\tau'|\leq\delta_1}\left|m\left(\tau,W_i\right) - m\left(\tau',W_i\right)\right|^j \\
& +CE\left\{\left|\lambda_n^{-3}G^{*}\left(\frac{-Y_i + m\left(\tau,W_i\right)}{\lambda_n}\right)\right| \sup_{|\tau-\tau'|\leq\delta_1}\left|m\left(\tau,W_i\right) - m\left(\tau',W_i\right)\right|^3\right\} + C\delta_2^{r\nu} \\
\leq\ & C\delta_1 + C\delta_1^3 + C\delta_2^{r\nu} \leq 2C\delta^{1\wedge(rs_\gamma)} = 2C\delta^{\overline{\nu}},
\end{aligned}$$

where $\delta \leq 1$, $\overline{\nu} = \min(1, r\nu)$, and $G^*(\cdot)$ is defined in the proof of Proposition B.7. That is, $\mathcal{F}_2$ is a class of uniformly bounded functions satisfying L_r-continuity. The L_r-continuity implies that the bracketing number satisfies

$$N\left(\epsilon, \mathcal{F}_2, \|\cdot\|_{L_r(P)}\right) \leq C_2 \left(\frac{1}{\epsilon}\right)^{(1+d_\Gamma)/\overline{\nu}}.$$

By Theorem 2.2 of Andrews and Pollard (1994), this, together with Assumption B2(ii) and the result in the proof of Theorem 4.3, implies that $\{\mathbb{V}_{n12}(\tau,\gamma) : (\tau,\gamma) \in \mathcal{T} \times \Gamma\}$ is stochastically equicontinuous. It follows that

$$V_{n12} = \max_{1\leq l\leq \overline{n}_2} \max_{1\leq j\leq \overline{n}_1} \sup_{\gamma\in\Gamma_l} \sup_{\tau\in\mathcal{T}_j} |\mathbb{V}_{1n}(\tau,\gamma) - \mathbb{V}_{1n}(\tau_j,\gamma_l)| = o_P(1).$$

This completes the proof of the proposition. ■

Proposition B.5 *Under H_{1n}, $\mathbb{V}_{n2}(\tau,\gamma) \equiv n^{-1/2}h^{-d_X}\sum_{i=1}^n [\mathbf{1}(\varepsilon_{i\tau} \leq 0) - \mathbf{1}(u_{i\tau} \leq 0)]\varphi_{i\gamma}\mu_{ix}K_{ix} = n^{-1}h^{-d_X}\sum_{i=1}^n f(m_{i\tau}|W_i)\delta(\tau,W_i)\mu_{ix}K_{ix} + o_P(1)$ uniformly in $(\tau,\gamma) \in \mathcal{T}\times\Gamma$ for each $x \in \mathcal{X}$.*

Proof. Let $\mathbb{V}_{n2,1}(\tau,x) \equiv n^{-1/2}h^{-d_X}\sum_{i=1}^n \{\mathbf{1}(\varepsilon_{i\tau} \leq 0) - \mathbf{1}(u_{i\tau} \leq 0) - \tau + E_i[\mathbf{1}(u_{i\tau} \leq 0)]\}\varphi_{i\gamma}\mu_{ix}K_{ix}$, and $\mathbb{V}_{n2,2}(\tau,x) \equiv n^{-1/2}h^{-d_X}\sum_{i=1}^n \{\tau - E_i[\mathbf{1}(u_{i\tau} \leq 0)]\}\varphi_{i\gamma}\mu_{ix}K_{ix}$. Then $\mathbb{V}_{n2}(\tau,\gamma) = \mathbb{V}_{n2,1}(\tau,\gamma) + \mathbb{V}_{n2,2}(\tau,\gamma)$. By the proof of Theorem 3.1 (Lemma A.3 in particular), one can readily show that $\mathbb{V}_{n2,1}(\tau,\gamma) = o_P(1)$ uniformly in $(\tau,\gamma) \in \mathcal{T}\times\Gamma$. Let $\Delta_n(\tau,\gamma) \equiv n^{-1}h^{-d_X}\sum_{i=1}^n f(m_{i\tau}|W_i)\delta(\tau,W_i)\varphi_{i\gamma}\mu_{ix} \times K_{ix}$. By the Taylor expansion

$$\begin{aligned}
&\sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} \|\mathbb{V}_{n2,2}(\tau,\gamma) - \Delta_n(\tau,\gamma)\| \\
=& \sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} \left\| n^{-1/2}h^{-d_X}\sum_{i=1}^n [F(m_{i\tau}|W_i) - F(m(\tau,X_i)|W_i)]\varphi_{i\gamma}\mu_{ix}K_{ix} - \Delta_n(\tau,\gamma)\right\| \\
=& \sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} \left\| n^{-1}h^{-d_X}\sum_{i=1}^n [f(m_{i\tau}^*|W_i) - f(m_{i\tau}|W_i)]\delta(\tau,W_i)\varphi_{i\gamma}\mu_{ix}K_{ix}\right\| \\
\leq& \sup_{(\tau,\gamma)\in\mathcal{T}\times\Gamma} n^{-3/2}h^{-d_X}\sum_{i=1}^n C_2(W_i)\delta^2(\tau,W_i)\varphi_{i\gamma}\|\mu_{ix}\|K_{ix} \\
\leq& Cn^{-3/2}h^{-d_X}\sum_{i=1}^n \|\mu_{ix}\|K_{ix} = o_P(1),
\end{aligned}$$

where $m_{i\tau}^*$ lies between $m_{i\tau}$ and $m(\tau,X_i)$. ■

Proposition B.6 *Under H_{1n}, $\mathbb{V}_{n3}(\tau,\gamma) \equiv n^{-1/2}\lambda_n^{-1}\sum_{i=1}^n G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau} - \varepsilon_{i\tau})\varphi_{i\gamma} = -c_0 n^{-1/2} \times \sum_{i=1}^n b(X_i,\gamma)\psi_\tau(\varepsilon_{i\tau}) + \mathbf{\Delta}(\tau,\gamma) + o_P(1)$ uniformly in (τ,γ).*

Proof. Let $g_{ni\tau} \equiv \lambda_n^{-1}G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)$. Using Proposition B.1(iii) and $u_{i\tau} = \varepsilon_{i\tau} + n^{-1/2}\delta(\tau,W_i)$ yields

$$\begin{aligned}
\mathbb{V}_{n3}(\tau,\gamma) &= n^{-1/2}\sum_{i=1}^n g_{ni\tau}(\widehat{u}_{i\tau} - u_{i\tau})\varphi_{i\gamma} + n^{-1/2}\sum_{i=1}^n g_{ni\tau}\delta(\tau,W_i)\varphi_{i\gamma} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n f(m_{i\tau}|W_i)(\widehat{u}_{i\tau} - u_{i\tau})\varphi_{i\gamma} + \frac{1}{\sqrt{n}}\sum_{i=1}^n [g_{ni\tau} - E(g_{ni\tau}|W_i)](\widehat{u}_{i\tau} - u_{i\tau})\varphi_{i\gamma} \\
&\quad + \frac{1}{n}\sum_{i=1}^n g_{ni\tau}\delta(\tau,W_i)\varphi_{i\gamma} + o_P(1) \\
&\equiv \mathbb{V}_{n3,1}(\tau,\gamma) + \mathbb{V}_{n3,2}(\tau,\gamma) + \mathbb{V}_{n3,3}(\tau,\gamma) + o_P(1) \text{ uniformly in } (\tau,\gamma). \qquad \text{(B.2)}
\end{aligned}$$

By Corollary 3.2, we have

$$\begin{aligned}\widehat{u}_{i\tau}-u_{i\tau} &= -e_1^T H(\tau,X_i)^{-1}\frac{1}{nh^{d_X}}\sum_{j=1}^n \psi_\tau(u_{j\tau})\mu_{j,i}K_{ij}[1+o_P(1)]+o_P(n^{-1/2}) \\ &= [-\alpha_{1i}(\tau)-\alpha_{2i}(\tau)][1+o_P(1)]+o_P(n^{-1/2}),\end{aligned}$$

where both $o_P(1)$ and $o_P(n^{-1/2})$ hold uniformly in i and τ, $H(\tau,x)=f(m(\tau,x)|x)f(x)\mathbb{H}$, $\alpha_{1i}(\tau)=e_1^T H(\tau,X_i)^{-1}\frac{1}{nh^{d_X}}\sum_{j=1}^n\psi_\tau(\varepsilon_{j\tau})\mu_{j,i}K_{ij}$, and $\alpha_{2i}(\tau)=e_1^T H(\tau,X_i)^{-1}\frac{1}{nh^{d_X}}\sum_{j=1}^n[\mathbf{1}(\varepsilon_{j\tau}\le 0)-\mathbf{1}(u_{j\tau}\le 0)]\mu_{j,i}K_{ij}$. It follows that

$$\begin{aligned}\mathbb{V}_{n3,1}(\tau,\gamma) &\simeq -\frac{1}{\sqrt{n}}\sum_{i=1}^n f(m_{i\tau}|W_i)\varphi_{i\gamma}\alpha_{1i}(\tau)-\frac{1}{\sqrt{n}}\sum_{i=1}^n f(m_{i\tau}|W_i)\varphi_{i\gamma}\alpha_{2i}(\tau)+o_P(1) \\ &\equiv -\mathbb{V}_{n3,1}^{(1)}(\tau,\gamma)-\mathbb{V}_{n3,1}^{(2)}(\tau,\gamma)+o_P(1). \qquad \text{(B.3)}\end{aligned}$$

We first study $\mathbb{V}_{n3,1}^{(1)}(\tau,\gamma)$. As before, partition $\mathcal{T}$ as before by n_1 points $\underline{\tau}=\tau_0<\tau_1<\cdots<\tau_{n_1}=\overline{\tau}$ and cover the compact set Γ by $\Gamma_k\equiv\{\gamma:\|\gamma-\gamma_k\|\le\delta_{2n}\}$, but we now require $\tau_{s+1}-\tau_s=h^{d_X/2}/\log n$ and $\delta_{2n}=n^{-1/2}/\log n$. Fix $\gamma_l\in\Gamma_l$. Let $\mathcal{T}_s=[\tau_s,\tau_{s+1}]$ for $s=1,\cdots,n_1$ and let $(\tau,\gamma)\in\mathcal{T}_s\times\Gamma_l$. Then

$$\begin{aligned}\sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}\left|\mathbb{V}_{n3,1}^{(1)}(\tau,\gamma)\right| &\le \max_{1\le l\le n_2}\max_{1\le s\le n_1}\left|\mathbb{V}_{n3,1}^{(1)}(\tau_s,\gamma_l)\right| \\ &\quad+\max_{1\le l\le n_2}\sup_{\gamma\in\Gamma_l}\max_{1\le s\le n_1}\sup_{\tau\in\mathcal{T}_s}\left|\mathbb{V}_{n3,1}^{(1)}(\tau,\gamma)-\mathbb{V}_{n3,1}^{(1)}(\tau_s,\gamma_l)\right|. \qquad \text{(B.4)}\end{aligned}$$

Let $V_i\equiv(W_i^T,Y_i)^T$ and $\varsigma_{ij}(\tau,\gamma)\equiv h^{-d_X}f(m_{i\tau}|X_i)\varphi_{i\gamma}e_1^T H(\tau,X_i)^{-1}\mu_{j,i}K_{ij}$. Introducing $\Phi_{1\tau\gamma}(V_i,V_j)\equiv[\varsigma_{ij}(\tau,\gamma)\psi_\tau(\varepsilon_{j\tau})+\varsigma_{ji}(\tau,\gamma)\psi_\tau(\varepsilon_{i\tau})]$, we can write $\mathbb{V}_{n3,1}^{(1)}(\tau,\gamma)$ as $n^{-3/2}\sum_{i=1}^n\Phi_{1\tau\gamma}(V_i,V_i)$ plus the U-statistic

$$\mathbb{U}_n(\tau,\gamma)\equiv\frac{1}{n^{3/2}}\sum_{1\le i<j\le n}^n\Phi_{1\tau\gamma}(V_i,V_j).$$

It is straightforward to show that $n^{-3/2}\sum_{i=1}^n\Phi_{1\tau\gamma}(V_i,V_i)=O_P(n^{-1/2}h^{-d_X})=o_P(1)$ uniformly in (τ,γ). By the Hoeffding decomposition we have

$$\mathbb{U}_n(\tau,\gamma)=\frac{n-1}{n}\left\{\mathbb{U}_n^{(1)}(\tau,\gamma)+\mathbb{U}_n^{(2)}(\tau,\gamma)\right\}, \qquad \text{(B.5)}$$

where

$$\begin{aligned}\mathbb{U}_n^{(1)}(\tau,\gamma) &= \frac{1}{n^{1/2}}\sum_{j=1}^n\int\Phi_{1\tau\gamma}(V_i,V_j)dP(V_i), \\ \mathbb{U}_n^{(2)}(\tau,\gamma) &= \frac{1}{n^{1/2}(n-1)}\sum_{1\le i<j\le n}^n\left[\Phi_{1\tau\gamma}(V_i,V_j)-\int\Phi_{1\tau\gamma}(V_i,V_j)dP(V_i)-\int\Phi_{1\tau\gamma}(V_i,V_j)dP(V_j)\right],\end{aligned}$$

and $P(V_i)$ denotes the distribution of V_i. Noting that $\int\Phi_{1\tau\gamma}(V_i,V_j)dP(V_i)=h^{-d_X}E_i[f(m_{i\tau}|W_i)e_1^T H(\tau,X_i)^{-1}\mu_{j,i}K_{ij}\varphi_{i\gamma}]\psi_\tau(\varepsilon_{j\tau})$, it is straightforward to show that

$$\begin{aligned}\mathbb{U}_n^{(1)}(\tau,\gamma) &= \frac{1}{n^{1/2}h^{d_X}}\sum_{j=1}^n\int f(m_{i\tau}|W_i)e_1^T H(\tau,X_i)^{-1}\mu_{j,i}K_{ij}\varphi_{i\gamma}dP(V_i)\psi_\tau(\varepsilon_{j\tau}) \\ &= \frac{1}{n^{1/2}}\sum_{j=1}^n\int\int f(m(\tau,(X_j+hv,z'))|X_j+hv,z')e_1^T H(\tau,X_j+hv)^{-1}\mu(v)K(v) \\ &\quad\times\varphi((X_j+hv,z'),\gamma)f(X_j+hv,z')dvdz'\ \psi_\tau(\varepsilon_{j\tau})\end{aligned}$$

$$\simeq \frac{1}{n^{1/2}} \sum_{j=1}^{n} \int f\left(m\left(\tau,\left(X_j, z'\right)\right)|X_j, z'\right) f\left(X_j, z'\right) e_1^T H\left(\tau, X_j\right)^{-1} \mathbb{B}_0 \varphi\left(\left(X_j, z'\right), \gamma\right) dz' \psi_\tau\left(\varepsilon_{j\tau}\right)$$

$$\simeq \frac{c_0}{n^{1/2}} \sum_{j=1}^{n} \int f\left(X_j, z'\right) f\left(X_j\right)^{-1} \varphi\left(\left(X_j, z'\right), \gamma\right) dz' \psi_\tau\left(\varepsilon_{j\tau}\right)$$

$$= \frac{c_0}{n^{1/2}} \sum_{j=1}^{n} b\left(X_j, \gamma\right) \psi_\tau\left(\varepsilon_{j\tau}\right), \tag{B.6}$$

where we use the fact that $H\left(\tau, X_j\right) = f\left(m\left(\tau, X_j\right)|X_j\right) f\left(X_j\right) \mathbb{H}$ and that $m\left(\tau, W_j\right) = m\left(\tau, X_j\right) + n^{-1/2}\delta\left(\tau, W_j\right)$. For $\Phi_{1\tau w}\left(V_i, V_j\right)$ define M_{n1s} $(s = 1, 2, 3, 4)$ and M_{n2s} $(s = 1, 2, 3)$ as in Lemma C.2. It is easy to verify that

$$\begin{aligned} M_{n11} &= M_{n12} = O\left(h^{-d_X \eta}\right),\ M_{n13} = M_{n14} = O(h^{-d_X(1+\eta)}), \\ M_{n21} &= O(h^{-2d_X}),\ M_{n22} = O(h^{-2d_X}),\ \text{and } M_{n23} = O(h^{-3d_X}), \end{aligned}$$

which implies that $E\left[\mathbb{U}_n^{(2)}\left(\tau_s, \gamma_l\right)\right]^4 = O\left(n^{-2}(h^{-4d_X\eta/(4+\eta)} + h^{-2d_X})\right)$. Fix $\epsilon > 0$. By Lemma C.2(i) and the Markov inequality,

$$\begin{aligned} P\left(\max_{1\le l\le n_2} \max_{1\le s\le n_1} \left|\mathbb{U}_n^{(2)}\left(\tau_s, \gamma_l\right)\right| \ge \epsilon\right) &\le n_1 n_2 \max_{1\le l\le n_2} \max_{1\le s\le n_1} P\left(\mathbb{U}_n^{(2)}\left(\tau_s, \gamma_l\right) \ge n^{3/2}\epsilon\right) \\ &\le n_1 n_2 O\left(n^{-2}(h^{-4d_X\eta/(4+\eta)} + h^{-2d_X})\right) \\ &= O\left(n^{-1}(h^{-4d_X\eta/(4+\eta)} + h^{-2d_X})h^{-d_X/2}\log n\right) = o(1). \end{aligned}$$

Thus

$$\max_{1\le l\le n_2} \max_{1\le s\le n_1} \left|\mathbb{U}_n^{(2)}\left(\tau_s, \gamma_l\right)\right| = o_P(1). \tag{B.7}$$

Next, write

$$\begin{aligned} &\left|\mathbb{V}_{n3,1}^{(1)}(\tau, \gamma) - \mathbb{V}_{n3,1}^{(1)}\left(\tau_s, \gamma_l\right)\right| \\ =\ & \frac{1}{n^{3/2}h^{d_X}} \sum_{j=1}^{n} \left[\psi_\tau\left(\varepsilon_{j\tau}\right) - \psi_{\tau_s}\left(\varepsilon_{j\tau_s}\right)\right] \sum_{i=1}^{n} f\left(m_{i\tau_s}|X_i\right) e_1^T H\left(\tau_s, X_i\right)^{-1} \mu_{j,i} K_{ij} \varphi_{i\gamma} \\ &+ \frac{1}{n^{3/2}h^{d_X}} \sum_{i=1}^{n} \left[f\left(m_{i\tau}|X_i\right) e_1^T H\left(\tau, X_i\right)^{-1} - f\left(m_{i\tau_s}|X_i\right) e_1^T H\left(\tau_s, X_i\right)^{-1}\right] \sum_{j=1}^{n} \psi_\tau\left(\varepsilon_{j\tau}\right) \mu_{j,i} K_{ij} \varphi_{i\gamma} \\ &+ \frac{1}{n^{3/2}h^{d_X}} \sum_{i=1}^{n} \sum_{j=1}^{n} \psi_{\tau_s}\left(\varepsilon_{j\tau_s}\right) f\left(m_{i\tau_s}|X_i\right) e_1^T H\left(\tau_s, X_i\right)^{-1} \mu_{j,i} K_{ij} \left(\varphi_{i\gamma} - \varphi_{i\gamma_l}\right) \\ \equiv\ & V_{n31}\left(\tau, \tau_s; \gamma, \gamma_l\right) + V_{n32}\left(\tau, \tau_s; \gamma, \gamma_l\right) + V_{n33}\left(\tau, \tau_s; \gamma, \gamma_l\right). \end{aligned}$$

First, by the boundedness of φ, the absolute value of $V_{n31}\left(\tau, \tau_s; \gamma, \gamma_l\right)$ is no bigger than

$$\frac{c_\varphi}{n^{1/2}} \sum_{j=1}^{n} \left|\psi_\tau\left(\varepsilon_{j\tau}\right) - \psi_{\tau_s}\left(\varepsilon_{j\tau_s}\right)\right| \times \frac{1}{nh^{d_X}} \sum_{i=1}^{n} \left|f\left(m_{i\tau_s}|X_i\right) e_1^T H\left(\tau_s, X_i\right)^{-1} \mu_{j,i} K_{ij}\right|.$$

The second term in the last expression is $O_P(1)$ uniformly in j whereas the first term is $o_P(1)$ uniformly in τ such that $|\tau - \tau_s| = o\left(n^{-1/2}\right)$ by the stochastic equicontinuity of $\{n^{-1/2}\sum_{j=1}^n \psi_\tau\left(\varepsilon_{j\tau}\right) : \tau \in \mathcal{T}\}$ (cf. the proof of Theorem 4.3). It follows that $\max_{1\le l\le n_2} \sup_{\gamma\in\Gamma_l} \max_{1\le s\le n_1} \sup_{\tau\in\mathcal{T}_s} \left|V_{n31}\left(\tau, \tau_s; \gamma, \gamma_l\right)\right|$

$= o_P(1)$. Similarly, by the fact that that $H(\tau, X_i) = f(m(\tau, X_i)|X_i) f(X_i)\mathbb{H}$ and $m(\tau, W_i) = m(\tau, X_i) + n^{-1/2}\delta(\tau, W_i)$ under H_{1n}, we can readily show that $\max_{1\leq l\leq n_2}\sup_{\gamma\in\Gamma_l}\max_{1\leq s\leq n_1}\sup_{\tau\in\mathcal{T}_s}$ $|V_{n32}(\tau,\tau_s;\gamma,\gamma_l)| = o_P(1)$. By the boundedness of the conditional and marginal densities, we have

$$|V_{n33}(\tau,\tau_s;\gamma,\gamma_l)| \leq \frac{C}{n^{1/2}}\sum_{i=1}^{n}\left|\varphi_{i\gamma}-\varphi_{i\gamma_l}\right| \times \frac{1}{nh^{d_X}}\sum_{j=1}^{n} e_1^T\mathbb{H}\mu_{j,i}K_{ij}.$$

The second term is $O_P(1)$ uniformly in j whereas the first term is $o_P(1)$ uniformly in γ such that $|\gamma-\gamma_l| = o\left(n^{-1/2}\right)$. Hence $\max_{1\leq l\leq n_2}\sup_{\gamma\in\Gamma_l}\max_{1\leq s\leq n_1}\sup_{\tau\in\mathcal{T}_s}|V_{n33}(\tau,\tau_s;\gamma,\gamma_l)| = o_P(1)$. These results, together with (B.4) and the analysis of $\mathbb{U}_n(\tau,\gamma)$ (esp. (B.5)-(B.7)), imply that uniformly in (τ,γ)

$$\mathbb{V}_{n3,1}^{(1)}(\tau,\gamma) = -\frac{c_0}{n^{1/2}}\sum_{j=1}^{n} b(X_j,\gamma)\psi_\tau(\varepsilon_{j\tau}) + o_P(1). \tag{B.8}$$

Now we study $\mathbb{V}_{n3,1}^{(2)}(\tau,\gamma)$ defined in (B.3). By Proposition B.5, uniformly in (τ,γ)

$$\begin{aligned}
\mathbb{V}_{n3,1}^{(2)}(\tau,\gamma) &= \frac{1}{n^2h^{d_X}}\sum_{i=1}^{n}\sum_{j=1}^{n} f(m_{i\tau}|W_i)\varphi_{i\gamma}e_1^T H(\tau,X_i)^{-1} f(m_{j\tau}|W_j)\delta(\tau,W_j)\mu_{j,i}K_{ij} + o_P(1)\\
&\simeq \frac{1}{n^2h^{d_X}}\sum_{i=1}^{n}\sum_{j=1}^{n} E_i\left[f(m_{i\tau}|W_i)\varphi_{i\gamma}e_1^T H(\tau,X_i)^{-1}\mu_{j,i}K_{ij}\right] f(m_{j\tau}|W_j)\delta(\tau,W_j) + o_P(1)\\
&= \frac{c_0}{n}\sum_{j=1}^{n} f(m_{j\tau}|W_j)\delta(\tau,W_j) b(X_j,\gamma) + o_P(1),
\end{aligned} \tag{B.9}$$

where we have used the fact that $f(m_{i\tau}|W_i) = f(m(\tau,X_i)|X_i) + O_P(n^{-1/2})$ under H_{1n}. Combining (B.3), (B.8), and (B.9) yields

$$\mathbb{V}_{n3,1}(\tau,\gamma) = -\frac{c_0}{n^{1/2}}\sum_{i=1}^{n} b(X_i,\gamma)\psi_\tau(\varepsilon_{i\tau}) - \frac{c_0}{n}\sum_{j=1}^{n} f(m_{j\tau}|W_j)\delta(\tau,W_j) b(X_j,\gamma) + o_P(1), \tag{B.10}$$

where $o_P(1)$ holds uniformly in (τ,γ).

Analogously to the proof of $\mathbb{V}_{n3,1}^{(1)}(\tau,\gamma)$ but with the application of Lemma C.2(ii) in place of Lemma C.2(i), we can show that

$$\sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}|\mathbb{V}_{n3,2}(\tau,\gamma)| = O_P\left(n^{-2}(\lambda_n h^{d_X})^{-6\eta/(\eta+6)} + n^{-2}(\lambda_n h^{d_X})^{-3})h^{-d_X/2}\log n\right) = o_P(1). \tag{B.11}$$

Now, by Proposition B.1 and Assumption B6, uniformly in (τ,γ)

$$\mathbb{V}_{n3,3}(\tau,\gamma) = \frac{1}{n}\sum_{i=1}^{n} f(m_{i\tau}|W_i)\delta(\tau,W_i)\varphi_{i\gamma} + O_P(\lambda_n^q). \tag{B.12}$$

Combining (B.2) and (B.10)-(B.12) yields the desired result. ■

Proposition B.7 *Under H_{1n}, $\mathbb{V}_{n4}(\tau,\gamma) \equiv n^{-1/2}\sum_{i=1}^{n}[G(-\varepsilon_{i\tau}/\lambda_n) - G(-\widehat{u}_{i\tau}/\lambda_n)]\varphi_{i\gamma} = -c_0 n^{-1/2}\sum_{i=1}^{n}$ $b(X_i,\gamma)\psi_\tau(\varepsilon_{i\tau}) + \mathbf{\Delta}(\tau,\gamma) + o_P(1)$ uniformly in (τ,γ).*

Proof. By the Taylor expansion

$$
\begin{aligned}
\mathbb{V}_{n4}(\tau,\gamma) &= \frac{1}{\sqrt{n}\lambda_n}\sum_{i=1}^{n} G^{(1)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau}-\varepsilon_{i\tau})\varphi_{i\gamma} + \frac{1}{2\sqrt{n}\lambda_n^2}\sum_{i=1}^{n} G^{(2)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau}-\varepsilon_{i\tau})^2\varphi_{i\gamma} \\
&\quad + \frac{1}{6\sqrt{n}\lambda_n^3}\sum_{i=1}^{n} G^{(3)}(-\varepsilon_{i\tau}/\lambda_n)(\widehat{u}_{i\tau}-\varepsilon_{i\tau})^3\varphi_{i\gamma} + R_{n1}(\tau,\gamma) \\
&\equiv V_{n41}(\tau,\gamma) + V_{n42}(\tau,\gamma) + V_{n43}(\tau,\gamma) + R_{n1}(\tau,\gamma), \qquad \text{(B.13)}
\end{aligned}
$$

where $R_{n1}(\tau,\gamma) \equiv (1/6)n^{-1/2}\lambda_n^{-3}\sum_{i=1}^{n}\left[G^{(3)}(-\overline{u}_{i\tau}/\lambda_n) - G^{(3)}(-u_{i\tau}/\lambda_n)\right](\widehat{u}_{i\tau}-\varepsilon_{i\tau})^3\varphi_{i\gamma}$ with $\overline{u}_{i\tau}$ lying between $\widehat{u}_{i\tau}$ and $\varepsilon_{i\tau}$. By Proposition B.6, it suffices to show the last three terms in (B.13) are uniformly $o_P(1)$.

By Proposition B.3(ii) and the boundedness of φ,

$$
\begin{aligned}
\sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}|V_{n42}(\tau,\gamma)| &\le Cn^{1/2}\max_{1\le i\le n}\sup_{\tau\in\mathcal{T}}|\widehat{u}_{i\tau}-\varepsilon_{i\tau}|^2 \sup_{\tau\in\mathcal{T}}\left\{\frac{1}{2n\lambda_n^2}\sum_{i=1}^{n}\left|G^{(2)}(-\varepsilon_{i\tau}/\lambda_n)\right|\right\} \\
&= O_P(n^{-1/2}h^{-d_X}\log n)O_P(1+n^{-1/2}\lambda_n^{-3/2}\sqrt{\log n}) = o_P(1).
\end{aligned}
$$

Similarly, we have $\sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}|V_{n43}(\tau,\gamma)| = O_P\left(n^{-1}h^{-3d_X/2}(\log n)^{3/2}\right)O_P(1+n^{-1/2}\lambda_n^{-5/2}(\log n)^{1/2})$ $= o_P(1)$. Assumption B5 implies that for all $|\varepsilon-\varepsilon^*| \le \delta \le A_G$, $|G^{(3)}(\varepsilon^*) - G^{(3)}(\varepsilon)| \le \delta G^*(\varepsilon)$ .In fact, one chooses $G^*(\varepsilon) = c_G\mathbf{1}(|\varepsilon| \le 2A_G)$ if $G^{(3)}(\varepsilon)$ has compact support and is Lipschitz continuous, and chooses $G^*(\varepsilon) = c_G\mathbf{1}(|\varepsilon| \le 2A_G) + |\varepsilon - A_G|^{-\gamma_0}\mathbf{1}(|\varepsilon| > 2A_G)$. In each case, $G^*(\varepsilon)$ is bounded and integrable and behaves like the kernel function $K(\cdot)$. Let $\vartheta_n \equiv \max_{1\le i\le n}\sup_{\tau\in\mathcal{T}}|\widehat{u}_{i\tau}-\varepsilon_{i\tau}|$. Then $\vartheta_n = O_P(n^{-1/2}h^{-d_X/2}\sqrt{\log n} + n^{-1/2}) = o(\lambda_n)$ so that $\vartheta_n/\lambda_n \le A_G$ with probability approaching one (w.p.a. 1). It follows that w.p.a. 1 $\left|G^{(3)}(-\overline{u}_{i\tau}/\lambda_n) - G^{(3)}(-\varepsilon_{i\tau}/\lambda_n)\right| \le \vartheta_n\lambda_n^{-1}G^*(-\varepsilon_{i\tau}/\lambda_n)$ and

$$
\sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}|R_{n1}(\tau,\gamma)| \le Cn^{1/2}\vartheta_n^4\lambda_n^{-3}\sup_{\tau\in\mathcal{T}}\frac{1}{n\lambda_n}\sum_{i=1}^{n}G^*(-\varepsilon_{i\tau}/\lambda_n) = O_P\left(n^{-3/2}\lambda_n^{-3}h^{-2d_X}(\log n)^2\right) = o_P(1)
$$

because $\sup_{\tau\in\mathcal{T}}\frac{1}{n\lambda_n}\sum_{i=1}^{n}G^*(-\varepsilon_{i\tau}/\lambda_n) = O_P(1)$ following the proof of Proposition B.3(i). ■

Proposition B.8 $\max_{1\le j\le n}\sup_{\gamma\in\Gamma}\left|\widehat{b}(X_j,\gamma) - b(X_j,\gamma)\right| = O_P(n^{-1/2}h_b^{-d_X/2}\sqrt{\log n} + h_b^{p_b+1})$.

Proof. Masry (1996) proved that the almost sure uniform convergence result holds (uniformly in X_j) for general local polynomial estimates under strong mixing conditions. It is straightforward to extend his result to allow the result also to hold uniformly in γ by the standard chaining argument. Note here we only need convergence in probability. See also Hansen (2008). ■

Proposition B.9 $L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - G(-\varepsilon_{i\tau}/\lambda_n)]\widetilde{\varphi}_{i\gamma} = L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - \mathbf{1}(\varepsilon_{i\tau}\le 0)]\overline{\varphi}_{i\gamma} + o_P(1)$ *uniformly in* $(\tau,\gamma)\in\mathcal{T}\times\Gamma$, *where we recall that* $\widetilde{\varphi}_{i\gamma} \equiv \varphi(W_i,\gamma) - c_0\widehat{b}(X_i,\gamma)$ *and* $\overline{\varphi}_{i\gamma} \equiv \varphi(W_i,\gamma) - c_0 b(X_i,\gamma)$.

Proof. Write $L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - G(-\varepsilon_{j\tau}/\lambda_n)]\widetilde{\varphi}_{j\gamma}$ as

$$
\begin{aligned}
&L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - \mathbf{1}(\varepsilon_{j\tau}\le 0)]\overline{\varphi}_{j\gamma} + L^{-1/2}\sum_{j=i}^{i+L-1}[\mathbf{1}(\varepsilon_{j\tau}\le 0) - G(-\varepsilon_{j\tau}/\lambda_n)]\overline{\varphi}_{j\gamma} \\
&+c_0L^{-1/2}\sum_{j=i}^{i+L-1}[\tau - G(-\varepsilon_{i\tau}/\lambda_n)][\widehat{b}(X_j,\gamma) - b(X_j,\gamma)]. \qquad \text{(B.14)}
\end{aligned}
$$

It suffices to show that the last two terms are $o_P(1)$ uniformly in $(\tau,\gamma)\in\mathcal{T}\times\Gamma$. The last term is no bigger than

$$c_0 L^{-1/2}\sum_{j=i}^{i+L-1}\left|\widehat{b}(X_j,\gamma)-b(X_j,\gamma)\right| = L^{1/2}O_P(n^{-1/2}h_b^{-d_X/2}\sqrt{\log n}+h_b^{p_b+1}) = o_P(1).$$

Now, write the second term in (B.14) as

$$L^{-1/2}\sum_{j=i}^{i+L-1}\{\mathbf{1}(\varepsilon_{j\tau}\le 0)-\tau-G(-\varepsilon_{j\tau}/\lambda_n)+E_j[G(-\varepsilon_{j\tau}/\lambda_n)]\}\overline{\varphi}_{j\gamma}$$
$$+L^{-1/2}\sum_{j=i}^{i+L-1}\{\tau-E_j[G(-\varepsilon_{j\tau}/\lambda_n)]\}\overline{\varphi}_{j\gamma}.$$

The second term is $O_P\left(L^{1/2}\lambda_n^q\right)=o_P(1)$ uniformly in (τ,γ) by Proposition B.1(i). Partition $(\mathcal{T}\times\Gamma)$ as in the proof of Proposition B.4. By the stochastic equicontinuity of $L^{-1/2}\sum_{j=i}^{i+L-1}[\mathbf{1}(\varepsilon_{j\tau}\le 0)-\tau]\overline{\varphi}_{j\gamma}$ and $L^{-1/2}\sum_{j=i}^{i+L-1}\{G(-\varepsilon_{j\tau}/\lambda_n)-E_j[G(-\varepsilon_{j\tau}/\lambda_n)]\}\overline{\varphi}_{j\gamma}$ as proved in Theorem 4.3 and Proposition B.4, we have

$$\sup_{\gamma\in\Gamma}\sup_{\tau\in\mathcal{T}}\left|L^{-1/2}\sum_{j=i}^{i+L-1}\{\mathbf{1}(\varepsilon_{j\tau}\le 0)-\tau-G(-\varepsilon_{j\tau}/\lambda_n)+E_j[G(-\varepsilon_{j\tau}/\lambda_n)]\}\overline{\varphi}_{j\gamma}\right|$$
$$\le \max_{1\le l\le n_2}\max_{1\le k\le n_1}\left|L^{-1/2}\sum_{j=i}^{i+L-1}\{\mathbf{1}(\varepsilon_{j\tau_k}\le 0)-\tau_k-G(-\varepsilon_{i\tau_k}/\lambda_n)+E_i[G(-\varepsilon_{j\tau_k}/\lambda_n)]\}\overline{\varphi}_{i\gamma_l}\right|+o_P(1).$$

Analogously to the proof of Proposition B.4, it is straightforward to show that the dominating term in the last expression is $o_P(1)$ by another application of Lemma C.3. ■

Proposition B.10 $L^{-1/2}\sum_{j=i}^{i+L-1}[G(-\varepsilon_{i\tau}/\lambda_n)-G(-\widehat{u}_{i\tau}/\lambda_n)]\widetilde{\varphi}_{i\gamma}=o_P(1)$ *uniformly in* $(\tau,\gamma)\in\mathcal{T}\times\Gamma$.

Proof. The proof is analogous to that of Proposition B.7. The difference is that one now needs to apply Proposition B.8 and the fact that $L=o\left(n^{1/2}\right)$. ■

C Some Technical Lemmas

This appendix presents some technical lemmas that are used in proving the main results.

Lemma C.1 *Let* $\{V_i, i\ge 1\}$ *be a v-dimensional strong mixing process with mixing coefficient* $\alpha(\cdot)$. *Let* $F_{i_1,\dots,i_m}$, *denote the distribution function of* $(V_{i_1},\dots,V_{i_m})$. *For any integer* $m>1$ *and integers* $(i_1,\dots,i_m)$ *such that* $1\le i_1<i_2<\cdots<i_m$, *let* θ *be a Borel measurable function such that* $\max\{\int|\theta(v_1,\cdots,v_m)|^{1+\widetilde{\eta}}$ $dF_{i_1,\cdots,i_j}(v_1,\cdots,v_j)\,dF_{i_j+1,\cdots,i_m}(v_{j+1},\cdots,v_m),\int|\theta(v_1,\cdots,v_m)|^{1+\widetilde{\eta}}dF_{i_1,\cdots,i_m}\}\le M_n$ *for some* $\widetilde{\eta}>0$. *Then* $|\int\theta(v_1,\cdots,v_m)\,dF_{i_1,\cdots,i_m}(v_1,\cdots,v_m)-\int\theta(v_1,\cdots,v_m)\,dF_{i_1,\cdots,i_j}(v_1,\cdots,v_j)\,dF_{i_j+1,\cdots,i_m}(v_{j+1},$ $\cdots,v_m)|\le 4M_n^{1/(1+\widetilde{\eta})}\alpha(i_{j+1}-i_j)^{\widetilde{\eta}/(1+\widetilde{\eta})}$.

Proof. See Lemma 2.1 of Sun and Chiang (1997). ■

Let $P(V_i)$ denote the probability law of a random variable V_i. Let $1\le i_1,i_2,\dots,i_k\le n$ be arbitrary positive integers. For any j $(1\le j\le k)$, define a collection of probability measures $\mathcal{P}_j^k$ by

$$\mathcal{P}_j^k(V_{i_1},\dots,V_{i_k}) \equiv \Big\{P_j^k(V_{i_1},\dots,V_{i_k})\equiv\Pi_{s=1}^j P(\underline{V}_s):\underline{V}_s \text{ is a subset of } \{V_{i_1},\dots,V_{i_k}\},$$
$$\cup_{s=1}^j\underline{V}_s=\{V_{i_1},\dots,V_{i_k}\},\text{ and } \underline{V}_t\cap\underline{V}_s=\varnothing \text{ for all } 1\le t\ne s\le j\Big\}.$$

In the following, we frequently suppress the arguments of P_j^k and $\mathcal{P}_j^k$ when no confusion can arise. For example, when $k=2$, we use $\max_{1\leq i_1,i_2\leq n}\max_{1\leq j\leq 2}\max_{P_j^2\in\mathcal{P}_j^2}\int_{\mathbb{R}^{3v}}|\varphi(v_{i_1},v_{i_2})|dP_j^2$ to denote

$$\max_{1\leq i_1,i_2\leq n}\max\left\{\int_{\mathbb{R}^{2v}}\{|\varphi(v_1,v_2)|\,dF_{i_1i_2}(v_1,v_2),\ \int_{\mathbb{R}^{2v}}|\varphi(v_1,v_2)|\,dF_{i_1}(v_1)dF_{i_2}(v_2)\right\}.$$

Let $\mathbb{U}_n=n^{-2}\sum_{1\leq i_1<i_2\leq n}\phi(v_{i_1},v_{i_2})$, where $\int\phi(v_1,v)\,dF_{v_i}(v)=\int\phi(v,v_2)\,dF_{v_i}(v)=0$ for all i. Let $M(i_1,\cdots,i_8)\equiv\Pi_{j=1}^4\phi(v_{i_{2j-1}},v_{i_{2j}})$ and $N(i_1,\cdots,i_{12})\equiv\Pi_{j=1}^6\phi(v_{i_{2j-1}},v_{i_{2j}})$. Let $I_1=\{i_1,\cdots,i_8\}$ and $I_2=\{i_1,\cdots,i_{12}\}$. Define

$$M_{n1s}\equiv\max_{\substack{1\leq i_{2k-1}<i_{2k}\leq n,1\leq k\leq 4\\ \text{exactly } 9-s \text{ indices}\\ \text{in } I_1 \text{ are distinct}}}\max_{1\leq j\leq 9-s}\max_{P_j^{9-s}\in\mathcal{P}_j^{9-s}}\int|M(i_1,\cdots,i_8)|^{1+\eta/4}dP_j^{9-s},\ s=1,2,3,4,$$

$$M_{n2s}\equiv\max_{\substack{1\leq i_{2k-1}<i_{2k}\leq n,1\leq k\leq 4\\ \text{exactly } 5-s \text{ indices}\\ \text{in } I_1 \text{ are distinct}}}\max_{1\leq j\leq 5-s}\max_{P_j^{5-s}\in\mathcal{P}_j^{5-s}}\left|\int M(i_1,\cdots,i_8)\,dP_j^{5-s}\right|,\ s=1,2,3,$$

$$N_{n1s}\equiv\max_{\substack{1\leq i_{2k-1}<i_{2k}\leq n,1\leq k\leq 6\\ \text{exactly } 13-s \text{ indices}\\ \text{in } I_2 \text{ are distinct}}}\max_{1\leq j\leq 13-s}\max_{P_j^{13-s}\in\mathcal{P}_j^{13-s}}\int|N(i_1,\cdots,i_{12})|^{1+\eta/6}dP_j^{13-s},\ s=1,\cdots,6,$$

$$N_{n2s}\equiv\max_{\substack{1\leq i_{2k-1}<i_{2k}\leq n,1\leq k\leq 6\\ \text{exactly } 7-s \text{ indices}\\ \text{in } I_2 \text{ are distinct}}}\max_{1\leq j\leq 7-s}\max_{P_j^{7-s}\in\mathcal{P}_j^{7-s}}\left|\int N(i_1,\cdots,i_{12})\,dP_j^{13-s}\right|,\ s=1,\cdots,5.$$

Lemma C.2 *Using the notation defined above,*

(i) if $\sum_{s=1}^{\infty}s^3\alpha(s)^{\eta/(4+\eta)}<\infty$ *for some* $\eta>0$, *then* $E\left[\mathbb{U}_n^4\right]=O(\sum_{s=1}^4 n^{-3-s}M_{n1s}^{4/(4+\eta)}+\sum_{s=1}^3 n^{-3-s}M_{n2s})$;

(ii) if $\sum_{s=1}^{\infty}s^5\alpha(s)^{\eta/(4+\eta)}<\infty$ *for some* $\eta>0$, *then* $E\left[\mathbb{U}_n^6\right]=O(\sum_{s=1}^4 n^{-5-s}N_{n1s}^{6/(6+\eta)}+\sum_{s=1}^3 n^{-5-s}N_{n2s})$.

Proof. Write

$$E\left[\mathbb{U}_n^4\right]=n^{-8}\sum_{1\leq i_1<i_2\leq n}\sum_{1\leq i_3<i_4\leq n}\sum_{1\leq i_5<i_6\leq n}\sum_{1\leq i_7<i_8\leq n}\phi(v_{i_1},v_{i_2})\,\phi(v_{i_3},v_{i_4})\,\phi(v_{i_5},v_{i_6})\,\phi(v_{i_7},v_{i_8}). \quad \text{(C.1)}$$

It is easy to show that the terms in the above summation constitute seven cases: for $s=1,2,\cdots,7$, in case (s) there are exactly $9-s$ distinct indices among $i_1,...,i_8$. We will use $EU_{n(s)}$ to denote these cases $(s=1,2,,\cdots 7)$. For case (1), following Yoshihara (1976), let $i_1,...,i_8$ be distinct integers with $1\leq i_j\leq n$. Let $1\leq k_1<...<k_8\leq n$ be the permutation of $i_1,...,i_8$ in ascending order and let d_c be the c-th largest difference among $k_{j+1}-k_j$, $j=1,...,7$. Define

$$H(k_1,...,k_8)=\phi(v_{i_1},v_{i_2})\,\phi(v_{i_3},v_{i_4})\,\phi(v_{i_5},v_{i_6})\,\phi(v_{i_7},v_{i_8}).$$

For any $1\leq j\leq 7$, put $P_0^{(8)}\left(E^{(8)}\right)=P\left((v_{i_1},..,v_{i_8})\in E^{(8)}\right)$, and

$$P_j^{(8)}\left(E^{(j)}\times E^{(8-j)}\right)=P\left((v_{i_1},..,v_{i_j})\in E^{(j)}\right)P\left((v_{i_{j+1}},..,v_{i_8})\in E^{(8-j)}\right),$$

where $E^{(j)}$ is a Borel set in $\mathbb{R}^{jv}$ and v is the dimension of v_i. Since $\int H(i_1,...,i_8)\,dP_j^{(8)} = 0$ for $j = 1, 7$, we have, by Lemma C.1 with $\widetilde{\eta} = \eta/4$,

$$\begin{aligned}
\sum_{\substack{1\leq k_1<...<k_8\leq n \\ k_2-k_1=d_1}} |E[H(k_1,...,k_8)]| &\leq 4M_{n11}^{4/(4+\eta)} \sum_{k_1=1}^{n-7} \sum_{k_2=k_1+\max_{j\geq 3}\{k_j-k_{j-1}\}}^{n-6} \sum_{k_3=k_2+1}^{n-5} \cdots \sum_{k_8=k_7+1}^{n} \alpha^{\frac{\eta}{4+\eta}}(k_2-k_1) \\
&\leq 4M_{n11}^{4/(4+\eta)} \sum_{k_1=1}^{n-7} \sum_{k_2=k_1+1}^{n-6} (k_2-k_1)^6 \alpha^{\frac{\eta}{4+\eta}}(k_2-k_1) \\
&\leq 4n^4 M_{n11}^{4/(4+\eta)} \sum_{j=1}^{n} j^3 \alpha^{\frac{\eta}{4+\eta}}(j),
\end{aligned}$$

and similarly $\sum_{\substack{1\leq k_1<...<k_8\leq n \\ k_8-k_7=d_1}} |E[H(k_1,...,k_8)]| \leq 4n^4 M_{n11}^{4/(4+\eta)} \sum_{j=1}^{n} j^3 \alpha^{\frac{\eta}{4+\eta}}(j)$. If for some j_α ($2 \leq j_\alpha \leq 6$, $1 \leq \alpha \leq 4$), $k_{j_\alpha+1} - k_{j_\alpha} = d_\alpha$, then

$$\sum_{\substack{1\leq k_1<...<k_8\leq n \\ k_{j_\alpha+1}-k_{j_\alpha}=d_\alpha (1\leq\alpha\leq 4)}} |E[H(k_1,...,k_8)]| \leq 4n^4 M_{n11}^{4/(4+\eta)} \sum_{j=1}^{n} j^3 \alpha^{\frac{\eta}{4+\eta}}(j).$$

It follows that $EU_{n(1)} \leq \sum_{1\leq k_1<...<k_8\leq n} |E[H(k_1,...,k_8)]| = O(n^4 M_{n11}^{4/(4+\eta)})$.

For cases (2)-(4), by using Lemma C.1 repeatedly, we can show that

$$EU_{n(2)} = O(n^3 M_{n12}^{4/(4+\eta)}),\ EU_{n(3)} = O(n^2 M_{n13}^{4/(4+\eta)}),\ \text{and } EU_{n(4)} = O(n M_{n14}^{4/(4+\eta)}).$$

For all other cases, we can calculate the expectations directly to obtain

$$EU_{n(5)} = O(n^4 M_{n21}),\ EU_{n(6)} = O(n^3 M_{n22}),\ \text{and } EU_{n(7)} = O(n^2 M_{n23}).$$

The result in (i) follows. The proof of (ii) is analogous and thus is omitted. ■

Lemma C.3 *Let $\{\xi_t \in \mathbb{R}^q, t = 1, 2, ...\}$ be a strong mixing process, not necessarily stationary, with the mixing coefficients $\alpha(t)$ satisfying $\sum_{t=1}^{\infty} \alpha(t) < \infty$. Suppose that $\varsigma_n : \mathbb{R}^q \to \mathbb{R}$ is a measurable function such that $E[\varsigma_n(\xi_t)] = 0$, and $|\varsigma_n(\xi_t)| \leq M_n$ for every $t = 1, 2, ...$ Then for any $\epsilon^* > 0$,*

$$P\left(\left|n^{-1}\sum_{i=1}^{n} \varsigma_n(\xi_t)\right| > \epsilon\right) \leq C_0 \exp\left(-\frac{np_n\epsilon^{*2}}{C_1\sigma^2(p_n) + C_2 M_n p_n (p_n+1)\epsilon^*}\right) + C_3\sqrt{\frac{M_n}{\epsilon}}\frac{n}{p_n}\alpha(p_n+1),$$

where $1 \leq p_n \leq n/2$, $\sigma^2(p_n) = \sup_{1\leq j\leq 2p_n} \max\left\{\sigma_{j,p_n}^2, \sigma_{j,p_n+1}^2\right\}$, $\sigma_{j,p_n}^2 = E\left[\sum_{t=1}^{p_n} \varsigma_n(\xi_{j+t})\right]^2$, and $C_i's$, $i = 0, 1, 2, 3$, are constants that do not depend on n, ϵ, M_n, and p_n.

Proof. See Lemma 5.2 in Shen and Huang (1998). ■