Testing monotonicity in unobservables with panel data

Authors: Liangjun Su, Stefan Hoderlein, Halbert White

Boston College Working Papers in Economics, 2016

Testing Monotonicity in Unobservables with Panel Data*

Liangjun Su
School of Economics
Singapore Management University

Stefan Hoderlein
Department of Economics
Boston College

Halbert White
Department of Economics
University of California, San Diego

April 15, 2013

Abstract

Monotonicity in a scalar unobservable is a crucial identifying assumption for an important class of nonparametric structural models accommodating unobserved heterogeneity. Tests for this monotonicity have previously been unavailable. This paper proposes and analyzes 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 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, unobserved heterogeneity

JEL Classification: C12, C14, C33.

*We sincerely thank the editor, Cheng Hsiao, an associate editor and two anonymous referees for their many insightful comments and suggestions that lead to a substantial improvement of the presentation. 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 Ph.D. dissertation, which Aren has used to implement our test. We are also indebted to Alan Bester, Xiaohong Chen, Kirill Evdokimov, Maxwell Stinchcombe, Quang Vuong, and the participants of the Penn State CAPCP Conference on Auctions and Procurement, the Brown University econometrics seminar, the Xiamen University econometrics seminar, the 2010 Cowles Foundation econometrics conference, and the 7th Symposium on Econometric Theory and Application (SETA 2011) in Melbourne, who provided helpful suggestions and discussion.

1 Introduction

Suppose an observable scalar Y is structurally generated as

$$Y = g(X, A), \tag{1.1}$$

where g is an unknown function, X is an observable $d \times 1$ vector, and A is an unobservable attribute vector. An important example occurs when Y represents the quantity of a good demanded by a consumer, X represents income and prices, and A represents a consumer's fixed taste parameter, as in Stigler and Becker (1977). Alternatively, Y can represent the quantity produced by a firm, X cost and demand shifters, and A a firm's fixed technology parameter.

Across economic models, unobservables can enter in many different ways. As a consequence, g is generally not point identified. However, when A is a scalar and $g(x, \cdot)$ is strictly monotone for all $x \in \mathcal{X}$, the support of X ("monotonicity in a scalar unobservable" or just "scalar monotonicity"), g does become identified. This is an important consequence of the "structural function and distribution" framework considered by Matzkin (2003, 2007) and others (in our exposition, we follow in particular Altonji and Matzkin (2005, section 4), henceforth **AM**). Such a monotonicity assumption has played 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, because it allows one to link heterogeneity in unobservables to regression quantiles. Recent studies relying on monotonicity are those of Imbens and Newey (2009), Evdokimov (2010), and Komunjer and Santos (2010). Monotonicity has also been assumed in structural auction models to ensure a symmetric Bayesian Nash equilibrium strategy (e.g., Guerre et al., 2000) and to nonparametrically identify certain distributional structures with endogenous participation and unobserved heterogeneity (e.g., Guerre et al., 2009).

As Hoderlein (2011) notes, monotonicity is a strong assumption in general. Further, monotonicity is crucial in this context, as key identification results fail when scalar monotonicity is violated, leading to meaningless estimates and illegitimate inferences. It is thus important to have tests for this. To the best of our knowledge, no tests for monotonicity in scalar unobservables are currently available. Accordingly, our goal here is to propose and analyze some straightforward methods for testing scalar monotonicity. We emphasize even for this scalar case, a direct test for the scalar monotonicity in (1.1) seems impossible under the standard identification conditions detailed in Proposition A.1 in the appendix. Since the functional form in (1.1) is assumed to be unknown, it is not restrictive to assume that A is uniformly distributed on $\mathbb{I} \equiv [0, 1]$. For simplicity, we consider the classical strictly exogenous case where X is independent of A ($X \perp A$). Suppose that the conditional cumulative distribution function (CDF) $F(\cdot|x)$ of Y given $X = x$ is strictly monotone for each $x \in \mathcal{X}$. Let $q(x, \tau) \equiv F^{-1}(\tau|x)$ denote the τth conditional quantile function of Y given $X = x$. Then the observables Y and X have an equivalent quantile representation

$$Y = q(X, U) \tag{1.2}$$

where U is uniformly distributed on $\mathbb{I}$ and $U \perp X$. By construction, $q(x, \cdot)$ is strictly monotone for each $x \in \mathcal{X}$. As a result, the observations on (Y, X) generated from (1.1) are observationally equivalent to those generated from (1.2), and there is no way to test whether the structural function $g(x, \cdot)$ is strictly monotone for each $x \in \mathcal{X}$.

Although it may be possible to construct tests for monotonicity in the strictly exogenous case using a single cross section of observations, any such test must necessarily be indirect, since the conditional distribution function of Y given X generally yields a representation $Y = q(X, U)$, where U is independent of X and $q(X, \cdot)$ is strictly monotone. We emphasize that this representation has no necessary *structural* content unless the structural relation is indeed monotone in the scalar unobservable. This indirectness can have serious adverse consequences for the power of such tests against many alternatives. This motivates us to consider the test of scalar monotonicity by relying on exogenous variations coming from repeated observations of a single individual.

For clarity, and to maintain a manageable scope for the analysis here, we focus on the classical strictly exogenous case, where X is independent of A ($X \perp A$). In particular, we will consider structures monotonic in A_i with separable time-varying unobservable ε_{it},

$$Y_{it} = g(X_{it}, A_i) + \varepsilon_{it}, \tag{1.3}$$

as well as fully general nonseparable structures with time-varying unobservable ε_{it},

$$Y_{it} = g(X_{it}, \varepsilon_{it}, A_i), \tag{1.4}$$

where $i = 1, ..., N$, and $t = 1, ..., T$. Here, we use A to denote the time-invariant unobservable, emphasizing that the unobservables are fixed attributes, conforming with notation of Hoderlein (2011), Hoderlein and Mammen (2007), and Hoderlein and White (2009). Evdokimov (2010) considers the former structure, discussing its 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 nonseparable structures are quite general; their only significant vulnerabilities to misspecification are failures of monotonicity or exogeneity. As we shall argue, having many time periods enables us to test for monotonicity in the presence of additional unobservables. Intuitively, the exogenous variation provided by multiple time periods help us recover the scalar time-invariant unobservable A_i, which is generally needed in the presence of time-varying unobservable ε_{it} in (1.3) or (1.4). Due to the need for the recovery of A_i, our asymptotic theory requires that the number of time periods T pass to infinity at sufficiently fast rate. On the other hand, we will also remark that in the special case where ε_{it} is absent from (1.3), there is no need to recover A_i so that just two time periods suffices to proceed directly. In a companion paper (Hoderlein et al., 2012), we study the endogenous case, where a conditional form of exogeneity ($X \perp A \mid Z$, for given covariates Z) permits recovery of effects of interest in (1.1) without the need for repeated observations for a single individual, and with no added time varying unobservable ε_{it}. Beyond the hypothesis at hand, the two approaches have no overlap.

As mentioned, other than regularity conditions, the only other major structural assumption we rely on is exogeneity in the sense that we assume that $\varepsilon_{it} \perp A_i | X_{it}$, which is similar to Evdokimov (2010), and, in particular, that $A_i \perp X_{it}$. Two remarks are in order: First, we can relax the latter assumption. Our test uses the comparison between at least two time periods, say, t and $t+1$. We could replace the assumption that $A_i \perp X_{it}$ and $A_i \perp X_{i,t+1}$, by assuming that $A_i \perp X_{it}|X_{i,t-1},...,X_{i0}$ or $A_i \perp X_{i,t+1}|X_{i,t-1},...,X_{i0}$, as in Altonji and Matzkin (2005). This would allow for correlation between A_i and X_{it} that run through past levels of the process $\{X_{it}, t \geq 0\}$. Since this would amount to additional conditioning, for brevity of exposition we do not elaborate on this further. Second, in some applications monotonicity is accepted, and we may hence use the test statistic to assess other parts of the specification, in particular the exogeneity assumption. Alternatively, it may be used to test both properties, monotonicity and exogeneity, jointly. To draw unequivocal conclusions about whether monotonicity alone or exogeneity alone is violated, one may be willing to impose prior knowledge. This is of course common in all areas of Econometrics; indeed, any t-, or F-test in the linear model is only valid under the maintained assumptions of linearity and exogeneity - a rejection of a null could always also be due to misspecification of this assumption. In this paper, we will hence consider our test largely as specification test for scalar monotonicity under the maintained assumption of exogeneity, however, we will interpret the rejections we obtain in one application as, in least in parts, possibly generated by failure of exogeneity as well.[1]

The structure of this paper is as follows: Section 2 presents a monotonicity test for structures of the forms (1.1), (1.3) and (1.4) but with a focus on (1.3). We first introduce and analyze a monotonicity test for structures with time-varying unobservables of the forms (1.3). The test is fully nonparametric, but here we require T to be large, so as to average out the influence of the ε_{it}'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 $O(N^{-1/2})$. Since the test is not asymptotically pivotal, we also propose an effective nonparametric bootstrap method to obtain p-values for our test and justify its asymptotic validity. Interestingly, the test and bootstrap method that work for the "partially nonseparable" case also work for the "fully nonseparable" case in (1.4). We also propose a test for the structure in (1.1) and remark that in the absence of time-varying unobservable, a fully nonparametric test can be constructed 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 at the usual nonparametric rate.

Although it would be appealing to have a procedure that works with fixed T rather than large T, the presence of multiple unobservables A_i and ε_{it} permits recovery only of the *distributions* and not the actual values of the unobservables, as in Evdokimov (2010). These distributions cannot reliably be exploited to construct tests with power; for example, the leading identical distribution case obviously yields distribution-based tests with power equal to level. By taking T large, however, we can average out the ε_{it}'s, making possible recovery of the actual values of the A_i's. This yields tests with power generally. Even though this rules out a number of applications with very short T, there are of course numerous application areas where large T is common, for example, in industrial organization, marketing, finance, and consumer demand. And, as our simulation studies show, values for T realistic in practice deliver reliable inference.

Section 3 reports the results of some Monte Carlo experiments designed to study the level and power

[1] Alternatively, one may use a nonparametric test for exogeneity (not dependent on the monotonicity assumption), as in Blundell and Horowitz (2007), in a first stage to isolate the monotonicity hypothesis.

properties of the tests. We find that our tests perform reasonably well for $(N, T) = 20$, 40, and 60 for the structures in (1.3) and (1.4). Section 4 uses our tests to study asset returns and demand for ready-to-eat cereals, and Section 5 concludes. The Mathematical Appendix contains formal proofs of our results, together with supplementary results supporting the discussion of the main text. In particular, the appendix reviews and extends available results on representation with scalar unobservables, providing the necessary foundations for our tests.

2 Testing monotonicity in unobservable time-invariant attributes

In this section we first motivate our test statistic for the structure in (1.3), present a set of assumptions, and report the asymptotic distributions of our test statistic under the null hypothesis of monotonicity and a sequence of Pitman local alternatives. Then we propose a bootstrap method for our test and justify its asymptotic validity. Finally we discuss tests for monotonicity in (1.1) and (1.4).

2.1 Test statistic

For notational simplicity, we suppress the individual index and write $(Y_t, X_t, A, \varepsilon_t)$ for $(Y_{it}, X_{it}, A_i, \varepsilon_{it})$. We consider the following partially nonseparable structure

$$Y_t = g(X_t, A) + \varepsilon_t, \tag{2.1}$$

where $t = 1, ..., T$, A is an individual's time-invariant attribute, and ε_t is a time varying idiosyncratic error term. Because the structure is partly but not fully nonseparable in unobservables, we call it "partially nonseparable".

Evdokimov (2010, **E**) studies such a structure extensively. Beyond the already discussed examples in finance, demand or industrial organization, **E** (2010) gives many salient examples and provides identification and estimation results. An important further example arises in finance, where Y_t is the return of an asset in period t, 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. As **E** shows, for fixed $T \geq 2$, one can use deconvolution to extract the distribution of $M_t \equiv g(X_t, A)$ given X_t.

Let $F_t(\,\cdot \mid x)$ denote the conditional CDF of M_t given $X_t = x$. Exogeneity ($X_t \perp A$) and the time-invariance of A jointly ensure that F_t is time invariant. **AM** and Imbens and Newey (2009), among others, assume scalar monotonicity for all $x \in \mathcal{X}$ (scalar monotonicity *a.s.*). Under this monotonicity assumption, we have $M_t = g(X_t, A) = F^{-1}(A \mid X_t)$, $t = 1, 2, ...$ (see Proposition A.1), which implies our null hypothesis

$$H_0: \; F(M_t \mid X_t) = A \;\text{ for all } t = 1, 2, ... \;\; a.s. \tag{2.2}$$

We call (2.2) full identification *a.s.* When exogeneity or scalar monotonicity *a.s.* fails, we generally have

$$P[F_t(M_t \mid X_t) = F_s(M_s \mid X_s)] < 1 \text{ for some } t \neq s. \tag{2.3}$$

Proposition A.2 of the appendix formally states and proves (2.2) and its converse, with a brief discussion of the mild additional conditions required for the converse.

Without ε_t, $Y_t = M_t$ and we can compare $F_t(Y_t \mid X_t)$ to $F_s(Y_s \mid X_s)$ for $t \neq s$. In the presence of ε_t, ideally we would like to compare $F_t(M_t \mid X_t)$ to $F_s(M_s \mid X_s)$ for $t \neq s$; both equal A given identification. However, since ε_t is unobservable, so is $M_t = Y_t - \varepsilon_t$. Thus, identifying A itself and directly comparing $F_t(M_t \mid X_t)$ to $F_s(M_s \mid X_s)$ is not possible.

Nevertheless, applying **E**'s approach does permit a comparison of the conditional distributions of M_t given X_t. That is, we can compare F_t to F_s for $t \neq s$. But this comparison yields tests with power equal to level when $\{Y_t, X_t\}$ is identically distributed (ID), a leading case. Also, using **E**'s results for inference based on estimators of $F_1 - F_2$ (say) is hindered by the fact that so far there is no asymptotic distribution theory available for his estimators; only convergence rates are available. Another consideration is that **E**'s approach relies crucially on the additive separability of ε_t; presently, there are no methods analogous to **E**'s approach that would permit a treatment of the fully nonseparable case. Consequently, constructing a general monotonicity test with ε_t and fixed T is currently not a viable option.

On the other hand, straightforward specification testing is possible when T is large. For convenience, we assume that $\{X_t, \varepsilon_t\}$ is ID. Let non-negative weight functions w_τ, $\tau = 1, ..., \mathcal{T}$, be defined on $\mathcal{X}$. Given sufficient moments, we use w_τ to define $\tilde{Y}_\tau = \tilde{Y}_{\tau t} \equiv E[Y_t w_\tau(X_t) \mid A]$; the equality holds by ID. Then

$$\begin{aligned}
\tilde{Y}_\tau &= E\left[g(X_t, A) w_\tau(X_t) | A\right] + E\left[\varepsilon_t w_\tau(X_t) | A\right] \\
&= \int g(x, A) w_\tau(x)\ dF(x) + E\left[\varepsilon_t w_\tau(X_t)\right] \\
&\equiv \bar{g}_\tau(A) + \tilde{\varepsilon}_\tau, \qquad (2.4)
\end{aligned}$$

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

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, $\varepsilon_t \perp A \mid w_1(X_t)$ and $\varepsilon_t \perp A \mid w_2(X_t)$ are equivalent.

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

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

Thus, A is identified as

$$A = \bar{g}_\tau^{-1}(\tilde{Y}_\tau - \tilde{\varepsilon}_\tau) = \tilde{F}_\tau(\tilde{Y}_\tau). \qquad (2.5)$$

In the finance context, where 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. In view of the fact that (2.5) holds for all $\tau = 1, ..., \mathcal{T}$, it motivates a specification test based on

$$\tilde{H}_0 : \tilde{F}_\tau(\tilde{Y}_\tau) = \tilde{F}_\varsigma(\tilde{Y}_\varsigma) \text{ for all } (\tau, \varsigma) \text{ with } \tau \neq \varsigma.$$

When T and N are large, we can consistently estimate $\tilde{Y}_{\tau i}$ and $\tilde{F}_\tau$, $\tau = 1, 2, ...$, yielding $\hat{A}_{NT,\tau,i} \equiv \hat{F}_{NT,\tau}(\bar{Y}_{T,\tau,i})$, where

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

Under strict monotonicity, the estimators $\hat{A}_{NT,\tau,i}$ are consistent for A_i as $N, T \to \infty$; otherwise, they differ under suitably strong monotonicity failures. Proposition A.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 A is identified, regardless of the non-monotonicity of $g(x, \cdot)$. Identification of A is often of interest in its own right, for example in modeling asset returns.

Here, the exogeneity assumptions $X_t \perp A$ and $\varepsilon_t \perp A \mid w_\tau(X_t)$, $\tau = 1, 2, ...$, permit inference on monotonicity of $\bar{g}_\tau$. Further, as we discuss preceding Proposition A.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 $E[g(X_t, A)\ w_\tau(X_t) \mid A]$ or $E[\varepsilon_t\ w_\tau(X_t) \mid A]$, or both. When $E[\varepsilon_t\ w_\tau(X_t) \mid A]$ is non-constant in A, as generally holds when either $X_t \perp A$ or $\varepsilon_t \perp A \mid w_\tau(X_t)$ fail, it is generically non-monotonic. Non-monotonicity of $E[g(X_t, A)\ w_\tau(X_t) \mid A]$ can arise either from the non-monotonicity of g or from the failure of exogeneity, $X_t \perp A$. The appendix contains further discussion.

These statistics now permit specification tests based on the following test statistic

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

2.2 Assumptions

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 A.0 Y_{it} is structurally generated according to $Y_{it} = g(X_{it}, A_i) + \varepsilon_{it}$, where $X_{it} \perp A_i$ and $\varepsilon_{it} \perp A_i \mid w_\tau(X_{it})$, $\tau = 1, 2, ..., \mathcal{T}$.

Assumption A.1 (i) Let $Z_{it} \equiv (\varepsilon_{it}, X'_{it})'$ and $Z_i \equiv \{Z_{i1}, Z_{i2}, ...\}$. The sequence $\{(Z_i, A_i)\}$ 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 A.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 A.3 (i) For $\tau = 1, 2, ..., \mathcal{T}$, the CDF $\tilde{F}_\tau$ of $\tilde{Y}_{\tau,i} \equiv E[Y_{it} w_\tau(X_{it}) \mid A_i]$ 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 $\overline{g}_\tau(A_i) \equiv E[g(X_{it}, A_i)w_\tau(X_{it}) \mid A_i]$.

Assumption A.4 Let $\xi_i \equiv (\tilde{Y}_{1,i}, ..., \tilde{Y}_{\mathcal{T},i})'$ and $\psi(u, \xi_j) \equiv \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} [\mathbf{1}\{\tilde{Y}_{\tau,j} \leq u_\tau\} - \tilde{F}_\tau(u_\tau) - \mathbf{1}\{\tilde{Y}_{\varsigma,j} \leq u_\varsigma\} + \tilde{F}_\varsigma(u_\varsigma)]$ where $u = (u_1, ..., u_{\mathcal{T}})'$. Let $\tilde{\psi}(u, v) \equiv \int \psi(\xi, u)\psi(\xi, v)\tilde{F}(d\xi)$ where $\tilde{F}$ denotes the CDF of ξ_i. The non-zero eigenvalues λ_j, $j = 1, 2, ...$, for $\tilde{\psi}(u, v)$ satisfy $\sum_{j=1}^{\infty} |\lambda_j| < \infty$.

Assumption A.5 As $N \to \infty$, $T/N \to \infty$.

Assumption A.0 specify the data generating process. Given the exogeneity assumptions $X_{it} \perp A_i$ and $\varepsilon_{it} \perp A_i \mid w_\tau(X_{it})$, $\tau = 1, ..., \mathcal{T}$, strict monotonicity of g in its second argument implies $\tilde{H}_0$, as discussed above. A sufficient condition for these assumption is $(X_{it}, \varepsilon_{it}) \perp A_i$, which strengthens the contemporaneous uncorrelatedness requirement in classical random effects panel data models. Assumption A.1 rules out cross-section dependence across individuals and nonstationarity across time. We can relax strict stationarity at the cost of more complicated notation. Assumption A.2 imposes some moment conditions. Assumption A.3(i) is weak. Assumption A.4 is used to establish the asymptotic distribution of a certain degenerate second-order U-statistic. The summability condition is required in order to apply the result in Chen and White (1998). Assumption A.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 not much different from N.

2.3 Asymptotic distributions

Define the bias term

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

where $\tilde{F}_{T\tau}$ denotes the CDF of $\tilde{Y}_{\tau,i}$ for $\tau = 1, 2, ..., \mathcal{T}$. We can now describe the asymptotic distribution of $\hat{D}_{NT}$ under $\tilde{H}_0$ as $N \to \infty$.

Theorem 2.1 *Suppose Assumptions A.0-A.5 hold. Then under* $\tilde{H}_0 : \tilde{F}_\tau(\tilde{Y}_{\tau,i}) = \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i})$ *for* $\tau, \varsigma = 1, 2, ..., \mathcal{T}$, $\hat{D}_{NT} - B_{NT} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j (\mathcal{Z}_j^2 - 1)$, *where* $\{\mathcal{Z}_j\}$ *is a sequence of IID* $N(0, 1)$ *random variables, and* $\{\lambda_j\}$ *is as defined in Assumption A.4.*

Remark 1. The proof shows that $\hat{D}_{NT}$ is asymptotically equivalent to an infeasible test statistic ($\tilde{D}_{NT}$) based on the unobservable $\tilde{Y}_{\tau,i}$'s. After centering with D_{NT}, $\tilde{D}_{NT}$ can be written as a second-order degenerate U-statistic. That is, $\tilde{D}_{NT} - B_{NT} = \bar{\mathcal{H}}_n + o_P(1)$, where

$$\bar{\mathcal{H}}_N \equiv \frac{2}{N} \sum_{1 \leq i < j \leq N} \tilde{\psi}(\xi_i, \xi_j)$$

and $\tilde{\psi}(u, v) \equiv \int \psi(\xi, u)\psi(\xi, v)\tilde{F}(d\xi)$. Note that $\tilde{\psi}$ is a symmetric function such that $E\left[\tilde{\psi}(\xi_i, v)\right] = 0$ and $E[\tilde{\psi}(\xi_1, \xi_2)^2] < \infty$. Let Ξ denote the support of ξ_i and $L_2(\Xi, \tilde{F})$ the space of all square-integrable functions on Ξ with respect to $\tilde{F}$. Define $T_{\tilde{\psi}} : L_2(\Xi, \tilde{F}) \to L_2(\Xi, \tilde{F})$ as $T_{\tilde{\psi}}\phi(u) \equiv E\left[\tilde{\psi}(\xi_1, u)\phi(\xi_1)\right]$ for all $\phi \in L_2(\Xi, \tilde{F})$. Then $T_{\tilde{\psi}}$ is a compact self-adjoint linear operator with eigenvalues $\{\lambda_j\}$ and eigenfunctions $\{\varphi_j\}$ satisfying $\int \tilde{\psi}(u, v)\varphi_j(v)\, d\tilde{F}(v) = \lambda_j\varphi_j(u)$, $\int \varphi_j(v)\, d\tilde{F}(v) = 0$, and $\int \varphi_j(v)\varphi_l(v)\, d\tilde{F}(v) = \delta_{jl}$, where $\delta_{jl} = \mathbf{1}\{j = l\}$. We can represent the kernel $\tilde{\psi}$ as

$$\tilde{\psi}(u, v) = \sum_{j=1}^{\infty} \lambda_j\varphi_j(u)\varphi_j(v) \text{ for all } u, v \in \Xi,$$

where the convergence of the infinite series has to be understood in the L_2-sense, that is,

$$E\left[\tilde{\psi}(\xi_1, \xi_2) - \sum_{l=1}^{L} \lambda_l\varphi_l(\xi_1)\varphi_l(\xi_2)\right]^2 \to 0 \text{ as } L \to \infty.$$

This ensures that $\bar{\mathcal{H}}_N$ can be approximated by $\bar{\mathcal{H}}_N^{(L)}$, which denotes the U-statistic based on the underlying sample and the kernel $\tilde{\psi}^{(L)}(\xi_1, \xi_2) = \sum_{l=1}^{L} \lambda_l\varphi_l(\xi_1)\varphi_l(\xi_2)$. Noting that

$$\bar{\mathcal{H}}_N^{(L)} = \sum_{l=1}^{L} \lambda_L \left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \varphi_l(\xi_i)\right)^2 - \frac{1}{N} \sum_{i=1}^{N} \varphi_l(\xi_i)^2\right],$$

its limiting distribution can be obtained by an application of a central limit theorem (CLT) and a law of large numbers (LLN) to the inner sums: $\bar{\mathcal{H}}_N^{(L)} \xrightarrow{d} \sum_{l=1}^{\infty} \lambda_l\left(\mathcal{Z}_l^2 - 1\right)$. See Serfling (1980, pp. 194-199), Chen and White (1999, Proposition 5.2), and Leucht and Neumann (2011, Theorem 2.1) for more discussions.

Remark 2. To implement the test, we consistently estimate B_{NT} with

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

Under Lemma B.1 in the Appendix, it is straightforward to show that $\hat{B}_{NT} - B_{NT} = o_P(1)$. Then we have

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

As the limiting distribution depends on nuisance parameters $\{\lambda_j\}$ that in turn depend on the unknown distributions of $\tilde{Y}_{\tau,i}$ for $\tau = 1, 2, ..., \mathcal{T}$. To obtain the critical values or p-values for our asymptotic test, one can rely on either a resampling method or the above asymptotic distribution result. Leucht and Neumann (2011) study both methods for the case where the kernel function $\tilde{\psi}$ is known up to a finite dimensional parameter. In different but relevant contexts, Escanciano and Jacho-Chávez (2010) propose a numerical approximation of the critical values of Cramér-von Mises (CvM) tests by estimating eigenvalues for the associated kernel functions, whereas Chen and Fan (1999) propose to approximate the critical values of a test by either a conditional Monte-Carlo approach or a bootstrap method. In our case, note that the kernel function $\tilde{\psi}$ depends on the infinite dimensional parameter $\tilde{F}$ and the test statistic J_{NT} depends on both the cross section dimension N and the time dimension T. We are not sure how well the finite sample distribution of our test statistic can be approximated by its asymptotic distribution for different combinations of N and T. For this reason, we are in favor of the bootstrap approximation to the finite sample distribution of our test statistic. We will propose a bootstrap method to obtain the bootstrap p-values and justify its asymptotic validity.

Note that our J_{NT} test can detect any failure of full identification $a.s.$, whether due to failure of strict monotonicity, failure of exogeneity, or both. But the test itself does not reveal the source of failure, see the Introduction for more discussions.

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. The following theorem establishes the asymptotic local power of the J_{NT} test.

Theorem 2.2 *Suppose Assumptions A.0-A.5 hold. Suppose that* $\mu \equiv \lim_{N\to\infty} \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} E[\delta_{N,\tau,\varsigma}(\tilde{Y}_{\tau,i}, \tilde{Y}_{\varsigma,i})]^2 < \infty$. *Then under* $\tilde{H}_1(N^{-1/2})$ *where* $N^{-1/2}$ *denotes the rate at which the Pitman local alternatives converge to zero,* $J_{NT} \xrightarrow{d} \sum_{j=1}^{\infty} \lambda_j (\mathcal{Z}_j^2 - 1) + \mu$.

Remark 3. Theorem 2.2 shows that the J_{NT} test detects local alternatives converging to the null at rate $N^{-1/2}$. Intuitively, the time dimension mainly serves to average out the variation in ε_{it} to help recover the time-invariant unobservable A_i. The power of our test only comes from the variation of pairwise difference of $\tilde{F}_\tau(\tilde{Y}_{\tau,i})$, $\tau = 1, 2, ..., \mathcal{T}$. This explains why the rate at which the Pitman local alternatives converge to zero depends only the cross sectional dimension N.

The next theorem establishes the consistency of the test.

Theorem 2.3 *Suppose Assumptions A.0-A.5 hold. Then under* $\tilde{H}_1 \equiv \tilde{H}_1(1)$, $N^{-1}J_{NT} = \mu + o_P(1)$, *where* $\mu \equiv \sum_{\tau=1}^{\mathcal{T}-1} \sum_{\varsigma=\tau+1}^{\mathcal{T}} E[\tilde{F}_\tau(\tilde{Y}_{\tau,i}) - \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i})]^2$, *so that* $P(J_{NT} > c_N) \to 1$ *under* $\tilde{H}_1$ *for any nonstochastic sequence* $c_N = o(N)$.

Remark 4. Clearly, our J_{NT} test relies heavily on the assumption that $T/N \to \infty$ as $N \to \infty$. However, this is just a convenient assumption for ease of asymptotic analysis; we don't need it to literally hold in practice. As our simulations below show, the bootstrap version of our test gives good performance in empirically relevant cases where T can be either smaller or larger than N.

Remark 5. With large T and N, it is possible to propose alternative monotonicity tests for $g(x, \cdot)$. One approach is to obtain a consistent estimator $\hat{A}_{NT,i}$ of A_i for each i under the null and test monotonicity using nonparametric regression of Y_{it} on $(X_{it}, \hat{A}_{NT,i})$. To estimate A_i, let $\bar{Y}_{T,i} \equiv T^{-1} \sum_{t=1}^{T} Y_{it}$ and $\hat{F}_{NT}(y) \equiv N^{-1} \sum_{j=1}^{N} \mathbf{1}\{\bar{Y}_{T,j} \leq y\}$. Then $\hat{A}_{NT,i} \equiv \hat{F}_{NT}(\bar{Y}_{T,i})$ is consistent for A_i. We conjecture that one can extend existing nonparametric tests of regression monotonicity with a scalar observable regressor to multivariate regression, testing monotonicity in a single (estimated) regressor, $\hat{A}_{NT,i}$. In particular, the test of Ghosal et al. (2000) seems promising in this direction, but the asymptotic theory is nonstandard. Alternatively, with separability, we can consider estimating $M_{it} \equiv g_i(X_{it}) \equiv g(X_{it}, A_i)$ by time series regression of Y_{it} on X_{it} for each i. Let $\hat{M}_{it} \equiv \hat{g}_i(X_{it})$ denote the estimator. It seems feasible to base a test on the null restriction $F(M_{it}|X_{it}) = F(M_{is}|X_{is}) \ \forall \ t, s = 1, \ldots, T$, but with $\hat{M}_{it}$ used in place of M_{it} in constructing the test statistic. The rate at which such a test is able to detect local alternatives will certainly depend on N, T, and a properly chosen bandwidth sequence. Deriving the asymptotic distribution of such a test statistic is challenging and is left for future research.

2.4 A bootstrap version of the test

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

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

3. For $i = 1, ..., N$ and $t = 1, ..., T$, randomly draw $(\varepsilon_{it}^*, X_{it}^*)$ from $\{(\hat{\varepsilon}_{js}, X_{js}), j = 1, ..., N, s = 1, ..., T\}$ with replacement. Generate Y_{it}^* according to $Y_{it}^* = \hat{g}(X_{it}^*, A_i^*) + \varepsilon_{it}^*$, where A_i^*'s are IID $\mathbb{U}(0, 1)$ and are independent of $\{(\varepsilon_{it}^*, X_{it}^*)\}$.

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 make several remarks. First, we obtain the estimate $\hat{A}_{NT,i}$ of A_i without using any weight functions. Lemma B.2(i) in the appendix establishes the uniform convergence rate of $\hat{A}_{NT,i}$ to A_i under $\tilde{H}_0$. Second, we generate A_i^* independently of $(\varepsilon_{it}^*, X_{it}^*)$, which ensures the exogeneity condition automatically. Note also that we do not take into account the potential serial dependence structure of $(\varepsilon_{it}^*, X_{it}^*)$ along the time dimension as it does not play any role in the first order asymptotics of our test statistic. By construction, $(A_i^*, \varepsilon_{it}^*, X_{it}^*)$ is IID along the individual dimension for any fixed t. Third, 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 therein. **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:

Step 1. Let n be a large integer such that $n \to \infty$ as $N \to \infty$. For $i = 1, ..., N$, $t = 1, ..., T$, and $j = 1, ..., n$, compute the conventional local linear estimate $\tilde{g}(X_{it}, j/n)$ of $g(X_{it}, j/n)$ by using the product of Gaussian kernels (k) and bandwidth ($h = (h_1, ..., h_d, h_{d+1})'$) chosen according to Silverman's rule of thumb.

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

Step 3. Compute the estimate $\hat{g}(X_{it}, \hat{A}_{NT,i}) = \inf\{a : \hat{g}^{-1}(X_{it}, a) \geq \hat{A}_{NT,i}\}$.

Under conditions similar to those of **DNP**, we can show that $\hat{g}(x, a)$ is asymptotically equivalent to $\tilde{g}(x, a)$, although only the former is guaranteed to be monotone in its second argument. Lemma B.2(ii) in the appendix establishes the uniform consistency of $\hat{g}$ under $\tilde{H}_0$.

We also remark that the above bootstrap testing procedure is computationally expensive as one has to generate the bootstrap observations $\{Y_{it}^*\}$ through the constrained estimate $\hat{g}$ and the latter is obtained via the unconstrained estimate $\tilde{g}$. To generate each bootstrap sample for $\{Y_{it}^*\}$, $\tilde{g}$ has to be evaluated at $N \times T \times n$ points, which can be huge for moderate sizes of N, T, and n. But the advantage is that we can justify the asymptotic validity of this bootstrap procedure and demonstrate that it delivers very accurate levels for our test in finite samples for a variety of data generating processes.

To show that the bootstrap statistic J_{NT}^* can be used to approximate the asymptotic null distribution of J_{NT}, we follow Li et al. (2003), Su and White (2010), and Su and Ullah (2013) and rely on the notion of *convergence in distribution in probability*, which generalizes the usual convergence in distribution to allow for conditional (random) distribution functions. As Li et al. (2003) remark, one can also describe the weak convergence in probability of the bootstrap test statistic using the dual bounded Lipschitz metric on probability measures as in Giné and Zinn (1990, Section 3), but their definition is easier to understand. The following theorem establishes the asymptotic validity of the above bootstrap procedure.

Theorem 2.4 *Suppose Assumptions A.0-A5 hold. Suppose that either the support of* X_{it} *or the supports of* w_τ, $\tau = 1, 2, ..., \tau$ *are compact. Suppose that* $N^{-1/2}\sqrt{\log N} = o(h_{d+1})$, $b/h_{d+1} \to 0$, *and* $nb \to \infty$ *as* $N \to \infty$. *Then under* $\tilde{H}_0$, $J_{NT}^* \xrightarrow{d^*} \sum_{j=1}^{\infty} \lambda_j (\mathcal{Z}_j^2 - 1)$ *where* d^* *denotes convergence in distribution in probability,* $\{\mathcal{Z}_j\}$ *and* $\{\lambda_j\}$ *are as defined in Assumption A.4.*

Theorem 2.4 shows that the bootstrap provides an asymptotic valid approximation to the limit null distribution of J_{NT}. Because we only establish the consistency of $\hat{g}$ under the null, we need to impose the null hypothesis here. Similarly, Li et al. (2003) also imposes the null hypothesis in order to study their bootstrap validity. We conjecture that under the local alternatives, our bootstrap statistic also converges to $\sum_{j=1}^{\infty} \lambda_j (\mathcal{Z}_j^2 - 1)$ in distribution in probability, but the proof will become much more involved. Under the global alternative, it is hard to study the asymptotic distribution of J_{NT}^*. But one can show that J_{NT}^* is bounded in probability whereas J_{NT} is divergent to infinity. Consequently, our bootstrap test has power to detect all global alternatives.

2.5 Extensions and alternative specifications

In this subsection we discuss another two models. One is a special case of the model in (2.1) and the other extends the model in (2.1) from a partially nonseparable structure to a fully nonseparable structure. Again, for succinctness in notation, we continue to suppress the i subscript.

2.5.1 Panel structure without time-varying unobservables

If ε_t is absent from (2.1), we obtain the following model without the time-varying unobservable

$$Y_t = g(X_t, A), \qquad t = 1, ..., T. \tag{2.6}$$

In this case, using the notation in 2.1, $M_t = Y_t$ and (2.2) becomes

$$H_0: \; F(Y_t \mid X_t) = A \;\; a.s. \text{ for all } t = 1, 2, ..., T. \tag{2.7}$$

When exogeneity or scalar monotonicity $a.s.$ fails, we generally have

$$P[F_t(Y_t \mid X_t) = F_s(Y_s \mid X_s)] < 1 \text{ for some } t \neq s. \tag{2.8}$$

Here, $F_t(\,\cdot \mid x)$ denote the conditional CDF of Y_t given $X_t = x$ and it is time-invariant under the null and thus can be rewritten as $F(\,\cdot \mid x)$.

In section 2.1, due to the presence of ε_t, M_t is not observable and one has to average out the time-variation in ε_t in order to recover the time-invariant attribute A. For that reason, we need $T \to \infty$ sufficiently fast as $N \to \infty$. In stark contrast, here we directly observe Y_t and can construct a test statistic for (2.7) based on suitable estimates of F_t for T as small as 2.

Testing the null hypothesis (2.7) is similar to testing the equality of two regressions in the case of $T = 2$, i.e., $H_0^* : r_1(\mathcal{Z}_1) = r_2(\mathcal{Z}_2)$ $a.s.$, where $r_t(\mathcal{Z}_t) = E[\mathcal{Y}_t \mid \mathcal{Z}_t]$ for some dependent variable $\mathcal{Y}_t$ and conditioning variables $\mathcal{Z}_t$. A natural test statistic for this is $D_n = \sum_{i=1}^n [\hat{r}_1(\mathcal{Z}_{i1}) - \hat{r}_2(\mathcal{Z}_{i2})]^2$, where for $t = 1, 2$, $\hat{r}_t(\mathcal{Z}_{it})$ is a nonparametric estimate of $r_t(\mathcal{Z}_{it})$ based on observations $\{\mathcal{Y}_{it}, \mathcal{Z}_{it}\}_{i=1}^n$. If $\mathcal{Z}_1$ is a subvector of $\mathcal{Z}_2$, one has a statistic similar to that of Aït-Sahalia et al. (2001) for testing nonparametric significance. Alternatively, let $U_{it} = \mathcal{Y}_{it} - r_t(\mathcal{Z}_{it})$. Then testing H_0^* can also be regarded as testing for poolability of panel data as studied in Lavergne (2001): $r_t(\mathcal{Z}_{it}) = r(\mathcal{Z}_{it})$ $a.s.$ for some function $r(\cdot)$ and for $t = 1, 2$. Let $\varepsilon_{it} = \mathcal{Y}_{it} - r(\mathcal{Z}_{it})$ denote the restricted error term. As in Lavergne and Vuong (2000), Lavergne's test statistic builds on the observation that $E[\varepsilon_{it} E[\varepsilon_{it} \mid \mathcal{Z}_{it}] a(\mathcal{Z}_{it})]$ is zero under the null and strictly positive under the alternative for some nonnegative weight function $a(\cdot)$. Here, we adopt the first approach and measure the departure of $F_t(Y_t \mid X_t)$ from $F_s(Y_s \mid X_s)$ using

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

where $\hat{F}_{Nt}$'s are suitable estimates of F_t.

In the early version of the paper, we proposed to obtain the estimates $\hat{F}_{Nt}$'s by using the method of local polynomial regression. Under a set of regularity conditions, we showed that after appropriate normalization, D_{NT} is asymptotically normally distributed under the null hypothesis and a sequence of Pitman local alternatives. We also established the consistency of the test. To save space, we do not report the results here.

2.5.2 Fully Nonseparable Structures

We now consider a fully nonseparable structure of the form

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

This structure has been analyzed by Hoderlein and White (2012) under exogeneity assumptions alternative to the strict exogeneity considered here and without imposing monotonicity in unobservables. 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 $(X_t, \varepsilon_t) \perp A$, 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_τ, $\tau = 1, ..., \mathcal{T}$, be defined on $\mathcal{X}$, such that $E[w_\tau(X_t)] = 1$. Let $\tilde{Y}_\tau = \tilde{Y}_{\tau t} \equiv E[Y_t \ w_\tau(X_t) \mid A]$. Then for

$$\tilde{Y}_\tau = E[g(X_t, \varepsilon_t, A) \ w_\tau(X_t) \mid A] = \int g(x, e, A) \ w_\tau(x) \ dF(x, e) \equiv \bar{g}_\tau(A),$$

where $\tau = 1, ..., \mathcal{T}$, and the second equality in each line holds given $(X_t, \varepsilon_t) \perp A$.

The development of the previous section applies immediately, with the obvious modifications, so that $\tilde{F}_\tau(\tilde{Y}_\tau) = A$ for all $\tau = 1, ..., \mathcal{T}$ given strict monotonicity. Thus, we again test

$$\tilde{H}_0 : \tilde{F}_\tau(\tilde{Y}_\tau) = \tilde{F}_\varsigma(\tilde{Y}_\varsigma) \text{ for all } (\tau, \varsigma) \text{ with } \tau \neq \varsigma.$$

The statistics and tests are identical. Propositions A.3 and A.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. The bootstrap method in Section 2.4 also works here provided $(X_t, \varepsilon_t) \perp A$. To see why, letting $\bar{g}(X_t, A) \equiv E(Y_t | X_t, A) = E[g(X_t, \varepsilon_t, A) | X_t, A]$, we can write $Y_t = \bar{g}(X_t, A) + \bar{\varepsilon}_t$, where $\bar{\varepsilon}_t \equiv Y_t - \bar{g}(X_t, A)$ 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_t using estimates of $\bar{g}$ for the fully nonseparable case.

To close this subsection, we briefly discuss identification. If indeed $g(x, e, \cdot)$ is strictly monotone and $(X_t, \varepsilon_t) \perp A$, then, as we have just seen, A is identified as, e.g., $A = \tilde{F}(\tilde{Y})$, with $\tilde{Y} = \tilde{Y}_t \equiv E(Y_t \mid A)$ and $\tilde{F}$ the CDF of $\tilde{Y}$. Thus, A can be consistently estimated when $T \to \infty$; in this sense, A is known asymptotically. One can then identify g using the results of Appendix A, treating X_t and A as the observables, with ε_t the sole scalar unobservable. Specifically, with $g(x, \cdot, a)$ strictly monotone and $(X_t, A) \perp \varepsilon_t$, Proposition A.1 identifies $g(x, \cdot, a)$ 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.

3 Monte Carlo simulations

In this section we conduct Monte Carlo experiments to evaluate the finite sample performance of our tests. We consider both the partially and fully nonseparable cases in the presence of time-varying unobservables.

3.1 Data generating processes

We first consider the following six data generating processes (DGPs):
DGP 1. $Y_{it} = 1 + X_{it} + A_i + \varepsilon_{it}$,
DGP 2. $Y_{it} = 1 + X_{it,1} - X_{it,2}^2 + A_i + \varepsilon_{it}$,
DGP 3. $Y_{it} = 1 + X_{it,1} - X_{it,2}^2 + A_i + \sqrt{0.1 + 0.2X_{it,1}^2 + 0.2X_{it,2}^2}\varepsilon_{it}$,
DGP 4. $Y_{it} = 1 + X_{it} + (1 + \delta X_{it}) A_i + \varepsilon_{it}$,
DGP 5. $Y_{it} = 1 + (1 + \delta A_i) X_{it,1} - (1 + \delta A_i) X_{it,2}^2 + A_i + \varepsilon_{it}$,
DGP 6. $Y_{it} = 1 + (1 + \delta A_i) X_{it,1} - (1 + \delta A_i) X_{it,2}^2 + A_i + \sqrt{0.1 + 0.2X_{it,1}^2 + 0.2X_{it,2}^2}\varepsilon_{it}$,
where $i = 1, ..., N$, $t = 1, ..., T$, A_i is IID $\mathbb{U}(0,1)$, X_{it} is IID $N(0,1)$ and independent of A_j for each i, j, t in DGPs 1 and 4 and similarly for X_{1it} and X_{2it} in DGPs 2-3 and 5-6 where X_{1it} and X_{2it} are also mutually independent, ε_{it} is IID $N(0,1)$ across both i and t in DGPs 1 and 4, and an AR(1) process ($\varepsilon_{it} = 0.5\varepsilon_{it-1} + \epsilon_{it}$) in DGPs 2-3 and 5-6 where ϵ_{it} is IID $N(0,1)$ across both i and t. In addition, ε_{it} is independent of (X_{js}, A_j) DGPs 1 and 4 and (X_{1js}, X_{2js}, A_j) in DGPs 2-3 and 5-6) for all i, t, j, s. For DGPs 2-3 and 5-6, we write $X_{it} = (X_{it,1}, X_{it,2})'$.

We make several remarks on the DGPs. First, we use δ to control the degree of violation of monotonicity in A_i in DGPs 4-6. When $\delta = 0$, these DGPs reduce to DGPs 1-3, respectively. In the simulation below we simply set $\delta = 1$ in these DGPs where the non-monotoncity of the structural function in A_i is introduced mainly through the interaction between A_i and X_{it} (or functions of X_{it}). Apparently, we use DGPs 1-3 and 4-6 to study the finite sample level and power properties of our test, respectively. Second, the structures in DGPs 1, 2, 4, and 5 are partially nonseparable so that it is possible to write $Y_{it} = g(X_{it}, A_i) + \varepsilon_{it}$ for some measurable function g, whereas those in DGPs 3 and 6 are fully nonseparable: $Y_{it} = g(X_{it}, \varepsilon_{it}, A_i)$. Third, X_{it} in DGPs 1 and 4 contain only one regressor whereas X_{it} in DGPs 2-3 and 5-6 contain two regressors. Fourth, DGPs 1-3 specify linear or nonlinear models with homoskedastic or conditional heteroskedastic errors that are typically used in practice. DGPs 2 and 4 allow for serial correlation in the time varying unobservable (error term), whereas DGPs 3 and 6 allow for both serial correlation and conditional heteroskedasticity.

To examine whether our test has power against endogeneity, we consider the following two DGPs:
DGP 7. $Y_{it} = 1 + X_{it} + A_i + \varepsilon_{it}$,
DGP 8. $Y_{it} = 1 + X_{it} + (1 + X_{it}) A_i + \sqrt{0.1 + 0.5X_{it}^2}\varepsilon_{it}$,
where for $i = 1, ..., N$, $t = 1, ..., T$, ε_{it} is generated as in DGP 2, A_i is IID $\mathbb{U}(0,1)$, $X_{it} = -A_i^2 + 0.5\eta_{it}$, and η_{it} is IID $N(0,1)$ and mutually independent of A_i and ε_{it}. Apparently, exogeneity is not satisfied in

DGP 7 and neither exogeneity nor monotonicity is satisfied in DGP 8.

3.2 Implementation

To construct our test statistic, we need to choose the weight functions $w_\tau(\cdot)$, $\tau = 1, 2, ..., \mathcal{T}$. When the dimension d of X_{it} is one, we generate w_τ simply by evenly partitioning the support $\mathcal{X}$ of X_{it} into $\mathcal{T}$ parts. Specifically, for fixed $\mathcal{T}$, let $\tilde{q}_0 = -\infty$ and $\tilde{q}_{\mathcal{T}} = \infty$, 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 $1 \leq \tau \leq \mathcal{T} - 1$.[2] 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 A.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. When $d > 1$, there are various flexible ways to define the weight function. In this paper, we perform the weighting as follows: we first calculate $(X_{it,k} - \bar{X}_k)^2$ for the k-th regressor where $\bar{X}_k$ denote the sample mean of $\{X_{it,k}\}$ for $k = 1, 2, ..., d$, then sum them over k to get a one dimensional object $\sum_{k=1}^d (X_{it,k} - \bar{X}_k)^2$; then we use the equal-quantile-partition weights based on this sum. The idea is to ensure $\bar{Y}_{T,\tau,i} \equiv T^{-1} \sum_{t=1}^T Y_{it} w_\tau(X_{it})$ has roughly equal number of effective observations across τ.

To conduct the bootstrap test, we need to choose the kernel function, the bandwidth sequence $h = (h_1, ..., h_d, h_{d+1})'$, tuning parameters b and n. To obtain $\tilde{g}$ and its monotone version $\hat{g}$, we choose the kernel function as the product Gaussian kernel, and the bandwidth sequences according to Silverman's rule of thumb, namely, $h_l = 1.06 s_l n^{-1/((d+1)+4)}$, where s_l denotes the sample standard deviation of $X_{it,l}$ for $l = 1, ..., d$, and s_{d+1} is the sample standard deviation of the estimated regressor $\hat{A}_{NT,i}$. For b and n, we follow **DNP** and set $b = h_{d+1}^2$ and $n = 40$ to save computation time. Our simulation indicates that our test is robust to the choice of n. For example, $n = 25 \sim 100$ yields similar level and power properties.

Below we consider eight combinations of (N, T) by setting N, $T = 20$, 40, and 60. To see whether our test is robust to the choice of $\mathcal{T}$, we consider five values for $\mathcal{T}$, namely, 2, 4, 6, 8, and 10. For each case we use 250 replications and consider 200 bootstrap resamples in each replication.

3.3 Test results

Tables 1-2 report the empirical rejection frequencies for the bootstrap-based J_{NT} test at the 5% and 10% nominal levels for DGPs 1-3 and 4-8, respectively. Table 1 reports the level properties of our test for strict monotonicity when the exogeneity condition is satisfied. Table 2 reports the power of our test against non-monotonicity when the exogeneity condition is satisfied in DGPs 4-6, against endogeneity when monotonicity is satisfied in DGP 7, or against both endogeneity and non-monotonicity in DGP 8.

We summarize some important findings from Tables 1-2. First, the choice of $\mathcal{T}$ has some impact on both the level and power of our test but the pattern is not clear. For example, in terms of level,

[2] This specification creates no problem for the justification of the bootstrap asymptotic validity when X_{it} has compact support. If X_{it} has infinite support, in theory we need $\tilde{q}_0 = c_{1NT}$ and $\tilde{q}_{\mathcal{T}} = c_{2NT}$ so that c_{1NT} and c_{2NT} are either bounded or pass to the negative and positive infinity, respectively, at a slow rate. In the simulations and applications, we simply set $\tilde{q}_0 = -\infty$ and $\tilde{q}_{\mathcal{T}} = \infty$.

Table 1: Finite sample rejection frequency for DGPs 1-3

DGP	N	T	5% test					10% test				
			$\mathcal{T}$=2	$\mathcal{T}$=4	$\mathcal{T}$=6	$\mathcal{T}$=8	$\mathcal{T}$=10	$\mathcal{T}$=2	$\mathcal{T}$=4	$\mathcal{T}$=6	$\mathcal{T}$=8	$\mathcal{T}$=10
1	20	20	0.048	0.056	0.060	0.048	0.056	0.120	0.132	0.164	0.124	0.124
	20	40	0.040	0.032	0.032	0.052	0.056	0.072	0.096	0.104	0.132	0.100
	20	60	0.020	0.020	0.024	0.016	0.028	0.044	0.028	0.044	0.048	0.044
	40	20	0.040	0.028	0.028	0.028	0.028	0.100	0.076	0.064	0.076	0.072
	40	40	0.036	0.028	0.036	0.028	0.032	0.048	0.048	0.064	0.052	0.052
	40	60	0.024	0.016	0.032	0.024	0.024	0.032	0.040	0.056	0.040	0.032
	60	20	0.028	0.032	0.028	0.036	0.032	0.084	0.092	0.104	0.076	0.120
	60	40	0.032	0.028	0.032	0.036	0.036	0.076	0.036	0.040	0.044	0.068
2	20	20	0.024	0.028	0.044	0.060	0.084	0.052	0.084	0.108	0.144	0.120
	20	40	0.012	0.024	0.036	0.064	0.088	0.040	0.056	0.092	0.124	0.152
	20	60	0.028	0.072	0.076	0.084	0.076	0.088	0.104	0.144	0.144	0.152
	40	20	0.044	0.048	0.060	0.060	0.076	0.076	0.100	0.104	0.120	0.124
	40	40	0.024	0.020	0.040	0.060	0.072	0.060	0.052	0.092	0.120	0.136
	40	60	0.012	0.024	0.044	0.064	0.072	0.040	0.068	0.096	0.144	0.152
	60	20	0.028	0.036	0.028	0.024	0.048	0.060	0.048	0.060	0.100	0.108
	60	40	0.016	0.016	0.052	0.048	0.076	0.048	0.044	0.088	0.120	0.116
3	20	20	0.008	0.020	0.040	0.048	0.064	0.040	0.068	0.088	0.104	0.108
	20	40	0.012	0.024	0.032	0.056	0.060	0.048	0.080	0.068	0.124	0.128
	20	60	0.024	0.068	0.068	0.076	0.068	0.092	0.112	0.140	0.144	0.152
	40	20	0.020	0.044	0.052	0.056	0.060	0.060	0.068	0.080	0.092	0.124
	40	40	0.020	0.012	0.024	0.040	0.048	0.036	0.056	0.068	0.084	0.116
	40	60	0.012	0.032	0.068	0.080	0.076	0.036	0.060	0.100	0.152	0.148
	60	20	0.016	0.032	0.012	0.016	0.036	0.036	0.052	0.044	0.064	0.072
	60	40	0.020	0.016	0.036	0.040	0.048	0.036	0.060	0.072	0.080	0.096

Table 2: Finite sample rejection frequency for DGPs 4-8

DGP	N	T	5% test					10% test				
			$\mathcal{T}$=2	$\mathcal{T}$=4	$\mathcal{T}$=6	$\mathcal{T}$=8	$\mathcal{T}$=10	$\mathcal{T}$=2	$\mathcal{T}$=4	$\mathcal{T}$=6	$\mathcal{T}$=8	$\mathcal{T}$=10
4	20	20	0.192	0.228	0.100	0.064	0.028	0.276	0.368	0.220	0.116	0.064
	20	40	0.228	0.388	0.336	0.136	0.084	0.368	0.568	0.500	0.304	0.184
	20	60	0.312	0.604	0.584	0.524	0.336	0.452	0.764	0.728	0.640	0.532
	40	20	0.240	0.284	0.144	0.068	0.020	0.376	0.428	0.244	0.136	0.064
	40	40	0.408	0.700	0.544	0.380	0.220	0.564	0.804	0.704	0.544	0.332
	40	60	0.568	0.908	0.888	0.832	0.700	0.736	0.964	0.960	0.920	0.840
	60	20	0.304	0.424	0.216	0.060	0.024	0.460	0.552	0.340	0.168	0.072
	60	40	0.548	0.844	0.736	0.556	0.360	0.688	0.920	0.840	0.700	0.534
5	20	20	0.096	0.096	0.072	0.084	0.088	0.208	0.176	0.124	0.164	0.152
	20	40	0.184	0.148	0.172	0.172	0.152	0.304	0.252	0.276	0.288	0.268
	20	60	0.324	0.312	0.236	0.180	0.156	0.472	0.428	0.392	0.332	0.292
	40	20	0.108	0.112	0.096	0.100	0.092	0.232	0.208	0.168	0.192	0.180
	40	40	0.228	0.244	0.200	0.168	0.156	0.388	0.392	0.320	0.296	0.256
	40	60	0.432	0.372	0.312	0.260	0.208	0.596	0.556	0.452	0.368	0.324
	60	20	0.140	0.080	0.080	0.100	0.108	0.220	0.192	0.152	0.176	0.164
	60	40	0.300	0.308	0.268	0.192	0.192	0.460	0.444	0.400	0.316	0.284
6	20	20	0.112	0.120	0.140	0.104	0.104	0.252	0.220	0.208	0.200	0.156
	20	40	0.252	0.276	0.248	0.252	0.192	0.360	0.400	0.364	0.360	0.344
	20	60	0.456	0.480	0.356	0.284	0.228	0.580	0.604	0.532	0.484	0.404
	40	20	0.156	0.188	0.112	0.128	0.120	0.260	0.284	0.204	0.232	0.184
	40	40	0.300	0.424	0.312	0.252	0.232	0.476	0.620	0.488	0.424	0.336
	40	60	0.568	0.680	0.600	0.488	0.336	0.728	0.832	0.728	0.640	0.480
	60	20	0.172	0.184	0.136	0.124	0.096	0.288	0.288	0.236	0.204	0.180
	60	40	0.416	0.496	0.464	0.336	0.264	0.544	0.656	0.648	0.488	0.432
7	20	20	0.052	0.348	0.632	0.748	0.816	0.092	0.476	0.752	0.880	0.904
	20	40	0.016	0.248	0.452	0.688	0.776	0.080	0.348	0.608	0.780	0.888
	20	60	0.012	0.180	0.478	0.660	0.786	0.036	0.306	0.618	0.778	0.876
	40	20	0.060	0.584	0.876	0.936	0.964	0.140	0.744	0.936	0.968	0.992
	40	40	0.024	0.400	0.804	0.936	0.976	0.060	0.596	0.872	0.968	0.984
	40	60	0.024	0.240	0.704	0.904	0.952	0.024	0.368	0.816	0.960	0.984
	60	20	0.120	0.732	0.956	0.992	0.996	0.224	0.840	0.988	1.000	1.000
	60	40	0.032	0.416	0.856	0.968	0.992	0.064	0.544	0.920	0.976	0.992
8	20	20	0.132	0.232	0.368	0.528	0.556	0.208	0.348	0.496	0.660	0.692
	20	40	0.496	0.592	0.680	0.736	0.816	0.588	0.712	0.772	0.824	0.876
	20	60	0.628	0.820	0.884	0.912	0.940	0.756	0.892	0.924	0.944	0.984
	40	20	0.196	0.300	0.584	0.684	0.752	0.276	0.404	0.664	0.800	0.868
	40	40	0.640	0.776	0.840	0.900	0.924	0.704	0.820	0.912	0.924	0.952
	40	60	0.884	0.936	0.988	0.988	0.996	0.924	0.964	0.996	0.996	0.996
	60	20	0.236	0.324	0.628	0.860	0.904	0.328	0.440	0.716	0.908	0.948
	60	40	0.756	0.844	0.928	0.968	0.976	0.832	0.880	0.972	0.980	0.988

we find that a small value of $\mathcal{T}$ (e.g., $\mathcal{T} = 2$) tends to yield a moderately undersized test whereas a large value of $\mathcal{T}$ (e.g., $\mathcal{T} = 10$) tends to result in a slightly oversized test for some DGPs (e.g., DGP 2). In terms of power, for DGP 4 an intermediate value of $\mathcal{T}$ (e.g., $\mathcal{T} = 4$, 6) tends to yield greater power than a large or small value of $\mathcal{T}$, whereas in DGPs 7-8 a large value of $\mathcal{T}$ would deliver a greater power than a small value of $\mathcal{T}$. Secondly, overally speaking, the level of our test is satisfactory despite the undersized issue for some DGPs and some choices of $\mathcal{T}$. In particular, for DGP 1 the test tends to be moderately undersized for a variety of choices of $\mathcal{T}$ when T is large, giving a conservative test. Third, our test has power against non-monotoncity alone, endogeneity alone, or both non-monotonicity and endogeneity. The power usually increases as either N or T increases; exception occurs when only endogeneity is present in DGP 7. Fourth, noticeably our test tends to have a larger power in DGPs 7-8 when exogeneity is violated or both exogeneity and monotonicity is violated and than in DGPs 4-6 when only monotonicity is violated. Nevertheless, we have to admit that this is not necessarily the general phenomenon as the power of our test fully depends on the degree of violation of either monotonicity, or exogeneity, or both, and there does not exist any metric to measure the degree of violation for either monotonicity or exogeneity.

4 Two applications

In this section we apply the methods to 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 more structural, and exogeneity is viewed as implausible. Nevertheless, since the seminal work of Berry et al. (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, otherwise we interpret it as a joint specification test for both hypotheses.

4.1 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

$$Y_{it} = \alpha_i + \beta_i' X_t + \eta_{it}, \tag{4.1}$$

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. In the original Fama and French (1993) framework, small and high book-to-market equity are compensations for higher risk. In this paper, we follow Daniel and Titman (1997), and merely take these factors as primitives.

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

$$Y_{it} = \alpha_i + \beta_{it}' X_t + \eta_{it}. \tag{4.2}$$

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,

$$Y_{it} = g(X_t, \varepsilon_{it}, A_i), \tag{4.3}$$

where ε_{it} corresponds to $(\eta_{it}, \beta_{it}')'$ and A_i corresponds to α_i. Our theory allows, but does not require, X_t to also vary with i. The exogeneity condition is that $(X_t, \varepsilon_{it}) \perp A_i$. This is plausible if we think of A_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 A_i.

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 A_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.(4.2) 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.

4.1.1 Data

Our factor data come from French's webpage[3] 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 is available upon request. 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

[3]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 3: Summary statistics for the financial return data

Variables	Min	Max	Mean	Median	Std dev	IQR	Skew	Kurt
			Subperiod 1 ($N = 50$, $T = 472$)					
Firm's excess rate of return (Y)	-80.773	241.538	0.260	0.155	5.909	5.288	3.188	129.208
Excess market return ($RMRF$)	-13.740	9.180	0.056	0.160	2.444	2.775	-0.549	6.552
Small minus big (SMB)	-9.330	6.430	0.083	0.120	1.553	1.705	-0.585	8.743
High minus low (HML)	-6.830	9.700	0.110	0.140	1.512	1.500	0.525	7.897
			Subperiod 2 ($N = 50$, $T = 138$)					
Firm's excess rate of return (Y)	-48.407	110.506	0.219	0.057	7.201	6.568	0.752	16.833
Excess market return ($RMRF$)	-18.400	13.020	-0.030	0.040	3.895	4.170	-0.447	6.941
Small minus big (SMB)	-3.400	3.680	0.095	-0.040	1.331	1.620	0.056	3.128
High minus low (HML)	-6.850	7.630	0.036	-0.145	2.051	1.700	0.253	5.670

Note: All data are weekly, not annualized.

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. To ameliorate the problem of structural change due to the financial crisis of 2007–2008, we divide the whole period into two subperiods: 7/17/1998 - 8/7/2007 and 8/13/1997-3/26/2010. We choose August 7, 2007 as the separating point because the active phase of the crisis, which manifested as a liquidity crisis, can be dated from August 7, 2007, when BNP Paribas terminated withdrawals from three hedge funds citing "a complete evaporation of liquidity".

For each firm i in each subperiod, we calculate the 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. Table 3 reports some summary statistics on the dependent variable (Y_{it}) and three factors $RMRF_t$, SMB_t, and HML_t. Apparently the returns and factors behave quite differently over the two subperiods.

4.1.2 Test results

To apply our test procedure to the data described above, we follow the exact implementation procedure detailed in Section 3.2. The test statistic is computed just as in the simulations, following exactly the same steps there.

The results are summarized in Table 4. In all instances, we soundly reject the strict monotonicity hypothesis at 1% level. 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 underlying risks of a stock, including firm specific effects. Our procedure permits a more stringent test of this aspect of market efficiency.

Note, however, that even with the failure of monotonicity, useful information about risk premia may

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

Bootstrap replications	Subperiod	N	T	p-values				
				$\mathcal{T}=2$	$\mathcal{T}=4$	$\mathcal{T}=6$	$\mathcal{T}=8$	$\mathcal{T}=10$
500	1	50	472	0.002	< 0.002	< 0.002	< 0.002	< 0.002
500	2	50	138	< 0.002	< 0.002	< 0.002	< 0.002	< 0.002

still be recovered from nonparametric specifications of the sort used here. Although monotonicity failure rules out identifying alpha, the further exogeneity condition $X_t \perp (A_i, \varepsilon_{it})$ permits recovery of expected risk premia, such as $E[D_k g(x, \varepsilon_{it}, A_i)]$, 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).

4.2 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 exogeneity is as least as questionable as scalar monotonicity. In what follows, we look at five individual goods. For the purpose of testing, we ignore the additional restriction that they form a demand system, and always look at the binary decision to buy or not to buy a good, assuming separability in the utility of this choice from all other goods. This assumption ensures that the nonparametric aggregate relationship retains the monotonicity in unobservables, if the original binary decision had a monotonically (in applied work typically additive) unobserved factor. 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, even if we assume linearity of the individual binary decisions within the indicator.

4.2.1 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 $T = 156$. To reduce computational burden, we pick $T = 50$, which roughly corresponds to 2005, but have obtained the same results for different time periods. 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 $N = 50$ supermarket-city pairs. We analyze the top-selling product for each of the five manufacturers.

Table 5: p-values for endogeneity test - RTE cereal

Product	N	T	p-values				
			$\mathcal{T}=2$	$\mathcal{T}=4$	$\mathcal{T}=6$	$\mathcal{T}=8$	$\mathcal{T}=10$
G MILLS CHEERIOS 15OZ	50	50	< 0.002	< 0.002	< 0.002	< 0.002	< 0.002
KELLOGG FROSTED FLAKES 20OZ	50	50	< 0.002	< 0.002	< 0.002	< 0.002	< 0.002
POST HNY BNCHS OATS REG 16OZ	50	50	< 0.002	< 0.002	< 0.002	< 0.002	< 0.002
QUAKER LIFE REGULAR 21OZ	50	50	< 0.002	< 0.002	< 0.002	< 0.002	< 0.002
STR BDS RAISIN BRAN 20OZ	50	50	< 0.002	< 0.002	< 0.002	< 0.002	< 0.002

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

4.2.2 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; 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 using 500 bootstrap replications:

As is obvious from these results, exogeneity is widely rejected. For all products the p-values are virtually zero. Note again that the results are always for the binary buy - don't buy decision, where the structural relation retains monotonicity, if it is was present in the individual level specification, as is commonly assumed. We can safely conclude that the current exogenous monotonic specification is rejected. If we follow the IO literature, we conclude that endogeneity is indeed the issue the demand and IO literatures believe it to be. However, it may well be that there monotonicity in a scalar unobservable is questionable. The results do not change in any appreciable way if we include price of the closest neighbor in product characteristic space and the quantity-weighted average price of all 150 cereals as additional regressors. In summary, our test statistic illustrates that a simple, exogenous demand model with monotonicity in a scalar unobservable is not a good description of actual behavior, because it does not properly address confounding effects and the simultaneity structure typical in this literature.

5 Conclusion

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 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. 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 present analysis, we focus throughout on the strictly exogenous case. Allowing endogeneity (e.g., dependence between X_t and A) is an important extension, as this supports a wider scope for specification testing. In a companion paper (Hoderlein et al. (2012)), we allow endogeneity by imposing a conditional form of exogeneity, where, 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. In addition, we abstract from panel dynamics. It is interesting is to examine whether and how tests for scalar monotonicity can be conducted in dynamic panel structures. Finally, there is a considerable variety of opportunities for applying these tests and their further extensions.

Mathematical Appendix

A Representation and Identification with Scalar Unobservables

Here, we review and extend available results on representation with scalar unobservables, providing suitable foundations for our tests. 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 $\mathbb{I} \equiv [0,1]$.

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

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

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

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

$$\begin{aligned} F(y \mid x) &\equiv P[Y \leq y \mid X = x] = P[g(X, A) \leq y \mid X = x] \\ &= P[g(x, A) \leq y] = \int_0^1 \mathbf{1}\{g(x, a) \leq y\} \, da \\ &= \lambda\{g_x^{-1}(-\infty, y]\}, \end{aligned}$$

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

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

Now suppose that $g(x, \cdot)$ is not strictly increasing. First, suppose that $g(x, \cdot)$ is invertible, and also suppose $F(y \mid x) = g^{-1}(x, y)$ 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 $F(y \mid x) \neq g^{-1}(x, y)$ for some y.

Finally, if $g(x, \cdot)$ is not invertible, then $g^{-1}(x, \cdot)$ 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 $y \in \mathcal{Y}$. ■

Remark. When $g(x, \cdot)$ is invertible, $g^{-1}(x, \cdot)$ represents the inverse function such that $a = g^{-1}(x, y)$ if and only if $y = g(x, a)$. More generally, $g^{-1}(x, \cdot)$ represents the correspondence defined by $g_x^{-1}(-\infty, y]$, the preimage in $\mathbb{I}$ of the half-ray $(-\infty, y]$ under $g(x, \cdot)$. Also, we adopt the convention suggested by **AM** that if $g(x, \cdot)$ is strictly decreasing, then we replace $g(x, \cdot)$ with $-g(x, \cdot)$. The key property is thus that $g(x, \cdot)$ is strictly monotone. Let $a = F(y \mid x)$; if eq.(A.1) holds, then $F(\cdot \mid x)$ is invertible and g is identified as $g(x, a) = F^{-1}(a \mid x)$. 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.*

The conditions in the above proposition are simpler than those of **AM**'s theorem 4.1, as we consider only the exogenous case in this paper. Also, we show that strict monotonicity of $g(x, \cdot)$ is *necessary* for full identification, not just sufficient. In addition, as we showed in the early version of the paper, representing Y using a scalar A in Proposition A.1 is much less restrictive than it might seem.

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 $P[Y_t = h(A)] < 1$ for all measurable h, $t = 1, ..., T$. This rules out the trivial case in which $Y_1 = \cdots = Y_T$ *a.s.*

Proposition A.2 *Suppose that* $Y_t = g(X_t, A)$ *and* A *is a random scalar distributed as* $\mathbb{U}[0,1]$. *Let the* X_t*'s have common minimal support* $\mathcal{X}$.

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

(ii) *Suppose that* $\mathcal{X}$ *contains at least two points, that* $P_1 P_2 \cdots P_T \ll P_{1,...,T}$, *and that* $P[Y_t = h(A)] < 1$ *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. (i) follows from Proposition A.1. For (ii), we give the proof for $T = 2$ as the proof for $T > 2$ is similar.

$(ii.1)$ First suppose that strict monotonicity *a.s.* (i.e., (a)) holds, but (b) fails, so that $(X_1, X_2) \not\perp A$. Then $A = g^{-1}(X_1, Y_1) = g^{-1}(X_2, Y_2)$. Then for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $t = 1, 2$,

$$\begin{aligned} F_t(y \mid x) &\equiv P[Y_t \leq y \mid X_t = x] = P[g(X_t, A) \leq y \mid X_t = x] \\ &= \int_0^1 \mathbf{1}\{g(x,a) \leq y\} \, d\bar{F}_t(a \mid x) \\ &\equiv \mu_t[g_x^{-1}(-\infty, y] \mid x], \end{aligned} \tag{A.2}$$

where $\bar{F}_t(\cdot \mid x)$ denotes the conditional CDF of A given $X_t = x$. It follows that

$$\begin{aligned} F_1(Y_1 \mid X_1) &= \mu_1[(0, A] \mid X_1] = \int_0^A d\bar{F}_1(a \mid X_1), \\ F_2(Y_2 \mid X_2) &= \mu_2[(0, A] \mid X_2] = \int_0^A d\bar{F}_2(a \mid X_2). \end{aligned}$$

Letting $\bar{F}_{1,2}(a \mid x_1, x_2)$ define the conditional CDF of A given $X_1 = x_1, X_2 = x_2$, we have

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

where $F(x_1, x_2)$ denotes the CDF of X_1 and X_2. The desired result follows if the integral in the expression above is positive.

To simplify notation, write $\mu_A(x_1,x_2) \equiv \int_0^1 \mathbf{1}\{\mu_1[(0,a] \mid x_1] \neq \mu_2[(0,a] \mid x_2]\}\ d\bar{F}_{1,2}(a \mid x_1,x_2)$. Then

$$\begin{aligned}\int_{\mathcal{X}\times\mathcal{X}} \int_0^1 \mathbf{1}\{\mu_1[(0,a] \mid x_1] \neq \mu_2[(0,a] \mid x_2]\}\ d\bar{F}_{1,2}(a \mid x_1,x_2)\ dF(x_1,x_2) \\ = \int \mu_A(x_1,x_2)\ dP_{1,2}(x_1,x_2).\end{aligned}$$

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 $\mu_A(x_1,x_2)$ 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 $\mu_A(x_1,x_2) > 0$ 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 $\int \mu_A(x_1,x_2)\ dP_{1,2}(x_1,x_2) = 0$. Then by Bartle (1966, corollary 4.10), $\mu_A(x_1,x_2) = 0\ P_{1,2} - a.e.$, which further implies $\mu_1[(0,a] \mid x_1] = \mu_2[(0,a] \mid x_2]$ for almost all a, 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 $\mu_1[(0,a] \mid x_1] = \mu_2[(0,a] \mid x_2] = \mu_0[(0,a]]$, for almost all a, x_1, and x_2. If $\mu_0 = \lambda$, this is a contradiction. If $\mu_0 \neq \lambda$, a further monotone transformation of A can be applied without loss of generality to ensure $\mu_0 = \lambda$. But this is again a contradiction. Thus, $\int \mu_A(x_1,x_2)\ dP_{1,2}(x_1,x_2) > 0$.

($ii.2$) 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 $P[F_1(Y_1 \mid X_1) \neq F_2(Y_2 \mid X_2)] > 0$. By (A.2), we have

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

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

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

When $P_t[\mathcal{X}_0^c] > 0$, $P[C_t = A \mid X_t \in \mathcal{X}_0^c]$ is defined, and we have $P[C_t = A \mid X_t \in \mathcal{X}_0^c] = 1$.

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

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

Now

$$
\begin{aligned}
&P_1P_2[(C_1 = A) \cap (A \neq C_2) \cap (X_1 \in \mathcal{X}_0^c) \cap (X_2 \in \mathcal{X}_0)] \\
&= P_1[(C_1 = A) \cap (X_1 \in \mathcal{X}_0^c)]\ P_2[(A \neq C_2) \cap (X_2 \in \mathcal{X}_0)] \\
&= P_1[\mathcal{X}_0^c]\ P_2[\mathcal{X}_0]\ (1 - P_2[A = C_2 \mid X_2 \in \mathcal{X}_0]) > 0,
\end{aligned}
$$

as $P_1[\mathcal{X}_0^c] > 0$, $P_2[\mathcal{X}_0] > 0$, and $P[A = C_2 \mid X_2 \in \mathcal{X}_0] < 1$. Because $P_1P_2 \ll P_{1,2}$, it follows that $P[(C_1 = A) \cap (X_1 \in \mathcal{X}_0^c) \cap (A \neq C_2) \cap (X_2 \in \mathcal{X}_0)] > 0$. 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 a, x_1, and x_2, we have $c_1(x_1, a) = c_2(x_2, a)$. Because $\mathcal{X}_0 = \mathcal{X}$ contains at least two values, this can hold only if $c_1(x_1, a) = c_2(x_2, a) = c_0(a)$, say, for all $(x_1, x_2, a) \in \mathcal{X} \times \mathcal{X} \times \mathcal{A}$, $\mathcal{A} \equiv \text{supp}(A)$. This can hold only if: (i) $X_t \perp A$, $t = 1, 2$; and, because for each $x \in \mathcal{X}$, $g(x, \cdot)$ is not strictly monotone, (ii) $g(x, a) = g_0(a)$, say, for all $(x, a) \in \mathcal{X} \times \mathcal{A}$, i.e., $P[Y_t = g_0(A)] = 1$, $t = 1, 2$. 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. ■

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.

Proposition A.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 $E[g(X, a)] < \infty$ for all $a \in \mathbb{I}$. 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 $0 \leq a_1^* < a_2^* \leq 1$ such that*

$$
\begin{aligned}
&\int [g(x, a_2^*) - g(x, a_1^*)]\ w(x)\ \mathbf{1}\{x \in \mathcal{X}_w^*\}\ dF_X(x) \\
&\qquad \leq -\int [g(x, a_2^*) - g(x, a_1^*)]\ w(x)\ \mathbf{1}\{x \notin \mathcal{X}_w^*\}\ dF_X(x).
\end{aligned}
$$

(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 $0 \leq a_1^* <$*

$a_2^* \leq 1$ *such that*

$$\int [g(x, a_2^*) - g(x, a_1^*)] \ \tilde{w}(x) \ \mathbf{1}\{x \in \tilde{\mathcal{X}}_w^*\} \ dF_X(x)$$
$$\leq - \int [g(x, a_2^*) - g(x, a_1^*)] \ \tilde{w}(x) \ \mathbf{1}\{x \notin \tilde{\mathcal{X}}_w^*\} \ dF_X(x).$$

Proof. (i) Under the conditions given, $|\int g(x,a) \ w(x) \ dF_X(x)| < \infty$ for all $a \in \mathbb{I}$. If $g(X, \cdot)$ is strictly increasing $a.s.$, then for all $0 \leq a_1 < a_2 \leq 1$,

$$\bar{g}_w(a_2) - \bar{g}_w(a_1) = \int [g(x, a_2) - g(x, a_1)] \ w(x) \ dF_X(x) > 0,$$

where the inequality follows from corollary 4.10 of Bartle (1966) as $[g(x, a_2) - g(x, a_1)] \ w(x)$ 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 $0 \leq a_1^* < a_2^* \leq 1$,

$$\bar{g}_w(a_2^*) - \bar{g}_w(a_1^*) = \int [g(x, a_2^*) - g(x, a_1^*)] \ w(x) \ dF_X(x)$$
$$= \int [g(x, a_2^*) - g(x, a_1^*)] \ w(x) \ \mathbf{1}\{x \in \mathcal{X}_w^*\} \ dF_X(x)$$
$$+ \int [g(x, a_2^*) - g(x, a_1^*)] \ w(x) \ \mathbf{1}\{x \notin \mathcal{X}_w^*\} \ dF_X(x) \leq 0,$$

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 a_1^*, a_2^*, then for all $0 \leq a_1 < a_2 \leq 1$, $\bar{g}_w(a_2) - \bar{g}_w(a_1) > 0$, 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 $0 \leq a_1^* < a_2^* \leq 1$, 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). $\blacksquare$

For succinctness in what follows, we continue to suppress the i subscript and write $Y_t = g(X_t, A) + \varepsilon_t$. Assume that $\{X_t, \varepsilon_t\}$ is identically distributed (ID). Provided the necessary moments exist, we have

$$\tilde{Y}_\tau \equiv E(Y_t \ w_\tau(X_t) \mid A) = \bar{g}_\tau(A) + \bar{\varepsilon}_\tau(A),$$

where now $\bar{g}_\tau(A) \equiv E[g(X_t, A) \ w_\tau(X_t) \mid A]$ and $\bar{\varepsilon}_\tau(A) \equiv E[\varepsilon_t \ w_\tau(X_t) \mid A]$. 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 $X_t \perp A$ (see (2.4)); here $\bar{g}_\tau$ explicitly does not rely on this.

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

$$\bar{\gamma}_\tau(a) \equiv P[\bar{g}_\tau(A) \leq \bar{g}_\tau(a)], \quad a \in \mathbb{I}.$$

This quantifies the departure of $\bar{g}_\tau$ from monotonicity. When $\bar{g}_\tau$ is strictly monotone, $\bar{\gamma}_\tau(a) = a$. 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 $\lambda[a : \bar{\gamma}_1(a) = \bar{\gamma}_{\tau^*}(a)] < 1$, 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 Proposition A.2. Other examples exist, but these are exceptional; we conjecture that they are shy. Shyness is the function space analog of being a subset of a set of Lebesgue measure zero; see Corbae et al. (2009, pp. 545-547).

In part (ii), we drop the requirements that $X_t \perp A$ and $\varepsilon_t \perp A \mid w_\tau(X_t)$. Now we write

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

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

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

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 $X_t \perp A$ and $\varepsilon_t \perp A \mid w_\tau(X_t)$ 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 $X_t \perp A$ generally introduces non-monotonicity into $\bar{g}_\tau$. For example, take $w_\tau(X_t) = 1$, and suppose that $g(X_t, A) = X_t + A$ and that $X_t \not\perp A$ holds because $X_t = -A^2 + \eta_t$, where $\eta_t \perp A$. (This choice is illustrative, as the relation between X_t and A is generically non-monotonic.) Then

$$\begin{aligned} \bar{g}_\tau(A) &\equiv E[g(X_t, A)\; w_\tau(X_t) \mid A] = E(X_t + A \mid A) = E(-A^2 + \eta_t + A \mid A) \\ &= A(1 - A) + E(\eta_t). \end{aligned}$$

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

Proposition A.4 *Suppose* $Y_t = g(X_t, A) + \varepsilon_t$ *and* $\{X_t, \varepsilon_t\}$ *is ID. For* $\mathcal{T} \geq 2$, *let* $w_\tau : \mathcal{X} \to \mathbb{R}^+$, $\tau = 1, ..., \mathcal{T}$ *be as in Proposition A.3. Suppose that* $E[g(X_t, a)] < \infty$ *for each* $a \in \mathbb{I}$ *and that* $E(\varepsilon_t) < \infty$.

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

Proof. (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 $\lambda[a : \bar{\gamma}_1(a) = \bar{\gamma}_\tau(a)] = 1$ 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

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

Given $X_t \perp A$ and $\varepsilon_t \perp A \mid w_\tau(X_t)$, $\tilde{\varepsilon}_\tau$ is constant. It follows that $\tilde{F}_\tau(\bar{g}_\tau(a)+\tilde{\varepsilon}_\tau) = P[\bar{g}_\tau(A)+\tilde{\varepsilon}_\tau \leq \bar{g}_\tau(a)+\tilde{\varepsilon}_\tau] = \bar{\gamma}_\tau(a)$. Thus, for all τ,

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

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 $\lambda[a : \bar{\gamma}_1(a) = \bar{\gamma}_{\tau^*}(a)] < 1$. 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

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

But $\lambda[a : \bar{\gamma}_1(a) = \bar{\gamma}_{\tau^*}(a)] < 1$, so $1 - \lambda[a : \bar{\gamma}_1(a) = \bar{\gamma}_{\tau^*}(a)] > 0$, 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 $\tilde{\varepsilon}_\tau$. ■

B Proofs of the main results in Section 2

Recall $\tilde{F}_\tau$ and $\tilde{F}_{T\tau}$ denote the CDF of $\tilde{Y}_{\tau,i}$ and $\bar{Y}_{T,\tau,i}$, respectively; and $\tilde{f}_\tau$ denotes the PDF of $\tilde{Y}_{\tau,i}$ Let $\tilde{f}_{T\tau}$ denote the PDF of $\bar{Y}_{T,\tau,i}$.To prove the main results in section 2, we first prove the following lemma.

Lemma B.1 *Suppose Assumptions A.0, A.1(ii), A.2, and A.3(i) hold. Then for* $\tau = 1, 2, ..., \mathcal{T}$, *(i)* $E(\bar{Y}_{T,\tau,i} - \tilde{Y}_{\tau,i})^2 = O\left(T^{-1}\right)$; *(ii)* $E|\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i}) - \tilde{F}_\tau(\tilde{Y}_{\tau,i})| = O\left(T^{-1/2}\right)$; *and (iii)* $\sup_y |\tilde{F}_{T\tau}(y) - \tilde{F}_\tau(y)| = O\left(T^{-1/2}\right)$.

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

For (ii), we have

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

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 $c_{\tau,i}$ 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 A.3(i) and the Jensen inequality; and the last equality follows from (i).

Noting that (i) implies that $|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|\leq MT^{-1/2}$ for sufficiently large constant M with probability approaching 1 as $T\to\infty$. Then by Assumption A.3(i)

$$
\begin{aligned}
\sup_y|\tilde{F}_{T\tau}(y)-\tilde{F}_\tau(y)| &= \sup_y\left|E\left[\mathbf{1}\{\tilde{Y}_{\tau,i}\leq y\}-\mathbf{1}\{\bar{Y}_{T,\tau,i}\leq y\}\right]\right| \\
&\leq \sup_y E\left[\mathbf{1}\{|y-\tilde{Y}_{\tau,i}|\leq|\tilde{Y}_{\tau,i}-\bar{Y}_{T,\tau,i}|\}\right]=O\left(T^{-1/2}\right).
\end{aligned}
$$

■

Proof of Theorems 2.1 and 2.2

We only prove Theorem 2.2 as the proof of Theorem 2.1 is a special case. For notational simplicity, we only prove the case where $\mathcal{T}=2$. Let $\bar{F}_{NT,\tau}(y)\equiv\frac{1}{N}\sum_{i=1}^N\mathbf{1}\{\tilde{Y}_{\tau,i}\leq y\}$ and $\tilde{D}_{NT}\equiv\sum_{i=1}^N[\hat{F}_{NT,1}(\tilde{Y}_{1,i})-\hat{F}_{NT,2}(\tilde{Y}_{2,i})]^2$. We prove Theorem 2.2 by showing that (i) $\hat{D}_{NT}-\tilde{D}_{NT}=o_P(1)$; (ii) $\tilde{D}_{NT}-B_{NT}-\mu\xrightarrow{d}\sum_{j=1}^\infty\lambda_j(\mathcal{Z}_j^2-1)=o_P(1)$; and (iii) $\hat{B}_{NT}-B_{NT}=o_P(1)$ 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

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

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

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

and Lemma B.1(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 B.1(ii) and Assumption A.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 moment calculations and Chebyshev inequality, $\bar{Y}_{T,\tau,i}-\tilde{Y}_{\tau,i}=o_P(1)$ under Assumptions A.1(ii) and A.2. This implies that as $T\to\infty$, the limiting distribution and support of $\bar{Y}_{T,\tau,i}$ will coincide with those of $\tilde{Y}_{\tau,i}$. By the continuity of $\overline{g}_\tau$ in Assumption A.3(ii), the support of $\tilde{Y}_{\tau,i}$ is compact. This implies that for sufficiently large T, with probability approaching one 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 B.1 and by the triangle and C_r inequalities, we can show that uniformly in $i,\ j=1,...,N$,

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

and

$$\begin{aligned}E\left(\beta_{4,i}^2\right)&\leq 4\sum_{\tau=1}^{2}E\{[\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})-\tilde{F}_{\tau}(\tilde{Y}_{\tau,i})]^2\}+2E\{[\tilde{F}_1(\tilde{Y}_{1,i})-\tilde{F}_2(\tilde{Y}_{2,i})]^2\}\\&=O\left(T^{-1}+N^{-1}\right)\ \text{under } \tilde{H}_1(N^{-1/2}).\end{aligned}\tag{B.3}$$

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

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

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 (B.2). It follows that $E[\theta^2_{NT2,1\tau}]=O\left(T^{-1/2}+N^{-1}\right)$ 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

$$\begin{aligned} & N^{-2}\sum_{i=1}^{N}\sum_{i'\neq i}^{N}\sum_{j\neq i,i'}^{N}\left[E(\beta_{1\tau,ij}^2\beta_{1\tau,i'j}^2)\right]^{1/2}\left[E(\beta_{4,i}^2)E(\beta_{4,i'}^2)\right]^{1/2} \\ \leq\ & 8^{1/2}N\sup_{i,j}\left\{E\left|\beta_{1\tau,ij}\right|\right\}^{1/2}E(\beta_{4,1}^2) = O\left(N\right)O(T^{-1/4})O\left(T^{-1}+N^{-1}\right) = o\left(1\right). \end{aligned}$$

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

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

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

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

$$\begin{aligned} \tilde{D}_{NT} &= \sum_{i=1}^{N}\left[\hat{F}_{NT,1}(\tilde{Y}_{1,i}) - \hat{F}_{NT,2}(\tilde{Y}_{2,i})\right]^2 \\ &= \sum_{i=1}^{N}\left[\tilde{F}_{T1}(\tilde{Y}_{1,i}) - \tilde{F}_{T2}(\tilde{Y}_{2,i})\right]^2 + \sum_{i=1}^{N}\left[\hat{F}_{NT,1}(\tilde{Y}_{1,i}) - \tilde{F}_{T1}(\tilde{Y}_{1,i}) - \hat{F}_{NT,2}(\tilde{Y}_{2,i}) + \tilde{F}_{T2}(\tilde{Y}_{2,i})\right]^2 \\ &\quad +2\sum_{i=1}^{N}\left[\hat{F}_{NT,1}(\tilde{Y}_{1,i}) - \tilde{F}_{T1}(\tilde{Y}_{1,i}) - \hat{F}_{NT,2}(\tilde{Y}_{2,i}) + \tilde{F}_{T2}(\tilde{Y}_{2,i})\right]\left[\tilde{F}_{T1}(\tilde{Y}_{1,i}) - \tilde{F}_{T2}(\tilde{Y}_{2,i})\right] \\ &\equiv \tilde{D}_{NT1} + \tilde{D}_{NT2} + 2\tilde{D}_{NT3},\ \text{say.} \end{aligned}$$

We further decompose $\tilde{D}_{NT1}$ as follows:

$$\begin{aligned} \tilde{D}_{NT1} &= \sum_{i=1}^{N}\left[\tilde{F}_1(\tilde{Y}_{1,i}) - \tilde{F}_2(\tilde{Y}_{2,i})\right]^2 + \sum_{i=1}^{N}\left[\tilde{F}_{T1}(\tilde{Y}_{1,i}) - \tilde{F}_{T2}(\tilde{Y}_{2,i}) - \tilde{F}_1(\tilde{Y}_{1,i}) + \tilde{F}_2(\tilde{Y}_{2,i})\right]^2 \\ &\quad +2\sum_{i=1}^{N}\left[\tilde{F}_{T1}(\tilde{Y}_{1,i}) - \tilde{F}_{T2}(\tilde{Y}_{2,i}) - \tilde{F}_1(\tilde{Y}_{1,i}) + \tilde{F}_2(\tilde{Y}_{2,i})\right]\left[\tilde{F}_1(\tilde{Y}_{1,i}) - \tilde{F}_2(\tilde{Y}_{2,i})\right] \\ &\equiv \tilde{D}_{NT1,1} + \tilde{D}_{NT1,2} + 2\tilde{D}_{NT1,3}. \end{aligned}$$

By the weak law of large numbers, $\tilde{D}_{NT1,1} \xrightarrow{P} \mu$ under $\tilde{H}_1\left(N^{-1/2}\right)$. By the C_r inequality and Lemma B.1(ii), $\tilde{D}_{NT1,2} \leq 2\sum_{\tau=1}^{2}\sum_{i=1}^{N}\left[\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i}) - \tilde{F}_{\tau}(\tilde{Y}_{\tau,i})\right]^2 = O_P\left(NT^{-1}\right) = o_P(1)$. Then $|\tilde{D}_{NT1,3}| \leq \{\tilde{D}_{NT1,1}\}^{1/2}\{\tilde{D}_{NT1,2}\}^{1/2} = o_P(1)$ by Cauchy-Schwarz inequality.

Now, let $\xi_i \equiv (\bar{Y}_{T,1,i}, \bar{Y}_{T,2,i}, \tilde{Y}_{1,i}, \tilde{Y}_{2,i})'$ and $\psi_T(\xi_i, \xi_j) \equiv \mathbf{1}\{\bar{Y}_{T,1,j} \leq \tilde{Y}_{1,i}\} - \tilde{F}_{T1}(\tilde{Y}_{1,i}) - \mathbf{1}\{\bar{Y}_{T,2,j} \leq \tilde{Y}_{2,i}\} + \tilde{F}_{T2}(\tilde{Y}_{2,i})$. Then we can decompose $\tilde{D}_{NT2}$ as follows

$$
\begin{aligned}
\tilde{D}_{NT2} &= N^{-2}\sum_{i=1}^{N}\left[\sum_{j=1}^{N}\psi_T(\xi_i,\xi_j)\right]^2 \\
&= N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq j,i}^{N}\psi_T(\xi_i,\xi_j)\,\psi_T(\xi_i,\xi_k) + N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi_T(\xi_i,\xi_j)^2 \\
&\quad +2N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi_T(\xi_i,\xi_i)\,\psi_T(\xi_i,\xi_j) + N^{-2}\sum_{i=1}^{N}\psi_T(\xi_i,\xi_i)^2 \\
&\equiv V_{NT} + B_{NT} + 2R_{NT1} + R_{NT2}, \text{ say.}
\end{aligned}
$$

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

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

where $\bar{V}_{NT} \equiv \frac{6}{N(N-1)(N-2)}\sum_{1\leq i<j<k\leq N}\bar{\psi}_T(\xi_i,\xi_j,\xi_k)$. By the Hoeffding decomposition (e.g., Lee (1990, p. 26)), $\bar{V}_{NT} = 3H_{NT}^{(2)} + H_{NT}^{(3)}$, where

$$
\begin{aligned}
H_{NT}^{(2)} &\equiv \frac{2}{N(N-1)}\sum_{1\leq i<j\leq N}\bar{\psi}_{2T}(\xi_i,\xi_j), \\
H_{NT}^{(3)} &\equiv \frac{6}{N(N-1)(N-2)}\sum_{1\leq i<j<k\leq N}\bar{\psi}_{3T}(\xi_i,\xi_j,\xi_k),
\end{aligned}
$$

$\bar{\psi}_{2T}(\xi_i,\xi_j) \equiv \int\bar{\psi}_T(\xi_i,\xi_j,\xi)\,\tilde{F}(d\xi) = \frac{1}{3}\int\psi_T(\xi,\xi_i)\,\psi_T(\xi,\xi_j)\,\tilde{F}(d\xi)$, $\bar{\psi}_{3T}(\xi_i,\xi_j,\xi_k) \equiv \bar{\psi}_T(\xi_i,\xi_j,\xi_k) - \bar{\psi}_{2T}(\xi_i,\xi_j) - \bar{\psi}_{2T}(\xi_i,\xi_k) - \bar{\psi}_{2T}(\xi_j,\xi_k)$, 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, Thus,

$$
\begin{aligned}
V_{NT} &= \frac{(N-1)(N-2)}{N}\left[3H_{NT}^{(2)} + H_{NT}^{(3)}\right] = \frac{N-2}{N}\left[3(N-1)H_{NT}^{(2)} + O_P\left(N^{-1/2}\right)\right] \\
&= \{1+o(1)\}\,\mathcal{H}_{NT} + O_P\left(N^{-1/2}\right),
\end{aligned}
$$

where $\mathcal{H}_{NT} \equiv \frac{2}{N}\sum_{1\leq i<j\leq N}\int\psi_T(\xi,\xi_i)\,\psi_T(\xi,\xi_j)\,\tilde{F}(d\xi)$ is a second order degenerate U-statistic whose kernel function is T-dependent. Let $\zeta_i \equiv (\tilde{Y}_{1,i}, \tilde{Y}_{2,i})'$ and $\psi(u,\zeta_j) \equiv \mathbf{1}\{\tilde{Y}_{1,j} \leq u_1\} - \tilde{F}_1(u_1) - \mathbf{1}\{\tilde{Y}_{2,j} \leq u_2\} + \tilde{F}_2(u_2)$ where $u = (u_1,u_2)'$. Let $\bar{\mathcal{H}}_N \equiv \frac{2}{N}\sum_{1\leq i<j\leq N}\int\psi(\zeta,\zeta_i)\,\psi(\zeta,\zeta_j)\,\tilde{F}_{\zeta}(d\zeta)$ where $\tilde{F}_{\zeta}$ denote the

CDF of ζ_i. Note that

$$\begin{aligned}
\mathcal{H}_{NT}-\bar{\mathcal{H}}_N &= \frac{2}{N}\sum_{1\leq i<j\leq N}\int\left[\psi_T\left(\xi,\xi_i\right)\psi_T\left(\xi,\xi_j\right)-\psi\left(\zeta,\zeta_i\right)\psi\left(\zeta,\xi_j\right)\right]\tilde{F}\left(d\xi\right)\\
&= \frac{2}{N}\sum_{1\leq i<j\leq N}\int\left[\psi_T\left(\xi,\xi_i\right)-\psi\left(\zeta,\zeta_i\right)\right]\psi_T\left(\zeta,\xi_j\right)\tilde{F}\left(d\xi\right)\\
&\quad+\frac{2}{N}\sum_{1\leq i<j\leq N}\int\psi\left(\xi,\xi_i\right)\left[\psi_T\left(\zeta,\zeta_j\right)-\psi\left(\zeta,\xi_j\right)\right]\tilde{F}\left(d\xi\right)\\
&\equiv H_{NT,1}+H_{NT,2},\text{ say.}
\end{aligned}$$

Using Lemma B.1, we can readily show that $E\left(H_{NT,s}^2\right)=O(T^{-1/2}+N^{-1/2})$ under $\tilde{H}_1(N^{-1/2})$ for both $s=1,2$. It follows that $\mathcal{H}_{NT}=\bar{\mathcal{H}}_N+o_P\left(1\right)$ by Chebyshev inequality. By Serfling (1980, p.194) or Proposition 5.2 of Chen and White (1998), $\bar{\mathcal{H}}_N\overset{d}{\to}\sum_{j=1}^{\infty}\lambda_j(\mathcal{Z}_j^2-1)$ where $\{\mathcal{Z}_j\}$ is a sequence of IID $N\left(0,1\right)$ random variables, and $\{\lambda_j\}$ is the sequence of nonzero eigenvalues for $\mathcal{K}\left(u,v\right)\equiv\int\psi\left(\zeta,u\right)\psi\left(\zeta,v\right)\tilde{F}_{\zeta}\left(d\zeta\right)$. Next, noting that $E\left(R_{NT1}^2\right)=O(N^{-1})$ and $E\left|R_{NT2}\right|=O(N^{-1})$, we have $R_{NT1}=O_P\left(N^{-1/2}\right)$ and $R_{NT2}=O_P\left(N^{-1}\right)$ by Chebyshev and Markov inequalities. Consequently $\tilde{D}_{NT2}-B_{NT}\overset{d}{\to}\sum_{j=1}^{\infty}\lambda_j\left(\mathcal{Z}_j^2-1\right)$.

For $\tilde{D}_{NT3}$, we have

$$\begin{aligned}
\tilde{D}_{NT3} &= \sum_{i=1}^{N}\left[\hat{F}_{NT,1}(\tilde{Y}_{1,i})-\tilde{F}_{T1}(\tilde{Y}_{1,i})-\hat{F}_{NT,2}(\tilde{Y}_{2,i})+\tilde{F}_{T2}(\tilde{Y}_{2,i})\right]\left[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})\right]\\
&= N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi\left(\xi_i,\xi_j\right)\left[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})\right]+N^{-1}\sum_{i=1}^{N}\psi\left(\xi_i,\xi_i\right)\left[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})\right]\\
&\equiv \tilde{D}_{NT3,1}+\tilde{D}_{NT32}.
\end{aligned}$$

By triangle inequality and Lemma B.1(ii), under $\tilde{H}_1\left(N^{-1/2}\right)$ we have

$$\begin{aligned}
E|\tilde{D}_{NT32}| &\leq 2N^{-1}\sum_{i=1}^{N}E|\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})|\\
&\leq 2N^{-1}\left\{\sum_{\tau=1}^{2}\sum_{i=1}^{N}\{E|\tilde{F}_{T\tau}(\tilde{Y}_{\tau,i})-\tilde{F}_{\tau}(\tilde{Y}_{\tau,i})|+\sum_{i=1}^{N}E|\tilde{F}_1(\tilde{Y}_{1,i})-\tilde{F}_2(\tilde{Y}_{2,i})|\right\}\\
&= O\left(T^{-1/2}+N^{-1/2}\right).
\end{aligned}$$

Thus $\tilde{D}_{NT32}=O_P\left(T^{-1/2}+T^{-1/2}\right)$ by Markov inequality. Letting $\chi\left(\xi_i,\xi_j\right)=\psi\left(\xi_i,\xi_j\right)[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]$, then $\tilde{D}_{NT3,1}=N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\chi\left(\xi_i,\xi_j\right)$. By the Hölder and C_r inequalities,

$$\begin{aligned}
E\left(\tilde{D}_{NT3,1}^2\right) &= N^{-2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\sum_{j\neq i,i'}^{N}E\left[\chi\left(\xi_i,\xi_j\right)\chi\left(\xi_{i'},\xi_j\right)\right]\\
&\leq N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left[\psi\left(\xi_i,\xi_j\right)^2[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]^2\right]\\
&\leq 2N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left\{\left[\mathbf{1}\{\bar{Y}_{T,1,j}\leq\tilde{Y}_{1,i}\}-\mathbf{1}\{\tilde{Y}_{T,2,j}\leq\tilde{Y}_{2,i}\}\right]^2[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&+2N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left\{\left[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})\right]^{4}\right\}\\
\equiv\ & 2ED_{N1}+2ED_{N2},\ \text{say.}
\end{aligned}
$$

For ED_{N1}, we have

$$
\begin{aligned}
ED_{N1} &\leq N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left[|\mathbf{1}\{\bar{Y}_{T,1,j}\leq\tilde{Y}_{1,i}\}-\mathbf{1}\{\bar{Y}_{T,2,j}\leq\tilde{Y}_{2,i}\}|[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]^{2}\right]\\
&= N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left[|\mathbf{1}\{\tilde{F}_{T1}(\tilde{Y}_{T,1,j})\leq\tilde{F}_{T1}(\tilde{Y}_{1,i})\}-\mathbf{1}\{\tilde{F}_{T2}(\bar{Y}_{T,2,j})\leq\tilde{F}_{T2}(\tilde{Y}_{2,i})\}|[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]^{2}\right]\\
&\leq N^{-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}E\left[\mathbf{1}\{|\tilde{F}_{T1}(\tilde{Y}_{T,1,j})-\tilde{F}_{T1}(\tilde{Y}_{1,i})|\leq|\alpha_{NT}|\}[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]^{2}\right]\\
&\to 0,
\end{aligned}
$$

where $\alpha_{NT}=\tilde{F}_{T1}(\tilde{Y}_{T,1,j})-\tilde{F}_{T2}(\bar{Y}_{T,2,j})-\tilde{F}_{T1}(\tilde{Y}_{1,i})+\tilde{F}_{T2}(\tilde{Y}_{2,i})=O_{P}(N^{-1/2}+T^{-1/2})$; the third line follows from the fact that $|\mathbf{1}\{z\leq 0\}-\mathbf{1}\{z\leq a\}|\leq\mathbf{1}\{|z|\leq|a|\}$; and the last line follows from the dominated convergence theorem (DCT) and the fact that $N\ E[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]^{2}=O(1)$ under $\tilde{H}_{1}\left(N^{-1/2}\right)$. Consequently, $\tilde{D}_{NT3,1}=o_{P}(1)$ by the Chebyshev inequality. Similarly, by the DCT and the fact that $NE[\tilde{F}_{T1}(\tilde{Y}_{1,i})-\tilde{F}_{T2}(\tilde{Y}_{2,i})]^{2}=O(1)$ under $\tilde{H}_{1}\left(N^{-1/2}\right)$, we have $ED_{N2}=o(1)$. It follows that $\tilde{D}_{NT3}=o_{P}(1)$.

Lastly, it is straightforward to prove (iii). ■

Proof of Theorem 2.3

Again, we focus on the case $\mathcal{T}=2$. Using the notation in the proof of Theorem 2.2, it is easy to show that $N^{-1}(\hat{D}_{NT}-\tilde{D}_{NT})=o_{P}(1)$ under $\tilde{H}_{1}(1)$. Further, $N^{-1}\tilde{D}_{NT}=N^{-1}\sum_{i=1}^{N}\left[\tilde{F}_{1}(\tilde{Y}_{1,i})-\tilde{F}_{2}(\tilde{Y}_{2,i})\right]^{2}+o_{P}(1)=\mu+o_{P}(1)$, and $N^{-1}\hat{B}_{NT}=O_{P}\left(N^{-1}\right)$. Consequently, $N^{-1}J_{NT}=N^{-1}(\hat{D}_{NT}-\hat{B}_{NT})=N^{-1}(\hat{D}_{NT}-\tilde{D}_{NT})+N^{-1}\tilde{D}_{NT}-N^{-1}\hat{B}_{NT}=\mu+o_{P}(1)$, and the conclusion follows. ■

To prove Theorem 2.4, we prove the following lemma first.

Lemma B.2 *Suppose the conditions in Theorem 2.4 hold. Then*

(i) $\max_{1\leq i\leq N}|\hat{A}_{NT,i}-A_{i}|=O_{P}(N^{-1/2}\sqrt{\log N})$;

(ii) $\sup_{(x,a)\in\tilde{\mathcal{X}}\times\mathbb{I}}|\hat{g}(x,a)-g(x,a)|=O_{P}(\|h\|^{p+1}+(NTh!)^{-1/2}\sqrt{\log(NT)}+N^{-1/2}\sqrt{\log N}+b^{2}+(nb)^{-1})=o_{P}(1)$, *where* $h!=\Pi_{l=1}^{d+1}h_{l}$ *and* $\tilde{\mathcal{X}}$ *denotes the intersection of the support* $\mathcal{X}$ *and the uniform of the supports of* $w_{\tau}(\cdot)$.

Proof. Let $\tilde{F}$ and $\tilde{F}_{T}$ denote the CDF of $\tilde{Y}_{i}$ and $\bar{Y}_{T,i}$, respectively, where $\tilde{Y}_{i}\equiv E(Y_{it}|A_{i})$ and $\bar{Y}_{T,i}=T^{-1}\sum_{t=1}^{T}Y_{it}$. To prove (i), we first decompose $\hat{A}_{NT,i}-A_{i}$ as follows

$$
\begin{aligned}
\hat{A}_{NT,i}-A_{i} &= \frac{1}{N}\sum_{j=1}^{N}\mathbf{1}\left\{\bar{Y}_{T,j}\leq\bar{Y}_{T,i}\right\}-A_{i}=\frac{1}{N_{1}}\sum_{j=1,j\neq i}^{N}\mathbf{1}\left\{\bar{Y}_{T,j}\leq\bar{Y}_{T,i}\right\}-A_{i}+O\left(N^{-1}\right)\\
&= \frac{1}{N_{1}}\sum_{j=1,j\neq i}^{N}\left[\mathbf{1}\left\{\tilde{Y}_{j}\leq\tilde{Y}_{i}\right\}-A_{i}\right]+\frac{1}{N_{1}}\sum_{j=1,j\neq i}^{N}\left[\mathbf{1}\left\{\bar{Y}_{T,j}\leq\tilde{Y}_{i}\right\}-\mathbf{1}\left\{\tilde{Y}_{j}\leq\tilde{Y}_{i}\right\}\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{1}{N_1}\sum_{j=1,j\neq i}^{N}\left[\mathbf{1}\left\{\bar{Y}_{T,j}\leq \bar{Y}_{T,i}\right\}-\mathbf{1}\left\{\bar{Y}_{T,j}\leq \tilde{Y}_i\right\}\right]+O\left(N^{-1}\right)\\
\equiv\ & S_{1i}+S_{2i}+S_{3i}+O\left(N^{-1}\right),
\end{aligned}
$$

where $N_1 = N-1$. Noting that $E\left(S_{1i}\right) = E[\tilde{F}(\tilde{Y}_i) - A_i] = 0$ and $\mathrm{Var}(S_{1i}) = O\left(N^{-1}\right)$, we have $S_{1i} = O\left(N^{-1/2}\right)$. By Boole's and Bernstein's inequalities (e.g., Serfling (1980, p. 95)), for any $\epsilon > 0$

$$
\begin{aligned}
P\left(\max_{1\leq i\leq N}|S_{1i}|\geq \epsilon N_1^{-1/2}\sqrt{\log N_1}\right) &\leq NP\left(|S_{1i}|\geq N^{-1/2}\log N\epsilon\right)\leq 2N\exp\left(-\frac{\epsilon^2 N_1\log N_1}{2N_1+\frac{2}{3}N_1\epsilon}\right)\\
&= 2\exp\left(-\frac{\epsilon^2 N_1\log N_1}{2N_1+\frac{2}{3}N_1\epsilon}+\log N\right)\to 0 \text{ for sufficiently large } \epsilon.
\end{aligned}
$$

It follows that $\max_{1\leq i\leq N}|S_{1i}| = O_P(N^{-1/2}\sqrt{\log N})$. For S_{2i}, we have

$$
\begin{aligned}
S_{2i} &= \frac{1}{N_1}\sum_{j=1,j\neq i}^{N}\left[\mathbf{1}\{\bar{Y}_{T,j}\leq \tilde{Y}_i\}-\tilde{F}_T(\tilde{Y}_i)+\mathbf{1}\{\tilde{Y}_j\leq \tilde{Y}_i\}+\tilde{F}(\tilde{Y}_i)\right]-\left[\tilde{F}_T(\tilde{Y}_i)-\tilde{F}(\tilde{Y}_i)\right]\\
&\equiv S_{2i,1}-S_{2i,2},\ \text{say.}
\end{aligned}
$$

Analogous to the study of S_{1i}, we can show that $\max_{1\leq i\leq N}|S_{2i,1}| = O_P(N^{-1/2}\sqrt{\log N})$. As in the proof of Lemma B.1(iii), we can show that $\max_{1\leq i\leq N}|S_{2i,2}| \leq \sup_y\left|\tilde{F}_T(y)-\tilde{F}(y)\right| = O_P(T^{-1/2})$. Thus $\max_{1\leq i\leq N}|S_{2i}| = O_P(N^{-1/2}\sqrt{\log N}+T^{-1/2}) = O_P(N^{-1/2}\sqrt{\log N})$. For S_{3i}, we have

$$
\begin{aligned}
S_{3i} &= \frac{1}{N_1}\sum_{j=1,j\neq i}^{N}\left[\mathbf{1}\{\bar{Y}_{T,j}\leq \bar{Y}_{T,i}\}-\tilde{F}_T(\bar{Y}_{T,i})-\mathbf{1}\{\bar{Y}_{T,j}\leq \tilde{Y}_i\}+\tilde{F}_T(\tilde{Y}_i)\right]-\left[\tilde{F}_T(\bar{Y}_{T,i})-\tilde{F}_T(\tilde{Y}_i)\right]\\
&\equiv S_{3i,1}-S_{3i,2},\ \text{say.}
\end{aligned}
$$

Analogous to the study of S_{1i}, we can show that $\max_{1\leq i\leq N}|S_{3i,1}| = O_P(N^{-1/2}\sqrt{\log N})$. In addition, by Boole's and Bernstein's inequalities, $\max_{1\leq i\leq N}|S_{3i,2}| \leq \sup_y\left|\tilde{f}_T(y)\right|\max_{1\leq i\leq N}|\bar{Y}_{T,i}-\tilde{Y}_i| = O_P(T^{-1/2}\sqrt{\log N})$. Thus $\max_{1\leq i\leq N}|S_{3i}| = O_P(N^{-1/2}\sqrt{\log N})$. Consequently, $\max_{1\leq i\leq N}|\hat{A}_{NT,i}-A_i| = O_P(N^{-1/2}\sqrt{\log N})$.

For (ii), we only give a sketchy proof. Recall $\hat{g}^{-1}(x,a)\equiv (nb)^{-1}\sum_{j=1}^{n}\int_{-\infty}^{a} k\left(b^{-1}[\tilde{g}(x,j/n)-\tilde{a}]\right)d\tilde{a}$. Let $g_n^{-1}(x,a)$ be defined as $\hat{g}^{-1}(x,a)$ with $\tilde{g}(x,j/n)$ being replaced by $g(x,j/n)$, i.e., $g_n^{-1}(x,a)\equiv (nb)^{-1}\sum_{j=1}^{n}\int_{-\infty}^{a} k\left(b^{-1}[g(x,j/n)-\tilde{a}]\right)d\tilde{a}$. Following the proofs of Lemmas 2.1 and 2.2 in Dette et al. (2006), we can show that

$$
g_n^{-1}(x,a) = g^{-1}(x,a)+O_P\left(b^2+(nb)^{-1}\right)\ \text{and}\ g_n(x,a) = g(x,a)+O_P\left(b^2+(nb)^{-1}\right) \tag{B.4}
$$

uniformly in $(x,a)\in\tilde{\mathcal{X}}\times\mathbb{I}$. By the uniform consistency result for local polynomial estimate with generated regressors (c.f., Mammen et al. (2012, Theorem 2) and Su and Ullah (2006, Lemmas A.2-A.5) for the cases of nonparametrically generated regressors) and the result in part (i), we have

$$
\tilde{g}(x,a) = g(x,a)+O_P\left[\|h\|^{p+1}+(NTh!)^{-1/2}\sqrt{\log(NT)}+N^{-1/2}\sqrt{\log N}\right] \tag{B.5}
$$

where the first two terms in the last expression are present even if we observe A_i, and the third term signals the cost of replacing A_i by $\hat{A}_i$. Using (B.5) and similar arguments as used in the proofs of Theorems

3.1 and 3.2 in Dette et al. (2006) in conjunction with Boole's and Bernstein's inequality, we can show that

$$\hat{g}(x,a) - g_J(x,a) = O\left(\|h\|^{p+1} + (NTh!)^{-1/2}\sqrt{\log(NT)} + N^{-1/2}\sqrt{\log N}\right) \tag{B.6}$$

uniformly in $(x,a) \in \tilde{\mathcal{X}} \times \mathbb{I}$. Combining (B.4) and (B.6) yields the desired result. ■

Proof of Theorem 2.4

Let P^* denote the probability distribution induced by the bootstrap resampling, with expectation and variance operators given by $E^*(\cdot)$ and $\mathrm{Var}^*(\cdot)$, respectively. In addition, we use $O_{P^*}(\cdot)$ and $o_{P^*}(\cdot)$ to denote the probability orders of magnitude according to the bootstrap-induced probability law; e.g., $a_{NT} = o_{P^*}(1)$ denotes that $P^*(|a_{NT}| \geq \epsilon) = o_P(1)$ for any positive $\epsilon > 0$. Note that $a_{NT} = o_P(1)$ implies that $a_{NT} = o_{P^*}(1)$. We use $\mathcal{W}_{NT}$ to denote the original sample.

Recall that $Y_{it}^* = \hat{g}(X_{it}^*, A_i^*) + \varepsilon_{it}^*$, where the monotonicity of $\hat{g}$ in its second argument is imposed and A_i^* is independent of $(X_{it}^*, \varepsilon_{it}^*)$ conditional on $\mathcal{W}_{NT}$. By construction, both monotonicity and exogeneity are satisfied in the bootstrap world. The bootstrap analogue of $\tilde{Y}_{\tau,i} = E[Y_{it}w_\tau(X_{it})|A_i]$ is now given by

$$\begin{aligned}
\tilde{Y}_{\tau,i}^* &\equiv E^*\left[\hat{g}(X_{it}^*, A_i^*)w_\tau(X_{it}^*)|A_i^*\right] + E^*[\varepsilon_{it}^* w_\tau(X_{it}^*)|A_i^*] \\
&= \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T \hat{g}(X_{js}, A_i^*)w_\tau(X_{js}) + \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T \hat{\varepsilon}_{js}w_\tau(X_{js}) \\
&\equiv \bar{g}_\tau^*(A_i^*) + \bar{\varepsilon}_\tau^*.
\end{aligned} \tag{B.7}$$

Since w_τ is nonnegative, $\bar{g}_\tau^*$ preserves the monotonicity of $\hat{g}$ in its second argument.

As in the proof of Theorem 2.2, we only prove the case where $\mathcal{T} = 2$. For $\tau = 1, 2$, let $\bar{Y}_{T,\tau,i}^*$, $\bar{F}_{NT,\tau}^*(\cdot)$, $\hat{F}_{NT,\tau}^*(\cdot)$, $\hat{A}_{NT,\tau,i}^*$, $\hat{D}_{NT}^*$, and $\tilde{D}_{NT}^*$ denote the bootstrap analogue of $\bar{Y}_{T,\tau,i}$, $\bar{F}_{NT,\tau}(\cdot)$, $\hat{F}_{NT,\tau}(\cdot)$, $\hat{A}_{NT,\tau,i}$, $\hat{D}_{NT}$, and $\tilde{D}_{NT}$, respectively. That is, $\bar{Y}_{T,\tau,i}^* \equiv T^{-1}\sum_{t=1}^T Y_{it}^* w_\tau(X_{it})$, $\bar{F}_{NT,\tau}^*(\cdot) \equiv \frac{1}{N}\sum_{i=1}^N \mathbf{1}\{\tilde{Y}_{\tau,i}^* \leq \cdot\}$, $\hat{F}_{NT,\tau}^*(\cdot) \equiv N^{-1}\sum_{j=1}^N \mathbf{1}\{\bar{Y}_{T,\tau,j}^* \leq \cdot\}$, $\hat{A}_{NT,\tau,i}^* \equiv \hat{F}_{NT,\tau}^*(\bar{Y}_{T,\tau,i}^*)$, $\hat{D}_{NT}^* \equiv \sum_{i=1}^N(\hat{A}_{NT,1,i}^* - \hat{A}_{NT,2,i}^*)^2$, and $\tilde{D}_{NT}^* \equiv \sum_{i=1}^N[\hat{F}_{NT,1}^*(\tilde{Y}_{1,i}^*) - \hat{F}_{NT,2}^*(\tilde{Y}_{2,i}^*)]^2$. Let $\tilde{F}_{T\tau}^*$ and $\tilde{f}_{T\tau}^*$ denote the CDF and PDF of $\bar{Y}_{T,\tau,i}^*$ given $\mathcal{W}_{NT}$, respectively. We prove Theorem 2.4 by showing that (i) $\hat{D}_{NT}^* - \tilde{D}_{NT}^* = o_{P^*}(1)$; (ii) $\bar{D}_{NT}^* - B_{NT}^* \xrightarrow{d^*} \sum_{j=1}^\infty \lambda_j(\mathcal{Z}_j^2 - 1)$; and (iii) $B_{NT}^* - \hat{B}_{NT}^* = o_{P^*}(1)$.

Noting that $\bar{Y}_{T,\tau,i}^* - \tilde{Y}_{\tau,i}^* = T^{-1}\sum_{t=1}^T \varepsilon_{it}^* w_\tau(X_{it})$, we can readily show that

$$E^*(\bar{Y}_{T,\tau,i}^* - \tilde{Y}_{\tau,i}^*)^2 = O_P(1/T) \text{ and } \sup_y |\tilde{F}_{T\tau}^*(y) - \tilde{F}_\tau^*(y)| = O_{P^*}(T^{-1/2}). \tag{B.8}$$

We can follow the proof of part (i) in the proof of Theorem Theorem 2.2 closely and show (i) analogously, now using (B.8) in place of Lemma B.1.

Now, we show (ii). We decompose $\tilde{D}^*_{NT}$ as follows

$$\begin{aligned}
\tilde{D}^*_{NT} &= \sum_{i=1}^{N}\left[\tilde{F}^*_{T1}(\tilde{Y}^*_{1,i})-\tilde{F}^*_{T2}(\tilde{Y}^*_{2,i})\right]^2 \\
&\quad+\sum_{i=1}^{N}\left[\hat{F}^*_{NT,1}(\tilde{Y}^*_{1,i})-\tilde{F}^*_{T1}(\tilde{Y}^*_{1,i})-\hat{F}^*_{NT,2}(\tilde{Y}^*_{2,i})+\tilde{F}^*_{T2}(\tilde{Y}^*_{2,i})\right]^2 \\
&\quad+2\sum_{i=1}^{N}\left[\hat{F}^*_{NT,1}(\tilde{Y}^*_{1,i})-\tilde{F}^*_{T1}(\tilde{Y}^*_{1,i})-\hat{F}^*_{NT,2}(\tilde{Y}^*_{2,i})+\tilde{F}^*_{T2}(\tilde{Y}^*_{2,i}\right]\left[\tilde{F}^*_{T1}(\tilde{Y}^*_{1,i})-\tilde{F}^*_{T2}(\tilde{Y}^*_{2,i}\right] \\
&\equiv \tilde{D}^*_{NT1}+\tilde{D}^*_{NT2}+2\tilde{D}^*_{NT3},\ \text{say.}
\end{aligned}$$

Noting that $\bar{g}^*_\tau$ in (B.7) is strictly monotone a.s.$-P^*$, $\tilde{F}^*_1(\tilde{Y}^*_{1,i})=A^*_i=\tilde{F}^*_2(\tilde{Y}^*_{2,i})$. It follows by (B.8) that

$$\begin{aligned}
\tilde{D}^*_{NT1} &= \sum_{i=1}^{N}\left[\tilde{F}^*_{T1}(\tilde{Y}^*_{1,i})-\tilde{F}^*_1(\tilde{Y}^*_{1,i})-\tilde{F}^*_{T2}(\tilde{Y}^*_{2,i})+\tilde{F}^*_2(\tilde{Y}^*_{2,i})\right]^2 \\
&\leq 2\sum_{\tau=1}^{2}\sum_{i=1}^{N}\left[\tilde{F}^*_{T\tau}(\tilde{Y}^*_{\tau,i})-\tilde{F}^*_\tau(\tilde{Y}^*_{\tau,i})\right]^2=O_{P^*}(N/T)=o_{P^*}(1).
\end{aligned}$$

Let $\xi^*_i\equiv(\bar{Y}^*_{T,1,i},\bar{Y}^*_{T,2,i},\tilde{Y}^*_{1,i},\tilde{Y}^*_{2,i})'$ and $\psi^*_T\left(\xi^*_i,\xi^*_j\right)\equiv\mathbf{1}\{\bar{Y}^*_{T,1,j}\leq\tilde{Y}^*_{1,i}\}-\tilde{F}_{T1}(\tilde{Y}^*_{1,i})-\mathbf{1}\{\bar{Y}^*_{T,2,j}\leq\tilde{Y}^*_{2,i}\}+\tilde{F}_{T2}(\tilde{Y}^*_{2,i})$. Then

$$\tilde{D}^*_{NT2}=\sum_{i=1}^{N}\left[\hat{F}^*_{NT,1}(\tilde{Y}^*_{1,i})-\tilde{F}^*_{T1}(\tilde{Y}^*_{1,i})-\hat{F}^*_{NT,2}(\tilde{Y}^*_{2,i})+\tilde{F}_{T2}(\tilde{Y}^*_{2,i})\right]^2=N^{-2}\sum_{i=1}^{N}\left[\sum_{j=1}^{N}\psi^*\left(\xi^*_i,\xi^*_j\right)\right]^2.$$

We can decompose $\tilde{D}^*_{NT2}$ as follows

$$\begin{aligned}
\tilde{D}^*_{NT2} &= N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\sum_{k\neq j,i}^{N}\psi^*_T\left(\xi^*_i,\xi^*_j\right)\psi^*_T\left(\xi^*_i,\xi^*_k\right)+N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi^*_T\left(\xi^*_i,\xi^*_j\right)^2 \\
&\quad+2N^{-2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\psi^*_T\left(\xi^*_i,\xi^*_i\right)\psi^*_T\left(\xi^*_i,\xi^*_j\right)+N^{-2}\sum_{i=1}^{N}\psi^*\left(\xi^*_i,\xi^*_i\right)^2\equiv V^*_{NT}+B^*_{NT}+2R^*_{NT1}+R^*_{NT2}.
\end{aligned}$$

Noting that $E^*(R^{*2}_{NT1})=O(N^{-1})$ and $E^*\left|R^*_{NT2}\right|=O(N^{-1})$, we have $R^*_{NT1}=O_P\left(N^{-1/2}\right)$ and $R^*_{NT2}=O_P\left(N^{-1}\right)$ by Chebyshev and Markov inequalities. Let $\zeta^*_i=(\tilde{Y}^*_{1,i},\tilde{Y}^*_{2,i})'$ and $\psi^*\left(u,\xi^*_j\right)\equiv\mathbf{1}\{\tilde{Y}^*_{1,j}\leq u_1\}-\tilde{F}_1(u_1)-\mathbf{1}\{\tilde{Y}^*_{2,j}\leq u_2\}+\tilde{F}_2(u_2)$ with $u=(u_1,u_2)'$. For V^*_{NT}, using arguments analogous to those used in the study of V_{NT}, we can readily show that $V^*_{NT}=\{1+o_{P^*}(1)\}\bar{\mathcal{H}}^*_{NT}+O_{P^*}\left(N^{-1/2}\right)$, where

$$\bar{\mathcal{H}}^*_{NT}\equiv\frac{2}{N}\sum_{1\leq i<j\leq N}\int\psi^*\left(\zeta,\zeta^*_i\right)\psi^*\left(\zeta,\zeta^*_j\right)\tilde{F}^*_\zeta\left(d\xi\right)$$

is a second order degenerate U-statistic and $\tilde{F}^*_\zeta$ denotes the CDF of $\zeta^*_i=(\tilde{Y}^*_{1,i},\tilde{Y}^*_{2,i})'$. By Proposition 5.2 of Chen and White (1998), $\bar{\mathcal{H}}^*_{NT}\xrightarrow{d^*}\sum_{j=1}^{\infty}\lambda^*_j(\mathcal{Z}^2_j-1)$ where $\{\mathcal{Z}_j\}$ is a sequence of IID $N(0,1)$ random variables, and $\left\{\lambda^*_j\right\}$ is the sequence of nonzero eigenvalues for $\text{plim}_{T\to\infty}\mathcal{K}^*_T(u,v)$ where $\mathcal{K}^*_T(u,v)=$

$\int \psi^*(\zeta,u)\psi^*(\zeta,v)\tilde{F}_\zeta^*(d\zeta)$. To show that $\{\lambda_j^*\}$ coincide with $\{\lambda_j\}$ so that $\bar{\mathcal{H}}_{NT}^*$ has the same limiting distribution as the asymptotic distribution of $\bar{\mathcal{H}}_N$, it suffices to show that $\mathcal{K}_T^*(u,v) = \mathcal{K}_T(u,v) + o_P(1)$ uniformly in (u,v). The last desired result is true provided $\tilde{F}_\tau^* \to \tilde{F}_\tau$ for $\tau = 1,2$.

Recall that $\tilde{F}_\tau$ is the CDF of $\tilde{Y}_{\tau,i} = E[g(X_{it},A_i)w_\tau(X_{it})|A_i] + E[\varepsilon_{it}w_\tau(X_{it})] = \bar{g}_\tau(A_i) + \bar{\varepsilon}_\tau$ and $\tilde{F}_\tau^*$ is the CDF of $\tilde{Y}_{\tau,i}^* = E^*[\hat{g}(X_{it}^*,A_i^*)w_\tau(X_{it}^*)|A_i^*] + E^*[\varepsilon_{it}^*w_\tau(X_{it}^*)|A_i^*] = \bar{g}_\tau^*(A_i^*) + \bar{\varepsilon}_\tau^*$ conditional on $\mathcal{W}_{NT}$. Noting that A_i^* and A_i are both $\mathbb{U}(0,1)$ and $\bar{\varepsilon}_\tau$ is a constant, it suffices to show that for $\tau = 1,2$: (1) $\bar{\varepsilon}_\tau^* = \bar{\varepsilon}_\tau + o_P(1)$ and (2) $\bar{g}_\tau^*(a) = \bar{g}_\tau(a) + o_P(1)$ uniformly in a. (1) follows because by the LLN, Lemmas B.2(i)-(ii), and the continuity of $\hat{g}(\cdot,\cdot)$, we have

$$\begin{aligned}
\bar{\varepsilon}_\tau^* &= \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T\left[Y_{js} - \hat{g}(X_{js},\hat{A}_{NT,j})\right]w_\tau(X_{js}) \\
&= \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T \varepsilon_{js}w_\tau(X_{js}) + \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T[g(X_{js},A_j) - \hat{g}(X_{js},A_j)]w_\tau(X_{js}) \\
&\quad + \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T\left[\hat{g}(X_{js},A_j) - \hat{g}(X_{js},\hat{A}_{NT,j})\right]w_\tau(X_{js}) \\
&= \bar{\varepsilon}_\tau + o_P(1) + o_P(1) = \bar{\varepsilon}_\tau + o_P(1).
\end{aligned}$$

For (2), we have

$$\begin{aligned}
\bar{g}_\tau^*(a) - \bar{g}_\tau(a) &= \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T \hat{g}(X_{js},a)w_\tau(X_{js}) - E[g(X_{it},a)w_\tau(X_{it})] \\
&= \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T\{g(X_{js},a)w_\tau(X_{js}) - E[g(X_{js},a)w_\tau(X_{js})]\} \\
&\quad + \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T[\hat{g}(X_{js},a) - g(X_{js},a)]w_\tau(X_{js}) \\
&\equiv G_1(a) + G_2(a), \text{ say.}
\end{aligned}$$

The pointwise convergence of $G_1(a)$ to 0 follows from the LLN. The uniform convergence follows by a simple application of Bernstein inequality. For $G_2(a)$, we have $\sup_{a\in\mathbb{I}}|G_2(a)| \le \sup_{(x,a)\in\tilde{\mathcal{X}}\times\mathbb{I}}|\hat{g}(x,a) - g(x,a)| \times \frac{1}{NT}\sum_{j=1}^N\sum_{s=1}^T w_\tau(X_{js}) = o_P(1)E[w_\tau(X_{js})] = o_P(1)$. (1) and (2) imply that $\tilde{Y}_{\tau,i}^*$ have the same limiting distribution as $\tilde{Y}_{\tau,i}$ and thus $\mathcal{K}_T^*(u,v) = \mathcal{K}_T(u,v) + o_P(1)$. Consequently, λ_j^*'s coincide with λ_j's and $\tilde{D}_{NT2}^* - B_{NT}^* \xrightarrow{d} \sum_{j=1}^\infty \lambda_j(\mathcal{Z}_j^2 - 1)$.

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

References

Aït-sahalia, Y., Bickel, P., Stoker, T., 2001. Goodness-of-fit tests for kernel regressions with an application to options implied volatilities. Journal of Econometrics 105, 363-412.

Altonji, J. G., Matzkin, R. L., 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., Berry, S., 2006. On the nonparametric identification of nonlinear simultaneous equations models: comment on Brown (1983) and Roehrig (1988). Econometrica 74, 1429-1440.

Berry, S., Levinsohn, J., Pakes, A., 1995. Automobile prices in market equilibrium. Econometrica 63, 841-890.

Berry, S., Haile, P., 2010. Identification in differentiated products markets using market level data. Cowles Foundation Discussion Papers 1744, Yale University.

Blundell, R., Horowitz, J. 2007. A non-parametric test of exogeneity. Review of Economic Studies 74, 1035-1058.

Chen, X., Fan, Y., 1999. Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series. Journal of Econometrics 91, 373-401.

Chen, X., White, H., 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., Stinchcombe, M. B., Zeman, J., 2009. An Introduction to Mathematical Analysis for Economic Theory and Econometrics. Princeton University Press.

Daniel, K. and S. Titman 1997, Evidence on the Characteristics of Cross Sectional Variation in Stock Returns, Journal of Finance, 52, 1-33.

Dette, H., Neumeyer, N., Pilz, K. F., 2006. A simple nonparametric estimator of a strictly monotone regression function. Bernoulli 12, 469-490.

Escanciano, J. C., Jacho-Chávez, D. T., 2010. Approximating the critical values of Cramér-von Mises tests in general parametric conditional specifications. Computational Statistics and Data Analysis 54, 625-636.

Evdokimov, K., 2010. Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working Paper, Dept. of Economics, Princeton University.

Fama, E. F., French, K. T., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3-56.

Ferson, W. E., Harvey, C. R., 1991. The variation of economic risk premiums. Journal of Political Economy 99, 285-315.

Ferson, W. E., Harvey, C. R., 1993. The risk and predictability of international equity returns. Review of Financial Studies 6, 527-566.

Ghosal, S., Sen, A., van der Vaart, W., 2000. Testing monotonicity of regression. Annals of Statistics 28, 1054-1082.

Ghysels, E., 1998. On stable factor structures in the pricing of risk: do time-varying betas help or hurt? Journal of Finance 53, 549-573.

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

Guerre, E., Perrigne, I., Vuong, Q., 2000. Optimal nonparametric estimation of first-price auctions. Econometrica 68, 525-574.

Guerre, E., Perrigne, I., Vuong, Q., 2009. Nonparametric identification of risk aversion in first-price auctions under exclusion restrictions. Econometrica, 77 1193-1227.

Harvey, C. R., 1989. Time-varying conditional covariances in tests of asset pricing models. Journal of Financial Economics 24, 289-317.

Hoderlein, S., 2011. How many consumer are rational? Journal of Econometrics 164, 294-309.

Hoderlein, S., Mammen, E., 2007. Identification of marginal effects in nonseparable models without monotonicity. Econometrica 75, 1513-1518.

Hoderlein, S., Su, L., White, H., 2012. Specification testing for nonparametric structural models with monotonicity in unobservables. Working Paper, Department of Economics, UCSD.

Hoderlein, S., White, H., 2012. Nonparametric identification in nonseparable panel data models with generalized fixed effects. Journal of Econometrics 168, 300-314.

Imbens, G. W., Newey, W. K., 2009. Identification and estimation of triangular simultaneous equations models without additivity. Econometrica 77, 1481-1512.

Jagannathan, R., Wang, Z., 1996. The conditional CAPM and the cross-section of expected returns. Journal of Finance 51, 3-53.

Komunjer, I., Santos, A., 2010. Semiparametric estimation of nonseparable models: a minimum distance from independence approach. Econometrics Journal 13, S28-S55.

Lavergne, P., 2001. An equality test across nonparametric regressions. Journal of Econometrics 103, 307-344.

Lavergne, P., Vuong, Q., 2000. Nonparametric significance testing. Econometric Theory 16, 576-601.

Lee, A. J., 1990. U-statistics: Theory and Practice. Marcel Dekker, New York.

Leucht, A., Neumann, M. H., 2011. Degenerate U- and V-statistics under ergodicity: asymptotics, bootstrap and applications in statistics. Working paper, Institut für Stochastik, Friedrich-Schiller-Universität Jena.

Li, Q., Hsiao, C., Zinn, J., 2003. Consistent specification tests for semiparametric/nonparametric models based on series estimation method. Journal of Econometrics 112, 295-325.

Mammen, E., Rothe, C., Schienle, M., 2012. Nonparametric regression with nonparametrically generated covariates. Annals of Statistics 40, 1132-1170.

Matzkin, R. L., 2003. Nonparametric estimation of nonadditive random functions. Econometrica 71, 1339-1375.

Matzkin, R. L., 2007. Heterogeneous choice. In R. Blundell, W. Newey, and T. Persson (eds), Advances in Economics and Econometrics, Theory and Applications, Ninth World Congress of the Econometric Society, Cambridge University Press.

Megerdichian, A., 2009. Identification of price effects in discrete choice models of demand for differentiated products. Ph.D. Dissertation Chapter, Department of Economics, UCSD.

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.

Su, L., Ullah, A., 2006. More efficient estimation in nonparametric regression with nonparametric autocorrelated errors. Econometric Theory 22, 98-126.

Su, L., Ullah, A., 2013. A nonparametric goodness-of-fit-based test for conditional heteroskedasticity. Econometric Theory 29, 187-212.

Su, L., White, H., 2010. Testing structural change in partially linear models. Econometric Theory 26, 1761-1806.

Stigler, G., Becker, G., 1977. De gustibus non est disputandum. American Economic Review 67, 76–90.