Testing Monotonicity in Unobservables with Panel Data

Liangjun Su
School of Economics
Singapore Management University

Stefan Hoderlein
Department of Economics
Brown University

Halbert White
Department of Economics
University of California, San Diego

April 27, 2010

Abstract

Monotonicity in a scalar unobservable is a crucial identifying assumption for an important class of nonparametric structural models accommodating unobserved heterogeneity, as in, for example, Altonji and Matzkin (2005) and Imbens and Newey (2009). Tests for this monotonicity have previously been unavailable. Here we propose and analyze tests for scalar monotonicity using panel data for structures with and without time-varying unobservables, either partially or fully nonseparable between observables and unobservables. As it turns out, our tests also have power against relevant failures of exogeneity. Our nonparametric tests are computationally straightforward, have well behaved limiting distributions under the null, are consistent against precisely specified alternatives, and have standard local power properties. We provide straightforward bootstrap methods for inference. Some Monte Carlo experiments show that, for empirically relevant sample sizes, these reasonably control the level of the test, and that our tests have useful power. We apply our tests to study asset returns and demand for ready-to-eat cereals.

Keywords: monotonicity, nonparametric, nonseparable, specification test, unoberved heterogeneity

Acknowledgements: We would like to express our appreciation to Aren Megerdichian, who provided outstanding research assistance. We especially thank Amjad Malik at Kellogg Corp. for providing Aren Megerdichian access to cereal scanner data for his UCSD PhD dissertation, which Aren has used to implement our test. We are also indebted to Maxwell Stinchcombe, who provided helpful suggestions and discussion.

1 Introduction

Suppose an observable scalar $Y_t = Y_{it}$ is structurally generated as

$$Y_t = \phi(X_t, A), \quad t = 1, ..., T, \tag{1}$$

where ϕ is an unknown function, $X_t = X_{it}$ is an observable $d \times 1$ vector, $d \in \mathbb{N}$, and $A = A_i$ is an unobservable attribute vector, heterogeneous across i. A leading example occurs when Y_t

represents the quantity of a good demanded by a consumer, X_t represents income and prices, and A represents consumer i's fixed taste parameters, as in Stigler and Becker (1977). Alternatively, Y_t can represent the quantity produced by a firm, X_t cost and demand shifters, and A firm i's fixed technology parameters.

When A is scalar and $\phi(x, \cdot)$ is strictly monotone ("monotonicity in a scalar unobservable" or just "scalar monotonicity"), this structure is an important special case of the "structural function and distribution" framework considered by Altonji and Matzkin (2005, section 4), henceforth **AM**. Such monotonicity assumptions play a key role in an important strand of flexible structural modeling, beginning with Roehrig (1988) and developed extensively by Matzkin (e.g., Matzkin, 2003; 2007) and Chesher (2003). Scalar monotonicity has gained increasing currency, being relied on recently by Imbens and Newey (2009), Evdokimov (2009), and Komunjer and Santos (2010).

As Hoderlein (2005) notes, monotonicity is a strong assumption. Further, monotonicity is crucial in this context, as key identification results fail when scalar monotonicity is violated. It is thus important to have tests for this. To the best of our knowledge, no such tests are currently available. Accordingly, our goal here is to propose and analyze some straightforward methods for testing scalar monotonicity. As it turns out, our tests are also sensitive to relevant failures of exogeneity.

We do not restrict attention to structures with a single unobservable, however. We also consider structures monotonic in A with separable time-varying unobservable ε_t,

$$Y_t = \phi(X_t, A) + \varepsilon_t, \quad t = 1, ..., T, \tag{2}$$

as well as fully general nonseparable structures with vector-valued ε_t,

$$Y_t = \phi(X_t, \varepsilon_t, A), \quad t = 1, ..., T. \tag{3}$$

Evdokimov (2009) considers the former structures, discussing their relevance to studying heterogeneous treatment effects, such as the effects of union membership on wages and the effects of wages on consumption. The latter can be used, among other things, to study price effects on consumer demand as well as nonlinear/nonparametric factor effects on asset returns in the presence of unobserved heterogeneity. The fully non-separable structures are quite general; their only significant vulnerabilities to misspecification are failures of monotonicity or exogeneity.

In Section 2, we provide some general results for structures monotonic in a scalar unobservable, reviewing and extending known representation and identification results. These results show the necessity of monotonicity for identification and provide the required foundations for our tests. For clarity, and to maintain a manageable scope for the analysis here, we focus on the classical strictly exogenous case, where X_t is independent of A ($X_t \perp A$) for all t. This also serves as a foundation for analysis under weaker conditions (see Hoderlein, Su, and White, 2010).

In Section 3, we propose a monotonicity test for structures of the form (1). The test is fully nonparametric and is available for T as small as 2. The test statistic is asymptotically normal under the null, is consistent against a precisely characterized set of alternatives, and can detect local alternatives with rate $O(N^{-1/2}h^{-d/4})$, where N is the number of cross-section observations and $h = h_N$ is a kernel bandwidth.

We introduce and analyze a monotonicity test for structures with time-varying unobservables of the forms (2) and (3) in Section 4. The test is again fully nonparametric, but here we require T to be large, so as to average out the influence of the ε_t's. The test statistic is asymptotically a mixture of chi-squares under the null, is consistent against a precisely characterized set of alternatives, and can detect local alternatives with rate $O(N^{-1/2})$. Interestingly, the same test works for both the "partially nonseparable" and the fully nonseparable cases.

Section 5 describes effective bootstrap methods for finding critical values and $p-$values for our tests and reports the results of some Monte Carlo experiments designed to study the level and power properties of the tests. We find that our tests perform reasonably well for $N = 100$ and $T = 2$ in the absence of time-varying unobservables and for $N, T = 50$, 100 with time-varying unobservables. Section 6 uses our tests to study asset returns and demand for ready-to-eat cereals, and Section 7 contains a summary and concluding remarks. The Mathematical Appendix contains formal proofs of our results, together with supplementary results supporting the discussion of the main text.

2 Representation and Identification with Scalar Unobservables

We begin with a version of an identification result of **AM**, their theorem 4.1, for the strictly exogenous case. We let $\mathbb{U}[0,1]$ denote the uniform distribution on $\mathbb{I} \equiv [0,1]$. We also write $\bar{\mathbb{N}} \equiv \mathbb{N} \cup \{\infty\}$.

Proposition 2.1 *Let X be a random $d \times 1$ vector, $d \in \bar{\mathbb{N}}$, let ε be a random scalar distributed as $\mathbb{U}[0,1]$, and suppose that $X \perp \varepsilon$. Let $m : \mathbb{R}^d \times \mathbb{I} \to \mathbb{R}$ be a measurable function, and suppose that $Y = m(X, \varepsilon)$. Let $F(y \mid x) \equiv P[Y \leq y \mid X = x]$. Then for given $x \in \mathcal{X} \equiv supp(X)$,*

$$F(y \mid x) = m^{-1}(x, y) \qquad \text{for all } y \in \mathcal{Y} \equiv \ supp(Y) \tag{4}$$

if and only if $m(x, \cdot)$ is strictly increasing.

When $m(x, \cdot)$ is invertible, $m^{-1}(x, \cdot)$ represents the inverse function such that $e = m^{-1}(x, y)$ if and only if $y = m(x, e)$. More generally, $m^{-1}(x, \cdot)$ represents the correspondence defined by $m_x^{-1}(-\infty, y]$, the preimage in $\mathbb{I}$ of the half-ray $(-\infty, y]$ under $m(x, \cdot)$. Also, we adopt the convention suggested by **AM** that if $m(x,\cdot)$ is strictly decreasing, then we replace $m(x,\cdot)$ with $-m(x,\cdot)$. The key property is thus that $m(x,\cdot)$ is strictly monotone.

Let $e = F(y \mid x)$; if eq.(4) holds, then $F(\cdot \mid x)$ is invertible and m is identified as $m(x,e) = F^{-1}(e \mid x)$. Because $F^{-1}(\cdot \mid x)$ is the conditional quantile function, we call this *full identification via conditional quantiles at* x or, for brevity, *full identification.*

These conditions are simpler than those of **AM**'s theorem 4.1, as we consider only the exogenous case. Also, we show that strict monotonicity of $m(x,\cdot)$ is necessary for full identification, not just sufficient.

Representing Y using a scalar ε in Proposition 2.1 is much less restrictive than it might seem. To show this, we formally specify a structural data generating process (DGP).

Assumption A.1 Let $(\Omega, \mathcal{F}, P)$ be a complete probability space on which are defined real-valued random vectors X and U of countable dimensions $d \times 1$ and $\ell \times 1$ respectively, where the distribution of U is nonatomic. Suppose $\phi : \mathbb{R}^d \times \mathbb{R}^\ell \to \mathbb{R}$ is a measurable function and that Y is structurally generated as $Y = \phi(X, U)$.

Note that both X and U can have a finite number or a countable infinity of elements. Usually, d is finite, and we observe X but not U. Requiring that U is nonatomic rules out cases where U has atoms, e.g., for some u, $P[U = u] > 0$. Nonatomicity holds when U is continuously distributed, but it is a weaker requirement. The lack of atoms is not necessarily restrictive, as ϕ can incorporate thresholding; for example, we can have $\phi(x,u) = x\ \mathbf{1}\{u < .5\}$, where $\mathbf{1}\{\cdot\}$ is the indicator function.

Proposition 2.2 *Let Assumption A.1 hold. Then* (i) *there exists a Borel isomorphism*[1] $\upsilon : \mathbb{R}^\ell \to \mathbb{I}$ *such that* $\varepsilon \equiv \upsilon(U)$ *is distributed as* $\mathbb{U}[0,1]$, *and* $Y = \phi(X, \upsilon^{-1}(\varepsilon))$ *a.s.* (ii) $X \perp U$ *holds if and only if* $X \perp \varepsilon$.

Thus, whenever Y is structurally determined by a countably dimensioned nonatomic unobservable, then it also has a representation involving a scalar unobservable. By (ii), this representation preserves exogeneity. Further, it preserves important information about effects of interest, as, for example, $D_x\phi(x,u) = D_x\phi(x, \upsilon^{-1}(e))$ for $e \equiv \upsilon(u)$. In this sense, there is no loss of generality in assuming a scalar unobservable compared to assuming any number of unobservables.

Which representation we use depends upon the purpose at hand. For thinking about economic relationships, it is appropriate to think in terms of the structure of A.1. If indeed there is only a single unobservable driver of Y, A.1 permits this. But even when this fails, the representation

[1] This Borel isomorphism is a one-to-one function υ from $\mathbb{R}^\ell$ onto $\mathbb{I}$ such that both υ and its inverse υ^{-1} are Borel measurable. See Dudley (2002, pp. 487-493) or Corbae, et. al. (2009, pp. 416-417).

with a scalar unobservable is convenient for specification testing. This claim is justified by the next result, which gives a structural characterization of full identification, not only in terms of strict monotonicity but also exogeneity.

Proposition 2.3 *Let Assumption A.1 hold, and suppose that for each $x \in \mathcal{X}$, $\phi(x, \cdot) = g(x, h(x, \cdot))$, where $h : \mathbb{R}^d \times \mathbb{R}^\ell \to \mathbb{I}$ and $g : \mathbb{R}^d \times \mathbb{I} \to \mathbb{R}$ are measurable and $V = h(X, U)$ is nonatomic on $\mathbb{I}$. (i) Suppose $V \perp X$. Let $x \in \mathcal{X}$ be given. Then $F(\cdot \mid x) = g^{-1}(x, \cdot)$ if and only if $g(x, \cdot)$ is strictly increasing. (ii) Suppose $V \not\perp X$. Then there exists $\mathcal{X}^* \subset \mathcal{X}$, $P[X \in \mathcal{X}^*] > 0$, such that for all $x \in \mathcal{X}^*$, $F(\cdot \mid x) \neq g^{-1}(x, \cdot)$.*

This supports specification testing: By part (i), if we maintain that V is independent of X and test $F(\cdot \mid x) = g^{-1}(x, \cdot)$, either (a) for given x or (b) for all x, then we can reject scalar monotonicity. Equivalently, full identification fails (a) at x or (b) on a set of positive probability. Part (ii) says that if we test $F(\cdot \mid x) = g^{-1}(x, \cdot)$ for all x, then rejection again implies the failure of full identification on a set of positive probability. Now, however, rejection is due to failure of exogeneity, of strict monotonicity, or both. In the following sections, we construct specification tests based on this result.

The allowed dependence of $h(x, \cdot)$ on x means that the representations of Propositions 2.2 and 2.3 need not coincide. Further, whereas $\upsilon^{-1}(\cdot)$ is by construction one-to-one, $h(x, \cdot)$ can be many-to-one. Of particular importance is that the broadly valid representation of Proposition 2.2 is generically non-monotonic in a precise sense: Proposition 2.2 ensures only that $\phi(x, \upsilon^{-1}(\cdot))$ is measurable on $\mathbb{I}$; but monotonic functions are *shy* in $L_p(\mathbb{I}, \mathcal{B}, \lambda)$, the space of Borel measurable functions on $\mathbb{I}$ with finite pth absolute moments, $p \in [1, \infty)$ (Stinchcombe, 2010). Shyness is the function space analog of being a subset of a set of Lebesgue measure zero; see Corbae, Stinchcombe, and Zeman (2009, pp. 545-547) (**CSZ**). In this precise sense, monotonicity is quite a strong necessary assumption.

It might seem counterintuitive that $V = h(X, U)$ can depend on X and yet be independent of X, as permitted by Proposition 2.3. Although V and X generally are dependent in this case, Benkard and Berry (2006) give an example where V and X are indeed independent. Our result shows that full identification is possible even in these cases.

3 Testing Monotonicity in Unobservable Attributes

Our first test applies to the fully nonseparable case where the unobservables vary across individuals but not time, with $2 \leq T < \infty$. That is, A.1 holds with

$$Y_t = \phi(X_t, A), \quad t = 1, ..., T.$$

Here $U_t = A$, emphasizing that the unobservables are fixed attributes, conforming with notation of Hoderlein (2005), Hoderlein and Mammen (2007), and Hoderlein and White (2009). As the examples in the introduction show, this is an important special of the structure considered by **AM**. Formally, we impose

Assumption B.0 Assumption A.1 holds with $\phi(x,a) = g(x, \beta(a))$, where $\beta : \mathbb{R}^\ell \to \mathbb{I}$ and $g : \mathbb{R}^d \times \mathbb{I} \to \mathbb{R}$ are measurable; and $Y_t = \phi(X_t, A)$, $t = 1, ..., T$, $2 \leq T < \infty$.

For simplicity in what follows, we let $T = 2$.

For concreteness, we refer to a time index t. Nevertheless, what follows applies to a wide variety of contexts. For example, t can index twins, siblings in a family, individuals in a geographic area, works by an artist, writer, composer, or director, or, in general, members of any group influenced by a common unobservable.

AM and Imbens and Newey (2009) assume scalar monotonicity for all $x \in \mathcal{X}$ (scalar monotonicity *a.s.*). This is the case typically of interest. With fixed A (i.e., for given i) and when exogeneity and scalar monotonicity *a.s.* hold, Proposition 2.3 ensures

$$H_0 : \quad F(Y_1 \mid X_1) = F(Y_2 \mid X_2) \quad a.s. \tag{5}$$

We call (5) full identification *a.s.* Significantly, exogeneity and the time-invariance of A jointly ensure that F_t, the conditional CDF of Y_t given X_t, is time invariant. When exogeneity or scalar monotonicity *a.s.* fails, we generally have

$$P[F_1(Y_1 \mid X_1) = F_2(Y_2 \mid X_2)] < 1.$$

Theorem 8.1 of the appendix formally states and proves (5) and its converse, with a brief discussion of the mild additional conditions required for the converse.

Thus, observing $F_1(Y_1 \mid X_1) \neq F_2(Y_2 \mid X_2)$ for a randomly drawn individual is sufficient to reject H_0. We gain power by considering a measure of the average departure of $F_1(Y_1 \mid X_1)$ from $F_2(Y_2 \mid X_2)$ in a random sample of size N. Thus, we measure the departures of $F_1(Y_1 \mid X_1)$ from $F_2(Y_2 \mid X_2)$ using

$$D_N \equiv \sum_{i=1}^{N} (\hat{F}_{N1}(Y_{i1} \mid X_{i1}) - \hat{F}_{N2}(Y_{i2} \mid X_{i2}))^2.$$

where $\hat{F}_{N1}$ and $\hat{F}_{N2}$ are suitable estimators of F_1 and F_2.

Specifically, we rely on local polynomial estimation of F_t, $t = 1, 2$. Following Masry (1996), we adopt the notation

$$\boldsymbol{j} \equiv (j_1, ..., j_d), \;\; |\boldsymbol{j}| \equiv \sum_{i=1}^{d} j_i, \;\; x^{\boldsymbol{j}} \equiv \boldsymbol{\Pi}_{i=1}^{d} x_i^{j_i}, \;\; \boldsymbol{j}! \equiv \boldsymbol{\Pi}_{i=1}^{d_1} j_i!, \quad \sum_{0 \leq |\boldsymbol{j}| \leq p} \equiv \sum_{k=0}^{p} \sum_{j_1=0}^{k} \cdots \sum_{\substack{j_d=0 \\ j_1+\cdots+j_d=k}}^{k},$$

where $j_1, ..., j_d$ are non-negative integers and $x = (x_1, ..., x_d)'$. Given observations $\{(Y_{it}, X_{it}),$ $i = 1, ..., N\}$, we estimate $F_t(Y_{it} \mid X_{it})$ by solving the weighted least squares minimization problem

$$\min_{\boldsymbol{\beta}} \sum_{j \neq i}^{N} \left[\mathbf{1}\{Y_{jt} \leq Y_{it}\} - \sum_{0 \leq |\boldsymbol{j}| \leq p} \beta'_{\boldsymbol{j}} ((X_{jt} - X_{it})/h)^{\boldsymbol{j}} \right]^2 K_h(X_{jt} - X_{it}),$$

where $\boldsymbol{\beta}$ stacks the $\beta_{\boldsymbol{j}}$'s $(0 \leq |\boldsymbol{j}| \leq p)$ in lexicographic order (with $\beta_{\mathbf{0}}$ in the first position, the element with index $(0, 0, ..., 1)$ next, etc.), $K_h(\cdot) \equiv K(\cdot/h)/h$, $K(\cdot)$ is a symmetric probability density function (PDF) on $\mathbb{R}^d$, and $h \equiv h(n)$ is a bandwidth parameter. Note that we use leave-one-out estimation here, which greatly facilitates the proofs of our results. Our estimator $\hat{F}_{Nt}(Y_{it} \mid X_{it})$ is the minimizing intercept term in the above problem.

Let $L_l \equiv (l+d-1)!/(l!(d-1)!)$ be the number of distinct d-tuples $\boldsymbol{j}$ with $|\boldsymbol{j}| = l$. This denotes the number of distinct lth order partial derivatives of $F_t(\cdot \mid \cdot)$ with respect to x. Let $L \equiv \sum_{l=0}^{p} L_l$. Let $\boldsymbol{\mu}$ be a stacking function such that $\mathbf{X}_{jt}(x) \equiv \boldsymbol{\mu}((X_{jt} - x)/h)$ denotes an $L \times 1$ vector that stacks $((X_{jt} - x)/h)^{\boldsymbol{j}}$, $0 \leq |\boldsymbol{j}| \leq p$, in lexicographic order (e.g., $\mathbf{X}_{jt}(x) = (1, ((X_{jt} - x)/h)')'$ when $p = 1$). Let $\mathbf{X}_{jt,-i} \equiv \mathbf{X}_{jt}(X_{it})$. Then it is easy to verify that

$$\hat{F}_{Nt}(Y_{it} \mid X_{it}) = e'_1 \, [\mathbf{S}_{Nt}(X_{it})]^{-1} \, \frac{1}{N-1} \sum_{j \neq i}^{N} K_{ji,t} \, \mathbf{X}_{jt,-i} \, \mathbf{1}\{Y_{jt} \leq Y_{it}\},$$

where $e'_1 \equiv (1, 0, ..., 0)$ is an L-vector, $K_{ji,t} \equiv K_h(X_{jt} - X_{it})$, and $\mathbf{S}_{Nt}(X_{it}) \equiv \frac{1}{N-1} \sum_{j \neq i}^{N} K_{ij,t}$ $\mathbf{X}_{jt,-i}\mathbf{X}'_{jt,-i}$.

To study the asymptotic properties of D_N under H_0, under a sequence of Pitman local alternatives, and under the global alternative, we impose the following assumptions:

Assumption B.1 Let $Z_{it} \equiv (Y_{it}, X'_{it})'$ and $Z_i \equiv (Z'_{i1}, Z'_{i2})'$. The sequence $\{Z_i\}_{i=1}^{N}$ is IID.

Assumption B.2 (*i*) For $t = 1, 2$, let f_{X_t} and f_{Z_t} denote the PDFs of X_{it} and Z_{it}, respectively. Let f_Z denote the joint PDF of Z_{i1} and Z_{i2}. All these PDFs exist and are uniformly bounded over their supports.

(*ii*) Let $\mathcal{X}_t$ and $\mathcal{Y}_t$ denote the supports of X_{it} and Y_{it}, respectively. Both $\mathcal{X}_t$ and $\mathcal{Y}_t$ are compact, and f_{X_t} is uniformly bounded away from 0 on $\mathcal{X}_t$ for $t = 1, 2$.

Assumption B.3 (*i*) For $t = 1, 2$, and for each $y \in \mathcal{Y}_t$, $F_t(y \mid \cdot)$ is Lipschitz continuous on $\mathcal{X}_t$ and has all partial derivatives up to order $p + 1$, $p \in \mathbb{N}$.

(*ii*) For $t = 1, 2$, and for each $y \in \mathcal{Y}_t$, the $(p+1)$th order partial derivatives with respect to x, $D^{\boldsymbol{k}} F_t(y \mid \cdot)$ with $|\boldsymbol{k}| = p + 1$, are uniformly bounded on $\mathcal{X}_t$, and are Hölder continuous on $\mathcal{X}_t$: for x, $\tilde{x} \in \mathcal{X}_t$, $|D^{\boldsymbol{k}} F_t(y \mid x) - D^{\boldsymbol{k}} F_t(y \mid \tilde{x})| \leq C \, ||x - \tilde{x}||$, where C is a generic finite constant and $\|\cdot\|$ is the Euclidean norm.

(*iii*) For $t = 1, 2$, for each $x \in \mathcal{X}_t$, and for each $y, \tilde{y} \in \mathcal{Y}_t$, $|D^{\boldsymbol{k}} F_t(y \mid x) - D^{\boldsymbol{k}} F_t(\tilde{y} \mid x)| \leq C$ $||y - \tilde{y}||$.

Assumption B.4 (*i*) The kernel function $K : \mathbb{R}^d \to \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF.

(*ii*) $x \to \|x\|^{2p+1} K(x)$ is integrable on $\mathbb{R}^d$ with respect to Lebesgue measure.

(*iii*) Let $\mathbf{K}_{\boldsymbol{j}}(x) \equiv x^{\boldsymbol{j}} K(x)$ for all $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p+1$. For some $C_1 < \infty$ and $C_2 < \infty$, either $K(\cdot)$ is compactly supported such that $K(x) = 0$ for $\|x\| > C_1$, and $|\mathbf{K}_{\boldsymbol{j}}(x) - \mathbf{K}_{\boldsymbol{j}}(\tilde{x})| \leq C_2 \|x - \tilde{x}\|$ for any $x, \tilde{x} \in \mathbb{R}^d$ and for all $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p+1$; or $K(\cdot)$ is differentiable, $\|D\mathbf{K}_{\boldsymbol{j}}(x)\| \leq C_1$, and for some $\iota_0 > 1$, $|D\mathbf{K}_{\boldsymbol{j}}(x)| \leq C_1 \|x\|^{-\iota_0}$ for all $\|x\| > C_2$ and for all $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p+1$.

Assumption B.5 As $N \to \infty$, $h \to 0$, $Nh^{2d}/(\log N)^3 \to \infty$, $Nh^{2(p+1)+d/2} \to 0$, and $h^{p+1-d/2} \to 0$.

Assumption B.1 is standard in conventional panel data modeling. Note that Z_{it} can be dependent across t. Assumption B.2 is standard for nonparametric local polynomial estimation. Assumptions B.3-B.4 are used to obtain uniform consistency for the local polynomial estimator of Masry (1996) and Hansen (2008). Assumption B.5 imposes appropriate conditions on the bandwidth.

Let $\bar{\mathbf{S}}_t(x) \equiv E\left[K_h(X_{jt} - x)\, \mathbf{X}_{jt}(x)\, \mathbf{X}'_{jt}(x)\right]$, $\boldsymbol{\mu}^*(x) \equiv \boldsymbol{\mu}(x) K(x)$, $\overline{\boldsymbol{\mu}}^*(x) \equiv \int \boldsymbol{\mu}^*(\tilde{x})$ $\times \boldsymbol{\mu}^*(x - \tilde{x})' \, d\tilde{x}$ and $\mathbf{K}_{t,x}(X_{jt} - x) \equiv h^{-d} e_1' [\bar{\mathbf{S}}_t(x)]^{-1} \boldsymbol{\mu}^*((X_{jt} - x)/h)$. Let $\bar{\mathbf{1}}_y(Z_{it}) \equiv \mathbf{1}\{Y_{it} \leq y\}$ $-F_t(y \mid X_{it})$, and $\sigma_t^2(y, \tilde{y}; x) \equiv E[\bar{\mathbf{1}}_y(Z_{it})\, \bar{\mathbf{1}}_{\tilde{y}}(Z_{it}) \mid X_{it} = x]$. Define

$$B_N \equiv \frac{h^{d/2}}{(N-1)^2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left[\mathbf{K}_{1,X_{i1}}(X_{j1} - X_{i1})\, \bar{\mathbf{1}}_{Y_{i1}}(Z_{j1}) - \mathbf{K}_{2,X_{i2}}(X_{j2} - X_{i2})\, \bar{\mathbf{1}}_{Y_{i2}}(Z_{j2})\right]^2 \tag{6}$$

$$\sigma_0^2 \equiv 2 \sum_{t=1}^{2} E\left[\int\!\!\int\!\!\int \eta_{it}(x)^2 \;\; \sigma_2^4(y, \tilde{y}; X_{it}) \;\; f_{Z_t}(y, X_{it}) \;\; f_{Z_t}(\tilde{y}, X_{it}) \;\; f_{X_t}(X_{it}) \;\; dy \; d\tilde{y} \; dx\right], \tag{7}$$

where $\eta_{it}(x) \equiv e_1'[\bar{\mathbf{S}}_1(X_{it})]^{-1}\, \overline{\boldsymbol{\mu}}^*(x)\, [\bar{\mathbf{S}}_2(X_{it})]^{-1} e_1$.

The next result says that after centering, $h^{d/2} D_N$ is asymptotically normal under H_0.

Theorem 3.1 *Suppose Assumptions B.1-B.5 hold. Then under* H_0, $h^{d/2} D_N - B_N \xrightarrow{d} N\left(0, \sigma_0^2\right)$.

Clearly, $B_N = B_{N1} + O_P\left(h^{d/2}\right)$, where $B_{N1} \equiv \frac{h^{d/2}}{(N-1)^2} \sum_{t=1}^{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} [\mathbf{K}_{t,X_{it}}(X_{jt} - X_{it})$ $\bar{\mathbf{1}}_{Y_{it}}(Z_{jt})]^2$. Thus, the result also holds if we replace B_N with B_{N1}.

To implement the test, we consistently estimate B_N and σ_0^2 using

$$\hat{B}_N \equiv \frac{h^{d/2}}{(N-1)^2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left[\hat{\alpha}_{ij,1} - \hat{\alpha}_{ij,2}\right]^2, \quad \text{and}$$

$$\hat{\sigma}_N^2 \equiv \frac{2h^d}{N(N-1)} \sum_{t=1}^{2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left[\frac{1}{N} \sum_{l=1}^{N} \hat{\alpha}_{li,t} \hat{\alpha}_{lj,t}\right]^2,$$

where $\hat{\alpha}_{ij,t} \equiv e_1'[\mathbf{S}_{Nt}(X_{it})]^{-1} K_{ji,t} \mathbf{X}_{jt,-i} \hat{\mathbf{1}}_{Y_{it}}(Z_{jt})$, $\hat{\mathbf{1}}_{Y_{it}}(Z_{jt}) \equiv \mathbf{1}\{Y_{jt} \leq Y_{it}\} - \tilde{F}_{Nt}(Y_{it} \mid X_{jt})$, and $\tilde{F}_{Nt}(y \mid x)$ is the pth order local polynomial estimate of $F_t(y \mid x)$ using all N observations $\{Y_{it}, X_{it}\}_{i=1}^{N}$, kernel K, and bandwidth h. We demonstrate below in Theorem 3.2 that $\hat{B}_N - B_N = o_P(1)$ and $\hat{\sigma}_N^2 - \sigma_0^2 = o_P(1)$. Then we can compare

$$J_N \equiv \left(h^{d/2} D_N - \hat{B}_N\right) / \sqrt{\hat{\sigma}_N^2}$$

to the one-sided critical value z_α, the upper α percentile from the $N(0,1)$ distribution. We reject the null at level α if $J_N > z_\alpha$.

To examine the asymptotic local power of the test, we consider the following sequence of Pitman local alternatives:

$$H_1(\gamma_N) : F_1(Y_{i1} \mid X_{i1}) - F_2(Y_{i2} \mid X_{i2}) = \gamma_N \; \delta_N(Z_i),$$

where $\gamma_N \to 0$ as $N \to \infty$ and $\delta_N(\cdot)$ is a continuous function such that $\mu_0 \equiv \lim_{N\to\infty} E[\delta_N(Z_i)]^2 < \infty$. The following theorem establishes the local power of the test.

Theorem 3.2 *Let Assumptions B.1-B.5 hold. Then under* $H_1\left(N^{-1/2} h^{-d/4}\right)$, $J_N \xrightarrow{d} N(\mu_0/\sigma_0, 1)$.

Thus, the test has nontrivial power against Pitman local alternatives that converge to zero at rate $N^{-1/2}h^{-d/4}$. The asymptotic local power function is given by $1 - \Phi(z_\alpha - \mu_0/\sigma_0)$, where Φ is the standard normal CDF.

The following theorem establishes the consistency of the test.

Theorem 3.3 *Suppose Assumptions B.1-B.5 hold. Then under* $H_1 \equiv H_1(1)$, $N^{-1}h^{-d/2} J_N = \mu_A/\sigma_0 + o_P(1)$, *where* $\mu_A \equiv E\left[F_1(Y_{i1} \mid X_{i1}) - F_2(Y_{i2} \mid X_{i2})\right]^2$, *so that* $P(J_N > c_N) \to 1$ *under* H_1 *for any nonstochastic sequence* $c_N = o\left(N h^{d/2}\right)$.

Thus, when the structural conditions of Theorem 8.1 hold, this test can detect any failure of full identification *a.s.*, whether due to failure of strict monotonicity, failure of exogeneity, or both.

Remark. When $T > 2$, the statistic D_N becomes

$$D_N \equiv \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{i=1}^{N} (\hat{F}_{Nt}(Y_{it} \mid X_{it}) - \hat{F}_{Ns}(Y_{is} \mid X_{is}))^2.$$

We consistently estimate the bias and variance by

$$\hat{B}_N = \frac{h^{d/2}}{(N-1)^2} \sum_{t=2}^{T-1} \sum_{s=t+1}^{T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left[\hat{\alpha}_{ij,t} - \hat{\alpha}_{ij,s}\right]^2,$$

$$\hat{\sigma}_N^2 \equiv \frac{2h^d}{N(N-1)} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \left\{ \left[\frac{1}{N} \sum_{l=1}^{N} \hat{\alpha}_{li,t} \hat{\alpha}_{lj,t}\right]^2 + \left[\frac{1}{N} \sum_{l=1}^{N} \hat{\alpha}_{li,s} \hat{\alpha}_{lj,s}\right]^2 \right\}.$$

It is easy to show under local alternatives $H_1\left(N^{-1/2}h^{-d/4}\right)$: $F_t(Y_{it} \mid X_{it}) - F_s(Y_{is} \mid X_{is}) = N^{-1/2}h^{-d/4}\delta_{N,ts}\left(Z_{it}, Z_{is}\right)$ for each pair (t, s), that $J_N \xrightarrow{d} N\left(\mu_0/\sigma_0, 1\right)$, where $\mu_0 \equiv \sum_{t=1}^{T-1}\sum_{s=t+1}^{T}$ $\lim_{N\to\infty} E\left[\delta_{N,ts}\left(Z_{it}, Z_{is}\right)\right]^2 < \infty$ and $\sigma_0^2 = \text{plim}_{N\to\infty}\,\hat{\sigma}_N^2$.

4 Structures with Time-Varying Unobservables

4.1 Partially Nonseparable Structures

We now extend our analysis to structures where there may also be time-varying unobservables, ε_t. We first treat the case where ε_t is additive:

$$Y_t = g(X_t, \beta(A)) + \varepsilon_t, \quad t = 1, ..., T.$$

Because these structures are partly but not fully nonseparable in unobservables, we call them "partially nonseparable". Below, we consider fully nonseparable structures, where $Y_t = g(X_t, \varepsilon_t, \beta(A))$, $t = 1, ..., T$.

For the partially nonseparable case, we impose

Assumption C.0 Assumption A.1 holds with $\phi(x, u) = g(x, \beta(a)) + \varphi(v)$, where $\beta : \mathbb{R}^{\ell_1} \to \mathbb{I}$, $\varphi : \mathbb{R}^{\ell_2} \to \mathbb{R}$, and $g : \mathbb{R}^d \times \mathbb{I} \to \mathbb{R}$ are measurable; and with $U_t = (A, V_t)$, $Y_t = \phi(X_t, U_t) = g(X_t, \beta(A)) + \varphi(V_t)$, $t = 1, 2, ...$.

Thus, $U_t = (A, V_t)$ contains a time-invariant attribute vector, A, and a time varying driver, V_t. We write $B \equiv \beta(A)$ and $\varepsilon_t \equiv \varphi(V_t)$. The latter is often called a "shock".

Evdokimov (2009) (**E**) studies such structures extensively. He gives many salient examples and provides identification and estimation results. An important further example arises in finance, where Y_t is the per period return of an asset, X_t represents market and other factors driving returns, A is *alpha*, the firm-specific return generating attribute, and ε_t is an idiosyncratic shock. This nonlinear asset return factor structure permits arbitrary interaction between alpha and the systematic factors driving returns; it may thus be useful not only for better understanding asset returns but also for improving portfolio allocation.

Just as for **AM**, a main goal for **E** is the identification of g. Although the presence of ε_t complicates matters, the main identification results are the same. As **E** shows, one can use deconvolution to extract the distribution of $M_t \equiv g(X_t, B)$ given X_t. Then strict monotonicity identifies $g(x, \cdot)$ as $F^{-1}(\cdot \mid x)$, where $F(\cdot \mid x)$ now denotes the CDF of M_t given $X_t = x$. In fact, Proposition 2.3 applies to show that either (i) strict monotonicity or (ii) strict monotonicity and exogeneity are necessary and sufficient for identification.

Without ε_t, we compared $F_1(Y_1 \mid X_1)$ to $F_2(Y_2 \mid X_2)$. Here, we would like to compare $F_1(M_1 \mid X_1)$ to $F_2(M_2 \mid X_2)$; both equal B given identification. But since ε_t is unobservable, so is

$M_t = Y_t - \varepsilon_t$. Directly comparing $F_1(M_1 \mid X_1)$ to $F_2(M_2 \mid X_2)$ is not possible. **E**'s results do permit a comparison of $F_1(m \mid x)$ to $F_2(m \mid x)$ for all (m, x), but this has no power when (Y_1, X_1) and (Y_2, X_2) are identically distributed (ID), a leading special case. Moreover, using **E**'s results for inference is challenging, as so far there is no asymptotic distribution theory available for his estimators; only convergence rates are available.

Nevertheless, straightforward specification testing is possible when T is large as well as N. For convenience, we assume that $\{X_t, \varepsilon_t\}$ is ID. Let non-negative weight functions w_1 and w_2 be defined on $\mathcal{X}$. Given sufficient moments, we use w_1 to define $\tilde{Y}_1 = \tilde{Y}_{1t} \equiv E(Y_t\ w_1(X_t) \mid B)$; the equality holds by ID. Then

$$\begin{aligned} \tilde{Y}_1 &= E(g(X_t, B)\ w_1(X_t) \mid B) + E(\varepsilon_t\ w_1(X_t) \mid B) \\ &= \int g(x, B)\ w_1(x)\ dF(x) + E(\varepsilon_t\ w_1(X_t)) \\ &\equiv \bar{g}_1(B) + \tilde{\varepsilon}_1, \end{aligned}$$

where the second line holds given exogeneity ($X_t \perp B$) and the further condition $\varepsilon_t \perp B \mid w_1(X_t)$. In particular, these conditions ensure that $\tilde{\varepsilon}_1 \equiv E(\varepsilon_t\ w_1(X_t))$ is a constant. Assuming that $\varepsilon_t \perp B \mid w_1(X_t)$ allows dependence between ε_t and X_t as well as ε_t and B. An alternative sufficient (but not necessary) condition giving $\tilde{Y}_1 = \bar{g}_1(B) + \tilde{\varepsilon}_1$ is $\varepsilon_t \perp B \mid X_t$. Together with $X_t \perp B$, this implies (and is implied by) $(X_t, \varepsilon_t) \perp B$.

Similarly, with $\tilde{Y}_2 = \tilde{Y}_{2t} \equiv E(Y_t\ w_2(X_t) \mid B)$, $X_t \perp B$, and $\varepsilon_t \perp B \mid w_2(X_t)$, we have

$$\tilde{Y}_2 = \bar{g}_2(B) + \tilde{\varepsilon}_2,$$

with

$$\bar{g}_2(b) \equiv \int g(x, b)\ w_2(x)\ dF(x) \qquad \text{and} \qquad \tilde{\varepsilon}_2 \equiv E(\varepsilon_t\ w_2(X_t)).$$

For example, let $\mathcal{X}_1$ be a subset of $\mathcal{X}$ with $0 < p_1 \equiv P[X_t \in \mathcal{X}_1] < 1$, let $\mathcal{X}_2 \equiv \mathcal{X} \backslash \mathcal{X}_1$, and take $w_1(x) = \mathbf{1}\{x \in \mathcal{X}_1\}/p_1$ and $w_2(x) = \mathbf{1}\{x \in \mathcal{X}_2\}/(1 - p_1)$. In this case, $\varepsilon_t \perp B \mid w_1(X_t)$ and $\varepsilon_t \perp B \mid w_2(X_t)$ are equivalent.

Strict monotonicity *a.s.* of $g(X_t, \cdot)$ directly ensures that $b \to \bar{g}_1(b)$ is strictly monotone in b. By Proposition 2.1 (with X absent), it follows that $B = \bar{g}_1^{-1}(\tilde{Y}_1 - \tilde{\varepsilon}_1)$ is the percentile of $\tilde{Y}_1 - \tilde{\varepsilon}_1$ in its distribution. But since $\tilde{\varepsilon}_1$ is a constant, this percentile is also that of $\tilde{Y}_1$ in its distribution, say $\tilde{F}_1$, defined by

$$\tilde{F}_1(y) \equiv P[\tilde{Y}_1 \leq y].$$

Thus, B is identified as

$$B = \bar{g}_1^{-1}(\tilde{Y}_1 - \tilde{\varepsilon}_1) = \tilde{F}_1(\tilde{Y}_1).$$

In the finance context, where $B = A$ is the firm's alpha, this has a natural interpretation: With $w_1(x) = 1$, this says that alpha is the firm's percentile in the distribution of unconditional expected firm-specific returns. An interesting question here is whether $\tilde{F}_1$ is degenerate, in which case there is no firm-specific heterogeneity.

We also have

$$B = \bar{g}_2^{-1}(\tilde{Y}_2 - \tilde{\varepsilon}_2) = \tilde{F}_2(\tilde{Y}_2),$$

where $\tilde{F}_2 \equiv P[\tilde{Y}_2 \leq y]$. This motivates a specification test based on

$$\tilde{H}_0 : \tilde{F}_1(\tilde{Y}_1) = \tilde{F}_2(\tilde{Y}_2).$$

When T and N are large, we can consistently estimate $\tilde{Y}_{\tau i}$ and $\tilde{F}_\tau$, $\tau = 1, 2$, yielding

$$\hat{B}_{N,T,1,i} \equiv \hat{F}_{N,T,1}(\bar{Y}_{T,1,i}) \qquad \text{and} \qquad \hat{B}_{N,T,2,i} \equiv \hat{F}_{N,T,2}(\bar{Y}_{T,2,i}),$$

where we define

$$\begin{aligned} \bar{Y}_{T,\tau,i} &\equiv T^{-1} \sum_{t=1}^{T} Y_{it}\ w_\tau(X_{it}), \qquad \text{and} \\ \hat{F}_{N,T,\tau}(y) &\equiv N^{-1} \sum_{j=1}^{N} \mathbf{1}\{\bar{Y}_{T,\tau,j} \leq y\}, \qquad \tau = 1, 2. \end{aligned}$$

Under strict monotonicity, the estimators $\hat{B}_{N,T,\tau,i}$ are consistent for B_i as $N, T \to \infty$; otherwise, they differ under suitably strong monotonicity failures. Lemma 8.3 provides a precise formal statement of the latter claim. An interesting situation arises here, as failures of strict monotonicity (hence identification of g) rendered undetectable by the weighted averaging (because $\bar{g}_1$ and $\bar{g}_2$ are nevertheless strictly monotone) are in fact cases where B is identified, regardless of the non-monotonicity of $g(x, \cdot)$. Identification of B is often of interest in its own right, for example in modeling asset returns.

Here, the exogeneity assumptions $X_t \perp B$ and $\varepsilon_t \perp B \mid w_\tau(X_t)$, $\tau = 1, 2$ permit inference on monotonicity of $\bar{g}_1$ and $\bar{g}_2$. Further, as we discuss preceding Theorem 8.4 in the appendix, dropping these conditions introduces multiple generic sources of non-monotonicity: rejecting $\tilde{H}_0$ may then be due to non-monotonicity of either $E(g(X_t, B)\ w_\tau(X_t) \mid B)$ or $E(\varepsilon_t\ w_\tau(X_t) \mid B)$, or both. When $E(\varepsilon_t\ w_\tau(X_t) \mid B)$ is non-constant in B, as generally holds when either $X_t \perp B$ or $\varepsilon_t \perp B \mid w_\tau(X_t)$ fail, it is generically non-monotonic. Non-monotonicity of $E(g(X_t, B)\, w_\tau(X_t) \mid B)$ can arise either from the non-monotonicity of g or from the failure of exogeneity, $X_t \perp B$. The appendix contains further discussion.

These statistics now permit specification tests based on an exact analog of D_N,

$$\hat{D}_{NT} \equiv \sum_{i=1}^{N} \ (\hat{B}_{N,T,1,i} - \hat{B}_{N,T,2,i})^2.$$

Use of multiple weighting functions w_τ, $\tau = 1, ..., \mathcal{T}$, leads to analogous test statistics

$$\hat{D}_{NT} \equiv \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} \sum_{i=1}^{N} (\hat{B}_{N,T,\tau,i} - \hat{B}_{N,T,\varsigma,i})^2.$$

To study the asymptotic properties of $\hat{D}_{NT}$ under $\tilde{H}_0$, we write $\|\mathcal{Z}\|_{2+\gamma} \equiv \{E|\mathcal{Z}|^{2+\gamma}\}^{1/(2+\gamma)}$ and impose the following assumptions:

Assumption C.1 (i) Let $Z_{it} \equiv (\varepsilon_{it}, X_{it}')'$ and $Z_i \equiv \{Z_{i1}, Z_{i2}, ...\}$. The sequence $\{(Z_i, B_i)\}$ is IID. (ii) For each i, $\{(X_{it}, \varepsilon_{it})\}$ is strictly stationary and strong mixing with mixing coefficient $\alpha(\cdot)$ satisfying $\sum_{s=1}^{\infty} \alpha(s)^{\gamma/(2+\gamma)} < \infty$ for some $\gamma > 0$.
Assumption C.2 Let $\mathcal{T} \in \mathbb{N}$. For $\tau = 1, 2, ..., \mathcal{T}$, $w_\tau : \mathcal{X} \to \mathbb{R}^+$ is a measurable function such that for some $C < \infty$, $\|g(X_{it}, B_i)\, w_\tau(X_{it})\|_{2+\gamma} < C$ and $\|\varepsilon_{it}\, w_\tau(X_{it})\|_{2+\gamma} < C$.
Assumption C.3 (i) For $\tau = 1, 2, ..., \mathcal{T}$, the CDF $\tilde{F}_\tau$ of $\tilde{Y}_{\tau,i} \equiv E[Y_{it}\, w_\tau(X_{it}) \mid B_i]$ admits a PDF $\tilde{f}_\tau$ that is uniformly bounded on its support. (ii) For $\tau = 1, 2, ..., \mathcal{T}$ and sufficiently large T, the CDF $\tilde{F}_{T\tau}$ of $\bar{Y}_{T,\tau,i}$ admits a PDF $\tilde{f}_{T\tau}$ that is continuous on its support, and $\overline{g}_\tau$ is continuous, where $\overline{g}_\tau(B_i) \equiv E(g(X_{it}, B_i)\, w_\tau(X_{it}) \mid B_i)$.
Assumption C.4 Let $\xi_i \equiv (\tilde{Y}_{1,i}, ..., \tilde{Y}_{\mathcal{T},i})'$, $\psi(\xi_i, \xi_j) \equiv \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} [\mathbf{1}\{\tilde{Y}_{\tau,j} \leq \tilde{Y}_{\tau,i}\} - \tilde{F}_\tau(\tilde{Y}_{\tau,i}) - \mathbf{1}\{\tilde{Y}_{\varsigma,j} \leq \tilde{Y}_{\varsigma,i}\} + \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i})]$, and $\psi^*(u,v) \equiv \int \psi(\xi, u)\, \psi(\xi, v)\, \tilde{F}(d\xi)$, where $\tilde{F}$ denotes the CDF of ξ_i. The non-zero eigenvalues[2] λ_j, $j = 1, 2, ...$, for $\psi^*(u,v)$ satisfy $\sum_{j=1}^{\infty} |\lambda_j| < \infty$.
Assumption C.5 As $N \to \infty$, $T/N \to \infty$.

Together, Assumptions C.0 and C.1 specify the data generating process. Given the exogeneity assumptions $X_t \perp B$ and $\varepsilon_t \perp B \mid w_\tau(X_t)$, $\tau = 1, ..., \mathcal{T}$, strict monotonicity implies $\tilde{H}_0$, as discussed above. Assumption C.1 rules out cross-section dependence across individuals and non-stationarity across time. We can relax strict stationarity at the cost of more complicated notation. Assumption C.2 imposes some moment conditions. Assumption C.3(i) is weak. Assumption C.4 is used to establish the asymptotic distribution of a certain degenerate second-order U-statistic. Assumption C.5 imposes conditions on (N, T) that greatly facilitate the asymptotic analysis. As we show below, however, suitable bootstrap methods deliver reliable finite sample inference even when T is a modest multiple of N.

Define the bias term

$$\mathcal{B}_{NT} \equiv N^{-2} \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} \sum_{i=1}^{N} \sum_{j \neq i}^{N} [\mathbf{1}\{\tilde{Y}_{\tau,j} \leq \tilde{Y}_{\tau,i}\} - \tilde{F}_\tau(\tilde{Y}_{\tau,i}) - \mathbf{1}\{\tilde{Y}_{\varsigma,j} \leq \tilde{Y}_{\varsigma,i}\} + \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i})]^2.$$

We can now describe the asymptotic distribution of $\hat{D}_{NT}$ under $\tilde{H}_0$ as $N \to \infty$.

[2] The eigenvalues depend on the choice of basis for the underlying Hilbert space; the specifics are not critical here. See Chen and White (1998) for details.

Theorem 4.1 *Suppose Assumptions C.0-C.5 hold. Then under* $\tilde{H}_0 : \tilde{F}_\tau(\tilde{Y}_{\tau,i}) = \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i})$ *for* $\tau, \varsigma = 1, 2, ..., \mathcal{T}$, $\hat{D}_{NT} - \mathcal{B}_{NT} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j (\mathcal{Z}_j^2 - 1)$, *where* $\{\mathcal{Z}_j\}$ *is a sequence of IID* $N(0,1)$ *random variables.*

The proof shows that $\hat{D}_{NT}$ is asymptotically equivalent to an infeasible test statistic ($\bar{D}_{NT}$) based on the unobservable $\tilde{Y}_{\tau,i}$'s. After centering with $\mathcal{B}_{NT}$, $\bar{D}_{NT}$ can be written as a second-order degenerate U-statistic whose asymptotic distribution has been well studied (see, e.g., Chen and White, 1998).

To implement the test, we consistently estimate $\mathcal{B}_{NT}$ with

$$\hat{\mathcal{B}}_{NT} \equiv N^{-2} \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} \sum_{i=1}^{N} \sum_{j \neq i}^{N} [\mathbf{1}\{\bar{Y}_{T,\tau,j} \leq \bar{Y}_{T,\tau,i}\} - \hat{F}_{N,T,\tau}(\bar{Y}_{T,\tau,i}) - \mathbf{1}\{\bar{Y}_{T,\varsigma,j} \leq \bar{Y}_{T,\varsigma,i}\} + \hat{F}_{N,T,\tau}(\bar{Y}_{T,\varsigma,i})]^2.$$

It is straightforward to show that $\hat{\mathcal{B}}_{NT} - \mathcal{B}_{NT} = o_P(1)$. Then we have

$$J_{NT} \equiv \hat{D}_{NT} - \hat{\mathcal{B}}_{NT} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \left(\mathcal{Z}_j^2 - 1 \right) \text{ under } \tilde{H}_0.$$

As the limiting distribution depends on the difficult to estimate nuisance parameters $\{\lambda_j\}$, we will propose a bootstrap method to obtain the needed p-values.

To examine the asymptotic local power of the J_{NT} test, we consider the sequence of Pitman local alternatives

$$\tilde{H}_1(\gamma_N) : \tilde{F}_\tau(\tilde{Y}_{\tau,i}) - \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i}) = \gamma_N \delta_{N,\tau,\varsigma}(\tilde{Y}_{\tau,i}, \tilde{Y}_{\varsigma,i}) \text{ for } 1 \leq \tau \neq \varsigma \leq \mathcal{T},$$

where $\gamma_N \to 0$ as $N \to \infty$ and the $\delta_{N,\tau,\varsigma}$'s are continuous functions such that $\mu \equiv \lim_{N\to\infty} \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} E[\delta_{N,\tau,\varsigma}(\tilde{Y}_{\tau,i}, \tilde{Y}_{\varsigma,i})]^2 < \infty$. The following theorem establishes the asymptotic local power of the J_{NT} test.

Theorem 4.2 *Suppose Assumptions C.0-C.5 hold. Then under* $\tilde{H}_1\left(N^{-1/2}\right)$, $J_{NT} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j (\mathcal{Z}_j^2 - 1) + \mu$.

Theorem 4.2 shows that the J_{NT} test detects local alternatives converging to the null at rate $N^{-1/2}$.

The next theorem establishes the consistency of the test.

Theorem 4.3 *Suppose Assumptions C.0-C.5 hold. Then under* $\tilde{H}_1 \equiv \tilde{H}_1(1)$, $N^{-1} J_{NT} = \mu + o_P(1)$, *where* $\mu \equiv \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} E\left[\tilde{F}_\tau(\tilde{Y}_{\tau,i}) - \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i})\right]^2$, *so that* $P(J_{NT} > c_N) \to 1$ *under* H_1 *for any nonstochastic sequence* $c_N = o(N)$.

4.2 Fully Nonseparable Structures

We now consider fully nonseparable structures of the form

$$Y_t = g(X_t, \varepsilon_t, \beta(A)), \quad t = 1, 2, \ldots .$$

Formally, we impose

Assumption D.0 Assumption A.1 holds with $\phi(x,u) = g(x, \varphi(v), \beta(a))$, where $\beta : \mathbb{R}^{\ell_1} \to \mathbb{I}$, $\varphi : \mathbb{R}^{\ell_2} \to \mathbb{R}$, and $g : \mathbb{R}^d \times \mathbb{R} \times \mathbb{I} \to \mathbb{R}$ are measurable; and with $U_t = (A, V_t)$, $Y_t = \phi(X_t, U_t) = g(X_t, \varphi(V_t), \beta(A))$, $t = 1, 2, \ldots$.

As above, $U_t = (A, V_t)$ contains a time-invariant attribute vector, A, and a time varying driver, V_t; we again write $B \equiv \beta(A)$ and $\varepsilon_t \equiv \varphi(V_t)$. Taking ε_t to be scalar is without loss of generality, in view of Proposition 2.2. We first discuss specification testing; we then briefly provide further discussion of identification.

The key step in treating this case is to view (X_t, ε_t) here as corresponding to X_t in the partially nonseparable case. Thus, we impose the exogeneity condition $(X_t, \varepsilon_t) \perp B$, and the monotonicity condition becomes that $g(x, e, \cdot)$ is strictly monotone. The only difference is that because ε_t is unobservable, we cannot directly construct weights using ε_t; instead, the weights are functions only of X_t. As above, let $\{X_t, \varepsilon_t\}$ be ID, and let non-negative weight functions w_1 and w_2 be defined on $\mathcal{X}$, such that $E(w_\tau(X_t)) = 1$, $\tau = 1, 2$. Let $\tilde{Y}_1 = \tilde{Y}_{1t} \equiv E(Y_t\ w_1(X_t) \mid B)$ and $\tilde{Y}_2 = \tilde{Y}_{2t} \equiv E(Y_t\ w_2(X_t) \mid B)$. Then

$$\begin{aligned} \tilde{Y}_1 &= E(g(X_t, \varepsilon_t, B)\ w_1(X_t) \mid B) = \int g(x, e, B)\ w_1(x)\ dF(x, e) \equiv \bar{g}_1(B), \qquad \text{and} \\ \tilde{Y}_2 &= E(g(X_t, \varepsilon_t, B)\ w_2(X_t) \mid B) = \int g(x, e, B)\ w_2(x)\ dF(x, e) \equiv \bar{g}_2(B), \end{aligned}$$

where the second equality in each line holds given $(X_t, \varepsilon_t) \perp B$.

The development of the previous section applies immediately, with the obvious modifications, so that $\tilde{F}_1(\tilde{Y}_1) = B = \tilde{F}_2(\tilde{Y}_2)$ given strict monotonicity. Thus, we again test

$$\tilde{H}_0 : \tilde{F}_1(\tilde{Y}_1) = \tilde{F}_2(\tilde{Y}_2).$$

The statistics and tests are identical. Lemma 8.3 and Theorem 8.4 apply with (X_t, ε_t) replacing X_t, so we do not repeat our previous discussion. The only real difference from the partially separable case is that here the test may lack power against certain alternatives that can only be revealed by using weights that depend on ε_t.

To close this subsection, we briefly discuss identification. If indeed $g(x, e, \cdot)$ is strictly monotone and $(X_t, \varepsilon_t) \perp B$, then, as we have just seen, B is identified as, e.g., $B = \tilde{F}(\tilde{Y})$, with $\tilde{Y} = \tilde{Y}_t \equiv$

$E(Y_t \mid B)$ and $\tilde{F}$ the CDF of $\tilde{Y}$. Thus, B can be consistently estimated when $T \to \infty$; in this sense, B is known asymptotically. One can then identify g using the results of Section 2, treating X and B as the observables, with ε the sole scalar unobservable. Specifically, with $g(x, \cdot, b)$ strictly monotone and $(X, B) \perp \varepsilon$, Proposition 2.1 identifies $g(x, \cdot, b)$ and e. These identifications may be useful for testing whether or not the structural function is partially nonseparable. Further, they may be helpful in refining estimation for the partially nonseparable case treated by **E**. As these topics are well beyond the scope of the present study, we leave them for future research.

5 Monte Carlo Simulations

In this section we conduct some Monte Carlo experiments to evaluate the finite sample performance of our tests. We first consider the nonseparable case where the unobservables are attributes varying across individuals but not time. Then we consider the nonseparable case where we have both time-invariant and time-varying unobservables.

5.1 Unobservable Attributes

We consider two data generating processes (DGPs):

DGP 1. $Y_{it} = 1 + X_{it} + (1 + \delta X_{it}) A_i$

DGP 2. $Y_{it} = \sqrt{0.4 + X_{it}^2} A_i - \delta A_i^2 \varphi_{(0,0.5)} (X_{it})$,

where for $i = 1, ..., N$, $t = 1,\ 2$, A_i is IID $U(0, 1)$; X_{it} is IID, computed as the sum of 48 independent $U(-0.25, 0.25)$ random variables; X_{it} is independent of A_j for each i, j, t; $\varphi_{(0,0.5)}$ is the normal PDF with mean 0 and variance 0.5; and δ controls the degree of violation of monotonicity in A_i. We use $\delta = 0$ and $\delta > 0$ to study the finite sample level and power properties of our test, respectively. Note that by construction, both X_{it} and Y_{it} have compact support for each t. On the other hand, according to the central limit theorem we can treat X_{it} as being nearly standard normal random but with compact support $[-12, 12]$.

To construct the test statistic D_N, we estimate the conditional CDF $F_t (Y_{it} \mid X_{it})$ using leave-one-out local linear regression ($p = 1$). We choose the Gaussian kernel $K(x) = (2\pi)^{-1/2} \exp\left(-\frac{1}{2}x^2\right)$. Since there is no data-driven procedure for the optimal choice of bandwidth for our test, we follow a rule of thumb, choosing the bandwidth as $h = 0.5 s_X N^{-1/4}$, where s_X is the geometric average of the sample standard deviations $\{X_{it}\}_{i=1}^N$, $t = 1, 2$. Note that we use undersmoothing to eliminate the effect of the finite sample bias of the CDF estimates.

It is well known that the asymptotic normal distribution typically does not give a good approximation to the finite sample distribution of many nonparametric tests. Thus, we suggest a bootstrap method, yielding bootstrap critical values or p-values. For this, we generate bootstrap data $\{(Y_{it}^*, X_{it}^*) : i = 1, ..., N,\ t = 1, 2\}$ as follows:

1. Set $X_{it}^{*} = X_{it}$ for each (i,t) and generate A_i as IID $U(0,1)$ random variables.

2. Generate Y_{it}^{*} as $\tilde{F}_{NT}^{-1}(A_i \mid X_{it}^{*})$ where $\tilde{F}_{NT}^{-1}(\tau \mid x)$ is a nonparametric estimate of the quantile function $F^{-1}(\tau \mid x)$ (since $F_1^{-1}(\tau \mid x) = F_2^{-1}(\tau \mid x)$ under the null).

We consider two types of nonparametric estimates for $F^{-1}(\tau \mid x)$: one is the local linear (LL) quantile regression estimate (e.g., Yu and Jones, 1998), and the other is obtained by inverting a weighted Nadaraya-Watson (WNW) estimate of $F(y \mid x)$ (e.g., Cai, 2002). Specifically, the former is defined as the minimizing intercept in the problem

$$\min_{\{\beta_0,\beta_1\}} \sum_{t=1}^{2}\sum_{i=1}^{N} \rho_\tau\left(Y_{it} - \beta_0 - \beta_1'(X_{it} - x)\right) K_{\tilde{h}}(X_{it} - x), \tag{8}$$

where $\rho_\tau(z) \equiv z(\tau - \mathbf{1}\{z \le 0\})$ is the "check" function, and $\tilde{h} \equiv \tilde{h}(N)$ is the bandwidth. It is well known that the solution to this problem is determined by fitting certain observations exactly and that the resulting quantile function estimate must be non-monotonic in its argument τ at some values of X_{it}. As a result, $Y_{it}^{*} = \tilde{F}_{NT}^{-1}(A_i \mid X_{it}^{*})$ may not be monotonic in A_i for some values of X_{it}^{*}; this might have some adverse impact on test performance. Further, the LL estimate does not constrain the corresponding CDF estimate to lie between zero and one. To avoid these potential problems, Cai (2002) proposed a WNW estimator for $F(y \mid x)$ and inverted it to obtain the estimate of $F^{-1}(\tau \mid x)$. Let

$$\hat{F}_{wnw}(y \mid x) \equiv \frac{\sum_{t=1}^{2}\sum_{i=1}^{N} p_{it}(x) K_{\tilde{h}}(X_{it} - x)\mathbf{1}\{Y_{it} \le y\}}{\sum_{t=1}^{2}\sum_{i=1}^{N} p_{it}(x) K_{\tilde{h}}(X_{it} - x)}, \tag{9}$$

where the nonnegative weight functions, $p_{it}(x)$, are chosen such that

$$\sum_{t=1}^{2}\sum_{i=1}^{N} p_{it}(x) = 1, \text{ and } \sum_{t=1}^{2}\sum_{i=1}^{N} p_{it}(x)(X_{it} - x)K_{\tilde{h}}(X_{it} - x) = 0. \tag{10}$$

Cai (2002) proposed choosing $\{p_{it}(x)\}$ using empirical likelihood, i.e., to maximize $\sum_{t=1}^{2}\sum_{i=1}^{N} \log\{p_{it}(x)\}$ subject to the constraints specified in (10). Then $\tilde{F}_{NT}^{-1}(\tau \mid x)$ is given by $\inf\{y \in \mathbb{R} : \hat{F}_{wnw}(y \mid x) \ge \tau\}$. This ensures the monotonicity of $Y_{it}^{*} = \tilde{F}_{NT}^{-1}(A_i \mid X_{it}^{*})$ in A_i. For our bootstrap, we take $\tilde{h} = 2\tilde{s}_X N^{-1/6}$ where $\tilde{s}_X$ is the sample standard deviation of $\{X_{it}, 1 \le i \le N,\ t = 1, 2\}$.

We consider two sample sizes, $N = 100$, 200 in our simulation study. Due to the considerable computational burden for the bootstrap, we use 250 replications for each sample size N and 100 bootstrap resamples in each replication.

Table 1 reports the empirical rejection frequencies for our test at various nominal levels for DGPs 1-2. The bootstrap p-values are obtained in two ways: one is based on the LL conditional quantile regression, and the other is based on Cai's inverse CDF estimator. When $\delta = 0$, the

Table 1: Finite sample rejection frequency for DGPs 1-2

DGP	N	δ	Bootstrap based on LL estimate			Bootstrap based on inverse CDF		
			1%	5%	10%	1%	5%	10%
1	100	0	0.008	0.032	0.088	0.012	0.056	0.130
		0.5	0.052	0.164	0.256	0.208	0.376	0.488
		1	0.752	0.864	0.908	0.904	0.960	0.968
	200	0	0.008	0.036	0.076	0.020	0.048	0.112
		0.5	0.192	0.412	0.528	0.496	0.716	0.796
		1	0.996	1.000	1.000	0.996	1.000	1.000
2	100	0	0.004	0.036	0.080	0.012	0.052	0.104
		0.5	0.020	0.092	0.168	0.032	0.128	0.220
		1	0.728	0.884	0.944	0.744	0.916	0.956
	200	0	0.004	0.048	0.080	0.004	0.052	0.112
		0.5	0.044	0.136	0.216	0.040	0.156	0.240
		1	0.988	1.000	1.000	1.000	1.000	1.000

corresponding rows report the empirical level. We summarize the main findings from Table 1. First, the level of our test is fairly well behaved, and it can be close to the nominal level for sample sizes as small as $N = 100$. When N increases, the level generally improves somewhat. Second, the power of our test is reasonably good. It increases quickly as either δ increases or the sample size doubles. Third, the inverse CDF method generally outperforms the LL method.

5.2 Time-varying Unobservables and Unobservable Attributes

We consider the following three DGPs:

DGP 3. $Y_{it} = 1 + X_{it} + (1 + \delta X_{it}) B_i + \sigma\varepsilon_{it}$

DGP 4. $Y_{it} = \sqrt{0.4 + X_{it}^2} B_i - \delta B_i^2 \varphi_{(0,0.5)} (X_{it}) + \sigma\varepsilon_{it}$

DGP 5. $Y_{it} = \varphi_{(0,0.5)} (X_{it}) + (0.6 + 0.15 X_{it}) \varepsilon_{it} + (1 + \delta X_{it}) B_i$,

where for $i = 1, ..., N$, $t = 1, ..., T$, $B_i = A_i$ and X_{it} are generated as in DGPs 1-2; ε_{it} is IID $N(0, 1)$ across i and t, and independent of X_{js} and B_j for all i, t, j, s; and σ is taken to ensure that the signal-to-noise ratio in DGPs 3-4 is 1 across all simulations. The structures in DGPs 3-4 are partially nonseparable, whereas that in DGP 5 is fully nonseparable.

To construct our test statistic, we need to choose the weight functions $w_\tau(\cdot)$, $\tau = 1, 2, ..., \mathcal{T}$. For fixed $\mathcal{T}$, let $\tilde{q}_0 = -\infty$, $\tilde{q}_{\mathcal{T}} = \infty$, and let $\tilde{q}_\tau$ denote the sample $\tau/\mathcal{T}$-quantile of $\{X_{it}, 1 \leq i \leq N, 1 \leq t \leq T\}$ for $1 \leq \tau \leq \mathcal{T} - 1$. Then let

$$w_\tau (X_{it}) = \mathbf{1} \{\tilde{q}_{\tau-1} \leq X_{it} \leq \tilde{q}_\tau\}, \ \tau = 1, 2, ..., \mathcal{T}.$$

Under Assumption C.2(i) we can show that the sample quantiles estimate their population analog at the rate $(NT)^{-1/2}$, so this estimation error plays an asymptotically negligible role in our analysis.

As remarked earlier, the asymptotic distribution of J_{NT} depends on the sequence of eigenvalues $\{\lambda_j\}$, which is difficult to estimate accurately in practice. Further, our asymptotic theory relies on $T/N \to \infty$ as $N \to \infty$, which may appear too strong for many applications. Nevertheless, we can circumvent both issues using a suitable bootstrap method. Specifically, we propose the following procedure to obtain bootstrap p-values for the J_{NT} test:

1. For $i = 1, ..., N$, set $\hat{B}_{N,T,i} \equiv \hat{F}_{N,T}(\bar{Y}_{T,i})$, where $\bar{Y}_{T,i} \equiv T^{-1}\sum_{t=1}^{T} Y_{it}$, and $\hat{F}_{N,T}(\cdot) \equiv N^{-1}\sum_{i=1}^{N} \mathbf{1}\{\bar{Y}_{T,i} \leq \cdot\}$.

2. For $i = 1, ..., N$, and $t = 1, ..., T$, estimate $g(X_{it}, \hat{B}_{N,T,i})$ using the local linear regression of Y_{it} on $(X_{it}, \hat{B}_{N,T,i})$ and by imposing the monotonicity of $g(x, b)$ in b (details given below). Let $\hat{g}(X_{it}, \hat{B}_{N,T,i})$ denote the estimate.

3. Let $\hat{\varepsilon}_i \equiv (\hat{\varepsilon}_{i1}, ..., \hat{\varepsilon}_{iT})'$, where $\hat{\varepsilon}_{it} \equiv Y_{it} - \hat{g}(X_{it}, \hat{B}_{N,T,i})$. For $i = 1, ..., N$, randomly draw ε_i^* from $\{\hat{\varepsilon}_1, ..., \hat{\varepsilon}_N\}$ with replacement. Let ε_{it}^* denote the tth element of ε_i^*. Generate Y_{it}^* according to[3]
$$Y_{it}^* = \hat{g}(X_{it}, \hat{B}_{N,T,i}) + \varepsilon_{it}^*.$$

4. Compute the bootstrap test statistic J_{NT}^* in the same way as J_{NT} using $\{(X_{it}, Y_{it}^*), 1 \leq i \leq N, 1 \leq t \leq T\}$.

5. Repeat steps 3 and 4 B times to obtain B bootstrap test statistics $\{J_{NT,j}^*\}_{j=1}^{B}$. Calculate the bootstrap p-values $p^* \equiv B^{-1}\sum_{j=1}^{B} \mathbf{1}\{J_{NT,j}^* \geq J_{NT}\}$ and reject $\tilde{H}_0$ if $p^*$is smaller than the prescribed level of significance.

We impose the null hypothesis of monotonicity in Step 2. There exists a vast literature on the problem of estimating a monotone regression function. See, e.g., Dette, Neumeyer, and Pilz (2006, **DNP**) and the references there. **DNP** consider kernel estimation of a monotone regression function that is a function of a single variable. Compared to other approaches, theirs has the great advantage of simplicity, as it does not require constrained optimization; further, it is asymptotically equivalent to the unconstrained kernel estimate. Here we modify their procedure to allow another variable (X_{it} here) to enter the regression function non-monotonically. This procedure has three steps:

[3]This method also works for the fully nonseparable structure, where $Y_{it} = g(X_{it}, \varepsilon_{it}, B_i)$ with $(X_{it}, \varepsilon_{it}) \perp B_i$, as in DGP 5. Let $\bar{g}(X_{it}, B_i) \equiv E(Y_{it}|X_{it}, B_i) = E[g(X_{it}, \varepsilon_{it}, B_i)|X_{it}, B_i]$. Then $Y_{it} = \bar{g}(X_{it}, B_i) + \bar{\varepsilon}_{it}$, where $\bar{\varepsilon}_{it} \equiv Y_{it} - \bar{g}(X_{it}, B_i)$, and $\bar{g}(x, \cdot)$ is monotone for all x provided $g(x, \varepsilon, \cdot)$ is monotone for all (x, ε). This ensures that we can generate the bootstrap analog of Y_{it} using estimates of $\bar{g}$ for the fully nonseparable case.

Table 2: Finite sample rejection frequency for DGPs 3-5 ($\delta = 0$)

DGP	N	T	5% test				10% test			
			$\mathcal{T}=2$	$\mathcal{T}=3$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=2$	$\mathcal{T}=3$	$\mathcal{T}=5$	$\mathcal{T}=10$
3	50	50	0.116	0.080	0.080	0.084	0.176	0.144	0.144	0.160
	50	100	0.084	0.060	0.044	0.032	0.180	0.140	0.120	0.128
	100	50	0.068	0.072	0.036	0.028	0.124	0.132	0.116	0.096
	100	100	0.048	0.064	0.024	0.028	0.112	0.112	0.092	0.072
4	50	50	0.156	0.072	0.068	0.036	0.248	0.156	0.136	0.064
	50	100	0.084	0.080	0.084	0.036	0.180	0.160	0.148	0.116
	100	50	0.088	0.044	0.020	0.016	0.180	0.120	0.052	0.052
	100	100	0.076	0.036	0.032	0.016	0.124	0.084	0.060	0.028
5	50	50	0.100	0.012	0.024	0.032	0.184	0.048	0.052	0.084
	50	100	0.092	0.028	0.032	0.032	0.180	0.056	0.076	0.116
	100	50	0.084	0.024	0.008	0.024	0.140	0.060	0.028	0.060
	100	100	0.080	0.012	0.012	0.028	0.132	0.032	0.024	0.080

Step 1. Let J be a large integer such that $J \to \infty$ as $N \to \infty$. For $i = 1, ..., N$, $t = 1, ..., T$, and $j = 1, ..., J$, compute the conventional local linear estimate $\tilde{g}\left(X_{it}, j/J\right)$ of $g\left(X_{it}, j/J\right)$ by using the product of Gaussian kernels (k) and bandwidth ($h = (h_x, h_b)$) chosen according to Silverman's rule of thumb.

Step 2. For $i = 1, ..., N$ and $t = 1, ..., T$, obtain the estimate $\hat{g}^{-1}\left(X_{it}, z\right) = (Jh_d)^{-1} \sum_{j=1}^{J} \int_{-\infty}^{z} k\left(h_d^{-1}[\tilde{g}\left(X_{it}, j/J\right) - b]\right) db$, which estimates the inverse function $g^{-1}\left(X_{it}, \cdot\right)$ at z, where the inverse is taken with respect to the second argument of g for fixed X_{it}.

Step 3. Compute the estimate $\hat{g}(X_{it}, \hat{B}_{N,T,i}) = \inf\{z : \hat{g}^{-1}\left(X_{it}, z\right) \geq \hat{B}_{N,T,i}\}$.

Under conditions similar to those of **DNP**, we can show that $\hat{g}(X_{it}, \hat{B}_{N,T,i})$ is asymptotically equivalent to $\tilde{g}(X_{it}, \hat{B}_{N,T,i})$, although only the former is guaranteed to be monotone in its second argument. In the simulations, $h_d = h_b^2$, and we choose $J = 40$ to save computation time.

Tables 2-3 report the empirical rejection frequencies for the J_{NT} test at the 5% and 10% nominal levels for $\delta = 0$ and 1, respectively. Here, we use 250 replications for each sample size (N, T) and 200 bootstrap resamples in each replication. From Table 2, we see that the choice of $\mathcal{T}$ and thus the weight function w_τ ($\tau = 1, 2, ..., \mathcal{T}$) is important for the level behavior of the test. For small values of $\mathcal{T}$ (say 2), the test tends to be oversized, but the size distortion becomes less severe as either N or T increases. On the other hand, the test is undersized when $\mathcal{T}$, N, and T are all large, giving a conservative test. Table 3 indicates that our J_{NT} test has useful power in detecting departures from monotonicity in unobservables. The power performance also depends on the choice of $\mathcal{T}$. Choices of $\mathcal{T}$ that are too small or too large may have adverse effects on power performance. Also, both N and T affect the power: for DGPs 3 and 5, as N doubles, the

Table 3: Finite sample rejection frequency for DGPs 3-5 ($\delta = 1$)

DGP	N	T	5% test				10% test			
			$\mathcal{T}=2$	$\mathcal{T}=3$	$\mathcal{T}=5$	$\mathcal{T}=10$	$\mathcal{T}=2$	$\mathcal{T}=3$	$\mathcal{T}=5$	$\mathcal{T}=10$
3	50	50	0.804	0.908	0.900	0.668	0.876	0.956	0.964	0.800
	50	100	0.768	0.928	0.928	0.832	0.880	0.980	0.976	0.964
	100	50	0.944	0.972	0.960	0.760	0.980	1	0.992	0.908
	100	100	1	1	1	1	1	1	1	1
4	50	50	0.216	0.196	0.216	0.176	0.344	0.352	0.340	0.292
	50	100	0.184	0.252	0.268	0.236	0.292	0.412	0.408	0.384
	100	50	0.192	0.268	0.200	0.144	0.332	0.444	0.320	0.292
	100	100	0.364	0.552	0.452	0.360	0.512	0.666	0.596	0.548
5	50	50	0.848	0.964	0.992	1	0.924	0.988	0.996	1
	50	100	0.692	0.984	1	1	0.800	1	1	1
	100	50	0.964	1	1	1	0.992	1	1	1
	100	100	1	1	1	1	1	1	1	1

power increases regardless of whether T doubles or not; whereas for DGP 4, the power noticeably increases only when both N and T increase.

6 Two Applications

In this section we apply the methods put forward here to two applications, one from finance and one from consumer demand. They are meant to illustrate the power of our test to detect model deviations from exogeneity and scalar monotonicity. We have selected these two examples, because they are in a sense polar cases: In the finance literature, since Fama and French's (1993) seminal contribution, the emphasis is on reduced form explanation. Exogeneity is taken as given; our test hence examines whether there is a single firm-specific "fourth factor" that impacts the firm's valuation. Commonly, such a factor would be associated with the firms's quality or reputation. Maintaining the assumption of exogeneity, our test becomes a test of scalar monotonicity.

In contrast, in consumer demand, the models are rather structural, and exogeneity is hence implausible. Nevertheless, since the seminal work of Berry, Levinsohn, and Pakes (1995), monotonicity in a scalar unobservable is commonly assumed. Typically, the unobservable is an unobserved product characteristic, most often associated with quality. A recent reference that discusses nonparametric identification with scalar monotonicity is Berry and Haile (2010). Maintaining scalar monotonicity, our test becomes a test of exogeneity of the own price.

6.1 An Application from Finance

A major advance in understanding asset return behavior is the Fama and French (1993) ("FF") factor model of asset returns, which can be written

$$Y_{it} = \alpha_i + \beta_i' X_t + \eta_{it}, \tag{11}$$

where Y_{it} is the excess return of asset i in period t (net returns minus the T-Bill return); $X_t = (RMRF_t, SMB_t, HML_t)'$ is a vector of returns factors, where $RMRF_t$ is the period t excess return on a value-weighted aggregate market proxy portfolio, and SMB_t and HML_t are period t returns on value-weighted, zero-investment factor-mimicking portfolios for size and book-to-market equity, respectively, η_{it} is an exogenous shock, α_i is the asset's idiosyncratic return ("alpha"), and the elements of β_i are risk premia associated with the corresponding risk factors.

An extension of this model permits time-varying risk premia, β_{it} :

$$Y_{it} = \alpha_i + \beta_{it}' X_t + \eta_{it}. \tag{12}$$

See, for example, Harvey (1989), Ferson and Harvey (1991, 1993), Jagannathan and Wang (1996), and Ghysels (1998) for discussion of the importance of time-varying risk premia.

Here, we apply our monotonicity test to stock returns following a nonparametric version of the time-varying Fama-French model,

$$Y_{it} = g(X_t, \varepsilon_{it}, B_i), \tag{13}$$

where ε_{it} corresponds to $(\eta_{it}, \beta_{it}')'$ and B_i corresponds to α_i. Our theory allows, but does not require, X_t to also vary with i. The exogeneity condition is that $(X_t, \varepsilon_{it}) \perp B_i$. This is plausible if we think of B_i ($\Leftrightarrow \alpha_i$) as a persistent attribute specific to firm i, say, its firm culture, while market factors X_t are unrelated to the firm's attributes, and we view ε_{it} ($\Leftrightarrow (\eta_{it}, \beta_{it}')$) as transitory shocks like changes in firm management and in investor risk preferences that drive risk premia. The other regularity conditions of our theory also plausibly apply to the stock returns data we describe below, so we interpret our test as a test for strict monotonicity in B.

Although the monotonicity property is straightforward, it is important to understand the possible reasons for rejection in the present context. One possibility is that a single B_i interacts with shocks, risk factors, and risk preferences determining risk premia in possibly complicated ways. Another is that there are multiple firm-specific factors influencing asset returns. If either possibility holds, then eq.(12) is not a correct description of the data generating process, so that linear FF models with time-varying risk premia are misspecified, and there is no single persistent factor that captures the firm's attributes in a way that allows attaching a single permanent quality factor to their returns.

6.1.1 Data

Our factor data come from French's webpage[4] and are merged with data from Yahoo! finance. We obtained weekly stock price data for $N = 50$ companies randomly chosen from the S&P 500; a list of the firms analyzed can be found in Table A.1 of the appendix. We limit ourselves to fifty firms to ensure that $T > N$, while keeping computation costs manageable.

The data span a period of $T = 610$ weeks between 7/17/1998 and 3/26/2010. Note that when querying Yahoo's "weekly" data, the listed date is for the beginning of the trading week (usually a Monday), but the reported price is that week's closing price (usually a Friday). The data from French's webpage reports a week's last trading day's data, and labels that observation with the date of that week's last trading day.

For each firm i, we calculate returns in period t as $Y_{it} = [(P_{it}/P_{i,t-1}) - 1] - RF_t$, where P_{it} is the closing price (adjusted for splits and dividends) and RF_t is the risk free return, also obtained from French's webpage.

6.1.2 Results

To apply our test procedure to the data described above, we use the following specifications: The estimation method is local linear regression; the bandwidth is chosen by Silverman's rule of thumb, as in the simulations, but adjusted according to the number of continuous regressors. The kernel is a product of Gaussian kernels. The weighting is performed as follows: we first calculate $(X_{it,k} - \bar{X}_k)^2$ for the k-th regressor (here $\bar{X}_k$ is the sample mean), then sum over the regressors (indexed by k); we then use the quantile-partition weights based on this sum. The test statistic is computed just as in the simulations, following exactly the same steps as in Section 5.2.

The results are summarized in Table 4. In all instances, we soundly reject the strict monotonicity hypothesis. This implies that there is no single persistent factor that captures firm differences in a way that corresponds to alpha. This calls into question the linear time-varying FF model and suggests that additional effort might be profitably directed toward gaining a better understanding of the relation between firms' stock returns, firm characteristics, market factors, and investor risk preferences. This also resolves a puzzle: why do countless studies find statistically significant non-zero alphas if the market is in fact efficient? These results suggest a compelling reason: the linear FF model, even with time-varying risk premia, is not an accurate description of the DGP. Our procedure permits a more stringent test of this aspect of market efficiency.

[4] We obtained weekly Fama-French factor data from Ken French's website: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
The precise definitions of the factors can also be found here:
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

Table 4: p-values for monotonicity test - asset returns

Bootstrap Replications	N	T	p-values			
			$\mathcal{T}=2$	$\mathcal{T}=3$	$\mathcal{T}=5$	$\mathcal{T}=10$
$B=200$	50	610	<0.005	<0.005	<0.005	<0.005
$B=5000$	50	610	0.0002	0.0002	<0.0002	<0.0002
$B=10000$	50	610	0.0001	0.0001	<0.0001	<0.0001

Note, however, that even with the failure of monotonicity, useful information about risk premia may still be recovered from nonparametric specifications of the sort used here. Although monotonicity failure rules out identifying alpha, the further exogeneity condition $X_t \perp (B_i, \varepsilon_{it})$ permits recovery of expected risk premia, such as $E(D_k g(x, \varepsilon_{it}, B_i))$, where $D_k \equiv \partial/\partial x_k$, even in the absence of strict monotonicity, as implied by results of **AM**. Certain quantile effects may also be of interest; these are identified by results of Hoderlein and Mammen (2007).

6.2 An Application from Consumer Demand

In contrast to finance, in consumer demand exogeneity is a frequently criticized assumption, for instance due to simultaneity (the firms base their price-setting behavior on expected demand, but demand depends on prices), or due to omitted characteristics of the product. However, it is often argued that this endogeneity is due to a product-specific factor that may in fact enter monotonically (Berry, Levinsohn, Pakes (1992); Berry and Haile (2010)). Hence, for the rest of this section, we maintain the assumption that scalar monotonicity holds. Note that our general nonseparable approach is ideally suited to this problem: as we are considering an aggregate consumption relationship, we face, in general, a highly nonlinear relationship.

6.2.1 Data

The data are supermarket scanner data collected by Information Resources, Inc. (IRI). The scanner data consist of variables measuring price, quantity, and promotional variables for the full range of available RTE cereal products on a weekly basis, for three years beginning January 2005 and ending December 2007, so that $T = 156$. The data have a panel structure, where the cross-section dimension is a particular supermarket retail chain operating in a particular geographic market. For example, San Diego is represented by three major chains; these are three distinct cross-section units. The cross-section dimension is $N = 70$ supermarket-city pairs. We analyze the top-selling product for each of the five manufacturers. Table A.2 presents a variety of summary statistics for quantity-weighted market share, price, and promotional variables.

Although there are some differences, IRI's definition of a geographic market is roughly equiva-

Table 5: p-values for endogeneity test - cereal

Product	N	T	p-values			
			$\mathcal{T}=2$	$\mathcal{T}=3$	$\mathcal{T}=5$	$\mathcal{T}=10$
G MILLS CHEERIOS 15OZ	70	156	0.0002	< 0.0001	< 0.0001	< 0.0001
KELLOGG FROSTED FLAKES 20OZ	70	156	< 0.0001	0.0001	< 0.0001	< 0.0001
POST HNY BNCHS OATS REG 16OZ	70	156	0.0004	< 0.0001	< 0.0001	< 0.0001
QUAKER LIFE REGULAR 21OZ	70	156	0.0090	0.1834	< 0.0001	< 0.0001
STR BDS RAISIN BRAN 20OZ	70	156	0.0006	0.0005	< 0.0001	< 0.0001

lent to the Census Bureau's metropolitan statistical area (MSA) or combined metropolitan statistical area (CMSA). This is convenient for merging income or demographics data with the scanner data. Here, we merge income data from the Bureau of Labor Statistics (BLS). Specifically, we obtain average weekly wage data for each geographic market from the BLS's Quarterly Census of Employment and Wages (QCEW) database. Wage data are collected quarterly, so although the scanner data contains data at a weekly frequency, the QCEW wage data is only updated quarterly. Although we could merge additional demographic information from the Census Bureau, due to the nonparametric setup, we focus only on those explanatory variables that have the strongest impact in Megerdichian's (2009) parametric study.

6.2.2 Results

In implementing the test, we have applied specifications nearly identical to those of the finance application. The dependent variable is quantity-weighted market share and the explanatory variables are: own price; price of the closest neighbor in product characteristic space; the quantity-weighted average price of all 150 cereals; promotions (an intensity index ranging between zero and one); and weekly wage. See Megerdichian (2009) for details about the data and construction of variables. Table 5 gives the test results.

As is obvious from these results, exogeneity is widely rejected. For all but one product the p-values are virtually zero. Only for "Quaker Life" do we have some evidence that endogeneity might not be an issue, though at higher values of $\mathcal{T}$ we also obtain rejections. A Bonferroni-Hochberg test (Hochberg, 1988) of multiple hypotheses applied to all results in this row gives $p < 0.0001$. Assuming monotonicity and using our general nonparametric test, we conclude that endogeneity is indeed the issue the demand and IO literatures believe it to be. This simple model thus does not properly address confounding effects and the simultaneous structure typical in this literature. We leave a more elaborate approach, closer to the structural IO models now common in the literature, for future analysis.

7 Summary and Concluding Remarks

Monotonicity in a scalar unobservable is a crucial identifying assumption for an important class of nonparametric structural specifications accommodating unobserved heterogeneity. Tests for this monotonicity have previously been unavailable. Here we propose and analyze tests for scalar monotonicity using panel data for structures with and without time-varying unobservables, either partially or fully nonseparable between observables and unobservables. Our nonparametric tests are computationally straightforward, have well behaved limiting distributions under the null, are consistent against relevant and precisely specified alternatives, and have standard local power properties. We provide straightforward bootstrap methods for inference. Some Monte Carlo experiments show that these reasonably control the level of the test, and that our tests have useful power. We apply our tests to study asset returns and demand for ready-to-eat cereals.

For clarity, and to maintain a manageable scope for the analysis here, we focus throughout on the strictly exogenous case. As we show, when exogeneity is not a maintained assumption, then our tests detect either non-monotonicity or exogeneity failure. Thus, it is important to explore whether rejection may be due to the latter failure. For this, one can relax exogeneity to conditional exogeneity, where one maintains that, e.g., X_t is independent of A, given control variables, Z_t ($X_t \perp A \mid Z_t$). The analysis of this case is rather more involved, but, as we show in Hoderlein, Su, and White (2010), analogous results and methods apply in this case as well. As in the analysis of this paper, we abstract there from panel dynamics. An interesting topic for further study is to examine whether and how tests for scalar monotonicity can be conducted in dynamic panel structures.

Another interesting topic for future research is to pursue our suggestions at the end of Section 4 about testing whether the structural function is partially nonseparable and refining estimation for the partially nonseparable case treated by **E**. Finally, there is a considerable variety of opportunities for applying these tests and their further extensions.

8 Mathematical Appendix

Proof of Proposition 2.1 For all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have

$$\begin{aligned} F(y \mid x) &\equiv P[Y \leq y \mid X = x] = P[m(X, \varepsilon) \leq y \mid X = x] \\ &= P[m(x, \varepsilon) \leq y] = \int_0^1 \mathbf{1}\{m(x, e) \leq y\}\, de \\ &= \lambda\{m_x^{-1}(-\infty, y]\}, \end{aligned}$$

where λ denotes Lebesgue measure and $m_x^{-1}(-\infty, y]$ is the preimage in $\mathbb{I}$ of the half-ray $(-\infty, y]$ under $m(x, \cdot)$. The second line uses $X \perp \varepsilon$ and $\varepsilon \sim \mathbb{U}[0, 1]$.

Let x be given. If $m(x,\cdot)$ is strictly increasing, $m_x^{-1}(-\infty, y] = (0, m^{-1}(x,y)]$ and $F(y \mid x) = m^{-1}(x,y)$ for all y. By our convention, this also covers $m(x,\cdot)$ strictly decreasing.

Now suppose that $m(x,\cdot)$ is not strictly increasing. First, suppose that $m(x,\cdot)$ is invertible, and also suppose $F(y \mid x) = m^{-1}(x,y)$ for all y. The monotonicity of $F(\cdot \mid x)$ and the invertibility of $m^{-1}(x,\cdot)$ imply that $m^{-1}(x,\cdot)$ is strictly increasing. But this implies that $m(x,\cdot)$ is strictly increasing, a contradiction, so $F(y \mid x) \neq m^{-1}(x,y)$ for some y.

Finally, if $m(x,\cdot)$ is not invertible, then $m^{-1}(x,\cdot)$ is a correspondence, not a function. But $F(\cdot \mid x)$ is a function, so $F(y \mid x) = m^{-1}(x,y)$ cannot hold for all $y \in \mathcal{Y}$. ■

Proof of Proposition 2.2 (i) The existence of the Borel isomorphism $\upsilon : \mathbb{R}^\ell \to \mathbb{I}$ such that $\varepsilon \equiv \upsilon(U) \sim \mathbb{U}[0,1]$ is a well known straightforward consequence of the Borel isomorphism theorem[5] (see, e.g., **CSZ**, p.417, or Dudley (2002, theorem 13.1.1)). It follows immediately that $Y = \phi(X,U) = \phi(X, \upsilon^{-1}[\upsilon(U)]) = \phi(X, \upsilon^{-1}(\varepsilon))$ *a.s.* (ii) If $X \perp U$, it follows from Dawid (1979, lemma 4.2(i)) and $\varepsilon = \upsilon(U)$ that $X \perp \varepsilon$. Conversely, if $X \perp \varepsilon$, it follows from Dawid (1979, lemma 4.2(i)) and $U = \upsilon^{-1}(\varepsilon)$ that $X \perp U$. ■

Proof of Proposition 2.3 (i) Because V is nonatomic on $\mathbb{I}$, there exists a Borel isomorphism φ such that $\varepsilon = \varphi(V)$ and $\varepsilon \sim \mathbb{U}[0,1]$. Thus, ε and $m(X,\varepsilon) = g(X, \varphi^{-1}(\varepsilon))$ satisfy the conditions of Proposition **2.1**, and the result follows.

(ii) Given A.1, we can apply Proposition **2.2** to verify that $\phi(x, \upsilon^{-1}(\cdot))$ has the measurability properties required for $g(x, \upsilon(x,\cdot))$, and that $V = \upsilon(U)$ is uniformly distributed on $\mathbb{I}$, hence nonatomic. Accordingly, we take $g(x,\cdot) = \phi(x, \upsilon^{-1}(\cdot))$. Then for all $(x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} F(y \mid x) &\equiv P[Y \leq y \mid X = x] = P[g(X,V) \leq y \mid X = x] \\ &= \int_0^1 \mathbf{1}\{g(x,v) \leq y\}\, dF(v \mid x) \\ &\equiv \mu[g_x^{-1}(-\infty, y] \mid x]. \end{aligned}$$

As V is not independent of X, there exists $\mathcal{X}^* \subset \mathcal{X}$, $P[X \in \mathcal{X}^*] > 0$, such that for all $x^* \in \mathcal{X}^*$, $V \mid X = x^*$ is not $\mathbb{U}[0,1]$. Let any $x^* \in \mathcal{X}^*$ be given. If $g(x^*,\cdot)$ is strictly increasing, then for all $y \in \mathcal{Y}$, $g_{x^*}^{-1}(-\infty, y] = (0, g^{-1}(x^*, y)]$. But $F(\cdot \mid x^*) = \mu[(0, g^{-1}(x^*,\cdot)] \mid x^*] \neq g^{-1}(x^*,\cdot)$, as $\mu[\cdot \mid x^*]$ is not Lebesgue measure on $\mathbb{I}$.

Next, suppose that $g(x^*,\cdot)$ is not strictly increasing. First, suppose that $g(x^*,\cdot)$ is invertible, and also suppose $F(y \mid x^*) = g^{-1}(x^*,\cdot)$ for all y. The monotonicity of $F(\cdot \mid x^*)$ and the invertibility of $g^{-1}(x^*,\cdot)$ imply that $g^{-1}(x^*,\cdot)$ is strictly increasing. But this implies that $g(x^*,\cdot)$ is strictly increasing, a contradiction, so $F(\cdot \mid x^*) \neq g^{-1}(x^*,\cdot)$. Finally, if $g(x^*,\cdot)$ is not invertible, then

[5] We thank Max Stinchcombe for pointing this out.

$g^{-1}(x^*, \cdot)$ is a correspondence, not a function. But $F(\cdot \mid x^*)$ is a function, so $F(y \mid x^*) = g^{-1}(x^*, y)$ cannot hold for all $y \in \mathcal{Y}$. ■

For any Borel set G of $\mathbb{R}^d$ we define $P_t[G] \equiv P[X_t \in G]$, $t = 1, ..., T$. For any Borel set H of $\times_{t=1}^{T} \mathbb{R}^d$, we define $P_{1,...,T}[H] \equiv P[(X_1, ..., X_T) \in H]$. The requirement imposed in (ii) below that the product measure $P_1 P_2 \cdots P_T$ is absolutely continuous ($\ll$) with respect to the joint measure $P_{1,...,T}$ ensures that sets with positive $P_1 P_2 \cdots P_T$ measure have positive $P_{1,...,T}$ measure. This rules out extreme forms of dependence (e.g., $X_1 = X_2$ $a.s.$). In (ii), we also require that $P[Y_t = h(B)] < 1$ for all measurable h, $t = 1, ..., T$, where B is a function only of A. This rules out the trivial case in which $Y_1 = \cdots = Y_T$ $a.s.$

Theorem 8.1 *Suppose Assumption B.0 holds. Let the X_t's have common minimal support X, and write $B \equiv \beta(A)$.*

(i) Suppose (a) $g(X_t, \cdot)$ is strictly monotone a.s., $t = 1, ..., T$; and (b) $X_t \perp B$, $t = 1, ..., T$. Then $B = F(Y_t \mid X_t)$ a.s., $t = 1, ..., T$.

(ii) Suppose that X contains at least two points, that $P_1 P_2 \cdots P_T \ll P_{1,...,T}$, and that $P[Y_t = h(B)] < 1$ for all measurable h, $t = 1, \cdots, T$. Suppose either $(i.a)$ or $(i.b)$ does not hold. Then $P[F_1(Y_1 \mid X_1) = \cdots = F_T(Y_T \mid X_T)] < 1$.

Proof of Theorem 8.1 We give the proof for $T = 2$. The proof for $T > 2$ is similar.

(i) Suppose (a) and (b) hold. Without loss of generality, we treat the case where $g(X_t, \cdot)$ is strictly increasing $a.s.$; the strictly decreasing case is handled by replacing $g(X_t, \cdot)$ with $-g(X_t, \cdot)$. Given (a) and (b), Proposition 2.3(i) applies, as the measurability of β and the nonatomicity of A ensure that $B \equiv \beta(A)$ is nonatomic, and the other conditions hold by assumption. This gives $g^{-1}(X_1, Y_1) = F_1(Y_1 \mid X_1)$ $a.s.$ and $g^{-1}(X_2, Y_2) = F_2(Y_2 \mid X_2)$ $a.s.$ From (a), we have $g^{-1}(X_1, Y_1) = B = g^{-1}(X_2, Y_2)$, so

$$F_1(Y_1 \mid X_1) = F_2(Y_2 \mid X_2) \quad a.s.$$

Further, $X_t \perp B$, $t = 1, 2$, ensures that for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$,

$$\begin{aligned} F_1(y \mid x) &= P[g(X_1, B) \le y \mid X_1 = x] = P[g(x, B) \le y] \\ &= P[g(X_2, B) \le y \mid X_2 = x] = F_2(y \mid x), \end{aligned}$$

so $F_1 = F_2 = F$, say, which implies

$$B = F(Y_1 \mid X_1) = F(Y_2 \mid X_2) \quad a.s.$$

$(ii.1)$ First suppose that strict monotonicity $a.s.$ (i.e., (a)) holds, but (b) fails, so that $(X_1, X_2) \not\perp B$. Again we explicitly treat only the strictly increasing case. Then $B = g^{-1}(X_1, Y_1) = g^{-1}(X_2, Y_2)$.

By the the proof of Proposition 2.3(ii), we also have

$$\begin{aligned} F_1(Y_1 \mid X_1) &= \mu_1[(0,B] \mid X_1] = \int_0^B dF_1(b \mid X_1) \\ F_2(Y_2 \mid X_2) &= \mu_2[(0,B] \mid X_2] = \int_0^B dF_2(b \mid X_2), \end{aligned}$$

where $F_t(b \mid x)$ defines the conditional CDF of B given $X_t = x$. Letting $F_{1,2}(b \mid x_1, x_2)$ define the conditional CDF of B given $X_1 = x_1, X_2 = x_2$, we have

$$\begin{aligned} &P[F_1(Y_1 \mid X_1) = F_2(Y_2 \mid X_2)] = P[\ \mu_1[(0,B] \mid X_1] = \mu_2[(0,B] \mid X_2]\ ] \\ &\quad = 1 - P[\ \mu_1[(0,B] \mid X_1] \neq \mu_2[(0,B] \mid X_2]\ ] \\ &\quad = 1 - \int_{\mathcal{X}\times\mathcal{X}} [\int_0^1 \mathbf{1}\{\mu_1[(0,b] \mid x_1] \neq \mu_2[(0,b] \mid x_2]\}\ dF_{1,2}(b \mid x_1, x_2)]\ dF(x_1, x_2). \end{aligned}$$

The desired result follows if the integral in the expression above is positive.

To simplify notation, write

$$\mu_B(x_1, x_2) \equiv \int_0^1 \mathbf{1}\{\mu_1[(0,b] \mid x_1] \neq \mu_2[(0,b] \mid x_2]\}\ dF_{1,2}(b \mid x_1, x_2),$$

Then

$$\begin{aligned} &\int_{\mathcal{X}\times\mathcal{X}} \int_0^1 \mathbf{1}\{\mu_1[(0,b] \mid x_1] \neq \mu_2[(0,b] \mid x_2]\}\ dF_{1,2}(b \mid x_1, x_2)\ dF(x_1, x_2) \\ &\qquad = \int \mu_B(x_1, x_2)\ dP_{1,2}(x_1, x_2). \end{aligned}$$

The desired result follows from corollary 4.10 of Bartle (1966) (i.e., for integrable $f \geq 0$, $\int f\, d\mu = 0$ iff $f = 0$ $\mu - a.e.$), provided $\mu_B(x_1, x_2)$ is positive on a set of positive $P_{1,2}-$measure.

To show this, let $\mathcal{X}_t \equiv \{x \in \mathcal{X} : \mu_t[\ \cdot \mid x] \neq \lambda(\cdot)\}$ and $\mathcal{X}_t^c \equiv \mathcal{X}\backslash\mathcal{X}_t$. By assumption, $P_1[\mathcal{X}_1] > 0$ or $P_2[\mathcal{X}_2] > 0$. Without loss of generality, take $P_2[\mathcal{X}_2] > 0$; then $0 \leq P_1[\mathcal{X}_1] \leq 1$. Two cases exhaust the possibilities: either $P_1[\mathcal{X}_1] = P_2[\mathcal{X}_2] = 1$ or not. First, suppose not; we take $P_1[\mathcal{X}_1^c] > 0$. This covers the cases $0 \leq P_1[\mathcal{X}_1] < 1$ and $0 < P_2[\mathcal{X}_2] \leq 1$. Then $\mu_B(x_1, x_2) > 0$ on $\mathcal{X}_1^c \times \mathcal{X}_2$. (If not, $x_2 \notin \mathcal{X}_2$.) Because $P_1P_2 \ll P_{1,2}$, $P_1P_2(\mathcal{X}_1^c \times \mathcal{X}_2) = P_1(\mathcal{X}_1^c)\ P_2(\mathcal{X}_2) > 0$ implies $P_{1,2}(\mathcal{X}_1^c \times \mathcal{X}_2) > 0$, as was to be shown.

The remaining case is $P_1[\mathcal{X}_1] = 1$ and $P_2[\mathcal{X}_2] = 1$, i.e. $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$. Suppose $\int \mu_B(x_1, x_2)\, dP_{1,2}(x_1, x_2) = 0$. Then by Bartle (1966, corollary 4.10), $\mu_B(x_1, x_2) = 0$ $P_{1,2} - a.e.$, which further implies $\mu_1[(0,b] \mid x_1] = \mu_2[(0,b] \mid x_2]$ for almost all b, x_1, and x_2. Since $\mathcal{X}$ contains at least two points, this can only hold if there exists μ_0, say, such that $\mu_1[(0,b] \mid x_1] = \mu_2[(0,b] \mid x_2] = \mu_0[(0,b]]$, for almost all b, x_1, and x_2. If $\mu_0 = \lambda$, this is a contradiction. If $\mu_0 \neq \lambda$, a further monotone transformation of B can be applied without loss of generality to ensure $\mu_0 = \lambda$. But this is again a contradiction. Thus, $\int \mu_B(x_1, x_2)\ dP_{1,2}(x_1, x_2) > 0$.

$(ii.2)$ Now suppose that (a) fails. Since

$$P[F_1(Y_1 \mid X_1) = F_2(Y_2 \mid X_2)] = 1 - P[F_1(Y_1 \mid X_1) \neq F_2(Y_2 \mid X_2)],$$

the desired result follows if $P[F_1(Y_1 \mid X_1) \neq F_2(Y_2 \mid X_2)] > 0$.

From the proof of Proposition 2.3(ii), we have

$$\begin{aligned} F_t(Y_t \mid X_t) &= \mu_t\{g_{X_t}^{-1}(-\infty, Y_t] \mid X_t\} = \mu_t\{g_{X_t}^{-1}(-\infty, g(X_t, B)] \mid X_t\} \\ &\equiv c_t(X_t, B) \equiv C_t. \end{aligned}$$

Since (a) fails, there exists a set $\mathcal{X}_0 \subset \mathcal{X}$ with $P_1[\mathcal{X}_0] > 0$ or $P_2[\mathcal{X}_0] > 0$ such that when $X_t \in \mathcal{X}_0$, $g(X_t, \cdot)$ is not strictly monotone. When $P_t[\mathcal{X}_0] > 0$, $P[C_t = B \mid X_t \in \mathcal{X}_0]$ is defined, and we have

$$0 \leq P[C_t = B \mid X_t \in \mathcal{X}_0] < 1.$$

When $P_t[\mathcal{X}_0^c] > 0$, $P[C_t = B \mid X_t \in \mathcal{X}_0^c]$ is defined, and we have

$$P[C_t = B \mid X_t \in \mathcal{X}_0^c] = 1.$$

Without loss of generality, take $P_2[\mathcal{X}_0] > 0$; then $0 \leq P_1[\mathcal{X}_0] \leq 1$. Two cases exhaust the possibilities: either $P_1[\mathcal{X}_0] = P_2[\mathcal{X}_0] = 1$ or not. First, suppose not; we take $P_1[\mathcal{X}_0^c] > 0$. This covers the cases $0 \leq P_1[\mathcal{X}_0] < 1$ and $0 < P_2[\mathcal{X}_0] \leq 1$. We have

$$\begin{aligned} &P[F(Y_1 \mid X_1) \neq F(Y_2 \mid X_2)] = P[C_1 \neq C_2] \\ &\qquad \geq P[(C_1 \neq C_2) \cap (X_1 \in \mathcal{X}_0^c) \cap (X_2 \in \mathcal{X}_0)] \\ &\qquad = P[(C_1 = B) \cap (B \neq C_2) \cap (X_1 \in \mathcal{X}_0^c) \cap (X_2 \in \mathcal{X}_0)]. \end{aligned}$$

Now

$$\begin{aligned} &P_1 P_2[(C_1 = B) \cap (B \neq C_2) \cap (X_1 \in \mathcal{X}_0^c) \cap (X_2 \in \mathcal{X}_0)] \\ &\quad = P_1[(C_1 = B) \cap (X_1 \in \mathcal{X}_0^c)]\ P_2[(B \neq C_2) \cap (X_2 \in \mathcal{X}_0)] \\ &\quad = P_1[\mathcal{X}_0^c]\ P_2[\mathcal{X}_0]\ (1 - P_2[B = C_2 \mid X_2 \in \mathcal{X}_0]) \\ &\quad > 0, \end{aligned}$$

as $P_1[\mathcal{X}_0^c] > 0$, $P_2[\mathcal{X}_0] > 0$, and $P[B = C_2 \mid X_2 \in \mathcal{X}_0] < 1$. Because $P_1 P_2 \ll P_{1,2}$, it follows that $P[(C_1 = B) \cap (X_1 \in \mathcal{X}_0^c) \cap (B \neq C_2) \cap (X_2 \in \mathcal{X}_0)] > 0$. Thus, $P[F(Y_1 \mid X_1) \neq F(Y_2 \mid X_2)] > 0$, as was to be shown.

The remaining case is $P_1[\mathcal{X}_0] = 1$ and $P_2[\mathcal{X}_0] = 1$, i.e., $\mathcal{X}_0 = \mathcal{X}$. Again, we must show $P[C_1 \neq C_2] > 0$. Suppose not. Then for almost all b, x_1, and x_2, we have $c_1(x_1, b) = c_2(x_2, b)$.

Because $\mathcal{X}_0 = \mathcal{X}$ contains at least two values, this can hold only if $c_1(x_1, b) = c_2(x_2, b) = c_0(b)$, say, for all $(x_1, x_2, b) \in \mathcal{X} \times \mathcal{X} \times \mathcal{B}$, $\mathcal{B} \equiv \text{supp}(B)$. This can hold only if: (i) $X_t \perp B$, $t = 1, 2$; and, because for each $x \in \mathcal{X}$, $g(x, \cdot)$ is not strictly monotone, (ii) $g(x, b) = g_0(b)$, say, for all $(x, b) \in \mathcal{X} \times \mathcal{B}$, i.e., $P[Y_t = g_0(B)] = 1$, $t = 1, 2$. But this contradicts our assumption that there is no such g_0. Thus, $P[C_1 \neq C_2] > 0$, as was to be shown. ■

Let $D^{\boldsymbol{r}} F_t(y|x) \equiv \partial^{|\boldsymbol{r}|} F_t(y \mid x) / \partial^{r_1} x_1 \cdots \partial^{r_d} x_d$ for $\boldsymbol{r} = (r_1, ..., r_d)$ with $|\boldsymbol{r}| = p + 1$. Let $\Delta_{jt}(y, x) \equiv F_t(y|X_{jt}) - \sum_{0 \leq |\boldsymbol{j}| \leq p} \frac{1}{\boldsymbol{j}!} D^{\boldsymbol{j}} F_t(y|x) (X_{jt} - x)^{\boldsymbol{j}} = \sum_{|\boldsymbol{j}| = p+1} \frac{1}{\boldsymbol{j}!} \int_0^1 D^{\boldsymbol{j}} F_t(y|x + v(X_{jt} - x)) (1 - v)^p dv (X_{jt} - x)^{\boldsymbol{j}}$, and $\Delta_{jt,-i} \equiv \Delta_{jt}(Y_{it}, X_{it})$. Recall $\mathbf{S}_{Nt}(X_{it}) \equiv \frac{1}{N-1} \sum_{j \neq i}^{N} K_{ji,t} \mathbf{X}_{jt,-i} \mathbf{X}'_{jt,-i}$. Define

$$\begin{aligned}
\mathbf{V}_{Nt}(Z_{it}) &\equiv \frac{1}{N-1} \sum_{j \neq i}^{N} K_{ji,t} \mathbf{X}_{jt,-i} \bar{\mathbf{1}}_{Y_{it}}(Z_{jt}) \text{ and} \\
\mathbf{B}_{Nt}(Z_{it}) &\equiv \frac{1}{N-1} \sum_{j \neq i}^{N} K_{ji,t} \mathbf{X}_{jt,-i} \Delta_{jt,-i}.
\end{aligned}$$

Similarly, let $\tilde{\mathbf{S}}_{Nt}(x) \equiv \frac{1}{N} \sum_{j=1}^{N} K_h(X_{jt} - x) \mathbf{X}_{jt}(x) \mathbf{X}_{jt}(x)'$, $\tilde{\mathbf{V}}_{Nt}(y, x) \equiv \frac{1}{N} \sum_{j=1}^{N} K_h(X_{jt} - x) \mathbf{X}_{jt}(x) \bar{\mathbf{1}}_y(Z_{jt})$, and $\tilde{\mathbf{B}}_{Nt}(y, x) \equiv \frac{1}{N} \sum_{j=1}^{N} K_h(X_{jt} - x) \mathbf{X}_{jt}(x) \Delta_{jt}(y, x)$. Note that $\bar{\mathbf{S}}_t(x) = E[\tilde{\mathbf{S}}_{Nt}(x)]$ and that $\bar{\mathbf{S}}_t(X_{it})$ gives the expectation of $\mathbf{S}_{Nt}(X_{it})$ with respect to all elements but X_{it}. Let $\bar{\mathbf{B}}_t(x) \equiv E[\tilde{\mathbf{B}}_{Nt}(x)]$. The following lemma establishes the consistency of $\hat{F}_{Nt}(Y_{it} \mid X_{it})$ uniformly in (i, t).

Lemma 8.2 *Suppose Assumptions B.1-B.5 hold. Then uniformly in (i, t) we have: (i)* $\hat{F}_{Nt}(Y_{it}|X_{it}) - F_t(Y_{it}|X_{it}) = e_1'[\bar{\mathbf{S}}_t(X_{it})]^{-1} \mathbf{V}_{Nt}(Z_{it}) + e_1'[\bar{\mathbf{S}}_t(X_{it})]^{-1} \bar{\mathbf{B}}_t(Z_{it}) + O_P(\nu_N^2 + \nu_N h^{p+1})$; *(ii)* $\hat{F}_{Nt}(Y_{it}|X_{it}) - F_t(Y_{it}|X_{it}) = O_P(\nu_N + h^{p+1})$.

Proof of Lemma 8.2 Let $\tilde{F}_{Nt}(y \mid x)$ denote the version of the pth order local polynomial estimator of $F_t(y \mid x)$ that uses all N observations $\{X_{it}, Y_{it}\}_{i=1}^{N}$, kernel K, and bandwidth h. Since $[\tilde{\mathbf{S}}_{Nt}(x)]^{-1} \tilde{\mathbf{S}}_{Nt}(x) = [\mathbf{S}_{Nt}(X_{it})]^{-1} \mathbf{S}_{Nt}(X_{it}) = I_L$, an $L \times L$ identity matrix, we obtain the following standard bias and variance decompositions:

$$\tilde{F}_{Nt}(y \mid x) - F_t(y \mid x) = e_1'[\tilde{\mathbf{S}}_{Nt}(x)]^{-1} \tilde{\mathbf{V}}_{Nt}(y, x) + e_1'[\tilde{\mathbf{S}}_{Nt}(x)]^{-1} \tilde{\mathbf{B}}_{Nt}(y, x), \tag{14}$$

and

$$\hat{F}_{Nt}(Y_{it} \mid X_{it}) - F_t(Y_{it} \mid X_{it}) = e_1'[\mathbf{S}_{Nt}(X_{it})]^{-1} \mathbf{V}_{Nt}(Z_{it}) + e_1'[\mathbf{S}_{Nt}(X_{it})]^{-1} \mathbf{B}_{Nt}(Z_{it}). \tag{15}$$

By Theorems 2 and 4 in Masry (1996),[6]

$$\tilde{\mathbf{S}}_{Nt}(x) = \bar{\mathbf{S}}_t(x) + O_P(\nu_N), \tilde{\mathbf{V}}_{Nt}(y, x) = O_P(\nu_N), \text{ and } \tilde{\mathbf{B}}_{Nt}(y, x) - \bar{\mathbf{B}}_t(y, x) = O_P(\nu_N h^{p+1}),$$

[6] The compact support of the kernel function K in Masry (1996) can be easily relaxed, following the line of proof in Hansen (2008, theorem 4).

where $\nu_N \equiv N^{-1/2}h^{-d/2}\sqrt{\log N}$ and the probability orders hold uniformly in $x \in \mathcal{X}_t$. With this and Assumptions B.4-B.5, it is easy to show that $\mathbf{S}_{Nt}(X_{it}) = \tilde{\mathbf{S}}_{Nt}(X_{it}) + O_P\left(N^{-1}h^{-d}\right) = \bar{\mathbf{S}}_t(X_{it}) + O_P(\nu_N)$ uniformly in (i,t), as $N^{-1}h^{-d} = o(\nu_N)$. This implies

$$[\mathbf{S}_{Nt}(X_{it})]^{-1} = \left\{\bar{\mathbf{S}}_t(X_{it}) + \left[\mathbf{S}_{Nt}(X_{it}) - \bar{\mathbf{S}}_t(X_{it})\right]\right\}^{-1} = [\bar{\mathbf{S}}_t(X_{it})]^{-1} + O_P(\nu_N). \quad (16)$$

Proof. By the same arguments as used in the proof of theorem 4.1 of Boente and Fraiman (1991), we can show that $\tilde{\mathbf{V}}_{Nt}(y,x) = O_P(\nu_N)$ uniformly in (y,x) under Assumption B.3. It follows that

$$\mathbf{V}_{Nt}(Y_{it}, X_{it}) = \tilde{\mathbf{V}}_{Nt}(Y_{it}, X_{it}) + O_P\left(N^{-1}h^{-d}\right) = O_P(\nu_N) \text{ uniformly in } (i,t). \quad (17)$$

Similarly,

$$\begin{aligned}\mathbf{B}_{Nt}(Y_{it}, X_{it}) - \bar{\mathbf{B}}_t(Y_{it}, X_{it}) &= \tilde{\mathbf{B}}_{Nt}(Y_{it}, X_{it}) - \bar{\mathbf{B}}_t(Y_{it}, X_{it}) + O_P\left(N^{-1}h^{-d}\right) \\ &= O_P\left(\nu_N h^{p+1} + N^{-1}h^{-d}\right). \quad (18)\end{aligned}$$

It follows that $\hat{F}_{Nt}(Y_{it} \mid X_{it}) - F_t(Y_{it} \mid X_{it}) = e_1'\left[[\bar{\mathbf{S}}_t(X_{it})]^{-1} + O_P(\nu_N)\right]\{\mathbf{V}_{Nt}(X_{it}) + [\bar{\mathbf{B}}_t(Y_{it}, X_{it}) + O_P\left(\nu_N h^{p+1} + N^{-1}h^{-d}\right)]\} = e_1'[\bar{\mathbf{S}}_t(X_{it})]^{-1}\mathbf{V}_{Nt}(Y_{it}, X_{it}) + e_1'[\bar{\mathbf{S}}_t(X_{it})]^{-1}\bar{\mathbf{B}}_t(Y_{it}, X_{it}) + O_P(\nu_N^2 + \nu_N h^{p+1})$. This proves (i). Noting that $\bar{\mathbf{S}}_t(X_{it})$ is *p.d. a.s.* and $\bar{\mathbf{B}}_t(y,x) = O\left(h^{p+1}\right)$ uniformly in (y,x), we obtain (ii), given (17). ■

Proof of Theorems 3.1 and 3.2

We only prove Theorem 3.2, as the proof of Theorem 3.1 is a special case. First, we decompose $h^{d/2}D_N$ as follows:

$$\begin{aligned}h^{d/2}D_N &= h^{d/2}\sum_{i=1}^N \left[\hat{F}_{N1}(Y_{i1} \mid X_{i1}) - \hat{F}_{N2}(Y_{i2} \mid X_{i2})\right]^2 \\ &= h^{d/2}\sum_{i=1}^N [F_1(Y_{i1} \mid X_{i1}) - F_2(Y_{i2} \mid X_{i2})]^2 \\ &\quad + h^{d/2}\sum_{i=1}^N \left[\hat{F}_{N1}(Y_{i1} \mid X_{i1}) - F_1(Y_{i1} \mid X_{i1}) - \hat{F}_{N2}(Y_{i2} \mid X_{i2}) + F_2(Y_{i2} \mid X_{i2})\right]^2 \\ &\quad + 2h^{d/2}\sum_{i=1}^N \left[\hat{F}_{N1}(Y_{i1} \mid X_{i1}) - F_1(Y_{i1} \mid X_{i1}) - \hat{F}_{N2}(Y_{i2} \mid X_{i2}) + F_2(Y_{i2} \mid X_{i2})\right] \\ &\qquad \times [F_1(Y_{i1} \mid X_{i1}) - F_2(Y_{i2} \mid X_{i2})] \\ &\equiv D_{N1} + D_{N2} + 2D_{N3}.\end{aligned}$$

Under $H_1\left(N^{-1/2}h^{-d/4}\right)$, we prove the theorem by showing that (i) $D_{N1} \xrightarrow{P} \mu_0$, (ii) $D_{N2} - B_N \xrightarrow{d} N\left(0, \sigma_0^2\right)$, (iii) $D_{N3} = o_P(1)$, (iv) $\hat{B}_N = B_N + o_P(1)$, and (v) $\hat{\sigma}_N^2 = \sigma_0^2 + o_P(1)$. For (i), $D_{N1} = N^{-1}\sum_{i=1}^N \delta_N(Z_i)^2 = \mu_0 + o_P(1)$ under $H_1\left(N^{-1/2}h^{-d/4}\right)$. It remains to show (ii)-(iv).

To show (ii), we first apply Lemma 8.2 to obtain

$$\begin{aligned}
D_{N2} &= h^{d/2}\sum_{i=1}^{N}\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\mathbf{V}_{N1}(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\mathbf{V}_{N2}(Z_{i2})\right.\\
&\quad \left.+\, e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\mathbf{B}}_1(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\bar{\mathbf{B}}_2(Z_{i2}) + O_P\left(\nu_N^2+\nu_N h^{p+1}\right)\right\}^2\\
&= h^{d/2}\sum_{i=1}^{N}\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\mathbf{V}_{N1}(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\mathbf{V}_{N2}(Z_{i2})\right\}^2\\
&\quad + 2h^{d/2}\sum_{i=1}^{N}\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\mathbf{V}_{N1}(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\mathbf{V}_{N2}(Z_{i2})\right\}\\
&\qquad\qquad \times\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\mathbf{B}}_1(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\bar{\mathbf{B}}_2(Z_{i2})\right\}\\
&\quad + h^{d/2}\sum_{i=1}^{N}\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\mathbf{B}}_1(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\bar{\mathbf{B}}_2(Z_{i2})\right\}^2\\
&\quad + Nh^{d/2}O_P\left(\nu_N^2+\nu_N h^{p+1}\right)O_P\left(\nu_N+h^{p+1}\right)\\
&\equiv D_{N21}+2D_{N22}+D_{N23}+o_P(1) \qquad (19)
\end{aligned}$$

where the definitions of D_{N21}, D_{N22}, and D_{N23} are self-evident. Using the notation defined above eq. (6), we have $D_{N21} = \frac{h^{d/2}}{(N-1)^2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}\varsigma(Z_i,Z_j)\right]^2$, where $\varsigma(Z_i,Z_j) \equiv \mathbf{K}_{1,X_{i1}}(X_{j1}-X_{i1})\,\bar{\mathbf{1}}_{Y_{i1}}(Z_{j1}) - \mathbf{K}_{2,X_{i2}}(X_{j2}-X_{i2})\,\bar{\mathbf{1}}_{Y_{i2}}(Z_{j2})$. Decompose

$$\begin{aligned}
D_{N21} &= \frac{h^{d/2}}{(N-1)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq i,j}^{N}\varsigma(Z_i,Z_j)\varsigma(Z_i,Z_k) + \frac{h^{d/2}}{(N-1)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\varsigma(Z_i,Z_j)^2\\
&\equiv V_N + B_N,\ \text{say}.
\end{aligned}$$

Let $\bar{\varsigma}(Z_i,Z_j,Z_k) \equiv [\varsigma(Z_i,Z_j)\varsigma(Z_i,Z_k)+\varsigma(Z_j,Z_i)\varsigma(Z_j,Z_k)+\varsigma(Z_k,Z_i)\varsigma(Z_k,Z_j)]/3$. Then

$$\begin{aligned}
V_N &= \frac{h^{d/2}}{(N-1)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq i,j}^{N}\bar{\varsigma}(Z_i,Z_j,Z_k)\\
&= \frac{6h^{d/2}}{(N-1)^2}\sum_{1\le i<j<k\le N}\bar{\varsigma}(Z_i,Z_j,Z_k) = \frac{N(N-2)}{N-1}\bar{V}_N,
\end{aligned}$$

where $\bar{V}_N \equiv \frac{6h^{d/2}}{N(N-1)(N-2)}\sum_{1\le i<j<k\le N}\bar{\varsigma}(Z_i,Z_j,Z_k)$. Note that for all $i\neq j\neq k$, $\theta \equiv E[\bar{\varsigma}(Z_i,Z_j,Z_k)] = 0$, $\bar{\varsigma}_1(z) \equiv E[\bar{\varsigma}(z,Z_j,Z_k)] = 0$, and $\bar{\varsigma}_2(z,\tilde{z}) \equiv E[\bar{\varsigma}(z,\tilde{z},Z_k)] = \frac{1}{3}E[\varsigma(Z_k,z)\ \varsigma(Z_k,\tilde{z})]$. Let $\bar{\varsigma}_3(z,\tilde{z},\bar{z}) \equiv \bar{\varsigma}(z,\tilde{z},\bar{z}) - \bar{\varsigma}_2(z,\tilde{z}) - \bar{\varsigma}_2(z,\bar{z}) - \bar{\varsigma}_2(\tilde{z},\bar{z})$. By the Hoeffding decomposition,

$$\bar{V}_N = 3H_N^{(2)} + H_N^{(3)},$$

where $H_N^{(2)} \equiv \frac{2h^{d/2}}{N(N-1)}\sum_{1\le i<j\le N}\bar{\varsigma}_2(Z_i,Z_j)$ and $H_N^{(3)} \equiv \frac{6h^{d/2}}{N(N-1)(N-2)}\sum_{1\le i<j<k\le N}\bar{\varsigma}_3(Z_i,Z_j,\ Z_k)$. Noting that $E[\bar{\varsigma}_3(z,\tilde{z},Z_i)] = 0$ and that $\bar{\varsigma}_3$ is symmetric in its arguments by construction,

it is straightforward to show that $E[H_N^{(3)}] = 0$ and $E[H_N^{(3)}]^2 = O\left(N^{-3}h^{-d}\right)$. Hence, $H_N^{(3)} = O_P\left(N^{-3/2}h^{-d/2}\right) = o_P\left(N^{-1}\right)$ by the Chebyshev inequality. It follows that $V_N = \frac{N(N-2)}{N-1}\bar{V}_N = \{1+o(1)\}\mathcal{H}_N + o_P(1)$, where

$$\mathcal{H}_N \equiv \frac{2h^{d/2}}{N}\sum_{1\le i\le j\le N} 3\,\bar{\varsigma}_2(Z_i,Z_j) = \frac{2h^{d/2}}{N}\sum_{1\le i<j\le N}\int \varsigma(z,Z_i)\,\varsigma(z,Z_j)\,dF(z).$$

Noting that $\mathcal{H}_N$ is a second order degenerate U-statistic, we can easily verify that all the conditions of theorem 1 of Hall (1984) are satisfied and a central limit theorem applies to $\mathcal{H}_N$:

$$\mathcal{H}_N \xrightarrow{d} N\left(0,\sigma_0^2\right),$$

where the asymptotic variance of $\mathcal{H}_N$ is given by $\sigma_0^2 \equiv \lim_{N\to\infty}\sigma_N^2$ and

$$\begin{aligned}
\sigma_N^2 &\equiv 2h^d E_iE_j\left[\int \varsigma(Z_i,z)\,\varsigma(Z_j,z)\,dF_Z(z)\right]^2 \\
&= 2h^d E_iE_j\Bigg[\int \{\mathbf{K}_{1,X_{i1}}(x_1-X_{i1})\,\bar{\mathbf{1}}_{y_1}(Z_{i1}) - \mathbf{K}_{2,X_{i2}}(x_2-X_{i2})\,\bar{\mathbf{1}}_{y_2}(Z_{i2})\} \\
&\quad\times\left\{\mathbf{K}_{1,X_{j1}}(x_1-X_{j1})\,\bar{\mathbf{1}}_{y_1}(Z_{j1}) - \mathbf{K}_{2,X_{j2}}(x_2-X_{j2})\,\bar{\mathbf{1}}_{y_2}(Z_{j2})\right\} f_Z(z)\,dz\Bigg]^2,
\end{aligned}$$

where $z = (y_1, x_1', y_2, x_2')'$ and E_i denotes expectation with respect to Z_i. By straightforward calculations, we can show that $\sigma_N^2 = \sum_{t=1}^2 \sigma_{Nt}^2 + O\left(h^d\right)$ where for $t = 1, 2$,

$$\sigma_{Nt}^2 \equiv 2h^d E_iE_j[\int \mathbf{K}_{t,X_{it}}(x_t-X_{it})\,\bar{\mathbf{1}}_{y_t}(Z_{it})\,\mathbf{K}_{t,X_{jt}}(x_t-X_{jt})\,\bar{\mathbf{1}}_{y_t}(Z_{jt})\,dF_{Z_t}(z_t)]^2.$$

Using the notation above eq.(6), we have

$$\begin{aligned}
\sigma_{N1}^2 &= 2h^{-3d}E_iE_j\int\Bigg[e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\boldsymbol{\mu}}^*\left(\frac{x_1-X_{i1}}{h}\right)\bar{\mathbf{1}}_{y_1}(Z_{i1}) \\
&\quad\times e_1'[\bar{\mathbf{S}}_1(X_{j1})]^{-1}\boldsymbol{\mu}^*\left(\frac{x_1-X_{j1}}{h}\right)\bar{\mathbf{1}}_{y_1}(Z_{j1})\,f_{Z_1}(y_1,x_1)\,dy_1dx_1\Bigg]^2 \\
&\simeq 2h^{-3d}E_iE_j\int\Bigg[e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\boldsymbol{\mu}^*(\tilde{x}_1)\,\boldsymbol{\mu}^*\left(\tilde{x}_1+\frac{X_{i1}-X_{j1}}{h}\right)'[\bar{\mathbf{S}}_1(X_{j1})]^{-1}e_1 \\
&\quad\times\bar{\mathbf{1}}_{y_1}(Z_{i1})\,\bar{\mathbf{1}}_{y_1}(Z_{j1})\,f_{Z_1}(y_1,X_{i1})\,dy_1d\tilde{x}_1\Bigg]^2 \\
&= 2h^{-d}E_iE_j[\int\int\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\boldsymbol{\mu}}^*\left(\frac{X_{j1}-X_{i1}}{h}\right)[\bar{\mathbf{S}}_1(X_{j1})]^{-1}e_1\right\}^2\bar{\mathbf{1}}_{y_1}(Z_{i1})\,\bar{\mathbf{1}}_{\tilde{y}_1}(Z_{i1}) \\
&\quad\times\bar{\mathbf{1}}_{y_1}(Z_{j1})\,\bar{\mathbf{1}}_{\tilde{y}_1}(Z_{j1})\,f_{Z_1}(y_1,X_{i1})\,f_{Z_1}(\tilde{y}_1,X_{i1})\,dy_1d\tilde{y}_1] \\
&= 2h^{-d}E_iE_j[\int\int\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\boldsymbol{\mu}}^*\left(\frac{X_{j1}-X_{i1}}{h}\right)[\bar{\mathbf{S}}_1(X_{j1})]^{-1}e_1\right\}^2\sigma_1^2(y_1,\tilde{y}_1;X_{i1}) \\
&\quad\times\sigma_1^2(y_1,\tilde{y}_1;X_{j1})\,f_{Z_1}(y_1,X_{i1})\,f_{Z_1}(\tilde{y}_1,X_{i1})\,dy_1d\tilde{y}_1] \\
&\simeq 2E\left[\int\int\int \eta_{i1}(x)^2\,\sigma_1^4(y_1,\tilde{y}_1;X_{i1})\,f_{Z_1}(y_1,X_{i1})\,f_{Z_1}(\tilde{y}_1,X_{i1})\,f_{X_1}(X_{i1})\,dy_1\,d\tilde{y}_1\,dx\right].
\end{aligned}$$

Similarly, $\sigma_{N2}^2 = 2E\left[\int\int\int \eta_{i2}(x)^2 \sigma_2^4(y,\tilde{y};X_{i2}) f_{Z_1}(y,X_{i2}) f_{Z_1}(\tilde{y},X_{i2}) f_{X_2}(X_{i2})\, dy\ d\tilde{y}\ dx\right] + o(1)$. It follows that $V_N \xrightarrow{d} N\left(0,\sigma_0^2\right)$ and

$$D_{N21} - B_N \xrightarrow{d} N\left(0,\sigma_0^2\right). \tag{20}$$

Let $b(Z_i) \equiv e_1'\left\{[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\mathbf{B}}_1(Z_{i1}) - [\bar{\mathbf{S}}_2(X_{i2})]^{-1}\bar{\mathbf{B}}_2(Z_{i2})\right\}$. Then $D_{N22} = D_{N22,1} - D_{N22,2}$, where $D_{N22,1} \equiv h^{d/2}\sum_{i=1}^N e_1'[\bar{\mathbf{S}}_t(X_{it})]^{-1}\mathbf{V}_{Nt}(Z_{it})\, b(Z_i)$ for $t = 1,2$. Write

$$\begin{aligned} D_{n22,1} &= N^{-1}h^{d/2}\sum_{i=1}^N\sum_{j\neq i}^N e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}K_{ji,1}\mathbf{X}_{j1,-i}\bar{\mathbf{1}}_{Y_{i1}}(Z_{j1})\, b(Z_i) \\ &\quad + N^{-1}h^{d/2}\sum_{i=1}^N e_1'[\bar{\mathbf{S}}_r(X_i)]^{-1}K_{ji,1}\mathbf{X}_{j1,-i}\bar{\mathbf{1}}_{Y_{i1}}(Z_{j1})\, b(Z_i) \\ &\equiv D_{n22,1a} + D_{n22,1b}, \text{ say.} \end{aligned}$$

Noting that $b(Z_i) = O_P(h^{p+1})$, it is straightforward to show that $D_{n22,1b} = O_P(h^{p+1-d/2}) = o_P(1)$. Noting that $E(D_{N22,1a}) = 0$ and $E(D_{N22,1a}^2) = O(Nh^{d+2(p+1)}) = o(1)$, we have $D_{N22,1a} = o_P(1)$ by the Chebyshev inequality. Similarly, we can show that $D_{N22,1b} = o_P(1)$ and thus $D_{N22,1} = o_P(1)$. By the same token, $D_{N22,2} = o_P(1)$. It follows that

$$D_{N22} = o_P(1). \tag{21}$$

By Lemma 8.2 and Assumption B.5, we have $D_{N23} = Nh^{d/2}O_P(h^{2(p+1)}) = O_P(Nh^{2(p+1)+d/2}) = o_P(1)$. This, in conjunction with (19), (20) and (21), implies that $D_{N2} - B_N \xrightarrow{d} N\left(0,\sigma_0^2\right)$.

Next, we show (*iii*). By Lemma 8.2, under $H_1\left(N^{-1/2}h^{-d/4}\right)$ we have

$$\begin{aligned} D_{N3} &= N^{-1/2}h^{d/4}\sum_{i=1}^N\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\mathbf{V}_{N1}(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\mathbf{V}_{N2}(Z_{i2})\right\}\delta_N(Z_i) \\ &\quad + N^{-1/2}h^{d/4}\sum_{i=1}^N\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\bar{\mathbf{B}}_1(Z_{i1}) - e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\bar{\mathbf{B}}_2(Z_{i2})\right\}\delta_N(Z_i) \\ &\quad + N^{1/2}h^{d/4}O_P\left(\nu_N^2 + \nu_N h^{p+1}\right) \\ &\equiv D_{N31} + D_{N32} + o_P(1), \text{ say.} \end{aligned}$$

For the first term, let $D_{N31} = D_{N31a} + D_{N32b}$, where $D_{N31a} \equiv N^{-1/2}h^{d/4}\sum_{i=1}^N e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}$ $\mathbf{V}_{N1}(Z_{i1})\,\delta_N(Z_i)$ and $D_{N32b} \equiv N^{-1/2}h^{d/4}\sum_{i=1}^N e_1'[\bar{\mathbf{S}}_2(X_{i2})]^{-1}\mathbf{V}_{N2}(Z_{i2})\,\delta_N(Z_i)$. By the leave-one-out property of our local polynomial estimator and since Z_i is IID, it is easy to show that $E[D_{N31a}] = 0$. Now, write $E[D_{N31a}]^2 = d_{N1} + d_{N2}$, where

$$\begin{aligned} d_{N1} &\equiv N^{-1}h^{d/2}\sum_{i=1}^N E\left[\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\mathbf{V}_{N1}(Z_{i1})\}^2\delta_N^2(Z_i)\right], \text{ and} \\ d_{N2} &\equiv N^{-1}h^{d/2}\sum_{i=1}^N\sum_{j\neq i}^N E\left\{e_1'[\bar{\mathbf{S}}_1(X_{i1})]^{-1}\mathbf{V}_{N1}(Z_{i1})\,\mathbf{V}_{N1}(Z_{j1})'\,[\bar{\mathbf{S}}_1(X_{j1})]^{-1}e_1\delta_N(Z_i)\delta_N(Z_j)\right\}. \end{aligned}$$

For d_{N1}, we have

$$d_{N1} = N^{-3}h^{d/2}\sum_{i=1}^{N}E\left[\left\{\sum_{k\neq i}^{N}\mathbf{K}_{1,X_{i1}}\left(X_{k1}-X_{i1}\right)\bar{\mathbf{1}}_{Y_{i1}}\left(Z_{k1}\right)\right\}^{2}\delta_N^2(Z_i)\right]$$
$$= N^{-3}h^{d/2}\sum_{i=1}^{N}\sum_{k\neq i}^{N}E\left[\left\{\mathbf{K}_{1,X_{i1}}\left(X_{k1}-X_{i1}\right)\bar{\mathbf{1}}_{Y_{i1}}\left(Z_{k1}\right)\right\}^{2}\delta_N^2(Z_i)\right] = O\left(N^{-1}h^{-d/2}\right).$$

Now $d_{N2} = N^{-3}h^{d/2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq i}^{N}\sum_{l\neq j}^{N}E[\mathbf{K}_{1,X_{i1}}\left(X_{k1}-X_{i1}\right)\bar{\mathbf{1}}_{Y_{i1}}\left(Z_{k1}\right)\mathbf{K}_{1,X_{i1}}(X_{l1}-X_{j1})\bar{\mathbf{1}}_{Y_{j1}}$ $(Z_{l1})\,\delta_N(Z_i)\delta_N(Z_j)]$. Noting that the term with all four indices (i,j,k,l) distinct vanishes in the last expression, it is is straightforward to show that $d_{N2} = O\left(h^{d/2}\right)$. Hence, $E\left[D_{N31a}\right]^2 = O\left(N^{-1}h^{-d/2}+h^{d/2}\right) = o\left(1\right)$ and $D_{N31a} = o_P\left(1\right)$ by the Chebyshev inequality. Similarly $D_{N31b} = o_P\left(1\right)$. It follows that $D_{N31} = o_P\left(1\right)$. Noting that $\bar{\mathbf{B}}_t\left(y,x\right) = O\left(h^{p+1}\right)$ uniformly in (y,x), we have

$$D_{N32} = N^{-1/2}h^{d/4}\sum_{i=1}^{N}\left\{e_1'[\bar{\mathbf{S}}_1\left(X_{i1}\right)]^{-1}\bar{\mathbf{B}}_1\left(Z_{i1}\right) - e_1'[\bar{\mathbf{S}}_2\left(X_{i2}\right)]^{-1}\bar{\mathbf{B}}_2\left(Z_{i2}\right)\right\}\delta_N(Z_i)$$
$$\leq N^{1/2}h^{d/4}\max_{t\in\{1,2\}}\sup_x\left|e_1'[\bar{\mathbf{S}}_t\left(x\right)]^{-1}\bar{\mathbf{B}}_t\left(y,x\right)\right|N^{-1}\sum_{i=1}^{N}\left|\delta_N(Z_i)\right|$$
$$= N^{1/2}h^{d/4}O\left(h^{p+1}\right)O_P\left(1\right) = O_P\left(N^{1/2}h^{p+1+d/4}\right) = o_P\left(1\right).$$

Consequently, $D_{N3} = o_P\left(1\right)$.

We now show (iv). Noting that $a^2-b^2 = (a-b)^2 + 2\left(a-b\right)b$, we have $\hat{B}_N - B_N = d_{N3}+2d_{N4}$, where $d_{N3} \equiv \frac{h^{d/2}}{(N-1)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left\{\hat{\alpha}_{ij,1}-\hat{\alpha}_{ij,2}\right\}^2$, $d_{N4} \equiv \frac{h^{d/2}}{(N-1)^2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}[\hat{\alpha}_{ij,1}-\hat{\alpha}_{ij,2}][\mathbf{K}_{1,X_{i1}}\left(X_{j1}-X_{i1}\right)\bar{\mathbf{1}}_{Y_{i1}}\left(Z_{j1}\right) - \mathbf{K}_{2,X_{i2}}\left(X_{j2}-X_{i2}\right)\bar{\mathbf{1}}_{Y_{i2}}\left(Z_{j2}\right)]$, and $\hat{\alpha}_{ij,t} = e_1'[\mathbf{S}_{Nt}\left(X_{it}\right)]^{-1}$ $K_{ji,t}\mathbf{X}_{jt,-i}\hat{\mathbf{1}}_{Y_{it}}\left(Z_{jt}\right) - \mathbf{K}_{t,X_{it}}\left(X_{jt}-X_{it}\right)\bar{\mathbf{1}}_{Y_{it}}\left(Z_{jt}\right)$. Noting that $[\mathbf{S}_{Nt}\left(X_{it}\right)]^{-1} = [\bar{\mathbf{S}}_t\left(X_{it}\right)]^{-1} + O_P\left(\nu_N\right)$ and $\hat{\mathbf{1}}_y\left(Z_{it}\right) - \bar{\mathbf{1}}_y\left(Z_{it}\right) = F_t\left(y \mid X_{it}\right) - \hat{F}_{Nt}\left(y \mid X_{it}\right) = O_P\left(\nu_N + h^{p+1}\right)$ uniformly in (i,t) and y, we have $\hat{\alpha}_{ij,t} = e_1'[\bar{\mathbf{S}}_t\left(X_{it}\right)]^{-1}K_{ji,t}\mathbf{X}_{jt,-i}\left\{\hat{\mathbf{1}}_{Y_{it}}\left(Z_{jt}\right) - \bar{\mathbf{1}}_{Y_{it}}\left(Z_{jt}\right)\right\} + O_P\left(\nu_N\right)$. It follows that

$$\left|d_{N3}\right| \leq O_P\left(\nu_N^2 + h^{2(p+1)}\right)\max_{t\in\{1,2\}}\sup_x\left\|[\bar{\mathbf{S}}_t\left(x\right)]^{-1}\right\|^2\frac{h^{d/2}}{\left(N-1\right)^2}\sum_{t=1}^{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left\{\left\|K_{ji,t}\mathbf{X}_{jt,-i}\right\|\right\}^2$$
$$= O_P\left(h^{-d/2}\left(\nu_N^2 + h^{2(p+1)}\right)\right) = o_P\left(1\right),$$

and

$$\left|d_{N4}\right| \leq O_P\left(\nu_N + h^{p+1}\right)\max_{t\in\{1,2\}}\sup_x\left\|[\bar{\mathbf{S}}_t\left(x\right)]^{-1}\right\|\frac{h^{d/2}}{\left(N-1\right)^2}\sum_{t=1}^{2}\sum_{s=1}^{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\left\|K_{ji,t}\mathbf{X}_{jt,-i}\right\|$$
$$\times\left|\mathbf{K}_{s,X_{is}}\left(X_{js}-X_{is}\right)\right|$$
$$= O_P\left(h^{-d/2}\left(\nu_N + h^{p+1}\right)\right) = o_P\left(1\right).$$

Consequently, $\hat{B}_N - B_N = o_P(1)$.

Lastly, to show (v) we define

$$\bar{\sigma}_{Nt}^2 \equiv \frac{2h^d}{N_2} \sum_{i \neq j} \left[\int \mathbf{K}_{t,X_{it}}(x_t - X_{it}) \bar{\mathbf{1}}_{y_t}(Z_{it}) \mathbf{K}_{t,X_{jt}}(x_t - X_{jt}) \bar{\mathbf{1}}_{y_t}(Z_{jt}) f_{Z_t}(z_t) dz_t \right]^2,$$

$$\tilde{\sigma}_{Nt}^2 \equiv \frac{2h^d}{N_2} \sum_{i \neq j} \left[\frac{1}{N} \sum_{l=1}^{N} e_1' [\mathbf{S}_{Nt}(X_{lt})]^{-1} K_{il,t} \mathbf{X}_{it,-l} \bar{\mathbf{1}}_{Y_{lt}}(Z_{it}) e_1' [\mathbf{S}_{Nt}(X_{lt})]^{-1} K_{jl,t} \mathbf{X}_{jt,-l} \bar{\mathbf{1}}_{Y_{lt}}(Z_{jt}) \right]^2,$$

$$\hat{\sigma}_{Nt}^2 \equiv \frac{2h^d}{N_2} \sum_{i \neq j} \left[\frac{1}{N} \sum_{l=1}^{N} e_1' [\mathbf{S}_{Nt}(X_{lt})]^{-1} K_{il,t} \mathbf{X}_{it,-l} \hat{\mathbf{1}}_{Y_{lt}}(Z_{it}) e_1' [\mathbf{S}_{Nt}(X_{lt})]^{-1} K_{jl,t} \mathbf{X}_{jt,-l} \hat{\mathbf{1}}_{Y_{lt}}(Z_{jt}) \right]^2,$$

where $z_t = (y_t, x_t')$, $N_2 \equiv N(N-1)$, and $\sum_{i \neq j} \equiv \sum_{i=1}^N \sum_{j \neq i}^N$. By the uniform consistency of $\hat{F}_{Nt}$, we have $\hat{\sigma}_{Nt}^2 = \tilde{\sigma}_{Nt}^2 + o_P(1)$. By (16) and the law of large numbers for U-statistics, we can show that $\tilde{\sigma}_{Nt}^2 = \overline{\sigma}_{Nt}^2 + o_P(1) = \sigma_{Nt}^2 + o_P(1)$. The result then follows by noticing that $\hat{\sigma}_N^2 = \sum_{t=1}^T \hat{\sigma}_{Nt}^2$ and $\sum_{t=1}^T \sigma_{Nt}^2 = \sigma_0^2 + o(1)$. ■

Proof of Theorem 3.3

Using the notation defined in the proof of Theorem 3.2, we again write $N^{-1}D_N = N^{-1}h^{-d/2}(D_{N1} + D_{N2} + 2D_{N3})$. Under the alternative, it is easy to show that $N^{-1}h^{-d/2}D_{N1} = E[F_1(Y_{i1}|X_{i1}) - F_2(Y_{i2}|\ X_{i2})]^2 + o_P(1)$, $N^{-1}h^{-d/2}D_{N2} = O_P\left(\nu_N^2 + h^{2(p+1)}\right) = o_P(1)$, and $N^{-1}h^{-d/2}D_{N3} = O_P\left(\nu_N + h^{p+1}\right) = o_P(1)$. On the other hand, $N^{-1}h^{d/2}\hat{B}_N = O_P\left(N^{-1}\right) = o_P(1)$ and $\hat{\sigma}_N^2 = \sigma_0^2 + o_P(1)$. It follows that $N^{-1}h^{d/2}J_N = (N^{-1}h^{d/2}D_N - N^{-1}h^{d/2} \times \hat{B}_N)/\sqrt{\hat{\sigma}_N^2} \xrightarrow{P} E[F_1(Y_{i1}|X_{i1}) - F_2(Y_{i2}|X_{i2})]^2/\sigma_0$, and the conclusion follows. ■

For the next result, let F_X denote the CDF of the random variable X, and let $\mathbb{R}^+ \equiv [0, \infty)$. Part (i) shows that strict monotonicity of $g(x, \cdot)$ is preserved by weighted averaging over x. Part (ii) shows that strict monotonicity of the weighted average can also occur when departures from strict monotonicity of $g(x, \cdot)$ are sufficiently mild. Together, results $(ii.1)$ and $(ii.2)$ show that when one weighting function places zero weight on the region where strict monotonicity of $g(x, \cdot)$ fails, there is another weighting function that can detect sufficient departures from strict monotonicity.

Lemma 8.3 *Let $g : \mathbb{R}^d \times \mathbb{I} \to \mathbb{R}$ be measurable, let X be a random element of $\mathbb{R}^d$, and suppose that $E[g(X, b)] < \infty$ for all $b \in \mathbb{I}$. Let $w : \mathcal{X} \to \mathbb{R}^+$ be a bounded measurable function with $\int w(x)$ $dF_X(x) = 1$.*

(i) *If $g(X, \cdot)$ is strictly increasing a.s., then $\bar{g}_w(\cdot)$ is strictly increasing, where $\bar{g}_w(\cdot) \equiv \int g(x, \cdot)$ $w(x)\ dF_X(x)$.*

(ii) *If $g(X, \cdot)$ is not strictly increasing a.s., there exists a set $\mathcal{X}^*$, $P[X \in \mathcal{X}^*] > 0$, such that for each $x \in \mathcal{X}^*$, $g(x, \cdot)$ is not strictly increasing. Let $\mathcal{X}_w^* \equiv \mathcal{X}^* \cap \mathcal{X}_w$, where $\mathcal{X}_w \equiv \{x \in \mathcal{X} : w(x) > 0\}$.*

(1) *Suppose* $P[X \in \mathcal{X}_w^*] > 0$. *Then* $\bar{g}_w(\cdot)$ *is not strictly increasing if and only if there exist* $0 \leq b_1^* < b_2^* \leq 1$ *such that*

$$\int [g(x, b_2^*) - g(x, b_1^*)] \ w(x) \ \mathbf{1}\{x \in \mathcal{X}_w^*\} \ dF_X(x)$$
$$\leq - \int [g(x, b_2^*) - g(x, b_1^*)] \ w(x) \ \mathbf{1}\{x \notin \mathcal{X}_w^*\} \ dF_X(x).$$

(2) *Suppose* $P[X \in \mathcal{X}_w^*] = 0$. *Then* $\bar{g}_w(\cdot)$ *is strictly increasing. Further,* $P[X \in \mathcal{X}_w] < 1$ *so* $P[X \notin \mathcal{X}_w^*] > 0$, *and, with* $\tilde{\mathcal{X}}_w \equiv \mathcal{X} \backslash \mathcal{X}_w$ *and* $\tilde{\mathcal{X}}_w^* \equiv \mathcal{X}^* \cap \tilde{\mathcal{X}}_w$, *we have* $P[X \in \tilde{\mathcal{X}}_w^*] > 0$. *Then there exists a bounded measurable function* $\tilde{w} : \mathcal{X} \to \mathbb{R}^+$*with* $\int \tilde{w}(x) \ dF_X(x) = 1$ *and* $\tilde{\mathcal{X}}_w = \mathcal{X}_{\tilde{w}} \equiv \{x \in \mathcal{X} : \tilde{w}(x) > 0\}$. *Let*

$$\bar{g}_{\tilde{w}}(\cdot) \equiv \int g(x, \cdot) \ \tilde{w}(x) \ dF_X(x).$$

Then $\bar{g}_{\tilde{w}}(\cdot)$ *is not strictly increasing if and only if there exist* $0 \leq b_1^* < b_2^* \leq 1$ *such that*

$$\int [g(x, b_2^*) - g(x, b_1^*)] \ \tilde{w}(x) \ \mathbf{1}\{x \in \tilde{\mathcal{X}}_w^*\} \ dF_X(x)$$
$$\leq - \int [g(x, b_2^*) - g(x, b_1^*)] \ \tilde{w}(x) \ \mathbf{1}\{x \notin \tilde{\mathcal{X}}_w^*\} \ dF_X(x).$$

Proof of Lemma 8.3 (i) Under the conditions given, $\int g(x, b) \ w(x) \ dF_X(x) < \infty$ for all $b \in \mathbb{I}$. If $g(X, \cdot)$ is strictly increasing *a.s.*, then for all $0 \leq b_1 < b_2 \leq 1$,

$$\begin{aligned} \bar{g}_w(b_2) - \bar{g}_w(b_1) &= \int g(x, b_2) \ w(x) \ dF_X(x) - \int g(x, b_1) \ w(x) \ dF_X(x) \\ &= \int [g(x, b_2) - g(x, b_1)] \ w(x) \ dF_X(x) \\ &> 0, \end{aligned}$$

where the inequality follows from corollary 4.10 of Bartle (1966) as $[g(x, b_2) - g(x, b_1)] \ w(x)$ is positive on a set of positive measure.

(ii)(1) By assumption, $g(X, \cdot)$ is not strictly increasing *a.s.*, so there exists $\mathcal{X}^*$, $P[X \in \mathcal{X}^*] > 0$, such that for each $x \in \mathcal{X}^*$, $g(x, \cdot)$ is not strictly increasing. Further, with $\mathcal{X}_w^* \equiv \mathcal{X}^* \cap \mathcal{X}_w$, we assume $P[X \in \mathcal{X}_w^*] > 0$. Then for the given $0 \leq b_1^* < b_2^* \leq 1$,

$$\begin{aligned} \bar{g}_w(b_2^*) - \bar{g}_w(b_1^*) &= \int [g(x, b_2^*) - g(x, b_1^*)] \ w(x) \ dF_X(x) \\ &= \int [g(x, b_2^*) - g(x, b_1^*)] \ w(x) \ \mathbf{1}\{x \in \mathcal{X}_w^*\} \ dF_X(x) \\ &\quad + \int [g(x, b_2^*) - g(x, b_1^*)] \ w(x) \ \mathbf{1}\{x \notin \mathcal{X}_w^*\} \ dF_X(x) \\ &\leq 0, \end{aligned}$$

where the final inequality follows from the assumed properties of g. This implies that $\bar{g}_w$ is not strictly increasing. Conversely, if there exist no such b_1^*, b_2^*, then for all $0 \leq b_1 < b_2 \leq 1$, $\bar{g}_w(b_2) - \bar{g}_w(b_1) > 0$, so $\bar{g}_w$ is strictly increasing. (2) If $P[X \in \mathcal{X}_w^*] = 0$, then the argument of part (i) gives that $\bar{g}_w$ is strictly increasing. Further, $p_w \equiv P[X \in \mathcal{X}_w] < 1$, as otherwise it must be that $P[X \in \mathcal{X}^*] = 0$, violating our assumption. Then $1 - p_w = P[X \notin \mathcal{X}_w] > 0$, and we can let $\tilde{w}(x) \equiv \mathbf{1}\{x : x \in \tilde{\mathcal{X}}_w\}/(1 - p_w)$. This choice for $\tilde{w}$ is measurable, bounded, and $\int \tilde{w}(x)\, dF_X(x) = 1$, ensuring that $\bar{g}_{\tilde{w}}$ is well defined, that $\tilde{\mathcal{X}}_w = \mathcal{X}_{\tilde{w}} \equiv \{x \in \mathcal{X} : \tilde{w}(x) > 0\}$, and that $P[X \in \tilde{\mathcal{X}}_w^*] > 0$. For the given $0 \leq b_1^* < b_2^* \leq 1$, the argument of part (1) now applies to give that $\bar{g}_{\tilde{w}}$ is not strictly increasing. The converse argument is also identical to part (1). ■

For the next result, we impose Assumption C.0 and write $B \equiv \beta(A)$. For convenience, let $\{X_t, \varepsilon_t\}$ be identically distributed. For succinctness in what follows, we drop the t subscript. Provided the necessary moments exist, we have

$$\tilde{Y}_\tau \equiv E(Y\ w_\tau(X) \mid B) = \bar{g}_\tau(B) + \bar{\varepsilon}_\tau(B), \qquad \text{where now}$$

$$\bar{g}_\tau(B) \equiv E(g(X, B)\ w_\tau(X) \mid B) \qquad \text{and} \qquad \bar{\varepsilon}_\tau(B) \equiv E(\varepsilon\ w_\tau(X) \mid B).$$

We let $\tilde{F}_\tau$ denote the CDF of $\tilde{Y}_\tau$. Note that for simplicity, we defined $\bar{g}_\tau$ in the text in a manner that incorporated $X \perp B$; here $\bar{g}_\tau$ explicitly does not rely on this.

In part (i) of the next result, we assume $X \perp B$ and $\varepsilon \perp B \mid w_\tau(X)$ for all $\tau \in \{1, ..., \mathcal{T}\}$, ensuring that $\tilde{\varepsilon}_\tau = \bar{\varepsilon}_\tau(B)$ is constant. We define the function $\bar{\gamma}_\tau : \mathbb{I} \to \mathbb{I}$ as

$$\bar{\gamma}_\tau(b) \equiv P[\bar{g}_\tau(B) \leq \bar{g}_\tau(b)], \qquad b \in \mathbb{I}.$$

This quantifies the departure of $\bar{g}_\tau$ from monotonicity. When $\bar{g}_\tau$ is strictly monotone, $\bar{\gamma}_\tau(b) = b$. Otherwise, $\bar{\gamma}_\tau$ exhibits variations reflecting those of $\bar{g}_\tau$. Part (i) of the next result shows that a test of $\tilde{H}_0$ has power if and only if there exists τ^* such that $\lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_{\tau^*}(b)] < 1$, where λ denotes Lebesgue measure. This holds with $\mathcal{T} = 2$ when $\bar{g}_1$ is strictly monotone and $\bar{g}_2$ is not strictly monotone on a set of positive $\lambda-$measure. Equivalently, the test has no power if and only if all the $\bar{\gamma}_\tau$'s coincide, except possibly on a set of $\lambda-$measure zero. This occurs when all $\bar{g}_\tau$'s are strictly monotone. It also occurs when $g(x, \cdot)$ does not depend on x, a case ruled out in Theorem 8.1. Other examples exist, but these are exceptional; we conjecture that they are shy.

In part (ii), we drop the requirements that $X \perp B$ and $\varepsilon \perp B \mid w_\tau(X)$. Now we write

$$\tilde{Y}_\tau = \tilde{g}_\tau(B) \equiv \bar{g}_\tau(B) + \bar{\varepsilon}_\tau(B),$$

and we define the functions $\tilde{\gamma}_\tau : \mathbb{I} \to \mathbb{I}$ as

$$\tilde{\gamma}_\tau(b) \equiv P[\tilde{g}_\tau(B) \leq \tilde{g}_\tau(b)], \qquad b \in \mathbb{I}.$$

Here, $\tilde{\gamma}_\tau$ measures the departure of $\tilde{g}_\tau$ from monotonicity. Non-monotonicity may come from $\bar{g}_\tau$, from $\bar{\varepsilon}_\tau$, or both.

Thus, maintaining $X \perp B$ and $\varepsilon \perp B \mid w_\tau(X)$ enables study of the monotonicity of the $\bar{g}_\tau$'s in isolation. Dropping this introduces generic non-monotonicity into $\tilde{g}_\tau$, as $\bar{\varepsilon}_\tau$ is then no longer constant and is thus generically non-monotonic. (Recall the shyness of monotone functions.) Further, the failure of $X \perp B$ generally introduces non-monotonicity into $\bar{g}_\tau$. For example, take $w_\tau(X) = 1$, and suppose that $g(X, B) = X + B$ and that $X \not\perp B$ holds because $X = -B^2 + \eta$, where $\eta \perp B$. (This choice is illustrative, as the relation between X and B is generically non-monotonic.) Then

$$\begin{aligned} \bar{g}_\tau(B) &\equiv E(g(X,B)\ w_\tau(X) \mid B) = E(X + B \mid B) = E(-B^2 + \eta + B \mid B) \\ &= B(1 - B) + E(\eta). \end{aligned}$$

Thus, although $g(x, \cdot)$ is monotone for each x, $\bar{g}_\tau$ is not monotone. Of course, if we instead have $X = B + \eta$, then $\bar{g}_\tau(B) = 2B + E(\eta)$, so the failure of $X \perp B$ is not guaranteed to induce non-monotonicity in $\bar{g}_\tau$. Such cases are exceptional, however. Moreover, when $X \not\perp B$, the role of $w_\tau(X)$ in defining $\bar{g}_\tau(B)$ further reinforces its generic non-monotonicity. ■

Theorem 8.4 *Suppose Assumption C.0 holds with $\{X_t, \varepsilon_t\}$ identically distributed. For $\mathcal{T} \geq 2$, let $w_\tau : \mathcal{X} \to \mathbb{R}^+$, $\tau = 1, ..., \mathcal{T}$ be as in Lemma 8.3. Suppose that $E(g(X, b)) < \infty$ for each $b \in \mathbb{I}$ and that $E(\varepsilon) < \infty$.*

(i) Suppose $X \perp B$ and $\varepsilon \perp B \mid w_\tau(X)$, $\tau = 1, ..., \mathcal{T}$. Then $P[\tilde{F}_1(\tilde{Y}_1) = \cdots = \tilde{F}_\mathcal{T}(\tilde{Y}_\mathcal{T})] = 1$ if and only if $\lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_\tau(b)] = 1$ for all τ.

(ii) $P[\tilde{F}_1(\tilde{Y}_1) = \cdots = \tilde{F}_\mathcal{T}(\tilde{Y}_\mathcal{T})] = 1$ if and only if $\lambda[b : \tilde{\gamma}_1(b) = \tilde{\gamma}_\tau(b)] = 1$ for all τ.

Proof of Theorem 8.4 *(i)* We have

$$P[\tilde{F}_1(\tilde{Y}_1) = \cdots = \tilde{F}_\mathcal{T}(\tilde{Y}_\mathcal{T})] = P[\cap_{\tau=2}^{\mathcal{T}} \{\tilde{F}_1(\tilde{Y}_1) = \tilde{F}_\tau(\tilde{Y}_\tau)\}],$$

so the implication rule gives

$$1 - P[\tilde{F}_1(\tilde{Y}_1) = \cdots = \tilde{F}_\mathcal{T}(\tilde{Y}_\mathcal{T})] \leq \sum_{\tau=2}^{\mathcal{T}} P[\tilde{F}_1(\tilde{Y}_1) \neq \tilde{F}_\tau(\tilde{Y}_\tau)].$$

The first result follows by showing that $\lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_\tau(b)] = 1$ implies $P[\tilde{F}_1(\tilde{Y}_1) \neq \tilde{F}_\tau(\tilde{Y}_\tau)] = 0$, so that $P[\tilde{F}_1(\tilde{Y}_1) = \cdots = \tilde{F}_\mathcal{T}(\tilde{Y}_\mathcal{T})] = 1$. Now

$$P[\tilde{F}_1(\tilde{Y}_1) = \tilde{F}_\tau(\tilde{Y}_\tau)] = \int_0^1 \mathbf{1}\{\tilde{F}_1(\bar{g}_1(b) + \tilde{\varepsilon}_1) = \tilde{F}_\tau(\bar{g}_\tau(b) + \tilde{\varepsilon}_\tau)\}\ db.$$

Given $X \perp B$ and $\varepsilon \perp B \mid w_\tau(X)$, $\tilde{\varepsilon}_\tau$ is constant. It follows that

$$\tilde{F}_\tau(\bar{g}_\tau(b) + \tilde{\varepsilon}_\tau) = P[\bar{g}_\tau(B) + \tilde{\varepsilon}_\tau \leq \bar{g}_\tau(b) + \tilde{\varepsilon}_\tau] = \bar{\gamma}_\tau(b).$$

Thus, for all τ,

$$P[\tilde{F}_1(\tilde{Y}_1) = \tilde{F}_\tau(\tilde{Y}_\tau)] = \int_0^1 \mathbf{1}\{\bar{\gamma}_1(b) = \bar{\gamma}_\tau(b)\}\ db = \lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_\tau(b)] = 1,$$

where the final equality holds by assumption. It follows that $P[\tilde{F}_1(\tilde{Y}_1) = \tilde{F}_\tau(\tilde{Y}_\tau)] = 1$, so $P[\tilde{F}_1(\tilde{Y}_1) \neq \tilde{F}_\tau(\tilde{Y}_\tau)] = 0$, as was to be shown.

For the converse, suppose $\lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_{\tau^*}(b)] < 1$. We have

$$P[\tilde{F}_1(\tilde{Y}_1) = \cdots = \tilde{F}_\mathcal{T}(\tilde{Y}_\mathcal{T})] = 1 - P[\cup_{\tau=2}^{\mathcal{T}}\{\tilde{F}_1(\tilde{Y}_1) \neq \tilde{F}_\tau(\tilde{Y}_\tau)\}].$$

Now

$$P[\cup_{\tau=2}^{\mathcal{T}}\{\tilde{F}_1(\tilde{Y}_1) \neq \tilde{F}_\tau(\tilde{Y}_\tau)\}] \geq P[\tilde{F}_1(\tilde{Y}_1) \neq \tilde{F}_{\tau^*}(\tilde{Y}_{\tau^*})] = 1 - \lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_{\tau^*}(b)].$$

But $\lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_{\tau^*}(b)] < 1$, so $1 - \lambda[b : \bar{\gamma}_1(b) = \bar{\gamma}_{\tau^*}(b)] > 0$, implying $P[\tilde{F}_1(\tilde{Y}_1) = \cdots = \tilde{F}_\mathcal{T}(\tilde{Y}_\mathcal{T})] < 1$.

(ii) Identical to (i), replacing $\bar{\gamma}_\tau$ with $\tilde{\gamma}_\tau$ and dropping $\tilde{\varepsilon}_\tau$. ■

Lemma 8.5 *Suppose Assumptions C.1(ii), C.2, and C.3(i) hold. Then for* $\tau = 1, 2, ..., \mathcal{T}$, *(i)* $E(\bar{Y}_{T,\tau,i} - \tilde{Y}_{\tau,i})^2 = O\left(T^{-1}\right)$; *and (ii)* $E|\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i}) - \tilde{F}_\tau(\tilde{Y}_{\tau,i})| = O\left(T^{-1/2}\right)$.

Proof of Lemma 8.5 Noting that $\tilde{Y}_{\tau,i} = E\left[Y_{it}w_\tau\left(X_{it}\right)|B_i\right] = E\left[g\left(X_{it}, B_i\right)w_\tau\left(X_{it}\right)|B_i\right] + E\left[\varepsilon_{it}w_\tau\left(X_{it}\right)|B_i\right] \equiv \overline{g}_\tau\left(B_i\right) + \overline{\varepsilon}_\tau\left(B_i\right)$, we have $\bar{Y}_{T,\tau,i} - \tilde{Y}_{\tau,i} = T^{-1}\sum_{t=1}^T\left[g\left(X_{it}, B_i\right)w_\tau\left(X_{it}\right) - \overline{g}_\tau\left(B_i\right)\right]$ $+T^{-1}\sum_{t=1}^T[\varepsilon_{it}w_\tau\left(X_{it}\right) - \overline{\varepsilon}_\tau\left(B_i\right)] \equiv \alpha_{NT1} + \alpha_{NT2}$, say. Let $\zeta_{i,t} \equiv g\left(X_{it}, B_i\right)w_\tau\left(X_{it}\right) - \overline{g}_\tau\left(B_i\right)$. Then $E\left[\alpha_{NT1}\right] = 0$, and $E\left[\alpha_{NT1}^2\right] = T^{-1}E\left[\zeta_{i,t}\right]^2 + 2T^{-1}\sum_{s=1}^T\text{Cov}\left(\zeta_{i,1}, \zeta_{i,1+s}\right) = O\left(T^{-1}\right)$ as $\sum_{s=1}^T\text{Cov}\ \left(\zeta_{i,1}, \zeta_{i,1+s}\right) \leq \left\|\zeta_{i,1}\right\|_{2+\gamma}^2\sum_{s=1}^\infty\alpha\left(s\right)^{\gamma/(2+\gamma)} < \infty$ by the Davydov inequality and Assumptions C.1(ii) and C.2. Similarly, $E\left[\alpha_{NT2}^2\right] = O_P\left(T^{-1}\right)$. Thus (i) follows.

For (ii), we have

$$\begin{aligned}
E\left|\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})-\tilde{F}_{\tau}(\tilde{Y}_{\tau,i})\right| &= \int\left|\tilde{F}_{\tau}(y)-\tilde{F}_{T\tau}(y)\right|\tilde{f}_{\tau}(y)\,dy \\
&= \int\left|E\left[\mathbf{1}\left\{\tilde{Y}_{\tau,i}\leq y\right\}-\mathbf{1}\left\{\bar{Y}_{T,\tau,i}\leq y\right\}\right]\right|\tilde{f}_{\tau}(y)\,dy \\
&\leq \int E\left|\mathbf{1}\left\{\tilde{Y}_{\tau,i}-y\leq 0\right\}-\mathbf{1}\left\{\tilde{Y}_{\tau,i}-y\leq \tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}\right\}\right|\tilde{f}_{\tau}(y)\,dy \\
&\leq \int E\left[\mathbf{1}\left\{|y-\tilde{Y}_{\tau,i}|\leq|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|\right\}\right]\tilde{f}_{\tau}(y)\,dy \\
&= E\left[\tilde{F}_{\tau}\left(\tilde{Y}_{\tau,i}+|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|\right)-\tilde{F}_{\tau}\left(\tilde{Y}_{\tau,i}-|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|\right)\right] \\
&= 2E\left[\tilde{f}_{\tau}(\tilde{Y}^{*}_{\tau,i})|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|\right] \\
&\leq C\left[E(\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i})^2\right]^{1/2}=O(T^{-1/2}),
\end{aligned}$$

where the first and second inequalities follow from the triangle inequality and the fact $|\mathbf{1}\{z<0\}-\mathbf{1}\{z<a\}|\leq\mathbf{1}\{|z|<|a|\}$, respectively; the third equality holds by the Fubini theorem; the next inequality holds by the mean value theorem , where $\tilde{Y}^{*}_{\tau,i}$ lies between $\tilde{Y}_{\tau,i}-|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|$ and $\tilde{Y}_{\tau,i}+|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|$; the last inequality follows from Assumption C.3(i) and the Jensen inequality; and the last equality follows from (i).■

Proof of Theorems 4.1 and 4.2

We only prove Theorem 4.2. For notational simplicity, we only prove the case where $\mathcal{T}=2$. Let $\tilde{F}_{T\tau}$ and $\tilde{f}_{T\tau}$ denote the CDF and PDF of $\bar{Y}_{T,\tau,i}$, $\tau=1,2$, respectively. Let $\bar{F}_{N,T,\tau}(y)\equiv\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\{\tilde{Y}_{\tau,i}\leq y\}$. Define

$$\tilde{D}_{NT}\equiv\sum_{i=1}^{N}\left[\hat{F}_{N,T,1}(\tilde{Y}_{1,i})-\hat{F}_{N,T,2}(\tilde{Y}_{2,i})\right]^2 \quad\text{and}\quad \bar{D}_{NT}\equiv\sum_{i=1}^{N}\left[\bar{F}_{N,T,1}(\tilde{Y}_{1,i})-\bar{F}_{N,T,2}(\tilde{Y}_{2,i})\right]^2.$$

We prove Theorem 4.2 by showing that (i) $\hat{D}_{NT}-\tilde{D}_{NT}=o_P(1)$; (ii) $\tilde{D}_{NT}-\bar{D}_{NT}=o_P(1)$; and (iii) $\bar{D}_{NT}-\hat{\mathcal{B}}_{NT}-\mu\xrightarrow{d}\sum_{j=1}^{\infty}\lambda_j(\mathcal{Z}_j^2-1)$ under $\tilde{H}_1\left(N^{-1/2}\right)$.

For (i), noting that $a^2-b^2=(a-b)^2+2(a-b)b$, we have

$$\begin{aligned}
\hat{D}_{NT}-\tilde{D}_{NT} &= \sum_{i=1}^{N}\left[\hat{F}_{N,T,1}(\bar{Y}_{T,1,i})-\hat{F}_{N,T,1}(\tilde{Y}_{1,i})-\hat{F}_{N,T,2}(\bar{Y}_{T,2,i})+\hat{F}_{N,T,2}(\tilde{Y}_{2,i})\right]^2 \\
&\quad +2\sum_{i=1}^{N}\left[\hat{F}_{N,T,1}(\bar{Y}_{T,1,i})-\hat{F}_{N,T,1}(\tilde{Y}_{1,i})-\hat{F}_{N,T,2}(\bar{Y}_{T,2,i})+\hat{F}_{N,T,2}(\tilde{Y}_{2,i})\right] \\
&\quad \times\left[\hat{F}_{N,T,1}(\tilde{Y}_{1,i})-\hat{F}_{N,T,2}(\tilde{Y}_{2,i})\right] \\
&\equiv \hat{\vartheta}_{NT1}+2\hat{\vartheta}_{NT2},\ \text{say.}
\end{aligned}$$

By the c_r inequality,

$$\begin{aligned}\hat{\vartheta}_{NT1} &\leq 2\sum_{\tau=1}^{2}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left[\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\bar{Y}_{T,\tau,i}\}-\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\tilde{Y}_{\tau,i}\}\right]\right]^{2}\\ &\leq 4\sum_{\tau=1}^{2}\sum_{i=1}^{N}\left[\frac{1}{N}\sum_{j=1}^{N}\left[\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\bar{Y}_{T,\tau,i}\}-\tilde{F}_{T\tau}\left(\bar{Y}_{T,\tau,i}\right)-\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\tilde{Y}_{\tau,i}\}+\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})\right]\right]^{2}\\ &\quad+4\sum_{\tau=1}^{2}\sum_{i=1}^{N}\left[\tilde{F}_{T\tau}\left(\bar{Y}_{T,\tau,i}\right)-\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})\right]^{2}.\end{aligned}$$

The first term in the last expression is $o_P(1)$ because by the stochastic equicontinuity (SE) of the empirical process

$$\eta_{NT}(\cdot)\equiv N^{-1/2}\sum_{j=1}^{N}[\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\cdot\}-\tilde{F}_{T\tau}(\cdot)] \tag{22}$$

and Lemma 8.5(i), $N^{-1/2}\sum_{j=1}^{N}[\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\bar{Y}_{T,\tau,i}\}-\tilde{F}_{T\tau}(\bar{Y}_{T,\tau,i})-\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\tilde{Y}_{\tau,i}\}+\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})]=o_P(1)$ uniformly in i. The second term is $o_P(1)$ because by Lemma 8.5(ii) and Assumption C.5, $\sum_{i=1}^{N}[\tilde{F}_{T\tau}\left(\bar{Y}_{T,\tau,i}\right)-\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})]^{2}=\sum_{i=1}^{N}\tilde{f}_{T\tau}(\tilde{Y}_{\tau,i}^{*})^{2}(\bar{Y}_{T,\tau,i}-\tilde{Y}_{\tau,i})^{2}\leq C\sum_{i=1}^{N}(\bar{Y}_{T,\tau,i}-\tilde{Y}_{\tau,i})^{2}=O_P\left(NT^{-1}\right)=o_P(1)$, provided $\tilde{f}_{T\tau}$ is uniformly bounded for sufficiently large T, where $\tilde{Y}_{\tau,i}^{*}$ lies between $\tilde{Y}_{\tau,i}$ and $\bar{Y}_{T,\tau,i}$. By the strong law of large number for strong mixing processes (e.g., Corollary 3.48 of White (2001)), $\bar{Y}_{T,\tau,i}-\tilde{Y}_{\tau,i}=o_{a.s.}(1)$ under Assumptions C.1(ii) and C.2. This implies that as $T\to\infty$, the limiting support of $\bar{Y}_{T,\tau,i}$ will coincide with that of $\tilde{Y}_{\tau,i}$. By the continuity of $\bar{g}_{\tau}$ in Assumption C.3(ii), the support of $\tilde{Y}_{\tau,i}$ is compact. This implies that for sufficiently large T, with probability 1 the support of $\bar{Y}_{T,\tau,i}$ is also compact, so that $\tilde{f}_{T\tau}$ is uniformly continuous on this support and must be bounded.

Let $\beta_{1\tau,ij}=\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\bar{Y}_{T,\tau,i}\}-\tilde{F}_{T\tau}\left(\bar{Y}_{T,\tau,i}\right)-\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\tilde{Y}_{\tau,i}\}+\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})$ and $\beta_{2\tau,i}=\tilde{F}_{T\tau}\left(\bar{Y}_{T,\tau,i}\right)-\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})$ for $\tau=1,2$ and $i,\ j=1,...,N$. Let $\beta_{3,ij}=\mathbf{1}\{\bar{Y}_{T,1,j}\leq\tilde{Y}_{1,i}\}-\tilde{F}_{T1}(\tilde{Y}_{1,i})-\mathbf{1}\{\bar{Y}_{T,2,j}\leq\tilde{Y}_{2,i}\}+\tilde{F}_{T2}(\tilde{Y}_{2,i})$, and $\beta_{4,i}=\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})$. Analogously to the proof of Lemma 8.5 and by the triangle and c_r inequalities, we can show that uniformly in $i,\ j=1,...,N$,

$$E|\beta_{1\tau,ij}|\leq E|\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\bar{Y}_{T,\tau,i}\}-\mathbf{1}\{\bar{Y}_{T,\tau,j}\leq\tilde{Y}_{\tau,i}\}|+E|\tilde{F}_{T\tau}\left(\bar{Y}_{T,\tau,i}\right)-\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})|=O(T^{-1/2}), \tag{23}$$

and

$$\begin{aligned}E\left(\beta_{4,i}^{2}\right) &\leq 4\sum_{\tau=1}^{2}E\{[\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})-\tilde{F}_{\tau}(\tilde{Y}_{\tau,i})]^{2}\}+2E\{[\tilde{F}_{1}(\tilde{Y}_{1,i})-\tilde{F}_{2}(\tilde{Y}_{2,i})]^{2}\}\\ &= O\left(T^{-1}+N^{-1}\right)\ \text{under}\ \tilde{H}_{1}(N^{-1/2}).\end{aligned} \tag{24}$$

Now decompose $\hat{\vartheta}_{NT2}$ as follows

$$\begin{aligned}
\hat{\vartheta}_{NT2} &= N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\left(\beta_{11,ij}-\beta_{12,ij}+\beta_{21,i}-\beta_{22,i}\right)\left(\beta_{3,ik}-\beta_{4,i}\right) \\
&= N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\left(\beta_{11,ij}-\beta_{12,ij}\right)\beta_{3,ik}+N^{-1}\sum_{i=1}^{N}\sum_{k=1}^{N}\left(\beta_{21,i}-\beta_{22,i}\right)\beta_{3,ik} \\
&\quad -N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(\beta_{11,ij}-\beta_{12,ij}\right)\beta_{4,i}-\sum_{i=1}^{N}\left(\beta_{21,i}-\beta_{22,i}\right)\beta_{4,i} \\
&\equiv \hat{\vartheta}_{NT2,1}+\hat{\vartheta}_{NT2,2}-\hat{\vartheta}_{NT2,3}-\hat{\vartheta}_{NT2,4},\ \text{say.}
\end{aligned}$$

Let $\hat{\vartheta}_{NT2,1\tau}=N^{-2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\beta_{1\tau,ij}\beta_{3,ik}$ for $\tau=1,2$. It is easy to show that $\hat{\vartheta}_{NT2,1\tau}=\theta_{NT2,1\tau}+o_P(1)$ under $\tilde{H}_1\left(N^{-1/2}\right)$, where $\theta_{NT2,1\tau}=N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq j,i}^{N}\beta_{1\tau,ij}\beta_{3,ik}$. Note that $E\left(\theta_{NT2,1\tau}\right)=0$, and

$$E[\theta_{NT2,1\tau}^2]=N^{-4}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq j,i}^{N}\sum_{i'=1}^{N}\sum_{j'\neq i'}^{N}\sum_{k\neq j',i'}^{N}E\left[\beta_{1\tau,ij}\beta_{3,ik}\beta_{1\tau,i'j'}\beta_{3,i'k'}\right].$$

If there are five or six distinct indices among $\{i,j,k,i',j',k'\}$, then the corresponding terms in the above summation drop out. For all other cases, it is straightforward to bound $|E[\beta_{1\tau,ij}\beta_{3,ik}\beta_{1\tau,i'j'}\beta_{3,i'k'}]|$ by a proportion of $E|\beta_{1\tau,ij}|=O\left(T^{-1/2}\right)$ by the uniform boundedness of $\beta_{1\tau,ij}$ and $\beta_{3,ik}$ and (23). It follows that $E[\theta_{NT2,1\tau}^2]=O\left(T^{-1/2}+N^{-1}\right)$ and $\hat{\vartheta}_{NT2,1\tau}=o_P(1)$. Then $\hat{\vartheta}_{NT2,1}=\hat{\vartheta}_{NT2,11}-\hat{\vartheta}_{NT2,12}=o_P(1)$. Similarly, we can show that $\hat{\vartheta}_{NT2,2}=o_P(1)$.

Let $\hat{\vartheta}_{NT2,3\tau}=N^{-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\beta_{1\tau,ij}\beta_{4,i}$ for $\tau=1,2$. Then we can show that $\hat{\vartheta}_{NT2,3\tau}=\theta_{NT2,3\tau}+O_P(N^{-1/2})$ under $\tilde{H}_1(N^{-1/2})$, where $\theta_{NT2,3\tau}=N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\beta_{1\tau,ij}\beta_{4,i}$. Note that $E\left[\theta_{NT2,3\tau}\right]=0$ and

$$\begin{aligned}
E[\theta_{NT2,3\tau}^2] &= N^{-2}\sum_{i=1}^{N}\sum_{i'\neq i}^{N}\sum_{j\neq i,i'}^{N}E\left[\beta_{1\tau,ij}\beta_{4,i}\beta_{1\tau,i'j}\beta_{4,i'}\right] \\
&\quad +N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left[\beta_{1\tau,ij}\beta_{4,i}\beta_{1\tau,ji}\beta_{4,j}+\left(\beta_{1\tau,ij}\beta_{4,i}\right)^2\right].
\end{aligned}$$

It is straightforward to show that the last term is $O\left(T^{-1/2}\right)$ under $\tilde{H}_1\left(N^{-1/2}\right)$. We can bound the first term by

$$\begin{aligned}
& N^{-2}\sum_{i=1}^{N}\sum_{i'\neq i}^{N}\sum_{j\neq i,i'}^{N}\left[E(\beta_{1\tau,ij}^2\beta_{1\tau,i'j}^2)\right]^{1/2}\left[E(\beta_{4,i}^2)E(\beta_{4,i'}^2)\right]^{1/2} \\
\leq\ & CN\sup_{i,j}\left\{E\left|\beta_{1\tau,ij}\right|\right\}^{1/2}E(\beta_{4,1}^2)=O(N)\,O(T^{-1/4})O\left(T^{-1}+N^{-1}\right)=o(1).
\end{aligned}$$

It follows that $\theta_{NT2,3\tau} = o_P(1)$ and $\hat{\vartheta}_{NT2,3\tau} = \hat{\vartheta}_{NT2,31} - \hat{\vartheta}_{NT2,32} = o_P(1)$. Similarly,

$$
\begin{aligned}
E|\hat{\vartheta}_{NT2,4}| &\leq \sum_{i=1}^{N} E\left|\left(\beta_{21,i} - \beta_{22,i}\right)\beta_{4,i}\right| \leq N\left\{E\left(\beta_{21,i} - \beta_{22,i}\right)^2\right\}^{1/2}\left\{E\left(\beta_{4,i}^2\right)\right\}^{1/2} \\
&= N\; O(T^{-1/2})O(T^{-1/2} + N^{-1/2}) = o(1).
\end{aligned}
$$

Consequently $\hat{\vartheta}_{NT2,4} = o_P(1)$. Thus, $\hat{\vartheta}_{NT2} = o_P(1)$.

For (ii), decompose $\tilde{D}_{NT} - \bar{D}_{NT}$ as follows:

$$
\begin{aligned}
\tilde{D}_{NT} - \bar{D}_{NT} &= \sum_{i=1}^{N}\left[\hat{F}_{N,T,1}(\tilde{Y}_{1,i}) - \bar{F}_{N,T,1}(\tilde{Y}_{1,i}) - \hat{F}_{N,T,2}(\tilde{Y}_{2,i}) + \bar{F}_{N,T,2}(\tilde{Y}_{2,i})\right]^2 \\
&\quad + 2\sum_{i=1}^{N}\left[\hat{F}_{N,T,1}(\tilde{Y}_{1,i}) - \bar{F}_{N,T,1}(\tilde{Y}_{1,i}) - \hat{F}_{N,T,2}(\tilde{Y}_{2,i}) + \bar{F}_{N,T,2}(\tilde{Y}_{2,i})\right] \\
&\quad \times\left[\bar{F}_{N,T,1}(\tilde{Y}_{1,i}) - \bar{F}_{N,T,2}(\tilde{Y}_{2,i})\right] \\
&\equiv \tilde{\vartheta}_{NT1} + 2\tilde{\vartheta}_{NT2}, \text{ say.}
\end{aligned}
$$

Note that $\tilde{\vartheta}_{NT1} \leq 2\sum_{\tau=1}^{2}\sum_{i=1}^{N}\left[N^{-1}\sum_{j=1}^{N}[\mathbf{1}\{\bar{Y}_{T,\tau,j} \leq \tilde{Y}_{\tau,i}\} - \mathbf{1}\{\tilde{Y}_{\tau,j} \leq \tilde{Y}_{\tau,i}\}]\right]^2 = 2\sum_{\tau=1}^{2}\vartheta_\tau$ $+O_P\left(N^{-1}\right)$, where $\vartheta_\tau = N^{-2}\sum_{i=1}^{N}\left[\sum_{j\neq i}^{N}[\mathbf{1}\{\bar{Y}_{T,\tau,j} \leq \tilde{Y}_{\tau,i}\} - \mathbf{1}\{\tilde{Y}_{\tau,j} \leq \tilde{Y}_{\tau,i}\}]\right]^2$. Further,

$$
\begin{aligned}
\vartheta_\tau = N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq i,j}^{N}&[\mathbf{1}\{\bar{Y}_{T,\tau,j} \leq \tilde{Y}_{\tau,i}\} - \mathbf{1}\{\tilde{Y}_{\tau,j} \leq \tilde{Y}_{\tau,i}\}][\mathbf{1}\{\bar{Y}_{T,\tau,k} \leq \tilde{Y}_{\tau,i}\} - \mathbf{1}\{\tilde{Y}_{\tau,k} \leq \tilde{Y}_{\tau,i}\}] \\
&+ N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}[\mathbf{1}\{\bar{Y}_{T,\tau,j} \leq \tilde{Y}_{\tau,i}\} - \mathbf{1}\{\tilde{Y}_{\tau,j} \leq \tilde{Y}_{\tau,i}\}]^2 = \vartheta_{\tau1} + \vartheta_{\tau2}, \text{ say.}
\end{aligned}
$$

By the proof of Lemma 8.5(ii), we can show that $E(\vartheta_{\tau1}) = O(N/T) = o(1)$ and $E(\vartheta_{\tau2}) = O\left(T^{-1/2}\right) = o(1)$. It follows that $E|\vartheta_\tau| = E(\vartheta_{\tau1}) + E(\vartheta_{\tau2}) = o(1)$. So $\vartheta_\tau = o_P(1)$ by the Markov inequality, and $\tilde{\vartheta}_{NT1} = o_P(1)$. Analogous to the determination of the probability order of $\hat{\vartheta}_{NT2}$, we can show that $\tilde{\vartheta}_{NT2} = o_P(1)$. Consequently, $\tilde{D}_{NT} - \bar{D}_{NT} = o_P(1)$.

Now, we show (iii). Decompose $\bar{D}_{NT}$ as follows

$$\begin{aligned}
\bar{D}_{NT} &= \sum_{i=1}^{N}\left[\bar{F}_{N,T,1}(\tilde{Y}_{1,i})-\bar{F}_{N,T,2}(\tilde{Y}_{2,i})\right]^2 \\
&= \sum_{i=1}^{N}\left\{\left[\bar{F}_{N,T,1}(\tilde{Y}_{1,i})-\tilde{F}_1(\tilde{Y}_{1,i})-\bar{F}_{N,T,2}(\tilde{Y}_{2,i})+\tilde{F}_2(\tilde{Y}_{2,i})\right]+\left[\tilde{F}_1(\tilde{Y}_{1,i})-\tilde{F}_2(\tilde{Y}_{2,i})\right]\right\}^2 \\
&= \sum_{i=1}^{N}\left[\tilde{F}_1(\tilde{Y}_{1,i})-\tilde{F}_2(\tilde{Y}_{2,i})\right]^2 \\
&\quad +\sum_{i=1}^{N}\left[\bar{F}_{N,T,1}(\tilde{Y}_{1,i})-\tilde{F}_1(\tilde{Y}_{1,i})-\bar{F}_{N,T,2}(\tilde{Y}_{2,i})+\tilde{F}_2(\tilde{Y}_{2,i})\right]^2 \\
&\quad +2\sum_{i=1}^{N}\left[\bar{F}_{N,T,1}(\tilde{Y}_{1,i})-\tilde{F}_1(\tilde{Y}_{1,i})-\bar{F}_{N,T,2}(\tilde{Y}_{2,i})+\tilde{F}_2(\tilde{Y}_{2,i})\right]\left[\tilde{F}_1(\tilde{Y}_{1,i})-\tilde{F}_2(\tilde{Y}_{2,i})\right] \\
&\equiv \bar{D}_{NT1}+\bar{D}_{NT2}+2\bar{D}_{NT3}. \qquad (25)
\end{aligned}$$

By the weak law of large numbers, $\bar{D}_{NT1} \xrightarrow{P} \mu$ under $\tilde{H}_1\left(N^{-1/2}\right)$. Let $\xi_i \equiv (\tilde{Y}_{1,i},\tilde{Y}_{2,i})'$ and $\psi\left(\xi_i,\xi_j\right) \equiv \mathbf{1}\{\tilde{Y}_{1,j}\leq\tilde{Y}_{1,i}\}-\tilde{F}_1(\tilde{Y}_{1,i})-\mathbf{1}\{\tilde{Y}_{2,j}\leq\tilde{Y}_{2,i}\}+\tilde{F}_2(\tilde{Y}_{2,i})$. Then

$$\begin{aligned}
\bar{D}_{NT2} &= N^{-2}\sum_{i=1}^{N}\left[\sum_{j=1}^{N}\psi\left(\xi_{T,i},\xi_{T,j}\right)\right]^2 \\
&= N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq j,i}^{N}\psi\left(\xi_{T,i},\xi_{T,j}\right)\psi\left(\xi_{T,i},\xi_{T,k}\right)+N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi\left(\xi_{T,i},\xi_{T,j}\right)^2 \\
&\quad +2N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi\left(\xi_{T,i},\xi_{T,i}\right)\psi\left(\xi_{T,i},\xi_{T,j}\right)+N^{-2}\sum_{i=1}^{N}\psi\left(\xi_{T,i},\xi_{T,i}\right)^2 \\
&\equiv V_{NT}+\mathcal{B}_{NT}+2R_{NT1}+R_{NT2}, \text{ say.}
\end{aligned}$$

Let $\bar{\psi}\left(\xi_i,\xi_j,\xi_k\right) \equiv [\psi\left(\xi_i,\xi_j\right)\psi\left(\xi_i,\xi_k\right)+\psi\left(\xi_j,\xi_i\right)\,\psi\left(\xi_j,\xi_k\right)+\psi\left(\xi_k,\xi_i\right)\psi\left(\xi_k,\xi_j\right)]/3$. Then

$$V_{NT}=6N^{-2}\sum_{1\leq i<j<k\leq N}\bar{\psi}\left(\xi_i,\xi_j,\xi_k\right)=\frac{(N-1)(N-2)}{N}\bar{V}_{NT},$$

where $\bar{V}_{NT} \equiv \frac{6}{N(N-1)(N-2)}\sum_{1\leq i<j<k\leq N}\bar{\psi}\left(\xi_i,\xi_j,\xi_k\right)$. By the Hoeffding decomposition, $\bar{V}_{NT} = 3H_{NT}^{(2)}+H_{NT}^{(3)}$, where

$$H_{NT}^{(2)} \equiv \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N}\bar{\psi}_2\left(\xi_i,\xi_j\right),\ H_{NT}^{(3)} \equiv \frac{6}{N(N-1)(N-2)}\sum_{1\leq i<j<k\leq N}\bar{\psi}_3\left(\xi_i,\xi_j,\xi_k\right),$$

$\bar{\psi}_2\left(\xi_i,\xi_j\right) \equiv \frac{1}{3}\int\psi\left(\xi,\xi_i\right)\psi\left(\xi,\xi_j\right)\tilde{F}\left(d\xi\right)$, $\bar{\psi}_3\left(\xi_i,\xi_j,\xi_k\right) \equiv \bar{\psi}\left(\xi_i,\xi_j,\xi_k\right)-\bar{\psi}_2\left(\xi_i,\xi_j\right)-\bar{\psi}_2\left(\xi_i,\xi_k\right)$ $-\bar{\psi}_2\left(\xi_j,\xi_k\right)$, and $\tilde{F}$ denotes the CDF of ξ_i. It is standard to show that $H_{NT}^{(3)}=O_P\left(N^{-3/2}\right)$.

Thus, $V_{NT} = \{1 + o(1)\} \mathcal{H}_{NT} + O_P\left(N^{-1/2}\right)$, where

$$\mathcal{H}_{NT} \equiv \frac{2}{N} \sum_{1 \le i < j \le N} \int \psi(\xi, \xi_i) \psi\left(\xi, \xi_j\right) \tilde{F}(d\xi)$$

is a second order degenerate U-statistic. By Serfling (1980, p.194) or Proposition 5.2 of Chen and White (1998), $\mathcal{H}_{NT} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j (\mathcal{Z}_j^2 - 1)$ where $\{\mathcal{Z}_j\}$ is a sequence of IID $N(0,1)$ random variables, and $\{\lambda_j\}$ is the sequence of nonzero eigenvalues for $\int \psi(\xi, u) \psi(\xi, v) \tilde{F}(d\xi)$. Next, $R_{NT1} = N^{-2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \psi(\xi_i, \xi_i) \psi\left(\xi_i, \xi_j\right) = O_P\left(T^{-1/2}\right)$, and $R_{NT2} = N^{-2} \sum_{i=1}^{N} \psi(\xi_i, \xi_i)^2 = O_P\left(N^{-1} T^{-1}\right) = o_P(1)$. Consequently $\bar{D}_{NT2} - \mathcal{B}_{NT} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j \left(\mathcal{Z}_j^2 - 1\right)$.

Letting $\chi\left(\xi_i, \xi_j\right) = [\mathbf{1}\{\tilde{Y}_{1,j} \le \tilde{Y}_{1,i}\} - \tilde{F}_1(\tilde{Y}_{1,i}) - \mathbf{1}\{\tilde{Y}_{2,j} \le \tilde{Y}_{2,i}\} + \tilde{F}_2(\tilde{Y}_{2,i})] \delta_{N,1,2}(\tilde{Y}_{1,i}, \tilde{Y}_{2,i})$, we have

$$\bar{D}_{NT3} = N^{-3/2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \chi\left(\xi_i, \xi_j\right) + N^{-3/2} \sum_{i=1}^{N} \chi(\xi_i, \xi_i).$$

It is easy to show that the second term is $O_P\left(N^{-1/2}\right)$. Let $R_N = N^{-3/2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} \chi\left(\xi_i, \xi_j\right)$. Then $E(R_N) = 0$ and by the Hölder and c_r inequalities,

$$\begin{aligned}
E\left(R_N^2\right) &= N^{-3} \sum_{i=1}^{N} \sum_{i'=1}^{N} \sum_{j \neq i, i'}^{N} E\left[\chi\left(\xi_i, \xi_j\right) \chi\left(\xi_{i'}, \xi_j\right)\right] + O\left(N^{-1}\right) \\
&\le N^{-2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} E\left[\chi\left(\xi_i, \xi_j\right)^2\right] + O\left(N^{-1}\right) \\
&\le 2N^{-2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} E\left\{\left[\mathbf{1}\{\tilde{Y}_{1,j} \le \tilde{Y}_{1,i}\} - \mathbf{1}\{\tilde{Y}_{2,j} \le \tilde{Y}_{2,i}\}\right]^2 \delta_{N,1,2}^2(\tilde{Y}_{1,i}, \tilde{Y}_{2,i})\right\} \\
&\quad + 2N^{-2} \sum_{i=1}^{N} \sum_{j \neq i}^{N} E\left\{\left[\tilde{F}_1(\tilde{Y}_{1,i}) - \tilde{F}_2(\tilde{Y}_{2,i})\right]^2 \delta_{N,1,2}^2(\tilde{Y}_{1,i}, \tilde{Y}_{2,i})\right\} + O\left(N^{-1}\right) \\
&\equiv 2R_{N1} + 2R_{N2} + O\left(N^{-1}\right), \text{ say.}
\end{aligned}$$

By the dominated convergence theorem (DCT), $R_{N2} = o(1)$ as $[\tilde{F}_1(\tilde{Y}_{1,i}) - \tilde{F}_2(\tilde{Y}_{2,i})]^2 \delta_{N,1,2}^2(\tilde{Y}_{1,i}, \tilde{Y}_{2,i}) \to 0$ a.s. under $\tilde{H}_1\left(N^{-1/2}\right)$. For R_{N1}, we have

$$\begin{aligned}
R_{N1} &\le E\left[|\mathbf{1}\{\tilde{Y}_{1,2} \le \tilde{Y}_{1,1}\} - \mathbf{1}\{\tilde{Y}_{2,2} \le \tilde{Y}_{2,1}\}| \delta_{N,1,2}^2(\tilde{Y}_{1,1}, \tilde{Y}_{2,1})\right] \\
&= E\left[|\mathbf{1}\{\tilde{F}_1(\tilde{Y}_{1,2}) \le \tilde{F}_1(\tilde{Y}_{1,1})\} - \mathbf{1}\{\tilde{F}_2(\tilde{Y}_{2,2}) \le \tilde{F}_2(\tilde{Y}_{2,1})\}| \delta_{N,1,2}^2(\tilde{Y}_{1,1}, \tilde{Y}_{2,1})\right] \\
&\le E\left[\mathbf{1}\{|\tilde{F}_1(\tilde{Y}_{1,2}) - \tilde{F}_1(\tilde{Y}_{1,1})| \le |\alpha_N|\} \delta_{N,1,2}^2(\tilde{Y}_{1,1}, \tilde{Y}_{2,1})\right] \\
&\to 0,
\end{aligned}$$

where $\alpha_N = \tilde{F}_1(\tilde{Y}_{1,2}) - \tilde{F}_2(\tilde{Y}_{2,2}) - \tilde{F}_1(\tilde{Y}_{1,1}) + \tilde{F}_2(\tilde{Y}_{2,1}) = N^{-1/2}[\delta_{N,1,2}(\tilde{Y}_{1,2}, \tilde{Y}_{2,2}) + \delta_{N,1,2}(\tilde{Y}_{1,1}, \tilde{Y}_{2,1})]$; the third line follows from the fact that $|\mathbf{1}\{z \le 0\} - \mathbf{1}\{z \le a\}| \le \mathbf{1}\{|z| \le |a|\}$; and the last line follows from the DCT. Consequently, $R_N = o_P(1)$ by the Chebyshev inequality and $\bar{D}_{NT3} = o_P(1)$. We complete the proof of (iii) by noting that $\hat{\mathcal{B}}_{NT} - \mathcal{B}_{NT} = o_P(1)$ under $\tilde{H}_1\left(N^{-1/2}\right)$ follows easily. ■

Proof of Theorems 4.3

Again, we focus on the case $\mathcal{T} = 2$. Using the notation in the proof of Theorem 4.2, it is easy to show that $N^{-1}(\hat{D}_{NT} - \tilde{D}_{NT}) = o_P(1)$ and $N^{-1}(\tilde{D}_{NT} - \bar{D}_{NT}) = o_P(1)$ under $\tilde{H}_1(1)$. Further, $N^{-1}\bar{D}_{NT} = N^{-1}\bar{D}_{NT1} + o_P(1) = N^{-1}\sum_{i=1}^{N}\left[\tilde{F}_1(\tilde{Y}_{1,i}) - \tilde{F}_2(\tilde{Y}_{2,i})\right]^2 + o_P(1) = \mu + o_P(1)$, and $N^{-1}\hat{\mathcal{B}}_{NT} = O_P\left(N^{-1}\right)$. Consequently, $N^{-1}J_{NT} = N^{-1}(\hat{D}_{NT} - \hat{\mathcal{B}}_{NT}) = N^{-1}\bar{D}_{NT} + N^{-1}\left(\hat{D}_{NT} - \tilde{D}_{NT}\right) + N^{-1}(\tilde{D}_{NT} - \bar{D}_{NT}) - N^{-1}\hat{\mathcal{B}}_{NT} = \mu + o_P(1)$, and the conclusion follows. ■

9 Data Appendix

This appendix contains two tables. One is the list of the 50 S&P500 firms used in Section 6.1, and the other contains the summary statistics for the 70 city-retailers over 156 weeks used in Section 6.2.

Table A.1: List of the 50 S&P500 firms used in the sample

Ticker symbol	Company	GICS Sector
ABT	Abbott Laboratories	Health Care
AGN	Allergan Inc	Health Care
AMZN	Amazon Corp	Consumer Discretionary
AN	AutoNation Inc	Consumer Discretionary
BBY	Best Buy Co. Inc	Consumer Discretionary
CAM	Cameron International Corp	Energy
CBE	Cooper Industries Ltd	Industrials
CINF	Cincinnati Financial	Financials
CLX	Clorox Co.	Consumer Staples
CSCO	Cisco Systems	Information Technology
EMC	EMC Corp	Information Technology
FLS	Flowserve Corporation	Industrials
FMC	FMC Corporation	Materials
GENZ	Genzyme Corp.	Health Care
GWW	Grainger (W.W.) Inc	Industrials
HCP	HCP Inc	Financials
HD	Home Depot	Consumer Discretionary
HP	Helmerich & Payne	Energy
HSY	The Hershey Company	Consumer Staples
IBM	International Bus. Machines	Information Technology
INTU	Intuit Inc.	Information Technology
IVZ	Invesco Ltd	Financials
JCI	Johnson Controls	Consumer Discretionary
JPM	JPMorgan Chase & Co	Financials
KLAC	KLA-Tencor Corp	Information Technology
LTD	Limited Brands Inc	Consumer Discretionary
MCD	McDonald's Corp	Consumer Discretionary
MHP	McGraw-Hill	Consumer Discretionary
MO	Altria Group Inc	Consumer Staples
MOLX	Molex Inc	Information Technology
MTB	M&T Bank Corp	Financials
MYL	Mylan Inc	Health Care
NBL	Noble Energy Inc	Energy
NWL	Newell Rubbermaid Co.	Consumer Discretionary
ODP	Office Depot	Consumer Discretionary
PSA	Public Storage	Financials
SII	Smith International	Energy
SRE	Sempra Energy	Utilities
STJ	St Jude Medical	Health Care
STT	State Street Corp	Financials
SYY	Sysco Corp	Consumer Staples
TGT	Target Corp.	Consumer Discretionary
THC	Tenet Healthcare Corp.	Health Care
VLO	Valero Energy	Energy
VNO	Vornado Realty Trust	Financials
WDC	Western Digital	Information Technology
WMB	Williams Cos.	Energy
WPI	Watson Pharmaceuticals	Health Care
X	United States Steel Corp.	Materials
YHOO	Yahoo Inc	Information Technology

Table A.2: Summary Statistics for 70 City-Retailers, 156 Weeks

		Mean	Median	StDev	Obs
G MILLS CHEERIOS BOX 15OZ	Mkt share	0.026	0.018	0.024	10920
	Price	3.466	3.416	0.707	10920
	Nbr price	3.533	3.580	0.571	10920
	Promo	0.107	0.000	0.158	10920
KELLOGG FROSTED FLAKES BOX 20OZ	Mkt share	0.020	0.012	0.025	10920
	Price	2.854	2.744	0.747	10920
	Nbr price	2.577	2.561	0.569	10920
	Promo	0.110	0.012	0.149	10920
POST HNY BNCHS OATS REG BOX 16OZ	Mkt share	0.015	0.010	0.015	10920
	Price	2.956	2.892	0.695	10920
	Nbr price	2.816	2.880	0.542	10920
	Promo	0.098	0.000	0.144	10920
QUAKER LIFE REGULAR BOX 21OZ	Mkt share	0.012	0.009	0.010	10920
	Price	2.839	2.804	0.585	10920
	Nbr price	3.743	3.663	0.832	10920
	Promo	0.036	0.000	0.099	10920
STR BDS RAISIN BRAN BOX 20OZ	Mkt share	0.013	0.010	0.010	10920
	Price	1.798	1.802	0.383	10920
	Nbr price	2.466	2.430	0.517	10920
	Promo	0.033	0.000	0.086	10920
All 150 Cereals	Price	2.930	2.906	0.300	10920
Wage		832.5	812.5	125.4	456

REFERENCES

ALTONJI, J. G. and R. L. MATZKIN (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica,* 73, 1053-1102.

BARTLE, R. (1966): *The Elements of Integration.* New York: Wiley.

BENKARD, C. L. and S. BERRY (2006): "On the Nonparametric Identification of Nonlinear Simultaneous Equations Models: Comment on Brown (1983) and Roehrig (1988)," *Econometrica,* 74, 1429-1440.

BERRY, S., J. LEVINSOHN, and A. PAKES (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

BERRY, S. and P. HAILE (2010): "Identification in Differentiated Products Markets Using Market Level Data," *Cowles Foundation Discussion Papers* 1744, Yale University.

BOENTE, G. and R. FRAIMAN (1991): "Strong Uniform Convergence Rates for Some Robust Equivariant Nonparametric Regression Estimates for Mixing Processes," *International Statistical Review,* 59, 355-372.

CAI, Z. (2002): "Regression Quantiles for Time Series," *Econometric Theory,* 18, 169-192.

CHEN, X. and H. WHITE (1998): "Central Limit Theorem and Functional Central Limit Theorems for Hilbert-Valued Dependent Heterogeneous Arrays with Applications," *Econometric Theory,* 14, 260-284.

CHESHER, A. (2003): "Identification in Nonseparable Models," *Econometrica,* 71, 1405-1441.

CORBAE, D., M. B. STINCHCOMBE, and J. ZEMAN (2009): *An Introduction to Mathematical Analysis for Economic Theory and Econometrics.* Princeton University Press.

DAWID, A. D. (1979): "Conditional Independence in Statistical Theory," *Journal of the Royal Statistical Society, Series B,* 41, 1-31.

DETTE, H., N., NEUMEYER, and K. F. PILZ (2006): "A Simple Nonparametric Estimator of a Strictly Monotone Regression Function," *Bernoulli,* 12, 469-490.

DUDLEY, R. M. (2002): *Real Analysis and Probability.* Cambridge University Press.

EVDOKIMOV, K. (2009): "Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity," *Working Paper,* Dept. of Economics, Yale University.

FAMA, E. F. and K. T. FRENCH (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

FERSON, W. E. and C. R. HARVEY (1991): "The Variation of Economic Risk Premiums," *Journal of Political Economy,* 99, 285–315.

FERSON, W. E. and C. R. HARVEY (1993): "The Risk and Predictability of International Equity Returns," *Review of Financial Studies,* 6, 527–566.

GHYSELS, E. (1998): "On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?" *The Journal of Finance*, 53, 549-573T

HALL, P. (1984): "Central Limit Theorem for Integrated Square Error Properties of Multivariate Nonparametric Density Estimators," *Journal of Multivariate Analysis,* 14, 1-16.

HANSEN, B. E. (2008): "Uniform Convergence Rates for Kernel Estimation with Dependent Data," *Econometric Theory,* 24, 726-748.

HARVEY, C. R. (1989): "Time-varying Conditional Covariances in Tests of Asset Pricing Models," *Journal of Financial Economics,* 24, 289–317.

HOCHBERG, Y. (1988): "A Sharper Bonferroni Procedure for Multiple Tests of Hypotheses," *Biometrika,* 75, 800-802.

HODERLEIN, S. (2005): "Nonparametric Demand Systems, Instrumental Variables and a Heterogeneous Population," *Working Paper*, Dept. of Economics, Brown University.

HODERLEIN, S. and E. MAMMEN (2007): "Identification of Marginal Effects in Nonseparable Models without Monotonicity," *Econometrica,* 75, 1513-1518.

HODERLEIN, S., L. SU, and H. WHITE (2010): "Testing Monotonicity in Unobservables with Panel Data using Control Variables", *Working Paper*, Dept. of Economics, UCSD.

HODERLEIN, S. and H. WHITE. (2009): "Nonparametric Identification in Nonseparable Panel Data Models with Generalized Fixed Effects," *Working Paper*, Dept. of Economics, Brown University.

IMBENS, G. W. and W. K. NEWEY (2009): "Identification and Estimation of Triangular Simultaneous Equations Models without Additivity," *Econometrica,77*, 1481-1512.

JAGANNATHAN, R. and Z. WANG (1996), "The Conditional CAPM and the Cross-section of Expected Returns," *Journal of Finance,* 51, 3-53.

KOMUNJER, I. and A. SANTOS (2010): "Semiparametric Estimation of Nonseparable Models: a Minimum Distance from Independence Approach," forthcoming in *Econometrics Journal*.

MASRY, E. (1996): "Multivariate Local Polynomial Regression for Time series: Uniform Strong Consistency Rates," *Journal of Time Series Analysis* 17, 571-599.

MATZKIN, R. L. (2003): "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica,* 71, 1339-1375.

MATZKIN, R. L. (2007): "Heterogeneous Choice," in *Advances in Economics and Econometrics, Theory and Applications*, *Ninth World Congress of the Econometric Society*, edited by R. Blundell, W. Newey, and T. Persson, Cambridge University Press.

MEGERDICHIAN, A.(2009): "Identification of Price Effects in Discrete Choice Models of Demand for Differentiated Products," UC San Diego, PhD Dissertation Chapter.

ROEHRIG, C. S. (1988): "Conditions for Identification in Nonparametric and Parametric Models," *Econometrica*, 56, 433-447.

SERFLING, R. J. (1980): *Approximation Theorems of Mathematical Statistics.* New York: John Wiley & Sons.

STIGLER, G. and G. BECKER (1977): "De Gustibus Non Est Disputandum," *American Economic Review*, 67, 76–90.

STINCHCOMBE, M. B. (2010): Personal communication.

WHITE, H. (2001): *Asymptotic Theory for Econometricians.* San Diego: Academic Press.

YU, K. and M. C. JONES (1998): "Local Linear Quantile Regression," *Journal of American Statistical Association,* 93*, 228-237.*