11-2011

Specification Testing for Nonparametric Structural Models with Monotonicity in Unobservables

Stefan HODERLEIN
Boston College

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

Halbert WHITE
University of California - San Diego

Specification Testing for Nonparametric Structural Models with Monotonicity in Unobservables

Stefan Hoderlein
Department of Economics
Boston College

Liangjun Su
School of Economics
Singapore Management University

Halbert White
Department of Economics
University of California, San Diego

November 7, 2011

Abstract

Monotonicity in a scalar unobservable is a now common assumption in economic theory and applications. Among other things, it allows one to recover the underlying structural function from certain conditional quantiles of observables. Nevertheless, monotonicity is a strong assumption, and its failure can have substantive adverse consequences for structural inference. So far, there are no generally applicable nonparametric specification tests designed to detect monotonicity failure. This paper provides such a test for cross-section data. We show how to exploit an exclusion restriction together with a conditional independence assumption, plausible in a variety of applications, to construct a test. Our statistic is asymptotically normal under local alternatives and consistent against nonparametric alternatives violating the conditional quantile representation. Monte Carlo experiments show that a suitable bootstrap procedure yields tests with reasonable level behavior and useful power. We apply our test to study the role of unobserved ability in determining Black-White wage differences and to study whether Engel curves are monotonically driven by a scalar unobservable.

Keywords: control variables, covariates, endogenous variables, exogeneity, monotonicity, nonparametric, nonseparable, specification test, unobserved heterogeneity

Acknowledgements: We thank Steve Berry, Jim Heckman, Joel Horowitz, Lars Nesheim, and Maxwell Stinchcombe for helpful comments and discussion. We are indebted to Richard Blundell, Whitney Newey, and Derek Neal for providing us with their data and to Xun Lu for superlative computational assistance.

1 Introduction

Model misspecification – that is, the failure of the assumptions maintained to justify identification, estimation, and inference for the features of a given economic phenomenon – can have serious adverse consequences. At one extreme, the resulting estimators may provide no information whatever about the phenomenon of interest. Less extreme but still serious is that estimators may be informative, but inference may be flawed by use of an incorrect sampling distribution.

In recognition of these dangers, economists and econometricians have begun to rely increasingly on methods that require ever weaker maintained assumptions. The use of nonparametric methods for structural estimation and the use of bootstrap methods for inference are two prominent examples of these trends. Nevertheless, no matter how flexible methods of estimation and inference may become, identifying economic features of interest[1] always requires some assumed economic structure that may or may not be embodied in the data. The possibility of misspecification at this fundamental level must always be confronted.

Global identification of structural features of interest generically involves exclusion restrictions (i.e., that certain variables do not affect the dependent variable of interest) and some form of exogeneity condition (i.e., that certain variables are stochastically orthogonal to – e.g., independent of – unobservable drivers of the dependent variable, possibly conditioned on other observables). These assumptions permit identification of such important structural features as average marginal effects or various average effects of treatment. Seminal examples are the LATE of Imbens and Angrist (1994), the MTE of Heckman and Vytlacil (1999, 2005), or the control function model of Imbens and Newey (2009, IN) to name but a few.

In addition, there may be nonparametric restrictions placed on the structural function of interest, such as separability between observable and unobservable drivers of the dependent variable ("structural separability"), or, more generally, the assumption that the dependent variable depends monotonically on a scalar unobservable ("scalar monotonicity"). Although these assumptions need not be crucial to identifying and estimating average effects of interest, when they do hold, they permit recovery of the structural function itself. This line of work dates back to Roehrig (1988); see also Matzkin (2003) and IN. Thus, knowing whether scalar monotonicity holds is key to knowing whether one can access the economic relationship of interest in its entirety, with all that this entails for the resulting economic insight, or whether one must make do with knowing average or distributional properties of the structural relationship.

In recognition of the potentially adverse consequences of misspecification, nonparametric specification testing procedures have been receiving increasing attention. As White and Chalak (2010) discuss, there are now numerous nonparametric tests for various forms of exogeneity. Further,

[1] The identification considered here is that discussed by Hurwicz (1950), which entails the correspondence of structural features of economic interest to features of the distribution of observable data.

Hoderlein and Mammen (2009) and Lu and White (2011) have proposed convenient nonparametric tests for structural separability. But generally applicable specification tests for monotonicity as so far lacking. Su, Hoderlein, and White (2011) do provide a test for scalar monotonicity under a strict exogeneity assumption for panel data with large T, but its applicability is limited by these data requirements. Thus, our main goal and contribution here is to provide a new generally applicable test designed specifically to detect the failure of scalar monotonicity, adding a further weapon to the nonparametric specification testing arsenal.

We also contribute by complementing and extending the literature on the identification and estimation of nonseparable models with scalar monotonicity. Our results build on, complement, and extend those of Altonji and Matzkin (2005, AM). Our results also complement those of IN; for example, our new test can assess the validity of instruments that IN construct from a first-stage nonparametric structural equation, relying on scalar monotonicity.

A final contribution is the application of our new test to study the black-white earnings gap and to study consumer demand. For the former, we test the specification proposed by Neal and Johnson (1996), which includes unobserved ability as scalar monotonic factor, A, and the armed forces qualification test (AFQT) as a control variable. We fail to reject the null, providing support for Neal and Johnson's (1996) specification. That our test has power is illustrated by an analysis of Engel curves, where a scalar monotone unobservable is implausible (Hoderlein, 2011). In a control function setup virtually identical to that analyzed in IN, we find that indeed the null of a scalar monotone unobservable as a description of unobserved preference heterogeneity is rejected. This suggests a demand analysis that allows for heterogeneity in a more structural fashion.

The remainder of this paper is organized as follows: In Section 2, we discuss relevant aspects of the literature on nonparametric structural estimation with scalar monotonicity and motivate our testing approach. In Section 3, we give a detailed analysis of identification under monotonicity. Based on these results, we discuss the heuristics for our test in Section 4, turning to the formal asymptotics of our estimators and tests in the fifth and sixth sections. A Monte Carlo study occupies Section 7, and in Section 8 we present our two applications. Section 9 contains a summary and conclusion. Proofs of all results are gathered into a Mathematical Appendix

2 Scalar Monotonicity and Test Motivation

Monotonicity of a structural function in one important - yet unobservable - factor is an assumption widely invoked in economics. For instance, it is often postulated in labor economics that ability enters a returns-to-schooling model in a monotonic fashion: Other things equal, the higher the individual's ability, the higher her resulting wage. Similarly, monotonicity in unobservables has frequently been invoked in industrial organization, e.g., in the literature on production functions (see, e.g., Olley and Pakes, 1996) and the literature on auctions. In econometrics, monotonicity has been used to point identify single-equation nonseparable structures (Matzkin, 2003).

The appeal of monotonicity stems at least in part from the fact that it allows one to specify structural functions that allow for complicated interaction patterns between observables and unobservables without losing tractability. Indeed, monotonicity combined with other appropriate assumptions allows one to recover the unknown structural functional from the regression quantiles. Specifically, if $G^{-1}(\tau \mid x)$ denotes the τ-conditional quantile of Y given $X = x$, and $Y = m(X, A)$ is the structural equation, then strict monotonicity of $m(x, \cdot)$, combined with full independence of A and X (strict exogeneity of X) and a normalization, allows recovery of m as $m(x, a) = G^{-1}(a \mid x)$ for all (a, x).

Clearly, however, scalar monotonicity is a strong assumption. As Hoderlein and Mammen (2007) argue, some of its implications in certain applications, such as consumer demand, may be unpalatable. In particular, monotonicity implies that the conditional rank order of the individual must be preserved under interventions to x. For example, under independence, if individual j attains the conditional median food consumption $G^{-1}(.5 \mid x_j)$, then he would remain at the conditional median for all other values of x.

The existence of the regression quantile representation makes it impossible to test for monotonicity without further information. One source of such information is that provided by panel data, as exploited by Su, Hoderlein, and White (2011). Here, we follow a different strategy, using additional cross-section information. In particular, we assume there are random variables Z that are excluded from the structural function, and conditional on which X is independent of A (for this, we use the shorthand $X \perp A \mid Z$). Chalak and White (2011a) call such variables Z "conditioning instruments," and say that X is "conditionally exogenous" with respect to A given Z. For example, X could be randomly assigned given Z, or Z could be a (generalized) proxy for A. Alternatively, Z could be the unobservable in a first-stage equation that relates X to an instrument S. For instance, the first stage could be $X = \phi(S, Z)$, where $S \perp (A, Z)$ and ϕ is strictly monotonic in Z, with Z appropriately estimated, as in IN.

Our test is based on the fact that, under mild assumptions, the availability of Z enables one to construct multiple consistent estimators of A. If scalar monotonicity holds, then these estimators will be close to one another; otherwise, they will diverge. We develop asymptotic distribution theory and propose bootstrap methods suitable for testing whether the differences between multiple estimators of A accord with the null of scalar monotonicity or whether this null must be rejected.

3 Identification Under Monotonicity

In this section, we first review a pioneering identification result of AM, their theorem 4.1. This relies on a certain exogeneity condition plausible for panel data. We then present a complement to AM's result, relying on a related but different exogeneity condition often plausible not only for panels, but also for pure cross-sections or time-series. This motivates structural estimators complementary to those of AM, as well as natural specification testing procedures.

We begin by stating the needed assumptions. We modify AM's notation somewhat, but maintain the content. First, we make explicit the data generating process (DGP).

Assumption A.0 $(\Omega, \mathcal{F}, P)$ is a complete probability space on which are defined the finitely dimensioned random vectors X and Z and the random scalar A.

Typically, X and Z are observable, and A is unobservable. We write the supports of X, Z, and A as $\mathcal{X}$, $\mathcal{Z}$, and $\mathcal{A}$, respectively. Below we require that A has a continuous distribution. We permit, but do not require, X and Z to be continuously distributed; either or both may have a finite or countable discrete distribution for now.

Assumptions A.1 - A.5 below correspond to AM's Assumptions 4.1 - 4.5. To specify the structural relationship, we say m is *product measurable* if $m : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ is a measurable function on the product measurable space $(\mathcal{X} \times \mathcal{A}, \sigma(X) \otimes \sigma(A))$. This ensures that $m(x, \cdot)$ is measurable$-\sigma(A)$ for each x in $\mathcal{X}$ and $m(\cdot, a)$ is measurable$-\sigma(X)$ for each a in $\mathcal{A}$.

Assumption A.1 There exists a product measurable function $m : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ such that Y is structurally determined as $Y = m(X, A)$.

AM also assume that $A = \Upsilon(\varepsilon)$, for some measurable function Υ and random vector ε, but as Su, Hoderlein, and White (2011) show, this imposes essentially no further restrictions on the structure generating Y. Observe that Z is excluded as a driver of Y. The product measurability of m ensures that Y is a random variable. We let $\mathcal{Y}$ denote the support of Y.

We call Y the "structural response" or simply the *response*. Similarly, we call m the "structural response function", or simply the "structural function" or the "response function".

Next, we impose monotonicity.

Assumption A.2 For all $x \in \mathcal{X}$, $m(x, \cdot)$ is strictly increasing.

AM refer to the next assumption as a normalization. Below, we discuss this further.

Assumption A.3 There exists $\bar{x}$ such that $m(\bar{x}, a) = a$ for all $a \in \mathcal{A}$.

In general, X and A can be correlated or otherwise dependent. That is, X can be endogenous. AM's next condition accommodates endogeneity by imposing a certain conditional form of exogeneity. Following AM, we let $f(a \mid x, z)$ define the conditional probability density function (PDF) of A given $X = x$, $Z = z$.

Assumption A.4 There exist measurable functions ζ_1 and ζ_2 such that $f(a \mid x, \zeta_1(x)) = f(a \mid \bar{x}, \zeta_2(x))$ for all $(a, x) \in \mathcal{A} \times \mathcal{X}$.

Finally, AM require that A is conditionally (and unconditionally) continuously distributed.

Assumption A.5 For all $(a, x, z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{Z}$, $f(a \mid x, z)$ is strictly positive.

The conditional cumulative distribution function (CDF) of Y given $X = x$, $Z = z$ is defined by $G(y \mid x, z) \equiv P[Y \leq y \mid X = x, Z = z]$. A.5 ensures that G is invertible. We also let $G_{Y|X}(\cdot \mid x)$ be the conditional CDF for Y given $X = x$ and $F_{A|X}(\cdot \mid x)$ the conditional CDF of A given $X = x$.

3.1 Identification via Conditional Quantiles

AM's identification result represents m using the conditional quantiles of Y:

Theorem 3.1 *Suppose Assumptions A.0 - A.5 hold. Then*

$$m(x, a) = G^{-1}(G(a \mid \bar{x}, \zeta_2(x)) \mid x, \zeta_1(x)) \qquad \forall (x, a) \in \mathcal{X} \times \mathcal{A},$$
$$F_{A|X}(a \mid x) = G_{Y|X}[\, G^{-1}(G(a \mid \bar{x}, \zeta_2(x)) \mid x, \zeta_1(x)) \mid x] \qquad \forall (x, a) \in \mathcal{X} \times \mathcal{A},$$
$$A = G^{-1}[G(Y \mid X, \zeta_1(X)) \mid \bar{x}, \zeta_2(X)]).$$

The final equality, providing a representation for A, is implicit in AM. We make this explicit here, as recovering A facilitates important analyses. Specifically, this representation could be used for estimating the control variables used in IN or for conducting specification tests.

Although AM provide discussion at the bottom of p.1073 that seems to suggest that $A.3$ imposes additional structure, $A.3$ is in fact redundant in a precise sense. Specifically, $A.1$ and $A.2$ ensure that for *every* $\bar{x}$ in $\mathcal{X}$, there is a function, say $\bar{m}$, for which $A.1 - A.3$ hold. This is a consequence of the following result.

Theorem 3.2 *Let Assumption* $A.1$ *hold. Suppose there exists* $\bar{x} \in \mathcal{X}$ *such that* $m(\bar{x}, \cdot)$ *is strictly increasing, and let* $\mathcal{V} := \{y : y = m(\bar{x}, a),\ a \in \mathcal{A}\}$. *Then there exists a product measurable function* $\bar{m} : \mathcal{X} \times \mathcal{V} \to \mathbb{R}$ *such that*

(*a*) *for each* v *in* $\mathcal{V}$, *there exists* a *in* $\mathcal{A}$ *such that* $m(x, a) = \bar{m}(x, v)$ *for every* x *in* $\mathcal{X}$; *and for each* a *in* $\mathcal{A}$, *there exists* v *in* $\mathcal{V}$ *such that* $m(x, a) = \bar{m}(x, v)$ *for every* x *in* $\mathcal{X}$;

(*b*) *for any* x *in* $\mathcal{X}$ *such that* $m(x, \cdot)$ *is strictly increasing on* $\mathcal{A}$, $\bar{m}(x, \cdot)$ *is strictly increasing on* $\mathcal{V}$. *In particular,* $\bar{m}(\bar{x}, \cdot)$ *is strictly increasing on* $\mathcal{V}$;

(*c*) *for each* v *in* $\mathcal{V}$, $\bar{m}(\bar{x}, v) = v$.

It follows that if $m(x, \cdot)$ is strictly increasing for all x in $\mathcal{X}$, as ensured by $A.2$, then *any* point in $\mathcal{X}$ can play the role of $\bar{x}$ in $A.3$. That is, once Assumptions $A.1$ and $A.2$ hold, then for each $\bar{x}$ in $\mathcal{X}$ the function $\bar{m}(\cdot, v) \equiv m(\cdot, \bar{v}^{-1}(v))$ also satisfies $A.1 - A.3$, where $\bar{v}(\cdot) \equiv m(\bar{x}, \cdot)$. The proof of this result remains valid if we replace "strictly increasing" with "invertible". We focus on the strictly increasing case, as we will not need the greater generality of invertibility.

Theorem 3.2 implies that the main role of $\bar{x}$ is to ensure that $A.4$ holds. Thus, $A.3$ and $A.4$ in Theorem 3.1 can be replaced by

Assumption A.4′ There exist $\bar{x}$ in $\mathcal{X}$ and measurable functions ζ_1 and ζ_2 such that $f(a \mid x, \zeta_1(x)) = f(a \mid \bar{x}, \zeta_2(x))$ for all a, x in $\mathcal{A} \times \mathcal{X}$.

Given this $\bar{x}$, one can replace m with $\bar{m}$, with $\bar{m}(\bar{x}, v) = v$ necessarily holding for all v in $\mathcal{V}$. With this normalization understood, one can drop reference to $\bar{m}$ and simply work with m.

AM provide detailed discussion of an exchangeability condition, useful for panel data, that makes plausible the particular choices $\zeta_1(x) = \bar{x}$, $\zeta_2(x) = x$, where the choice of $\bar{x}$ is natural and may differ across sample observations. With differing $\bar{x}$, however, recovered values for A are *not* comparable across observations, due to the lack of a common normalization for unobservables. This can create difficulties in interpretation and may pose other challenges in applications.

3.2 Characterizing the Conditional Quantile Representation

In pure cross-section or time-series data, exchangeability is not a natural assumption. Further, even if they exist, choices for $\bar{x}$, ζ_1, and ζ_2 are often not obvious. Nor is attempting to estimate $\bar{x}$, ζ_1, and ζ_2 appealing, due to difficulties with their identification.

Fortunately, however, there is an alternate assumption that permits X to be endogenous and that can be plausible in panels and elsewhere. This is a version of AM's conditional exogeneity Assumption 2.1, that X is independent of A given the "covariates" or "control variables" Z :

Assumption B.1 $X \perp A \mid Z$, where Z is not measurable-$\sigma(X)$.

An advantage of $B.1$ is that it allows a choice of an analog of $\bar{x}$ in $A.3$ that is common across sample observations. This makes A comparable across observations. We will exploit this comparability to identify A and to construct computationally feasible specification tests. Further, by requiring Z not to be solely a function of X, as it is in $A.4'$, we permit important flexibility for recovering objects of interest. See, for example, Hoderlein and Mammen (2007, 2009) and IN. White and Lu (2011) and Chalak and White (2011b) explicitly discuss structures ensuring B.1 where Z is not a function of X.

We also impose a more direct variant of Assumption A.5.

Assumption B.2 For each (x, z) in $\mathcal{X} \times \mathcal{Z}$, $G(\cdot \mid x, z)$ is invertible.

Our first main result characterizes the conditional quantile representation (CQR) of the structural function. We state this without assuming monotonicity. Thus, for $(x, y) \in \mathcal{X} \times \mathcal{Y}$, let $m_x^{-1}\{y\} = \{\alpha \in \mathcal{A} : m(x, \alpha) \leq y\}$ denote the pre-image of the interval $(-\infty, y]$ under $m(x, \cdot)$, and let $\bar{m}_x^{-1}\{y\} = \{a \in \mathcal{A} : m(x, a) = y\}$ denote the pre-image of the point y.

Theorem 3.3 *Suppose Assumptions* $A.0$, $A.1$, $B.1$, *and* $B.2$ *hold. For each* $(a, x, \tilde{x}, z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{X} \times \mathcal{Z}$, *the following are equivalent:*

(i) $m(x, a) = G^{-1}(G(m(\tilde{x}, a) \mid \tilde{x}, z) \mid x, z)$;

(ii) $a \in \bar{m}_x^{-1}\{G^{-1}[G(m(x, a) \mid x, z) \mid \tilde{x}, z]\}$;

(*iii*) $P[A \in m_x^{-1}\{m(x,a)\} \mid Z = z] = P[A \in m_{\tilde{x}}^{-1}\{m(\tilde{x},a)\} \mid Z = z]$.

Result (*i*) is the conditional quantile representation of the structural function. Result (*ii*) gives a partial identification result for the unobservable a. If the referenced set is a singleton, we have point identification. Result (*iii*) gives a necessary and sufficient condition for CQR and for partial identification of the unobservable. Assumption $A.2$ ensures this for all $(a, x, \tilde{x}, z)$, as it ensures $m_x^{-1}\{m(x,a)\} = m_{\tilde{x}}^{-1}\{m(\tilde{x},a)\} = (-\infty, a]$ for all $(a, x, \tilde{x})$. $A.2$ also ensures point identification of a. Nevertheless, there are non-monotone functions m for which (*iii*) and therefore CQR always hold. A non-monotone (indeed, non-invertible) example is

$$m(x,a) = (a - x)\,1\{0 \le a - x < .5\} + (1.5 - (a - x))\,1\{.5 \le a - x \le 1\},$$

where $x \in [0,1]$ and $A \mid Z \sim U[0,2]$, as some calculation verifies. Such non-monotone structures are clearly exceptional; indeed, we conjecture they are *shy* (see Corbae et al., 2009, pp.545-547). Shyness is the function-theoretic analog of being a subset of a set of measure zero.

Because economic theory often motivates or justifies strict monotonicity but is typically uninformative about the characterizing condition (*iii*), our focus here is on strict monotonicity. The exceptional alternatives to $A.2$ ensuring CQR become part of the "implicit null"; we further discuss this below and in the Appendix.

With the same $\tilde{x}$ for all sample observations, say[2] $\tilde{x} = x^*$, and the normalization $a = m(x^*, a)$, A becomes comparable across observations, in line with Theorem 3.2 and its discussion. For example, x^* can be the vector of medians of X. A complement to AM's theorem 4.1 is:

Corollary 3.4 *Let $A.0 - A.2$, $B.1$, and $B.2$ hold. Then with the normalization $a = m(x^*, a)$,*

$$m(x,a) = G^{-1}(G(a \mid x^*, z) \mid x, z) \qquad \forall (a,x,z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{Z}, \tag{3.1}$$

$$A = G^{-1}(G(Y \mid X, z) \mid x^*, z)) \qquad \forall z \in \mathcal{Z}, \tag{3.2}$$

and for all $(a,x,z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{Z}$,

$$F_{A|Z}(a \mid z) = F_{A|X,Z}(a \mid x, z) = G(\, m(x,a) \mid x, z), \tag{3.3}$$

$$F_{A|X}(a, x) = G_{Y|X}[\, G^{-1}(G(a \mid x^*, z) \mid x, z) \mid x].$$

The normalization thus ensures that the unconditional distribution of A is that of Y given $X = x^*$. In the strictly exogenous case, where covariates Z are absent (e.g., Su, Hoderlein, and White, 2011), the unconditional distribution of A is typically normalized to be standard uniform.

[2] In what follows, we always let x^* denote a choice common across sample observations, reserving $\bar{x}$ to denote a choice that may vary across observations, as in AM.

4 Heuristics of Estimation and Specification Testing

4.1 Estimation

Corollary 3.4 provides the basis for convenient estimators complementary to those proposed by AM. Because this result ensures that $m(x,a) = G^{-1}(G(a \mid x^*, z) \mid x, z)$ for given x^* and any z, one can estimate $m(x,a)$ as

$$\hat{m}_z(x,a) = \hat{G}^{-1}(\hat{G}(a \mid x^*, z) \mid x, z)),$$

for any choice of z, where $\hat{G}$ and $\hat{G}^{-1}$ are any convenient estimators of G and G^{-1} respectively. (One might, but need not, obtain $\hat{G}^{-1}$ from $\hat{G}$ by inversion or vice-versa.) Estimators dependent on z may exhibit undesirable variability; averaging over multiple z's may provide more reliable results. Such estimators have the form

$$\hat{m}_H(x,a) = \int \hat{G}^{-1}(\hat{G}(a \mid x^*, z) \mid x, z)) \, dH(z),$$

where H is a known or estimated distribution supported on $\mathcal{Z}_0 \subseteq \mathcal{Z}$, say, like the uniform or the sample distribution of Z. In the next section we examine the properties of $\hat{m}_H(x,a)$ constructed using p-th order local polynomial estimators $\hat{G}_{p,b}$ and $\hat{G}_{p,b}^{-1}$ using a bandwidth b.

As one should expect, even when monotonicity (A.2) fails, $\hat{m}_H$ is nevertheless generally consistent for[3]

$$m_H^*(x,a) \equiv \int G^{-1}(G(a \mid x^*, z) \mid x, z)) \, dH(z).$$

Thus, m_H^* is a "pseudo-true" value, meaningful regardless of misspecification, with $m = m_H^*$ under correct specification, i.e., when the conditions of Corollary 3.4 hold.

Similarly, one can estimate A as

$$\hat{A}_z = \hat{G}^{-1}(\hat{G}(Y \mid X, z) \mid x^*, z)),$$

for given x^* and any choice of z. Averaging over multiple z's gives estimators of the form

$$\hat{A}_H = \int \hat{G}^{-1}(\hat{G}(Y \mid X, z) \mid x^*, z)) \, dH(z).$$

Alternative estimators of A can be obtained by inverting $\hat{m}_H(X,A)$, yielding

$$\tilde{A}_H = \hat{m}_H^{-1}(X,Y) \equiv \inf \{a : \hat{m}_H(X,a) \geq Y\}.$$

We analyze $\hat{A}_H$ and $\tilde{A}_H$ constructed using $\hat{G}_{p,b}$ and $\hat{G}_{p,b}^{-1}$ in the next section.

[3] If $\hat{m}_H$ is defined using a sampling distribution, the distribution H appearing in the next expression is interpreted as the corresponding population distribution.

Parallel to the situation for $\hat{m}_H$, regardless of misspecification, $\hat{A}_H$ and $\tilde{A}_H$ are generally consistent for pseudo-true values

$$\begin{aligned} A_H^* &\equiv \int G^{-1}(G(Y \mid X, z) \mid x^*, z))\, dH(z), \\ A_H^\dagger &\equiv m_H^{*-1}(X, Y) \equiv \inf\{a : m_H^*(X, a) \geq Y\}. \end{aligned}$$

Under correct specification, $A_H^* = A_H^\dagger = A$.

4.2 Specification Testing

Corollary 3.4 motivates constructing specification tests by comparing various estimators of A, as there are multiple consistent estimators of A under correct specification. Under $A.0$, $A.1$, $B.1$, and $B.2$, the failure of these estimators to coincide asymptotically signals non-monotonicity. Su, Hoderlein, and White (2011) take a similar approach for the panel data case without covariates.

Below we will study the asymptotic properties of the test statistic

$$\begin{aligned} \hat{J}_n &\equiv b^{d_X} \sum_{i=1}^{n} (\hat{A}_{1,i} - \hat{A}_{2,i})^2 \pi(X_i, Y_i) \\ &= b^{d_X} \sum_{i=1}^{n} \left\{ \int \hat{G}^{-1}\left(\hat{G}(Y_i | X_i, z) | x^*, z\right) d\Delta(z) \right\}^2 \pi_i, \end{aligned} \tag{4.1}$$

where $b \equiv b_n$ is a suitable bandwidth; d_X is the dimension of X; $\hat{A}_{j,i} \equiv \int \hat{G}^{-1}(\hat{G}(Y_i \mid X_i, z) \mid x^*, z))$ $dH_j(z)$, $j = 1, 2$; $\hat{G}$ and $\hat{G}^{-1}$ are based on a sample of observations $\{X_i, Y_i, Z_i\}_{i=1}^n$ distributed identically to (X, Y, Z); and $\pi(\cdot, \cdot)$ is a nonnegative weight function with support on a compact subset $\mathcal{X}_0 \times \mathcal{Y}_0$ of $\mathcal{X} \times \mathcal{Y}$, $\pi_i \equiv \pi(X_i, Y_i)$. The weight π_i downweights observations for which $\hat{G}(Y_i \mid X_i, z)$ is close either to 0 or 1, so that $\hat{G}^{-1}$ can not be accurately obtained.

Finally, $\Delta(z) \equiv H_1(z) - H_2(z)$, where H_1 and H_2 are distinct distribution functions having supports $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, each a subset of a compact subset $\mathcal{Z}_0$ of $\mathcal{Z}$. These supports can be disjoint. As it turns out, we can allow H_1 and H_2 to depend on the data without altering the first-order asymptotic distribution of the test statistic $\hat{J}_n$. Thus, for notational simplicity, we do not distinguish between estimated and population H's. Different choices for Δ focus the power of the test in different directions. Our power analysis below suggests ways of ensuring that $\hat{J}_n$ can consistently detect violations of Theorem 3.3(iii). In fact, multiple choices of Δ may well be of interest, as the nonparametric context rules out a globally optimal test. We leave treatment of multiple Δ's aside here, as this is straightforward given the theory for a single choice of Δ.

As we show below, T_n, a standardized version of $\hat{J}_n$ constructed using $\hat{G} = \hat{G}_{p,b}$ and $\hat{G}^{-1} = \hat{G}_{p,b}^{-1}$, is asymptotically standard normal under correct specification. If T_n is incompatible with this distribution, we have evidence against correct specification. As we also show, this test has power against Pitman local alternatives converging to zero at rate $n^{-1/2}b^{-d_X/2}$ and is consistent against the class of global alternatives: those structures violating Theorem 3.3(iii).

Despite having a standard normal asymptotic distribution, T_n requires use of the bootstrap to compute useful critical values. This is an extreme but feasible computational task, making this statistic relevant for practical applications.[4]

4.3 Multiple Tests for Misspecification

Our test is explicitly designed to detect failures of monotonicity ($A.2$). Nevertheless, if conditional exogeneity ($B.1$) fails, then T_n may also detect this failure. If indeed conditional exogeneity is in question, we suggest applying the proper tool for the job. Specifically, one can and should test $B.1$ directly by applying any of a variety of available tests (e.g., White and Chalak, 2010) that do not rely on (and are not sensitive to) the monotonicity assumption. For those cases where both $B.1$ and $A.2$ are in question, one can use T_n together with a conditional exogeneity test to perform a *multiple* test of misspecification.

The joint null hypothesis for the multiple test is $\mathbb{H}_0' : \mathbb{H}_{01}'$ and $\mathbb{H}_{02}'$, where $\mathbb{H}_{01}'$ is that $B.1$ holds and $\mathbb{H}_{02}'$ is that $A.2$ holds. Let p_1 be the p-value associated with a test for $B.1$ that does not maintain monotonicity, and let p_2 be the p-value associated with our T_n statistic. By the Bonferroni inequality, we reject $\mathbb{H}_0'$ at level $p \leq 2\min(p_1, p_2)$. Alternatively, one can apply recent methods of King, Zhang, and Akram (2011) to directly estimate p for this multiple test.

Under rejection, the pattern of values of (p_1, p_2) provides diagnostic information about the plausible sources of misspecification. Specifically, if p_2 is small but not p_1, this indicates non-monotonicity only. If p_1 and p_2 are both small, this indicates failure of conditional exogeneity, but is silent about non-monotonicity. If p_1 is small, but not p_2, this again indicates failure of conditional exogeneity, but is silent about non-monotonicity. In this last case, the failure of p_2 to be small reflects the plausibly lower power associated with the T_n statistic in this context, as the power of this test is now spread across failures of both $B.1$ and $A.2$, whereas the conditional exogeneity test is more tightly focused. The fact that the multiple test is silent about non-monotonicity in the latter two cases is not particularly significant. When $B.1$ fails, the adverse consequences are much more severe than when monotonicity fails, as the discussion of the introduction suggests.

5 Asymptotics for Estimation and Inference

5.1 Local polynomial estimators

Throughout, we rely on local polynomial regression to estimate various unknown population objects. Let $u \equiv (x', z')' = (u_1, ..., u_d)'$ be a $d \times 1$ vector, $d \equiv d_X + d_Z$, where x is $d_X \times 1$ and z is $d_Z \times 1$. Let $\boldsymbol{j} \equiv (j_1, ..., j_d)$ be a $d \times 1$ vector of non-negative integers. Following Masry

[4] The invariance properties of Theorem 3.3 and Corollary 3.4 support construction of other specification tests. Nevertheless, each requires its own analysis and involves significant computational challenges. We thus leave their study to other work.

(1996), we adopt the notation: $u^{\boldsymbol{j}} \equiv \mathbf{\Pi}_{i=1}^{d} u_i^{j_i}$, $\boldsymbol{j}! \equiv \mathbf{\Pi}_{i=1}^{d} j_i!$, $|\boldsymbol{j}| \equiv \sum_{i=1}^{d} j_i$, and $\sum_{0\leq|\boldsymbol{j}|\leq p} \equiv \sum_{k=0}^{p} \sum_{j_1=0}^{k} \cdots \sum_{j_d=0}^{k}{}_{j_1+\cdots+j_d=k}$.

We first describe the p-th order local polynomial estimator $\hat{G}_{p,b}(y|x,z)$ of $G(y|x,z)$. The subscript $b = b_n$ is a bandwidth parameter. Let $U_i \equiv (X_i', Z_i')'$ so that $U_i - u = ((X_i - x)', (Z_i - z)')'$. Given observations $\{(Y_i, U_i), i = 1, ..., n\}$, $\hat{G}_{p,b}(y|x,z)$ can be obtained as the minimizing intercept term in the following minimization problem

$$\min_{\boldsymbol{\beta}} n^{-1} \sum_{i=1}^{n} \left[\mathbf{1}\{Y_i \leq y\} - \sum_{0\leq|\boldsymbol{j}|\leq p} \beta_{\boldsymbol{j}}' \left((U_i - u)/b\right)^{\boldsymbol{j}} \right]^2 K_b(U_i - u), \tag{5.1}$$

where $\boldsymbol{\beta}$ stacks the $\beta_{\boldsymbol{j}}$'s $(0 \leq |\boldsymbol{j}| \leq p)$ in lexicographic order (with $\beta_{\mathbf{0}}$, indexed by $\mathbf{0} \equiv (0, ..., 0)$, in the first position, the element with index $(0, 0, ..., 1)$ next, etc.) and $K_b(\cdot) \equiv K(\cdot/b)/b$, with $K(\cdot)$ a symmetric probability density function (PDF) on $\mathbb{R}^d$.

Let $N_{p,l} \equiv (l+d-1)!/(l!(d-1)!)$ be the number of distinct d-tuples $\boldsymbol{j}$ with $|\boldsymbol{j}| = l$. In the above estimation problem, this denotes the number of distinct lth order partial derivatives of $G(y|u)$ with respect to u. Let $N_p \equiv \sum_{l=0}^{p} N_{p,l}$. Let $\mu_p(\cdot)$ be a stacking function such that $\mu_p((U_i - u)/b)$ denotes an $N_p \times 1$ vector that stacks $((U_i - u)/b)^{\boldsymbol{j}}$, $0 \leq |\boldsymbol{j}| \leq p$, in lexicographic order (e.g., $\mu_p(u) = (1, u')'$ when $p = 1$). Let $\mu_{p,b}(u) \equiv \mu_p(u/b)$. Then $\hat{G}_{p,b}(y|u) = e_{1,p}' \hat{\boldsymbol{\beta}}(y|u)$ where

$$\hat{\boldsymbol{\beta}}(y|u) = [\mathbf{S}_{p,b}(u)]^{-1} n^{-1} \sum_{i=1}^{n} K_b(U_i - u) \mu_{p,b}(U_i - u) \ \mathbf{1}\{Y_i \leq y\}, \tag{5.2}$$

$$\mathbf{S}_{p,b}(u) \equiv n^{-1} \sum_{i=1}^{n} K_b(U_i - u) \mu_{p,b}(U_i - u) \mu_{p,b}(U_i - u)', \tag{5.3}$$

and $e_{1,p} \equiv (1, 0, ..., 0)'$ is an $N_p \times 1$ vector with 1 in the first position and zeros elsewhere.

We also use p-th order local polynomial estimation to estimate $G^{-1}(\tau|u)$, the τth conditional quantile function of Y_i given $U_i = u$. We denote this $\hat{G}_{p,b}^{-1}(\tau|u)$. Let $\rho_\tau(u) \equiv u(\tau - \mathbf{1}\{u \leq 0\})$ be the "check" function. We obtain $\hat{G}_{p,b}^{-1}(\tau|u)$ as the minimizing intercept term in the weighted quantile estimation problem

$$\min_{\boldsymbol{\alpha}} n^{-1} \sum_{i=1}^{n} \rho_\tau \left(Y_i - \sum_{0\leq|\boldsymbol{j}|\leq p} \alpha_{\boldsymbol{j}}' \left((U_i - u)/b\right)^{\boldsymbol{j}} \right) K_b(U_i - u), \tag{5.4}$$

where $\boldsymbol{\alpha}$ stacks the $\alpha_{\boldsymbol{j}}$'s $(0 \leq |\boldsymbol{j}| \leq p)$ in lexicographic order. Alternatively, one can invert $\hat{G}_{p,b}(\cdot|u)$ to obtain an estimator of $G^{-1}(\cdot|u)$, as in Cai (2002). We do not pursue this here.

In the next subsection, we study the asymptotic properties of the estimators $\hat{m}_H$, $\hat{A}_H$, and $\tilde{A}_H$ defined above, constructed using the local polynomial estimators $\hat{G}_{p,b}$ and $\hat{G}_{p,b}^{-1}$ just defined.

5.2 Asymptotic properties of $\hat{m}_H(x,a)$, $\hat{A}_H$, and $\tilde{A}_H$

We use the following assumptions.

Assumption C.1 Let $W_i \equiv (X_i', Y_i, Z_i')'$, $i = 1, 2, ...$ be IID random variables on $(\Omega, \mathcal{F}, P)$, with (X_i, Z_i) distributed identically to (X, Z) in Assumption A.0.

Let $g(u)$ and $g(y|u)$ denote the joint PDF of U_i and the conditional PDF of Y_i given $U_i = u$, respectively. Let $\mathcal{U} \equiv \mathcal{X} \times \mathcal{Z}$ and $\mathcal{U}_0 \equiv \mathcal{X}_0 \times \mathcal{Z}_0$. Let $\mathcal{Y}$ denote the (common) support of Y_i, and let $\mathcal{Y}_0 \equiv [\underline{y}, \bar{y}]$ for finite real numbers $\underline{y}$, $\bar{y}$.

Assumption C.2 (i) $g(u)$ is continuous in $u \in \mathcal{U}$, and $g(y|u)$ is continuous in $(y, u) \in \mathcal{Y} \times \mathcal{U}$.

(ii) There exist $C_1, C_2 \in (0, \infty)$ such that $C_1 \leq \inf_{u \in \mathcal{U}_0} g(u) \leq \sup_{u \in \mathcal{U}_0} g(u) \leq C_2$ and $C_1 \leq \inf_{(y,u) \in \mathcal{Y}_0 \times \mathcal{U}_0} g(y|u) \leq \sup_{(y,u) \in \mathcal{Y}_0 \times \mathcal{U}_0} g(y|u) \leq C_2$.

Assumption C.3 (i) There exist $\underline{\tau}, \bar{\tau} \in (0, 1)$ such that $\underline{\tau} \leq \inf_{u \in \mathcal{U}_0} G\left(\underline{y}|u\right) \leq \sup_{u \in \mathcal{U}_0} G(\bar{y}|u) \leq \bar{\tau}$ and $\underline{\tau} \leq \inf_{z \in \mathcal{Z}_0} G\left(\underline{y}|x^*, z\right) \leq \sup_{z \in \mathcal{Z}_0} G(\bar{y}|x^*, z) \leq \bar{\tau}$.

(ii) $G(\cdot|u)$ is equicontinuous: $\forall \epsilon > 0$, $\exists \delta > 0 : |y - \tilde{y}| < \delta \Rightarrow \sup_{u \in \mathcal{U}_0} |G(y|u) - G(\tilde{y}|u)| < \epsilon$. For each $y \in \mathcal{Y}_0$, $G(y \mid \cdot)$ is Lipschitz continuous on $\mathcal{U}_0$ and has all partial derivatives up to order $p + 1$, $p \in \mathbb{N}$.

(iii) Let $D^{\mathbf{j}} G(y|u) \equiv \partial^{|\mathbf{j}|} G(y|u) / \partial^{j_1} u_1 ... \partial^{j_d} u_d$. For each $y \in \mathcal{Y}_0$, $D^{\boldsymbol{j}} G(y \mid \cdot)$ with $|\boldsymbol{j}| = p+1$ is uniformly bounded and Lipschitz continuous on $\mathcal{U}_0$: for all u, $\tilde{u} \in \mathcal{U}_0$, $|D^{\boldsymbol{j}} G(y \mid u) - D^{\boldsymbol{j}} G(y \mid \tilde{u})| \leq C_3 ||u - \tilde{u}||$ for some $C_3 \in (0, \infty)$ where $\|\cdot\|$ is the Euclidean norm.

(iv) For each $u \in \mathcal{U}_0$ and for all y, $\tilde{y} \in \mathcal{Y}_0$, $|D^{\boldsymbol{j}} G(y \mid u) - D^{\boldsymbol{j}} G(\tilde{y} \mid u)| \leq C_4 \ |y - \tilde{y}|$ for some $C_4 \in (0, \infty)$ where $|\boldsymbol{j}| = p + 1$.

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

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

(iii) Let $\mathbf{K}_{\boldsymbol{j}}(u) \equiv u^{\boldsymbol{j}} K(u)$ for all $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p+1$. For some finite constants σ_K, $\bar{\sigma}_1$, and $\bar{\sigma}_2$, either $K(\cdot)$ is compactly supported such that $K(u) = 0$ for $\|u\| > \sigma_K$, and $|\mathbf{K}_{\boldsymbol{j}}(u) - \mathbf{K}_{\boldsymbol{j}}(\tilde{u})| \leq \bar{\sigma}_2 \|u - \tilde{u}\|$ for any $u, \tilde{u} \in \mathbb{R}^d$ and for all $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p + 1$; or $K(\cdot)$ is differentiable, $\|\partial \mathbf{K}_{\boldsymbol{j}}(u) / \partial u\| \leq \bar{\sigma}_1$, and for some $\iota_0 > 1$, $|\partial \mathbf{K}_{\boldsymbol{j}}(u) / \partial u| \leq \bar{\sigma}_1 \|u\|^{-\iota_0}$ for all $\|u\| > \sigma_K$ and for all $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p + 1$.

Assumption C.5 The distribution function $H(z)$ admits a PDF $h(z)$ continuous on $\mathcal{Z}_0$.

Assumption C.6 As $n \to \infty$, $b \to 0$, $b^{p+1-d_Z/2} \to 0$, and $nb^{2(p+1)+d_X/2} \to c_0 \in [0, \infty)$. There exists some $\epsilon^* > 0$ such that $n^{1-\epsilon^*} b^{d+2d_Z} \to \infty$.

The IID requirement of Assumption C.1 is standard in cross-section studies. Nevertheless, the asymptotic theory developed here can be readily extended to weakly dependent time series. To keep the results uncluttered, we leave the time-series case for study elsewhere. Assumption C.2 is standard for nonparametric local polynomial estimation of conditional mean and density. If U_i has compact support $\mathcal{U}$ and $g(u)$ is bounded away from zero on $\mathcal{U}$, it is possible to choose $\mathcal{U}_0 = \mathcal{U}$. Assumptions C.3-C.4 ensure the uniform consistency for our local polynomial estimators, based on results of Masry (1996) and Hansen (2008). Assumption C.5 makes Z continuously distributed, simplifying the analysis. Assumption C.6 appropriately restricts the choices of bandwidth sequence and the order of local polynomial regression.

To proceed, arrange the $N_{p,l}$ d-tuples as a sequence in lexicographical order, so that $\phi_l(1) \equiv (0,0,...,l)$ is the first element and $\phi_l(N_l) \equiv (l,0,...,0)$ is the last, and let ϕ_l^{-1} be the mapping inverse to ϕ_l. For each $\boldsymbol{j}$ with $0 \leq |\boldsymbol{j}| \leq 2p$, let $\mu_{\boldsymbol{j}} = \int_{\mathbb{R}^d} u^{\boldsymbol{j}} K(u)du$. Define the $N_p \times N_p$ matrix $\mathbb{S}_p$ and the $N_p \times N_{p,p+1}$ matrix $\mathbb{B}_p$ respectively by

$$\mathbb{S}_p \equiv \begin{bmatrix} \mathbb{M}_{0,0} & \mathbb{M}_{0,1} & \cdots & \mathbb{M}_{0,p} \\ \mathbb{M}_{1,0} & \mathbb{M}_{1,1} & \cdots & \mathbb{M}_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{M}_{p,0} & \mathbb{M}_{p,1} & \cdots & \mathbb{M}_{p,p} \end{bmatrix} \text{ and } \mathbb{B}_p = \begin{bmatrix} \mathbb{M}_{0,p+1} \\ \mathbb{M}_{1,p+1} \\ \vdots \\ \mathbb{M}_{p,p+1} \end{bmatrix}, \tag{5.5}$$

where $\mathbb{M}_{i,j}$ are $N_{p,i} \times N_{p,j}$ matrices whose (l,s) element is $\mu_{\phi_i(l)+\phi_j(s)}$. In addition, we arrange $D^{\mathbf{j}}G(a|u)/\mathbf{j}!$ with $|\mathbf{j}| = p+1$ as an $N_{p+1} \times 1$ vector, $\mathbf{G}_{p+1}(a|u)$, in lexicographical order.

Let $\mathcal{A}_H = \{a : m_H^*(x,a) = y,\, x \in \mathcal{X}_0, y \in \mathcal{Y}_0\}$. The asymptotic behavior of $\hat{m}_H(x,a)$ follows:

Theorem 5.1 *Suppose Assumptions C.1-C.6 hold. Let $x^* \in \mathcal{X}_0$ and $(x,a) \in \mathcal{X}_0 \times \mathcal{A}_H$. Then*

$$\sqrt{nb^{d_X}}\left(\hat{m}_H(x,a) - m_H^*(x,a) - B_m(x,a;x^*)\right) \xrightarrow{d} N\left(0, \sigma_m^2(x,a;x^*)\right), \quad \textit{where}$$

$$B_m(x,a;x^*) \equiv b^{p+1} e_{1,p}' \mathbb{S}_p^{-1} \mathbb{B}_p \int \left[\frac{\mathbf{G}_{p+1}(a|x^*,z)}{g\left(G^{-1}\left(G(a|x^*,z)|x,z\right)|x,z\right)} + \mathbf{G}_{p+1}^{-1}\left(G(a|x^*,z)|x,z\right)\right] dH(z), \tag{5.6}$$

$$\sigma_m^2(x,a;x^*) \equiv \kappa_{1p} \int \frac{G(a|x^*,z)\left[1-G(a|x^*,z)\right] h(z)^2}{g\left(G^{-1}\left(G(a|x^*,z)|x,z\right)|x,z\right)^2} \left[\frac{1}{g(x^*,z)} + \frac{1}{g(x,z)}\right] dz, \tag{5.7}$$

and $\kappa_{1p} \equiv \int e_{1,p}' \mathbb{S}_p^{-1} \mu_p(\tilde{x},\tilde{z}) \mu_p(\tilde{x},\tilde{z}-\bar{z})' \mathbb{S}_p^{-1} e_{1,p} K(\tilde{x},\tilde{z})\, K(\tilde{x},\tilde{z}-\bar{z})\, d(\tilde{x},\tilde{z},\bar{z})$. *In addition,*

$$\sup_{(x,a)\in\mathcal{X}_0\times\mathcal{A}_H} |\hat{m}_H(x,a) - m_H^*(x,a)| = O_P(n^{-1/2} b^{-d_X/2} \sqrt{\log n} + b^{p+1}). \tag{5.8}$$

This result does not impose correct specification. When this holds, we can replace m_H^* with m.

To obtain $\hat{m}_H(x,a)$, we estimate both $G(\cdot|x^*,z)$ and $G^{-1}(\cdot|x,z)$. Above, we use the same bandwidth and kernel for both, yielding nice expressions for $B_m(x,a;x^*)$ and $\sigma_m^2(x,a;x^*)$. Both the first-stage estimator $\hat{G}_{p,b}(a \mid x^*,z)$ and the second-stage estimator $\hat{G}_{p,b}^{-1}(\tau|x,z)$ with $\tau = \hat{G}_{p,b}(a \mid x^*,z)$ contribute to the asymptotic bias and variance. The terms involving $\frac{\mathbf{G}_{p+1}(a|x^*,z)}{g(G^{-1}(G(a|x^*,z)|x,z)|x,z)}$ in $B_m(x,a;x^*)$ and $\frac{1}{g(x^*,z)}$ in $\sigma_m^2(x,a;x^*)$ are due to the first stage, whereas those involving $\mathbf{G}_{p+1}^{-1}(G(a|x^*,z)|x,z)$ in $B_m(x,a;x^*)$ and $\frac{1}{g(x,z)}$ in $\sigma_m^2(x,a;x^*)$ are due to the second stage.

It is well known that in many nonparametric applications, the choice of kernel function is not so critical, but the choice of bandwidth may be crucial. Different choices of bandwidth for the first- and second-stage estimators may thus be important in practice. If we choose b_1 in the first-stage estimation of $G(a \mid x^*, z)$ to obtain $\hat{G}_{p,b_1}(\cdot \mid x^*, z)$ and b_2 in the second-stage estimation of $G^{-1}(\tau \mid x, z)$ with $\tau = \hat{G}_{p,b_1}(a \mid x^*, z)$ to obtain $\hat{G}^{-1}_{p,b_2}(\cdot \mid x, z)$, say, then the asymptotic bias and variance must be modified accordingly. In particular, if $b_1 \ll b_2$ in the sense that $b_1 = o(b_2)$, then the asymptotic bias will mainly be contributed by the second-stage estimation and the asymptotic variance by the first-stage estimation; and vice versa.

To state the next results, we define $A^*_{H,i}$ (and $A^\dagger_{H,i}$ below) in the obvious manner. Theorem 5.1 implies the following asymptotic properties for $\hat{A}_{H,i}$.

Corollary 5.2 *Suppose Assumptions C.1-C.6 hold. Then conditional on* $(X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\sqrt{nb^{d_X}}\left(\hat{A}_{H,i} - A^*_{H,i} - B_m(x^*, Y_i; X_i)\right) \xrightarrow{d} N\left(0, \sigma^2_m(x^*, Y_i; X_i)\right)$. *Further, for i such that* $(X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\hat{A}_{H,i} - A^*_{H,i} = O_P(n^{-1/2} b^{-d_X/2} \times \sqrt{\log n} + b^{p+1})$ *uniformly in i.*

The asymptotic properties of $\tilde{A}_{H,i}$ follow from the next theorem.

Theorem 5.3 *Suppose Assumptions C.1-C.6 hold. Then for any* $(x, y) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\hat{m}^{-1}_H(x, y) \xrightarrow{P} m^{*-1}_H(x, y)$ *and* $\sqrt{nb^{d_X}}\left(\hat{m}^{-1}_H(x, y) - m^{*-1}_H(x, y) - B_{m^{-1}}(x, y; x^*)\right) \xrightarrow{d} N\left(0, \sigma^2_{m^{-1}}(x, y; x^*)\right)$, *where*

$$\begin{aligned} B_{m^{-1}}(x, y; x^*) &\equiv -B_m\left(x, m^{*-1}_H(x, y); x^*\right) / D^*_H\left(x, m^{*-1}_H(x, y)\right), \\ \sigma^2_{m^{-1}}(x, y; x^*) &\equiv \sigma^2_m\left(x, m^{*-1}_H(x, y); x^*\right) / \left[D^*_H\left(x, m^{*-1}_H(x, y)\right)\right]^2, \end{aligned}$$

and $D^*_H(x, a) \equiv \int \frac{g(a|x^*,z)}{g(G^{-1}(G(a|x^*,z)|x,z)|x,z)} dH(z)$.

When correct specification holds, we can show that $D^*_H(x, a) = \int \frac{g(a|x^*,z)}{g(m(x,a)|x,z)} dH(z) = \partial m(x, a) / \partial a$ by Corollary 3.4, the fact that $\partial m(x, a) / \partial a = g(a|x^*, z)/g(m(x, a)|x, z)$, and that $\int dH(z) = 1$. Further, Theorem 5.3 implies that conditional on $(X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\sqrt{nb^{d_X}}(\tilde{A}_{H,i} - A^\dagger_{H,i} - B_{m^{-1}}(X_i, Y_i; x^*)) \xrightarrow{d} N\left(0, \sigma^2_{m^{-1}}(X_i, Y_i; x^*)\right)$.

6 Asymptotics for Specification Testing

In this section, we study the asymptotic behavior of the test statistic in (4.1).

6.1 Asymptotic distributions

To state the next result, we write $w_i \equiv (x_i', y_i, z_i')'$, and we introduce the following notation:

$$\begin{aligned} S_{p,b}(\tau; u) &\equiv n^{-1} \sum_{i=1}^n K_b(U_i - u)\, g(G^{-1}(\tau|U_i)|U_i) \mu_{p,b}(U_i - u)\, \mu_{p,b}(U_i - u)', \qquad (6.1) \\ \eta_{1k}(\tau; u) &\equiv e_{1,p}' \bar{\mathbf{S}}_{p,b}(u)\, \mu_{p,b}(U_k - u)\, K_b(U_k - u) / g\left(G^{-1}(\tau|x^*, z)\,|x^*, z\right), \\ \eta_{2k}(\tau; u) &\equiv e_{1,p}' \bar{S}_{p,b}(\tau; u)\, \mu_{p,b}(U_k - u)\, K_b(U_k - u), \\ \zeta_0(W_i, W_k; z) &\equiv \eta_{1k}(\tau_{iz}; X_i, z)\, \bar{\mathbf{1}}_{Y_i}(W_k) + \eta_{2k}(\tau_{iz}; x^*, z)\, \psi_{\tau_{iz}}\left(Y_k - G^{-1}(\tau_{iz}|U_k)\right), \quad \text{and} \\ \varphi(w_i, w_j) &\equiv E\left[\int\int \zeta_0(W_1, w_i; z)\, \zeta_0(W_1, w_j; \bar{z})\, d\Delta(z)\, d\Delta(\bar{z})\, \pi_i\right], \end{aligned}$$

where $\bar{\mathbf{S}}_{p,b}(u) \equiv E[\mathbf{S}_{p,b}(u)]$, $\bar{S}_{p,b}(\tau;u) \equiv E[S_{p,b}(\tau;u)]$, $\tau_{iz} \equiv G(Y_i|X_i,z)$, $\bar{\mathbf{1}}_{Y_i}(W_k) \equiv 1(Y_k \leq Y_i) - G(Y_i|U_k)$, and $\psi_\tau(u) \equiv \tau - 1(u \leq 0)$. The asymptotic bias and variance are respectively

$$\mathbb{B}_{J_n} \equiv n^{-2} b^{d_X} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\int \zeta_0(W_i, W_j; z)\, d\Delta(z) \right]^2 \pi_i \text{ and } \sigma_{J_n}^2 = 2 b^{2d_X} E[\varphi(W_1, W_2)^2].$$

To establish the asymptotic properties of $\hat{J}_n$, we add the following condition on the bandwidth.

Assumption C.7 As $n \to \infty$, $nb^{2d_X} \to \infty$, and $nb^{3d/2}/(\log n)^2 \to \infty$.

Assumptions C.6 and C.7 imply that a higher order local polynomial (i.e., $p \geq 2$) may be required in the case where d_X or d_Z is large in order to ensure that $p + 1 - d_Z/2 > 0$ and $2(p+1) + d_X/2 > \max(2d_X, d + 2d_Z, 3d/2)$. Intuitively, the use of higher order local polynomials helps to remove the asymptotic bias of nonparametric estimates.

We establish the asymptotic null distribution of the $\hat{J}_n$ test statistic as follows:

Theorem 6.1 *Suppose Assumptions C.1-C.7 hold. Then under A.1, A.2, B.1, and B.2 we have* $\hat{J}_n - \mathbb{B}_{J_n} \xrightarrow{d} N(0, \sigma_J^2)$, *where* $\sigma_J^2 \equiv \lim_{n\to\infty} \sigma_{J_n}^2$.

The key to obtaining the asymptotic bias and variance of the test statistic $\hat{J}_n$ is $\zeta_0(W_i, W_k; z)$. The first term, $\eta_{1k}(\tau_{iz}; X_i, z)\, \bar{\mathbf{1}}_{Y_i}(W_k)$, in the definition of ζ_0 reflects the influence of the first-stage estimator $\hat{G}_{p,b}(Y_i \mid X_i, z)$, whereas the second term $\eta_{2k}(\tau_{iz}; x^*, z)\, \psi_{\tau_{iz}}(Y_k - G^{-1}(\tau_{iz}|U_k))$ embodies the effect of the second-stage estimator $\hat{G}_{p,b}^{-1}(\tau \mid x^*, z))$ evaluated at $\tau = \hat{G}_{p,b}(Y_i \mid X_i, z)$. A careful analysis of $\mathbb{B}_{J_n}$ indicates that both terms contribute to the asymptotic bias of $\hat{J}_n$ to the order of $O(1)$. On the other hand, a detailed study of $\sigma_{J_n}^2$ shows that they contribute asymmetrically to the asymptotic variance: the asymptotic variance of $\hat{J}_n$ is mainly determined by the second-stage estimator, whereas the role played by the first-stage estimator is asymptotically negligible. The main reason for this is that the term X_i in $\eta_{1k}(\tau_{iz}; X_i, z)$ (but not the term x^* appearing in $\eta_{2k}(\tau_{iz}; x^*, z))$ is subject to a (smooth) expectation operator in the definition of $\varphi(w_i, w_j)$, which helps reduce the variation of the first-stage estimator. For the same reason, we need b^{d_X} instead of the usual term $b^{d_X/2}$ as the normalization constant in the definition of $\hat{J}_n$, which unavoidably reduces the size of the class of local alternatives that this test has power to detect.

To implement, we need consistent estimates of the asymptotic bias and variance. Let

$$\begin{aligned} & \hat{\zeta}_0(W_i, W_k; z) \\ \equiv\ & \frac{1}{\hat{g}_{iz}} \Big[e_{1,p}' \mathbf{S}_{p,b}(X_i, z)^{-1} \mu_b(X_k - X_i, Z_k - z) K_b(X_k - X_i, Z_k - z)\, \hat{\mathbf{1}}_{Y_i}(W_k) \\ & + e_{1,p}' \mathbf{S}_{p,b}(x^*, z)^{-1} \mu_b(X_k - x^*, Z_k - z) K_b(X_k - x^*, Z_k - z)\, \psi_{\hat{\tau}_{iz}}\left(Y_k - \hat{G}_{p,b}^{-1}(\hat{\tau}_{iz}|U_k) \right) \Big], \end{aligned}$$

where $\hat{g}_{iz} \equiv \hat{g}\left(\hat{G}_{p,b}^{-1}(\hat{\tau}_{zi}|x^*, z)\, |x^*, z \right)$ and $\hat{\mathbf{1}}_{Y_i}(W_k) \equiv \mathbf{1}\{Y_k \leq Y_i\} - \hat{G}_{p,b}(Y_i|U_k)$. We propose estimating $\mathbb{B}_{J_n}$ and $\sigma^2_{J_n}$ respectively by

$$\begin{aligned}\hat{\mathbb{B}}_{J_n} &= n^{-2}b^{d_X}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\int \hat{\zeta}_0\left(W_i, W_j; z\right)d\Delta\left(z\right)\right]^2 \pi_i \text{ and} \\ \hat{\sigma}^2_{J_n} &= \frac{2h^{d_X}}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\frac{1}{n}\sum_{l=1}^{n}\int \hat{\zeta}_0\left(W_l, W_i; z\right)d\Delta\left(z\right)\int \hat{\zeta}_0\left(W_l, W_j; \bar{z}\right)d\Delta\left(\bar{z}\right)\pi_l\right]^2.\end{aligned}$$

It is not hard to show $\hat{\mathbb{B}}_{J_n} - \mathbb{B}_{J_n} = o_P\left(1\right)$ and $\hat{\sigma}^2_{J_n} - \sigma^2_{J_n} = o_P\left(1\right).$ Then we can compare

$$T_n \equiv \left(\hat{J}_n - \hat{\mathbb{B}}_{J_n}\right)/\sqrt{\hat{\sigma}^2_{J_n}} \tag{6.2}$$

to the critical value z_α, the upper α percentile from the $N(0,1)$ distribution, as the test is one-sided; we reject the null when $T_n > z_\alpha$.

To study the local power of the T_n test, consider the sequence of Pitman local alternatives:

$$\mathbb{H}_1\left(\gamma_n\right): \int G_n^{-1}(G_n(y|x,z) \mid x^*, z))d\Delta\left(z\right) = \gamma_n \delta_n\left(x,y\right), \tag{6.3}$$

where $\gamma_n \to 0$ as $n \to \infty$, and δ_n is a non-constant measurable function with $\mu_0 \equiv \lim_{n\to\infty} E[\delta_n\left(X_1, Y_1\right)^2 \pi\left(X_1, Y_1\right)] < \infty$. Given $B.1$, such alternatives arise from non-monotonicity.

Theorem 6.2 *Suppose Assumptions C.1-C.7 hold. Then under* $\mathbb{H}_1\left(\gamma_n\right)$ *with* $\gamma_n = n^{-1/2}b^{-d_X/2}$, $T_n \xrightarrow{d} N\left(\mu_0/\sigma_J, 1\right).$

Theorem 6.2 implies that the T_n test has non-trivial power against Pitman local alternatives that converge to zero at rate $n^{-1/2}b^{-d_X/2}$, provided $0 < \mu_0 < \infty$. The asymptotic local power function of the test is given by $1 - \Phi\left(z_\alpha - \mu_0/\sigma_J\right)$, where Φ is the standard normal CDF.

The following theorem shows that the test is consistent for the class of global alternatives

$$\mathbb{H}_1: \mu_A \equiv E\left\{\left[\int G^{-1}(G(Y_1|X_1,z) \mid x^*, z))d\Delta\left(z\right)\right]^2 \pi\left(X_1, Y_1\right)\right\} > 0.$$

Theorem 6.3 *Suppose Assumptions C.1-C.7 hold. Given* $\mathbb{H}_1$, $P\left(T_n > \lambda_n\right) \to 1$ *for any nonstochastic sequence* $\lambda_n = o(nb^{d_X})$.

The Appendix discusses how proper choice of H_1 and H_2 assures $\mu_A > 0$ when CQR fails.

6.2 A bootstrap version of the test

It is well known that nonparametric tests based on their asymptotic normal null distributions may perform poorly in finite samples. Preliminary experiments showed this to be true here. Thus, we suggest using a bootstrap method to obtain bootstrap p-values.

Let $\mathcal{W}_n \equiv \left\{W_i = (X_i, Y_i, Z_i)\right\}_{i=1}^n$. Following Su and White (2008), we draw bootstrap resamples $\{X_i^*, Y_i^*, Z_i^*\}_{i=1}^n$ based on the following smoothed local bootstrap procedure:

1. For $i = 1, ..., n$, obtain a preliminary estimate of A_i as $\hat{A}_i = (\hat{A}_{1,i} + \hat{A}_{2,i})/2$, where $\hat{A}_{j,i} = \int \hat{G}_{p,b}^{-1}(\hat{G}_{p,b}(Y_i|X_i, z)|x^*, z))\ dH_j(z)$.

2. Draw a bootstrap sample $\{Z_i^*\}_{i=1}^n$ from the smoothed kernel density $\tilde{f}_Z(z) = n^{-1} \sum_{i=1}^n \phi_{\alpha_z}(Z_i - z)$, where $\phi_\alpha(z) = \alpha^{-d_Z}\phi(z/\alpha)$, $\phi(\cdot)$ is the standard normal PDF in the case where Z_t is scalar valued and becomes the product of univariate standard normal PDF otherwise, and $\alpha_z > 0$ is a bandwidth parameter.

3. For $i = 1, ..., n$, given Z_i^*, draw X_i^* and A_i^* independently from the smoothed conditional density $\tilde{f}_{X|Z}(x|Z_i^*) = \sum_{j=1}^n \phi_{\alpha_x}(X_j - x)\phi_{\alpha_z}(Z_j - Z_i^*) / \sum_{l=1}^n \phi_{\alpha_z}(Z_l - Z_i^*)$ and $\tilde{f}_{A|Z}(a|Z_i^*) = \sum_{j=1}^n \phi_{\alpha_a}(\hat{A}_j - a)\phi_{\alpha_z}(Z_j - Z_i^*) / \sum_{l=1}^n \phi_{\alpha_z}(Z_l - Z_i^*)$, respectively, where α_z, α_x, and α_a are given bandwidths.

4. For $i = 1, ..., n$, compute the bootstrap version of Y_i as $Y_i^* = (\hat{m}_{H_1}(X_i^*, A_i^*) + \hat{m}_{H_2}(X_i^*, A_i^*))/2$.

5. Compute a bootstrap statistic T_n^* in the same way as T_n, with $\mathcal{W}_n^* \equiv \{W_i^* = (X_i^*, Y_i^*, Z_i^*)\}_{i=1}^n$ replacing $\mathcal{W}_n$.

6. Repeat Steps 2-5 B times to obtain bootstrap test statistics $\left\{T_{nj}^*\right\}_{j=1}^B$. Calculate the bootstrap p-values $p^* \equiv B^{-1}\sum_{j=1}^B 1\left(T_{nj}^* \geq T_n\right)$ and reject the null hypothesis if p^* is smaller than the prescribed nominal level of significance.

Clearly, we impose conditional exogeneity (X_i^* and A_i^* are independent given Z_i^*) in the bootstrap world in Step 3. The null hypothesis of monotonicity is implicitly imposed in Step 4.

A full formal analysis of this procedure is lengthy and well beyond our scope here. Nevertheless, the Appendix sketches the main ideas needed to show that this bootstrap method is asymptotically valid under suitable conditions, that is,

$$(i)\ \ P(T_n^* \leq t|\mathcal{W}_n) \to \Phi(t)\ \text{ for all } t \in \mathbb{R},\ \text{ and }\ (ii)\ \ P(T_n > z_\alpha^*) \to 1 \text{ under } \mathbb{H}_1, \tag{6.4}$$

where z_α^* is the α-level bootstrap critical value based on B bootstrap resamples, i.e., z_α^* is the $1 - \alpha$ quantile of the empirical distribution of $\left\{T_{nj}^*\right\}_{j=1}^B$.

7 Estimation and Specification Testing in Finite Samples

In this section, we conduct some Monte Carlo simulations to evaluate the finite-sample performance of our estimators and tests. We first consider estimation of the response under correct specification. We then examine the behavior of the T_n test. The already significant computational burden of our statistics is substantially multiplied by the replications necessary for Monte Carlo study. Thus, we restrict attention to a modest number of judiciously chosen experiments.

7.1 Estimation of response

We begin by considering the following two DGPs:

DGP 1: $Y_i = (0.5 + 0.1X_i^2)A_i$,

DGP 2: $Y_i = \Phi\left(\left(X_i+1\right)A_i/4\right)\left(X_i+1\right)$,

where $i = 1, ..., n$, $\Phi(\cdot)$ is the standard normal CDF, $A_i = 0.5Z_i + \eta_{1i}$, $X_i = 0.25 + Z_i - 0.25Z_i^2 + \eta_{2i}$, and η_{1i}, η_{2i}, and Z_i are each IID $N(0,1)$ and mutually independent. Clearly, $m(x,a) = (0.5 + 0.1x^2)a$ in DGP 1 and $= \Phi\left((x+1)a/4\right)(x+1)$ in DGP 2. In either DGP, $m(x,\cdot)$ is strictly monotone for each x but does not satisfy the normalization condition $m(x_{med}, a) = a$ for all $a \in \mathcal{A}$, where $x_{med} \simeq 0.116$ is the population median of X_i.[5]

To illustrate how the normalization condition is met with $x^* = x_{med}$, we redefine the unobservable heterogeneity A_i and the functional form of m. For DGP 1, let $a^* = a/c_1$ and $m^*(x, a^*) = c_1(0.5 + 0.1x^2)a^*$ for some nonzero value c_1. To ensure $m^*(x_{med}, a^*) = c_1(0.5 + 0.1x_{med}^2)a^* = a^*$ for all $a^* \in \mathcal{A}^*$, where $\mathcal{A}^*$ is the support of A_i/c_1, we can solve for c_1 to obtain $c_1 = 1/\left(0.5 + 0.1x_{med}^2\right) = 1.9946$. For DGP 2, let $a^* = (x^*+1)\Phi\left(a(x^*+1)/4\right)$ (i.e., $a = 4\Phi^{-1}(a^*/(x^*+1))/(x^*+1)$). Then $m(x,a) = \Phi\left(\frac{x+1}{x^*+1}\Phi^{-1}\left(\frac{a^*}{x^*+1}\right)\right)(x+1) \equiv m^*(x,a^*)$. It is easy to verify that $m^*(x_{med}, a^*) = a^*$ for all $a^* \in \mathcal{A}^*$ provided $x^* = x_{med}$, where $\mathcal{A}^*$ is now the support of $(x_{med}+1)\,\Phi\left(A_i\left(x_{med}+1\right)/4\right)$. For notational simplicity, we continue to use $m(x,a)$ and A_i to denote $m^*(x,a)$ and A_i^*, respectively.

To estimate the response $m(x,a)$, we need to choose the local polynomial order p, the kernel function K, the bandwidth b, and the weight function H. Since $d_X = d_Z = 1$, it suffices to choose $p = 1$ to obtain the local linear estimates $\hat{G}_{p,b}(a|x^*, z)$ and $\hat{G}_{p,b}^{-1}\left(\hat{G}_{p,b}(a|x^*,z)\,|x,z\right)$, which we use to construct the estimator $\hat{m}_H(x,a)$. We choose K to be the product of univariate standard normal PDFs. To save time in computation, we choose b using Silverman's rule of thumb: $b = (1.06S_X n^{-1/6},\ 1.06S_Z n^{-1/6})$, where, e.g., S_X is the sample standard deviation of $\{X_i\}_{i=1}^n$. Note that we use different bandwidth sequences for X and Z. We consider two choices for H: H_1 is the CDF for the uniform distribution on $[\zeta_{\epsilon_0}, \zeta_{1-\epsilon_0}]$, and H_2 is a scaled beta(3,3) distribution on $[\xi_{\epsilon_0}, \xi_{1-\epsilon_0}]$, where ξ_{ϵ_0} is the ϵ_0-th sample quantile of $\{Z_i\}_{i=1}^n$ and $\epsilon_0 = 0.05$. For either H_1 or H_2, we choose $N = 30$ points for numerical integration.

We evaluate the estimates of $m(x,a)$ at prescribed points. We choose 15 equally spaced points on the interval $[-1.895,\ 1.750]$ for x, where -1.895 and 1.750 are the 10th and 90th quantiles of X_i, respectively. For a, we choose 15 equally spaced points on the interval $[-0.718, 0.718]$ for DGP 1, where -0.718 and 0.718 are the 10th and 90th quantiles of $A_i (= A_i^*)$. For DGP 2, we choose 15 equally spaced points on the interval $[0.384,\ 0.731]$, where 0.384 and 0.731 are the 10th and 90th quantiles of $A_i (= A_i^*)$ in DGP 2. Thus, (x,a) will take $15 \times 15 = 225$ possible values; we let (x_j, a_j), $j = 1, ..., 225$, denote these values. We obtain the estimates $\hat{m}_{H_l}(x,a)$, $l = 1, 2$, of

[5] For DGP 1, $m(x,a) = a$ for all a at $x = x^* = \pm\sqrt{5}$.

Table 1: Finite sample performance of the estimates of the response

DGP	n	H	MAD_H				$RMSE_H$			
			5^{th}	50^{th}	95^{th}	mean	5^{th}	50^{th}	95^{th}	mean
1	100	Uniform	0.186	0.267	0.390	0.276	0.246	0.350	0.506	0.360
		Beta	0.145	0.208	0.290	0.210	0.191	0.280	0.405	0.239
	400	Uniform	0.125	0.179	0.243	0.180	0.164	0.235	0.315	0.239
		Beta	0.085	0.118	0.160	0.121	0.116	0.166	0.229	0.168
2	100	Uniform	0.091	0.120	0.146	0.120	0.157	0.208	0.268	0.209
		Beta	0.083	0.112	0.136	0.112	0.140	0.193	0.246	0.174
	400	Uniform	0.082	0.101	0.117	0.100	0.138	0.174	0.213	0.174
		Beta	0.076	0.095	0.108	0.094	0.126	0.161	0.188	0.160

$m(x,a)$ at these 225 points, and calculate the corresponding mean absolute deviations (MADs) and root mean squared errors (RMSEs):

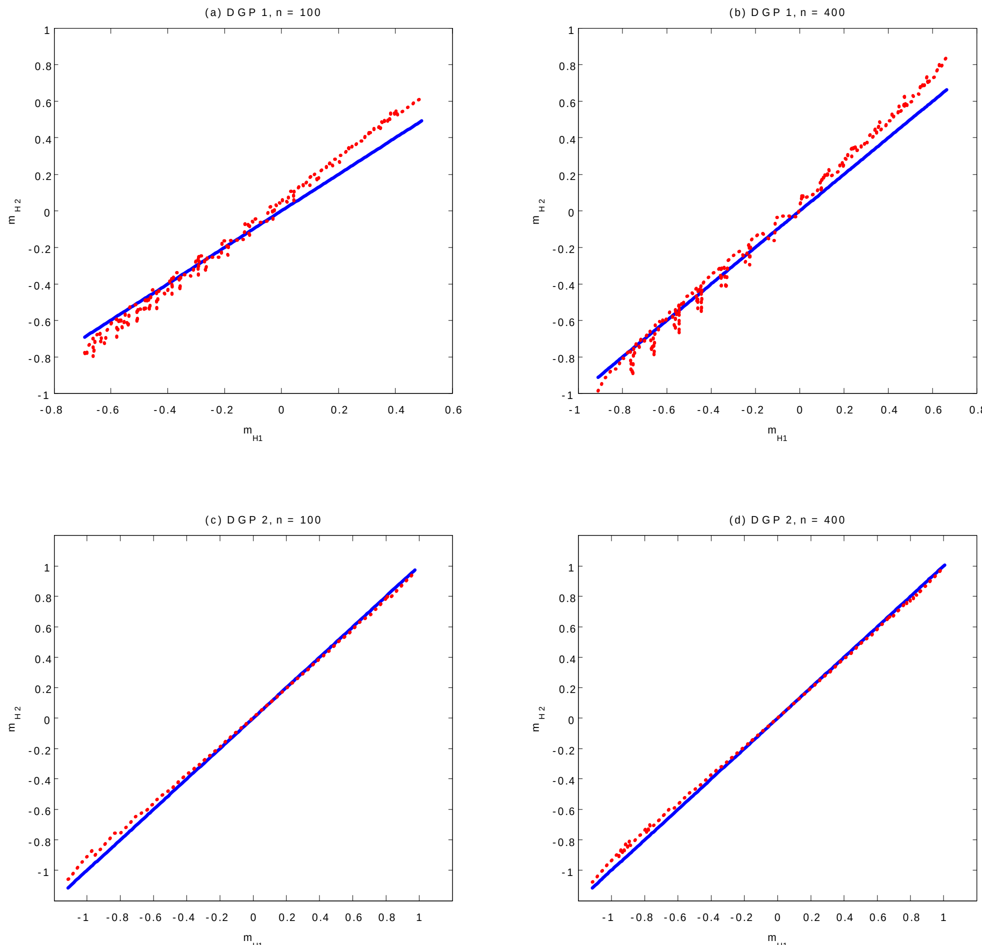
$$MAD_{H_l}^{(r)} = \frac{1}{225}\sum_{j=1}^{225}\left|m(x_j,a_j) - \hat{m}_{H_l}^{(r)}(x_j,a_j)\right|, \tag{7.1}$$

$$RMSE_{H_l}^{(r)} = \left\{\frac{1}{225}\sum_{j=1}^{225}\left[m(x_j,a_j) - \hat{m}_{H_l}^{(r)}(x_j,a_j)\right]^2\right\}^{1/2}, \tag{7.2}$$

where, for $r = 1, ..., 250$, $\hat{m}_{H_l}^{(r)}(x_j,a_j)$ is the estimate of $m(x_j,a_j)$ in the rth replication with weight function H_l, $l = 1, 2$. For feasibility in computation, we consider two sample sizes in our simulation study, namely, $n = 100$ and 400.

Table 1 reports the 5^{th}, 50^{th}, and 95^{th} percentiles of $MAD_{H_l}^{(r)}$ and $RMSE_{H_l}^{(r)}$ for the estimates of the $m(x,a)$, together with their means obtained by averaging over the 250 replications. We summarize the main findings from Table 1. First, for different choices of the distributional weights (H_1 or H_2), the MAD or RMSE performances of the response estimators may be quite different. In particular, we find that the estimators using the beta weight H_2 tend to have smaller MADs and RMSEs. This is especially true for DGP 1. Second, as n quadruples, both the MADs and RMSEs tend to improve, as expected. Third, and also as expected, the MADs and RMSEs improve at a rate much slower than the parametric rate $n^{-1/2}$.

Figure 1 plots the ratio of $\hat{m}_{H_2}(x_j,a_j)$ to $\hat{m}_{H_1}(x_j,a_j)$ for DGPs 1-2 and $n = 100$, 400, where $\hat{m}_{H_l}(x_j,a_j) \equiv \frac{1}{250}\sum_{r=1}^{250}\hat{m}_{H_l}^{(r)}(x_j,a_j)$ for $l = 1, 2$, and $j = 1, 2, ..., 225$. In theory, both $\hat{m}_{H_1}(x_j,a_j)$ and $\hat{m}_{H_2}(x_j,a_j)$ converge to $m(x_j,a_j)$ in probability so that the points $(\hat{m}_{H_1}(x_j,a_j), \hat{m}_{H_2}(x_j,a_j))$ should lie on the 45 degree line as $n \to \infty$. Nevertheless, we see some discrepancies between $\hat{m}_{H_1}(x_j,a_j)$ and $\hat{m}_{H_2}(x_j,a_j)$ for both DGPs and sample sizes. Clearly, the discrepancy is much larger for DGP 1 than that for DGP 2 for both sample sizes.



Plots of $\hat{m}_{H_2}$ over $\hat{m}_{H_1}$ for DGPs 1-2: the solid line is the 45 degree line, the points for ($\hat{m}_{H_1}$, $\hat{m}_{H_2}$) on the dotted line are obtained as averages of $(\hat{m}^{(r)}_{H_1}(x_j, a_j), \hat{m}^{(r)}_{H_2}(x_j, a_j))$, $r = 1, ..., 250$, over 250 replications

7.2 Specification Testing

To examine the finite-sample properties of the specification test, we consider the two DGPs:

DGP 3: $Y_i = (0.5 + 0.1X_i^2)A_i + 2\delta_0 X_i/(0.1 + e^{A_i^2/2})$,

DGP 4: $Y_i = \Phi\left((X_i + 1)A_i/4\right)(X_i + 1) - 0.5\delta_0 A_i/(1 + X_i^2)$,

where $i = 1, ..., n$, and A_i, X_i and Z_i are generated as in DGPs 1-2. Note that when $\delta_0 = 0$, DGPs 3 and 4 reduce to DGPs 1 and 2, respectively, permitting us to study the level behavior of our test. For other well-chosen values of δ_0, $m(x, a)$ as defined in DGP 3 or 4 is not strictly monotonic in a, permitting study of the test's power against non-monotone alternatives.

To construct the raw test statistic $\hat{J}_n$, we first obtain $\hat{G}_{p,b}(Y_i|X_i, z)$ and $\hat{G}^{-1}_{p,b}(\hat{G}_{p,b}(Y_i|X_i, z)\,|$

Table 2: Finite sample rejection frequency for DGPs 3-4

DGP	n	δ_0	Warp-speed bootstrap			Full bootstrap		
			1%	5%	10%	1%	5%	10%
3	100	0	0.018	0.042	0.124	0.008	0.080	0.140
		1	0.320	0.388	0.420	0.316	0.404	0.456
	200	0	0.020	0.060	0.148	0.020	0.076	0.140
		1	0.392	0.442	0.474	0.408	0.464	0.508
4	100	0	0.014	0.032	0.068	0.016	0.032	0.056
		1	0.288	0.576	0.660	0.456	0.656	0.724
	200	0	0.004	0.014	0.036	0.008	0.012	0.056
		1	0.356	0.564	0.688	0.476	0.712	0.792

$x^*, z)$ by choosing the order of the local polynomial regression, the kernel function K, and the bandwidth b. As when estimating the response, we choose $p = 1$ and let K be the product of univariate standard normal PDFs. Since we require undersmoothing for our test, we set $b = (c_2 S_X n^{-1/5}, c_2 S_Z n^{-1/5})$, where c_2 is a positive scalar that we use to check the sensitivity of our test to the choice of bandwidth. Next, we need to choose weight functions H_1, H_2, and π. We choose H_1 and H_2 as above and set $\pi(X_i, Y_i) = 1\left\{\xi_{\epsilon_0,X} \le X_i \le \xi_{1-\epsilon_0,X}\right\} \times 1\left\{\xi_{\epsilon_0,Y} \le Y_i \le \xi_{1-\epsilon_0,Y}\right\}$, where, e.g., $\xi_{\epsilon_0,X}$ is the ϵ_0th sample quantile of $\{X_i\}_{i=1}^n$ and $\epsilon_0 = 0.0125$. These sample quantiles converge to their population analogs at the parametric $\sqrt{n}$ rate, so they can be replaced by the latter in deriving the asymptotic theory. By construction, we trim $\hat{G}_{p,b}$ and $\hat{G}_{p,b}^{-1}$ in the tails.

In the bootstrap, we set $\alpha_z = S_Z n^{-1/6}$, $\alpha_x = S_X n^{-1/6}$, and $\alpha_a = S_A n^{-1/6}$, where, e.g., S_A is the sample standard deviation of $\hat{A}_i$. For computational feasibility, we consider two sample sizes, $n = 100$, 200, in our simulation study. Also for computational feasibility, for each sample size n, our "full" bootstrap experiments use 250 replications and $B = 100$ bootstrap resamples in each replication. Before performing the full bootstrap with $B = 100$, we study the sensitivity of the test to the bandwidth $b = (c_2 S_X n^{-1/5}, c_2 S_Z n^{-1/5})$ as suggested by Giacomini, Politis, and White (2007), using their warp-speed bootstrap. In this procedure, only one bootstrap resample is drawn in each replication. Because of its relatively low computational cost, we can use 500 replications for this study. We find that our test is not very sensitive to the choice of b as long as c_2 is not too big or small. For example, the test behaves reasonably well for c_2 lying between 1 and 2. Table 2 reports the warp-speed bootstrap results for the case $c_2 = 1.5$, together with the full bootstrap results for $c_2 = 1.5$ with $B = 100$.

Table 2 reports the empirical rejection frequencies for our test at various nominal levels for DGPs 3-4. The rows with $\delta_0 = 0$ report the empirical level of our test; those with $\delta_0 = 1$ show empirical power. We summarize the main findings from Table 2 as follows: First, the level of our test is fairly well behaved, and it can be close to the nominal level for sample sizes as small as $n = 100$. When n increases, the level generally improves somewhat. Second, the power of our test is reasonably good. It increases as the sample size doubles for both the warp-speed and full bootstrap methods. Finally, the full bootstrap tends to deliver somewhat better power performance than the warp-speed bootstrap, similar to findings of Cho and White (2011).

8 Empirical Applications

This section illustrates the usefulness of our tests with two examples. To show their broad applicability, we consider two very different applications. The first application analyzes an important question for policy analysis, namely the determinants of the Black-White earnings gap. The second application comes from a traditional area of economics: classical consumer demand using Engel curves. We discuss the economic background, provide details of the data and the test implementation, and discuss our findings.

8.1 The Black-White Earnings Gap: Just Ability?

8.1.1 Economic Background

The quest for the sources of the apparent differences in economic circumstances between the three major races in the US, i.e., Blacks, Hispanics, and Whites, has spurred an extensive and controversial debate over the last few decades. Starting with the seminal paper by Neal and Johnson (1996, NJ), a flourishing literature has emerged that focuses primarily on the sources of the Black-White earnings gap; obviously, a key concern in this is the potential existence of racial discrimination, i.e., the fact that people with the exact same ability get differential wages for the same task. Carneiro, Heckman, and Masterov (2005, CHM) give a recent overview of this literature, emphasizing the points important to our application.

As NJ argue, to obtain a measure of the full effect of discrimination in labor outcomes (e.g., wages) from a regression, one should not condition on variables that may indirectly channel discrimination, such as schooling, occupational choice, or years of work experience, as these may mask the full effects. As CHM aptly put it, the "full force" of discrimination would not be visible. Chalak and White (2011a) contains supporting discussion of the "included variable bias" arising by conditioning on variables indirectly channeling a cause of interest.

As argued convincingly in CHM, however, schooling is no longer a plausible channel of discrimination, given the extent of affirmative action. Indeed, CHM show that when conditioned just on schooling, the wage gap increases, rather than decreasing, as would be the case if schooling were an indirect channel of discrimination. Thus, we include years of schooling as a causal factor in our analysis. The structural relation is then

$$Y = m(X_1, X_2, A),$$

where A is work-related ability; X_1 is years of schooling; X_2 is race, a discrete variable taking three values; and Y is the wage an individual receives. As X and A are plausibly correlated, we seek a conditioning instrument Z such that X is independent of A given Z.

To find this variable, we go back to the literature. NJ suggest the 1980 AFQT score as a proxy for ability. Once NJ condition on the AFQT, which we now denote Z, the maintained hypothesis is that there is no relationship between X and A, that is, $X \perp A \mid Z$. This means that whatever is not exactly accounted for in A by using Z does not correlate with race or schooling, in line with NJ. Their finding (corroborated by the analysis in CHM, with the additional schooling variable) of an absence of a Black ($X_2 = 1$) $-$ White ($X_2 = 0$) earnings gap means in our notation that $E\left[m(X_1, 1, A)|X_1, Z\right] - E\left[m(X_1, 0, A)|X_1, Z\right] = 0$. This evidence is consistent with the absence of discrimination in the labor market.

Nevertheless, this is not the only testable implication of their hypotheses. We can now test the null hypothesis that there is indeed only a single unobservable that monotonically drives wages. Accordingly, we define ability as a scalar factor that drives up wages for all values of $X = x$. As such, scalar monotonicity is a natural assumption - the more able somebody is, the higher his wage is, *ceteris paribus*. The fact that we can apply this logic here hinges on the Z variable, AFQT80, which is chosen to ensure unconfoundedness. The alternative is that there is some more complex mechanism that generates wage outcomes. There are a number of reasons why there may be a more complex relationship. One is that discrimination acts through several unobserved channels; another is what CHM have argued, namely that there are unobserved (in their data, actually at least partially observed) factors in the early childhood of an individual that have a large impact on labor market outcomes, and that should be accounted for. With the data at hand, we cannot separate these two explanations; however, we can shed light on whether a scalar "ability" accounts for observed outcomes.

8.1.2 The Data

Our data come from NJ's original study,[6] which is based on the National Longitudinal Survey of Youth (NLSY). The NLSY is a panel data set of 12,686 youths born between 1957 and 1964. This data set provides us with information on schooling, race, and labor market outcomes. The Z variable is the normalized AFQT80 test score, i.e., the armed forces test in 1980. Individuals already in the labor market have been excluded. The test score is also year-adjusted and then normalized to have mean 0 and variance 1 as in NJ. After cleaning the data, we have 3,659 and 3,783 valid observations for the female and male subsamples, respectively, and following the literature, we analyze men and women separately. Since the data set is quite large and the computational requirements exceed what we can handle on our computers, we divide each subsample into three sub-subsamples to obtain 1,220 + 1,220 + 1,219 female observations, and

[6] Indeed, our data are exactly NJ's original data; we are indebted to Derek Neal for providing us with this.

1,261+1,261+1,261 male observations. These subpopulations were selected randomly by taking every third observation in each group. Since we are otherwise using exactly the same data as NJ, we refer to their paper for summary statistics and other data details.

8.1.3 Implementation Details

The details of the testing procedure we implement are largely identical to those for the simulation study of Section 6.2. The kernel is the product of univariate standard normal PDFs; the order of the local polynomial is 1. The bandwidth is chosen by the rule of thumb in Section 6.2, paragraph 2. We perform 200 bootstrap replications; with $n \approx 1,200$, each replication takes about two hours using Matlab code running on modern high-speed processors.

8.1.4 Empirical Results

The values of the test statistics we obtain using the procedures described above are 9.381, 12.633, and 85.119 for the three female subsamples, and 250.24, 140.55, and 246.85 for the male subsamples. The associated p-values are 0.995, 1, and 0.815 for the three female subsamples, and 0.510, 0.925, and 0.720 for the three male subsamples. Obviously, the fact that we have drawn subsamples reduces the power of our procedure. But increasing the sample size even moderately increases the processing time for each bootstrap replications to significant multiples of the two hours required with $n \approx 1,200$. Nevertheless, since the test statistic values are so far from the critical values and the p-values are so large, we believe it safe to conclude that the null of scalar monotonicity is not rejected. This is evidence consistent with the correct specification of the NJ/CHM model. Of course, further research with better data is required to analyze the importance of early childhood education as CHM suggest, but this is beyond our scope here.

Recall that we maintain conditional exogeneity, $B.1$. Without this, the test is a joint test for $A.2$ and $B.1$. Under this interpretation, we have no evidence against either $A.2$ or $B.1$. To illustrate use of a multiple test for misspecification, we also report the results of a pure test of $B.1$. By White and Chalak (2010, prop.2), we can test $B.1$ by testing $X \perp S \mid Z$, where $S = q(Z, A, V)$, with $X \perp V \mid (A, Z)$. A plausible candidate for S is another proxy for A, viewing V as a measurement error. Here, a natural choice for S is the 1989 AFQT score. To implement, we standardize AFQT89 in the same way as AFQT80, and we apply the conditional independence test of Huang and White (2010). The table below reports the results for each of the samples described above. For none of these do we reject $B.1$, consistent with our monotonicity test findings.

	Male			Female		
	Sample 1	Sample 2	Sample 3	Sample 1	Sample 2	Sample 3
p-value	.80	.28	.20	.39	.88	.85

8.2 Engel Curves in a Heterogeneous Population

8.2.1 Economic Background

Engel curves are among the oldest objects analyzed by economists.[7] Modern econometric Engel curve analysis assumes that

$$Y = m(X_1, X_2, A), \tag{8.1}$$

where Y is a K-vector of budget shares for K continuously-valued consumption goods; X_1 is wealth, represented by (log) total expenditure under the assumption that preferences are time-separable; X_2 denotes observable factors that reflect preference heterogeneity; and A denotes unobservable preference heterogeneity. Prices are absent here, as Engel curve analysis involves a single cross section only, and prices are assumed invariant. However, it is commonly thought that log total expenditure is endogenous[8] and is hence instrumented for, typically by labor income, say S. This is justified by the same intertemporal separability assumption.

Here, we follow IN, and write the X_1 structural equation as

$$X_1 = \phi(S, X_2, Z), \tag{8.2}$$

where the unobserved drivers of X_1 are denoted Z. For simplicity, we assume X_2 is exogenous. Following IN, we also assume S is exogenous, so we take $(S, X_2) \perp (A, Z)$, implying $(X_1, X_2) \perp A \mid Z$. With the usual normalization, $Z \mid (S, X_2)$ is $\mathcal{U}[0, 1]$, and Z is identified as

$$Z = F(X_1 \mid S, X_2). \tag{8.3}$$

IN's control function approach thus provides us with a variable Z that satisfies our assumptions. We are now able to test the hypothesis that there is a single unobservable A in equation (8.1) that enters monotonically. Put differently, due to the tight relationship between quantiles and nonseparable models with monotonicity, we can test whether in the conditional α-quantile regression of Y on X and Z, the parameter α can be given a structural interpretation. The alternative is that there is a more complex structure in the unobservables.

An example of a structural model that assumes monotonicity is provided in Blundell, Chen, and Kristensen (2007), who assume

$$Y = m(X_1, X_2) + A.$$

Nevertheless, to test this specification, we also must specify equation (8.2) as above.

[7] They were first analyzed in1857 by the Saxonian economist Ernst Engel (1821–1896), not to be confused with Friedrich Engels, the companion of Karl Marx.

[8] Nevertheless, the evidence is not strong; see Blundell, Horowitz, and Parey (2009) or Hoderlein (2010).

8.2.2 The Data

For our test, we use the British FES in exactly the form employed in IN.[9] The FES reports a yearly cross section of labor income, expenditures, demographic composition, and other characteristics of about 7,000 households in every year. Since we are considering Engel curves, we use only the cross section for 1995. We focus on households with two adults, where the adults are married or cohabiting, at least one is working, and the household head is aged between 20 and 55. We also exclude households with 3 or more children. This yields a sample with $n = 1,655$. This will be our operational subpopulation, not least because it is the one commonly used in the parametric demand system literature; see Lewbel (1999).

The expenditures for all goods are grouped into several categories. The first is related to food consumption and consists of the subcategories food bought, food out (catering), and tobacco. The second and third categories contain expenditures related to alcohol and catering. The alcohol category is probably mismeasured, so we do not employ it as dependent variable. The next group consists of transportation categories: motoring, fuel expenditures, and fares. Leisure goods and services are the last category. For brevity, we call these categories Food, Catering, Transportation, and Leisure. We work with these broader categories since more detailed accounts suffer from infrequent purchases (recall that the recording period is 14 days) and are thus often underreported. Together these account for approximately half of total expenditure, leaving a large fourth residual category. Labor income is as defined in the Household Below Average Income study (HBAI). Roughly, this is net labor income to the household head after taxes, but including state transfers.

The following table gives some summary statistics; for more details, see Hoderlein (2010).

Variable	Food	Catering	Alcohol	Transport	Leisure	LogExp	LogWage	nKids
Mean	0.2074	0.0805	0.0578	0.2204	0.1297	5.4215	5.8581	0.6205

8.2.3 Implementation Details

To apply the test, we let Y be the budget shares of Food, Catering, Transportation, or Leisure in (8.1).[10] In each case, we specify X_1 as the logarithm of total expenditure and X_2 as the number of kids in a family. The details of the testing procedure are again largely identical to those implemented in the simulation study in Section 6.2 and in the previous application. We use a product kernel and select the bandwidth by the same rule of thumb as in Section 6.2, paragraph 2. Again, we performed 200 bootstrap replications, each of which takes approximately 3.3 hours.

The major difference is that the instrument $Z = F(X_1 \mid S, X_2)$ must be estimated from the data in a first stage. To mitigate the bias from the first stage, we use local quadratic regression,

[9] We are grateful to Whitney Newey and Richard Blundell for providing us with the data.

[10] We did not consider the budget share for alcohol because there are too many 0 observations (258 out of 1,665) in the data.

employing a Gaussian kernel and Silverman's rule-of-thumb bandwidth. Standard U-statistic theory straightforwardly shows that the variance term in the decomposition of the difference between Z and its estimate will not affect the asymptotic null distribution of our test statistic. Consequently, the T_n statistic based on estimated Z is asymptotically equivalent to that based on the true unobservable Z under the null and some side conditions. This behavior is well known in the literature on nonparametrically estimated regressors, so to conserve space, we do not provide formal details. Using estimated Z also has implications for the bootstrap. Ideally, one would prefer a bootstrap method alternative to that applied here, based entirely on observables, while imposing the null. Developing and justifying this is a substantive undertaking, deserving of a paper in itself. We therefore leave this as a topic for future research.

8.2.4 Empirical Results

The following table summarizes our test results:

	Food	Catering	Transportation	Leisure
Value of Test Statistic	1.2895	0.7336	1.5905	1.1492
p-values	≤ 0.005	≤ 0.005	≤ 0.005	0.010

Two things are noteworthy. First, observe that rather small values of the test statistic are associated with small p-values. This indicates that the normal approximation is a poor description of the true finite-sample behavior, a result that is quite familiar in the nonparametric testing literature. Second, in all four categories analyzed, we soundly reject the null of monotonicity. The rejections are strongest in Food, Catering, and Transportation, and slightly less pronounced for Leisure. Whereas in the labor application above it seems conceivable that there is only one major omitted unobservable, i.e., ability, our test here suggests that this is not a valid description of the unobservables driving consumer behavior. This should not be surprising, given that consumer demand is usually thought to be a result of optimizing a rather complex preference ordering, given a budget set. Still, empirically establishing this fact, uniformly over a number of expenditure categories, is encouraging evidence of the ability of our test to produce economically interesting results in real-world applications.

9 Concluding Remarks

Monotonicity in unobservables is a now common assumption in applied economics. Its appeal arises from the fact that it allows one to recover the unknown structural function from the regression quantiles of the data. As we discuss, monotonicity is a strong assumption, and its failure has substantive consequences for structural inference.

So far, there are no generally applicable nonparametric specification tests designed to detect monotonicity failure. This paper provides such a test for cross-section data. We show how to exploit the power of an exclusion restriction together with a conditional independence assumption, plausible in a variety of applications, to construct a test statistic. We analyze the large-sample behavior of our estimators and tests and study their finite-sample behavior in Monte Carlo experiments. Our experiments show that a suitable bootstrap procedure yields tests with reasonably well behaved levels. Both theory and experiment show that the test has useful power.

When applied to data, the test exhibits these features. In a labor economics application where monotonicity in unobserved ability is plausible, we find that the test does not reject. In a consumer demand application, where monotonicity in a scalar unobserved preference parameter is less plausible, we find that the test clearly rejects. These two distinct applications also illustrate that our test applies to both observed and unobserved conditioning instrument cases and works well in both. Finally, we expect that our approach, or elements thereof, extends to tests for monotonicity in richer economic structures such as in Olley and Pakes (1996), but we leave this, in our opinion fascinating, extension for future research.

10 Mathematical Appendix

Proof of Theorem 3.2: Let $\upsilon(\cdot) \equiv m(\bar{x}, \cdot) : \mathcal{A} \to \mathbb{R}$. Clearly, $\upsilon(a)$ takes values in $\mathcal{V}$, so $\upsilon(\cdot) : \mathcal{A} \to \mathcal{V}$. By assumption, υ is measurable and strictly increasing, so its inverse function $\upsilon^{-1}(\cdot) : \mathcal{V} \to \mathcal{A}$ exists and is measurable and strictly increasing. Let $\bar{m} : \mathcal{X} \times \mathcal{V} \to \mathbb{R}$ be defined as $\bar{m}(\cdot, \cdot) \equiv m(\iota(\cdot), \upsilon^{-1}(\cdot))$, where $\iota(\cdot)$ is the identity map. This is the composition of m with the measurable map $\vartheta : \mathcal{X} \times \mathcal{V} \to \mathcal{X} \times \mathcal{A}$, defined by $\vartheta(x, v) = (\vartheta_1(x, v), \vartheta_2(x, v))$, where $\vartheta_1(x, v) = \iota(x)$, and $\vartheta_2(x, v) = \upsilon^{-1}(v)$. As compositions of measurable functions are measurable, $\bar{m} : \mathcal{X} \times \mathcal{V} \to \mathbb{R}$ is (product) measurable.

(a) Let $x \in \mathcal{X}$ be arbitrary. Take any $v \in \mathcal{V}$, and put $a = \upsilon^{-1}(v)$. Then $m(x, a) = m(x, \upsilon^{-1}(v)) = \bar{m}(x, v)$, as desired. The argument for the second claim is similar.

(b) By assumption, $m(x, \cdot)$ is strictly increasing on $\mathcal{A}$, and by construction, υ^{-1} is strictly increasing on $\mathcal{V}$. Because compositions of strictly increasing functions are strictly increasing, it follows that $\bar{m}(x, \cdot) = m(x, \upsilon^{-1}(\cdot))$ is strictly increasing on $\mathcal{V}$.

(c) For each v in $\mathcal{V}$, there exists $a = \upsilon^{-1}(v)$ such that $v = \upsilon(a) = m(\bar{x}, a) = m(\bar{x}, \upsilon^{-1}[\upsilon(a)]) = m(\bar{x}, \upsilon^{-1}(v)) = \bar{m}(\bar{x}, v)$. ■

Proof of Theorem 3.3 $A.0$, $A.1$, and $B.1$ ensure that for all (y, x, z)

$$G(y|x, z) \equiv P[Y \leq y|X = x, X = z] = P[m(X, A) \leq y|X = x, Z = z] = P[m(x, A) \leq y|Z = z].$$

Setting $y = m(x, a)$, it follows from $B.2$ that for all a, x, z

$$m(x, a) = G^{-1}(P[m(x, A) \leq m(x, a)|Z = z] \mid x, z). \tag{10.1}$$

Pick any $(a, x, \tilde{x}, z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{X} \times \mathcal{Z}$.

(a) We show that (i) implies $(ii) : a \in \bar{m}_{\tilde{x}}^{-1}\{G^{-1}[G(m(x,a) \mid x, z) \mid \tilde{x}, z]\}$. By $B.2$, $m(x,a) = G^{-1}(G(m(\tilde{x},a) \mid \tilde{x}, z) \mid x, z)$ implies $G(m(x,a) \mid x, z) = G(m(\tilde{x},a) \mid \tilde{x}, z)$, and

$$m(\tilde{x}, a) = G^{-1}[G(m(x,a) \mid x, z) \mid \tilde{x}, z].$$

Taking the pre-image $\bar{m}_{\tilde{x}}^{-1}$ gives the desired result.

(b) We show that (ii) implies (i): By the definition of the pre-image, $a \in \bar{m}_{\tilde{x}}^{-1}\{G^{-1}[G(m(x,a) \mid x, z) \mid \tilde{x}, z]\}$ implies

$$m(\tilde{x}, a) = G^{-1}[G(m(x,a) \mid x, z) \mid \tilde{x}, z].$$

Invoking $B.2$ to invert G twice, we obtain the desired result.

(c) We show that (i) implies (iii). Given (i), we have

$$\begin{aligned} m(x,a) &= G^{-1}(G(m(\tilde{x},a) \mid \tilde{x}, z) \mid x, z) \\ &= G^{-1}(P[Y \leq m(\tilde{x},a) \mid X = \tilde{x}, Z = z] \mid x, z) \\ &= G^{-1}(P[m(X,A) \leq m(\tilde{x},a) \mid X = \tilde{x}, Z = z] \mid x, z) \\ &= G^{-1}(P[m(\tilde{x},A) \leq m(\tilde{x},a) \mid X = \tilde{x}, Z = z] \mid x, z) \\ &= G^{-1}(P[m(\tilde{x},A) \leq m(\tilde{x},a) \mid Z = z] \mid x, z), \end{aligned}$$

where the last step follows from $B.1$. Using this result and (10.1) ensures that

$$G^{-1}(P[m(x,A) \leq m(x,a)|Z = z] \mid x, z) = G^{-1}(P[m(\tilde{x},A) \leq m(\tilde{x},a) \mid Z = z] \mid x, z).$$

By the invertibility of G^{-1} ensured by $B.2$, it follows that

$$\begin{aligned} P[m(x,A) \leq m(x,a)|Z = z] = P[m(\tilde{x},A) \leq m(\tilde{x},a) \mid Z = z] \quad \text{or} \\ P[A \in m_x^{-1}\{m(x,a)\}|Z = z] = P[A \leq m_{\tilde{x}}^{-1}\{m(\tilde{x},a)\} \mid Z = z], \end{aligned}$$

as desired.

(b) We show that (iii) implies (i). Thus, suppose that

$$\begin{aligned} P[A \in m_x^{-1}\{m(x,a)\}|Z = z] = P[A \leq m_{\tilde{x}}^{-1}\{m(\tilde{x},a)\} \mid Z = z] \quad \text{or} \\ P[m(x,A) \leq m(x,a)|Z = z] = P[m(\tilde{x},A) \leq m(\tilde{x},a) \mid Z = z]. \end{aligned}$$

Applying (10.1), the equality above, and $B.2$ gives

$$\begin{aligned} m(x,a) &= G^{-1}(P[m(x,A) \leq m(x,a)|Z = z] \mid x, z) \\ &= G^{-1}(P[m(\tilde{x},A) \leq m(\tilde{x},a) \mid Z = z] \mid x, z). \end{aligned}$$

Applying $B.1$ and simplifying gives

$$\begin{aligned} m(x,a) &= G^{-1}(P[m(\tilde{x},A) \leq m(\tilde{x},a) \mid X=\tilde{x}, Z=z] \mid x,z) \\ &= G^{-1}(P[m(X,A) \leq m(\tilde{x},a) \mid X=\tilde{x}, Z=z] \mid x,z) \\ &= G^{-1}(P[Y \leq m(\tilde{x},a) \mid X=\tilde{x}, Z=z] \mid x,z) \\ &= G^{-1}(G(m(\tilde{x},a) \mid \tilde{x},z) \mid x,z). \qquad \blacksquare \end{aligned}$$

Proof of Corollary 3.4 $A.2$ ensures Theorem 3.3(iii) for all $(a,x,\tilde{x},z)$. This ensures $m(x,a) = G^{-1}(G(m(\tilde{x},a) \mid \tilde{x},z) \mid x,z)$. Letting $\tilde{x} = x^*$ with $a = m(x^*,a)$ gives $m(x,a) = G^{-1}(a \mid x^*,z) \mid x,z)$, ensuring (3.1). Successively inverting $Y = G^{-1}(G(A \mid x^*,z) \mid X,z)$ for any z gives (3.2).

Next, $Y = m(X,A)$, $X \perp A \mid Z$, and strict monotonicity imply that for all y, x, and z,

$$\begin{aligned} G(y \mid x,z) &\equiv P[\, Y \leq y \mid X=x, Z=z] \\ &= P[\, m(X,A) \leq y \mid X=x, Z=z] = P[m(x,A) \leq y \mid X=x, Z=z] \\ &= P[\, m(x,A) \leq y \mid Z=z] = P[\, A \leq m^{-1}(x,y) \mid Z=z]. \end{aligned} \tag{10.2}$$

Setting $y = m(x,a)$ so that $a = m^{-1}(x,y)$ immediately gives

$$F_{A|Z}(a \mid z) = G(\, m(x,a) \mid x,z).$$

We have $F_{A|Z}(a \mid z) = F_{A|X,Z}(a \mid x,z)$ since $X \perp A \mid Z$. This establishes eq.(3.3).

Finally, for all $(a,x,z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{Z}$,

$$\begin{aligned} P[\, A &\leq a \mid X=x] = P[\, m(x,A) \leq m(x,a) \mid X=x] \\ &= P[\, m(X,A) \leq m(x,a) \mid X=x] \\ &= P[Y \leq m(x,a) \mid X=x] = G_{Y|X}(m(x,a) \mid x) \\ &= G_{Y|X}(\, G^{-1}[G(a \mid x^*,z) \mid x,z] \mid x). \qquad \blacksquare \end{aligned}$$

For the next results, recall that $\mathcal{U}_0 \equiv \mathcal{X}_0 \times \mathcal{Z}_0$, $U_i \equiv (X_i', Z_i')'$, $u \equiv (x', z')'$, $K_b(u) \equiv b^{-d} K(u/b)$, and $\mu_{p,b}(u) \equiv \mu_p(u/b)$. Let $W_i \equiv (Y_i, U_i')'$ and $w \equiv (y, u')'$. Let $\mathbf{S}_{p,b}(u)$ and $S_{p,b}(u)$ be as defined in (5.3) and (6.1), respectively. Define

$$\begin{aligned} \bar{V}_{p,b}(\tau;u) &\equiv \frac{1}{n}\sum_{i=1}^{n} K_b(U_i-u)\,\mu_{p,b}(U_i-u)\,\psi_\tau(Y_i - \boldsymbol{\beta}_b(\tau;u)), \\ V_{p,b}(\tau;u) &\equiv \frac{1}{n}\sum_{i=1}^{n} K_b(U_i-u)\,\mu_{p,b}(U_i-u)\,\psi_\tau\left(Y_i - G^{-1}(\tau|U_i)\right), \\ \mathbf{B}_{p,b}(y;u) &\equiv \frac{1}{n}\sum_{i=1}^{n} K_b(U_i-u)\,\mu_{p,b}(U_i-u)\Delta_{i,y}(u), \\ \mathbf{V}_{p,b}(y;u) &\equiv \frac{1}{n}\sum_{i=1}^{n} K_b(U_i-u)\,\mu_{p,b}(U_i-u)\,\bar{\mathbf{1}}_y(W_i), \end{aligned}$$

where $\psi_\tau(u) \equiv \tau - 1(u \leq 0)$, $\Delta_{i,y}(u) \equiv G(y|U_i) - G(y|u) - \sum_{1\leq|\mathbf{j}|\leq p} \frac{1}{\mathbf{j}!} G^{(\mathbf{j})}(y|u)(U_i - u)^{\mathbf{j}}$, and $\bar{\mathbf{1}}_y(W_i) \equiv \mathbf{1}\{Y_i \leq y\} - G(y|U_i)$. Let U_{is} and u_s denote the sth elements of U_i and u.

To prove Theorem 5.1, we first establish some technical lemmas. Let $\overline{\mathbf{S}}_{p,b}(u) \equiv E[\mathbf{S}_{p,b}(u)]$ and $\overline{\mathbf{B}}_{p,b}(y;u) \equiv E[\mathbf{B}_{p,b}(y;u)]$. The next lemma establishes uniform consistency of $\hat{\boldsymbol{\beta}}(y|u)$.

Lemma 10.1 *Suppose that Assumptions C.1-C.4 and C.6 hold. Then with* $\nu_b \equiv n^{-1/2} b^{-d/2}\sqrt{\log n}$, *we have that uniformly in* $(y,u) \in \mathbb{R} \times \mathcal{U}_0$,

(a) $\hat{\boldsymbol{\beta}}(y|u) - \boldsymbol{\beta}(y|u) = \overline{\mathbf{S}}_{p,b}(u)^{-1}[\mathbf{V}_{p,b}(y;u) + \overline{\mathbf{B}}_{p,b}(y;u)] + O_P(\nu_b^2 + \nu_b b^{p+1})$,

(b) $\hat{\boldsymbol{\beta}}(y|u) - \boldsymbol{\beta}(y|u) = O_P(\nu_b + b^{p+1})$.

Proof. Since $[\mathbf{S}_{p,b}(u)]^{-1}\mathbf{S}_{p,b}(u) = I_N$ where I_N is an $N \times N$ identity matrix, by (5.2) we obtain the following standard bias and variance decomposition:

$$\hat{\boldsymbol{\beta}}(y|u) - \boldsymbol{\beta}(y|u) = [\mathbf{S}_{p,b}(u)]^{-1}\mathbf{V}_{p,b}(y;u) + [\mathbf{S}_{p,b}(u)]^{-1}\mathbf{B}_{p,b}(y;u). \tag{10.3}$$

By Theorems 2 and 4 in Masry (1996) with some modification to account for the non-compact support of the kernel function,[11]

$$\mathbf{S}_{p,b}(u) = \overline{\mathbf{S}}_{p,b}(u) + O_P(\nu_b), \mathbf{V}_{p,b}(y;u) = O_P(\nu_b), \ \mathbf{B}_{p,b}(y;u) - \overline{\mathbf{B}}_{p,b}(y;u) = O_P(\nu_p b^{p+1}), \tag{10.4}$$

where the probability orders hold uniformly in $u \in \mathcal{U}_0$. By the same argument as used in the proof of Theorem 4.1 of Boente and Fraiman (1991), we can show that the last two results in (10.4) also hold uniformly in $y \in \mathbb{R}$ under Assumption C.3. In addition, by the Slutsky lemma,

$$\mathbf{S}_{p,b}(u)^{-1} = \left\{\overline{\mathbf{S}}_{p,b}(u) + \left[\mathbf{S}_{p,b}(u) - \overline{\mathbf{S}}_{p,b}(u)\right]\right\}^{-1} = [\overline{\mathbf{S}}_{p,b}(u)]^{-1} + O_P(\nu_b). \tag{10.5}$$

It follows that $\hat{\boldsymbol{\beta}}(y|u) - \boldsymbol{\beta}(y|u) = \{\overline{\mathbf{S}}_{p,b}(u)^{-1} + O_P(\nu_b)\}\{\mathbf{V}_{p,b}(y;u) + [\overline{\mathbf{B}}_{p,b}(y;u) + O_P(\nu_b b^{p+1})]\}$ $= \overline{\mathbf{S}}_{p,b}(u)^{-1}[\mathbf{V}_{p,b}(y;u) + \overline{\mathbf{B}}_{p,b}(y;u)] + O_P(\nu_b^2 + \nu_b b^{p+1}) = O_P(\nu_b + b^{p+1})$. ■

Recall that $\hat{G}_{p,b}(y|x,z) = e_{1,p}'\hat{\boldsymbol{\beta}}(y|u)$ where $e_{1,p}$ is defined after (5.3). Noting that uniformly in $(y,u) \in \mathbb{R}\times\mathcal{U}_0$, $\overline{\mathbf{S}}_{p,b}(u) = g(u)\mathbb{S}_p + O(b)$, and $\overline{\mathbf{B}}_{p,b}(y;u) = b^{p+1}g(u)\mathbb{B}_p\mathbf{G}_{p+1}(y|u) + o\left(b^{p+1}\right)$, with $\bar{\nu}_b \equiv \nu_b + b^{p+1}$, we have $\hat{G}_{p,b}(y|u) - G(y|u) = b^{p+1}e_1'\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}_{p+1}(y|u) + g(u)^{-1}e_1'\mathbb{S}_p^{-1}\mathbf{V}_{p,b}(y;u) + O_P(b\bar{\nu}_b)$.

Lemma 10.2 *Suppose that Assumptions C.1-C.4 and C.6 hold. Let* $\mathcal{T}$ *be any compact subset of* $(0,1)$. *Then uniformly in* $(\tau,u) \in \mathcal{T} \times \mathcal{U}_0$,

(a) $\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u) = e_{1,p}'\bar{S}_{p,b}(\tau;u)^{-1}\bar{V}_{p,b}(\tau;u) + O_P(\nu_b^2 + \nu_b b^{p+1}) + o_P\left(n^{-1/2}b^{-d_X/2}\right)$,

(b) $\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u) = e_{1,p}'\bar{S}_{p,b}(\tau;u)^{-1}V_{p,b}(\tau;u) + b^{p+1}e_{1,p}'\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}_{p+1}^{-1}(\tau|u) + O_P(\nu_b^2) + o_P\left(b^{p+1} + n^{-1/2}b^{-d_X/2}\right)$,

(c) $\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u) = e_{1,p}'S_p(\tau;u)^{-1}V_{p,b}(\tau;u)[1 + o_P(1)] + b^{p+1}e_{1,p}'\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}_{p+1}^{-1}(\tau|u) + o_P\left(b^{p+1} + n^{-1/2}b^{-d_X/2}\right)$,

where $S_p(\tau;u) \equiv \mathbb{S}_p g\left(G^{-1}(\tau|u)|u\right)g(u)$ *is the limit of* $\bar{S}_{p,b}(\tau;u) \equiv E[S_{p,b}(\tau;u)]$.

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

Proof. Noting that $S_{p,b}(\tau;u) - \bar{S}_{p,b}(\tau;u) = O_P(\nu_b)$ and $\bar{V}_{p,b}(\tau;u) = O_P(\nu_b + b^{p+1})$ by the proof of (b) below, (a) follows from Theorem 2.1 of Su and White (2011). To prove (b), write $\bar{V}_{p,b}(\tau;u) = V_{p,b}(\tau;u) + R_{p,b}(\tau;u)$, where $R_{p,b}(\tau;u) \equiv \frac{1}{n}\sum_{i=1}^{n}\{\mathbf{1}\left(Y_i \leq G^{-1}(\tau|U_i)\right) - \mathbf{1}(Y_i \leq \boldsymbol{\beta}_b(\tau;u)'\mu_{iu})\}K_b(U_i - u)\mu_{iu}$ and $\mu_{iu} \equiv \mu_{p,b}(U_i - u)$. Write $R_{p,b}(\tau;u)$ as $E[R_{p,b}(\tau;u)] + \{R_{p,b}(\tau;u) - E[R_{p,b}(\tau;u)]\}$. The first term is

$$\begin{aligned}
E[R_{p,b}(\tau;u)] &= E\left\{\left[G(G^{-1}(\tau|U_i)|U_i) - G(\boldsymbol{\beta}_b(\tau,u)'\mu_{iu}|U_i)\right]K_b(U_i-u)\mu_{iu}\right\} \\
&= E\left\{g(G^{-1}(\tau|U_i)|U_i)\left[G^{-1}(\tau|U_i) - \boldsymbol{\beta}_b(\tau,u)'\mu_{iu}\right]K_b(U_i-u)\mu_{iu}\right\}\{1+o(1)\} \\
&= b^{p+1}g(G^{-1}(\tau|u)|u)g(u)\mathbb{B}_p\mathbf{G}_{p+1}^{-1}(\tau|u)\{1+o(1)\}.
\end{aligned}$$

It is easy to show the second term is $o_P(b^{p+1})$ uniformly in (τ,u). Thus (b) follows. For (c), it suffices to show that $\sup_{(\tau,x)\in\mathcal{T}\times\mathcal{U}_0}\|S_{p,b}(\tau;u) - S_p(\tau;u)\| = O_P(n^{-1/2}b^{-d/2}\sqrt{\log n} + b) = o_P(1)$. The proof is similar to but simpler than that of Corollary 2 in Masry (1996) because we only need convergence in probability, whereas Masry proved almost sure convergence. ■

If $G(a|x^*,z) \in \mathcal{T}_0 = [\underline{\tau},\bar{\tau}] \subset (0,1)$ for $x^* \in \mathcal{X}_0$ and all $(a,z) \in \mathcal{A}_H \times \mathcal{Z}_0$, by Lemma 10.1, $\hat{G}_{p,b}(a|x^*,z) \in \mathcal{T}_0^\epsilon$ with probability approaching 1 for sufficiently large n, where $\mathcal{T}_0^\epsilon \equiv [\underline{\tau}-\epsilon,\bar{\tau}+\epsilon] \subset (0,1)$ for some $\epsilon > 0$. Then the result in Lemma 10.2 holds uniformly in $(\tau,u) \in \mathcal{T}_0^\epsilon \times \mathcal{U}_0$.

Lemma 10.3 *Suppose that Assumptions C.1-C.4 and C.6 hold. Then* $\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\leq Mv_b}\sup_{u\in\mathcal{U}_0}\sqrt{nb^{d_X}}\|V_{p,b}(\tilde{\tau};u) - V_{p,b}(\tau;u)\| = o_P(1)$.

Proof. Let $W(\tilde{\tau},\tau;u) = \omega'(V_{p,b}(\tilde{\tau};u) - V_{p,b}(\tau;u))$ where $\omega \in \mathbb{R}^{N_p}$ with $\|\omega\| = 1$. We need to show that

$$\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\leq Mv_b}\sup_{u\in\mathcal{U}_0}|W(\tilde{\tau},\tau;u)| = o_P(\vartheta_n) \text{ with } \vartheta_n = n^{-1/2}b^{-d_X/2}. \quad (10.6)$$

Let $a_{i,u} = K((U_i-u)/b)\omega'\mu_{p,b}(U_i-u)$, $a_{i,u}^+ = \max(a_{i,u},0)$ and $a_{i,u}^- = \max(-a_{i,u},0)$. Noting that $W(\tilde{\tau},\tau;u) = \left(nb^d\right)^{-1}\sum_{i=1}^n a_{i,u}\left[\tilde{\tau} - 1\left\{Y_i \leq G^{-1}(\tilde{\tau}|U_i)\right\} - \tau + 1\left\{Y_i \leq G^{-1}(\tau|U_i)\right\}\right]$, we can analogously define $W^+(\tilde{\tau},\tau;u)$ and $W^-(\tilde{\tau},\tau;u)$ by replacing $a_{i,u}$ in the definition of $W(\tilde{\tau},\tau;u)$ by $a_{i,u}^+$ and $a_{i,u}^-$, respectively. By the Minkowski inequality, (10.6) will hold if $\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\leq Mv_b}\sup_{u\in\mathcal{U}_0}|W^+(\tilde{\tau},\tau;u)| = o_P(\vartheta_n)$ and $\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\leq Mv_b}\sup_{u\in\mathcal{U}_0}|W^-(\tilde{\tau},\tau;u)| = o_P(\vartheta_n)$. We will only show the first part as the other case is similar.

Let $e_n \equiv n^{-1/2}$. By selecting $n_1 = O\left(e_n^{-1}\right)$ grid points, $\tau_1 < \tau_2 < \ldots < \tau_{n_1}$ with $\tau_j - \tau_{j-1} \leq e_n$, we can cover the compact set $\mathcal{T}_0^\epsilon$ by $\mathcal{T}_j = [\tau_{j-1},\tau_j]$ for $j = 1,\ldots,n_1$, where $\tau_0 = \underline{\tau} - \epsilon$ and $\tau_{n_1} = \bar{\tau}+\epsilon$. Similarly, we can select $n_2 = O\left(b^{-d}e_n^{-d}\right)$ grid points $u_1,\ldots,u_{n_2}$ to cover the compact set $\mathcal{U}_0$ by $\mathcal{U}_l = \{u : \|u-u_l\| \leq e_n b\}$, $l = 1,\ldots,n_2$. Observe that $\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\leq Mv_b}\sup_{u\in\mathcal{U}_0}|W^+(\tilde{\tau},\tau;u)| \leq W_{n1} + W_{n2}$, where

$$\begin{aligned}
W_{n1} &\equiv \max_{1\leq l\leq n_2}\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\leq Mv_b}\left|W^+(\tilde{\tau},\tau;u_l)\right|, \text{ and} \\
W_{n2} &\equiv \max_{1\leq l\leq n_2}\sup_{u\in\mathcal{U}_l}\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\leq Mv_b}\left|W^+(\tilde{\tau},\tau;u) - W^+(\tilde{\tau},\tau;u_l)\right|.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
W_{n1} &\leq \max_{1\leq l\leq n_2}\max_{1\leq k\leq n_1}\max_{1\leq j\leq n_1}\sup_{|\tau_j-\tau_k|\leq M\nu_b}\left|W^{+}(\tau_j,\tau_k;u_l)\right| \\
&\quad+\max_{1\leq l\leq n_2}\max_{1\leq j,k\leq n_1}\sup_{\tilde{\tau}\in\mathcal{T}_j}\sup_{\tau\in\mathcal{T}_k}\sup_{|\tilde{\tau}-\tau|\leq M\nu_b}\sup_{|\tau_j-\tau_k|\leq M\nu_b}\left|W^{+}(\tilde{\tau},\tau;u_l)-W^{+}(\tau_j,\tau_k;u_l)\right| \\
&\equiv W_{n11}+W_{n12},\text{ say.}
\end{aligned}$$

Let $\varsigma_{i,u_l}(\tau_j,\tau_k)=a^{+}_{i,u_l}\left[\tau_j-1\left\{Y_i\leq G^{-1}(\tau_j|U_i)\right\}-\tau_k+1\left\{Y_i\leq G^{-1}(\tau_k|U_i)\right\}\right]$. Noting that $|\varsigma_{i,u_l}(\tau_j,\tau_k)|\leq C$, $E[\varsigma_{i,u_l}(\tau_j,\tau_k)]=0$ and $E[\varsigma_{i,u_l}(\tau_j,\tau_k)^2]\leq Cb^d\nu_b$ as $|\tau_j-\tau_k|\leq M\nu_b$, we apply the Bernstein inequality (e.g., Serfling, 1980, p.95) and Assumption C6. to obtain

$$\begin{aligned}
P(W_{n11}>\vartheta_n\epsilon_0) &\leq C_1n_1n_2\nu_bn^{1/2}\max_{1\leq l\leq n_2}\max_{1\leq j,k\leq n_1:|\tau_j-\tau_k|\leq M\nu_b}P\left(W^{+}(\tau_j,\tau_k;u_l)>\vartheta_n\epsilon_0\right) \\
&\leq 2C_1n_1n_2\nu_bn^{1/2}\exp\left(-\frac{n^2b^{2d}\vartheta_n^2\epsilon_0^2}{2C_2nb^d\nu_b+\frac{2}{3}C_3nb^d\vartheta_n\epsilon_0}\right) \\
&= O\left(n_1n_2\nu_bn^{1/2}\right)\exp\left(-\frac{b^{d_Z}\epsilon_0^2}{C_4\left(n^{-1/2}b^{-d/2}\sqrt{\log n}+n^{-1/2}b^{-d_X/2}\epsilon_0\right)}\right)=o(1),
\end{aligned}$$

where C_i, $i=1,2,3,4$, are positive constants. Thus $W_{n11}=o_P(\vartheta_n)$. By the monotonicity of the indicator and quantile functions and the nonnegativity of a^{+}_{i,u_l}, we can readily show that

$$\begin{aligned}
W_{n12} &= \max_{1\leq l\leq n_2}\max_{\substack{1\leq j,k\leq n_1\\|\tau_j-\tau_k|\leq M\nu_b}}\sup_{\substack{\tilde{\tau}\in\mathcal{T}_j,\tau\in\mathcal{T}_k\\|\tilde{\tau}-\tau|\leq M\nu_b}}\Bigg|\frac{1}{nb^d}\sum_{i=1}^{n}a^{+}_{i,u_l}[\tilde{\tau}-1\left\{Y_i\leq G^{-1}(\tilde{\tau}|U_i)\right\}-\tau \\
&\quad+1\left\{Y_i\leq G^{-1}(\tau|U_i)\right\}]-a^{+}_{i,u_l}\left[\tau_j-1\left\{Y_i\leq G^{-1}(\tau_j|U_i)\right\}-\tau_k+1\left\{Y_i\leq G^{-1}(\tau_k|U_i)\right\}\right]\Bigg| \\
&\leq \max_{\substack{1\leq l\leq n_2,\\1\leq j\leq n_1}}\sup_{\tilde{\tau}\in\mathcal{T}_j}\left|\frac{1}{nh^d}\sum_{i=1}^{n}a^{+}_{i,u_l}\left[\tilde{\tau}-1\left\{Y_i\leq G^{-1}(\tilde{\tau}|U_i)\right\}-\tau_j+1\left\{Y_i\leq G^{-1}(\tau_j|U_i)\right\}\right]\right| \\
&\quad+\max_{\substack{1\leq l\leq n_2,\\1\leq j\leq n_1}}\sup_{\tau\in\mathcal{T}_k}\left|\frac{1}{nh^d}\sum_{i=1}^{n}a^{+}_{i,u_l}\left[\tau-1\left\{Y_i\leq G^{-1}(\tau|U_i)\right\}-\tau_k-1\left\{Y_i\leq G^{-1}(\tau_k|U_i)\right\}\right]\right| \\
&= O_P(n^{-1/2})=o_P(\vartheta_n).
\end{aligned}$$

We now study W_{n2}. Assumption C.4(iii) implies that for all $\|u_1-u_2\|\leq\delta\leq\sigma_K$,

$$|K(u_2)-K(u_1)|\leq\delta K^{*}(u_1), \tag{10.7}$$

where $K^{*}(u)=\overline{C}\mathbf{1}(\|u\|\leq 2d\sigma_K)$ for some constant $\overline{C}$ that depends on $\overline{\sigma}_1$ and $\overline{\sigma}_2$ in the assumption. For any $u\in\mathcal{U}_l$, $\|u-u_l\|/b\leq e_n$. It follows from (10.7) that $|K_{iu}-K_{iu_l}|\leq e_nK^{*}_{iu_l}$ where

$K_{iu} \equiv K\left((U_i - u_l)/b\right)$ and $K^*_{iu_l} \equiv K^*\left((U_i - u_l)/b\right)$, and

$$\begin{aligned}
&\left|\left(\frac{U_i - u}{b}\right)^{\boldsymbol{k}} K_{iu} - \left(\frac{U_i - u_l}{b}\right)^{\boldsymbol{k}} K_{iu_l}\right| \\
\leq\ & \left|\left(\frac{U_i - u}{b}\right)^{\boldsymbol{k}}\right| \left|K_{iu} - K_{iu_l}\right| + \left|\left(\frac{U_i - u}{b}\right)^{\boldsymbol{k}} - \left(\frac{U_i - u_l}{b}\right)^{\boldsymbol{k}}\right| K_{iu_l} \\
\leq\ & (2\sigma_k)^{|\boldsymbol{k}|} e_n K^*_{iu_l} + (2\sigma_k)^{|\boldsymbol{k}|-1} e_n K_{iu_l} \mathbf{1}\left(|\boldsymbol{k}| > 0\right) \leq C e_n (K^*_{iu_l} + K_{iu_l}).
\end{aligned}$$

With this, we can show that for any $u \in \mathcal{U}_l$ such that $\|u - u_l\|/b \leq e_n$, we have

$$\left|a^+_{i,u} - a^+_{i,u_l}\right| = \left|K_{iu}\omega' \mu_{p,b}\left(U_i - u\right) - K_{iu_l}\omega' \mu_{p,b}\left(U_i - u_l\right)\right| \leq C e_n (K^*_{iu_l} + K_{iu_l}).$$

It follows that

$$\begin{aligned}
W_{n2} &= \max_{1 \leq l \leq n_2} \sup_{u \in \mathcal{U}_l} \sup_{\tau \in \mathcal{T}_0^\epsilon} \sup_{\tilde{\tau}, |\tilde{\tau} - \tau| \leq M\nu_b} \left|W^+\left(\tilde{\tau}, \tau; u\right) - W^+\left(\tilde{\tau}, \tau; u_l\right)\right| \\
&\leq 2 \max_{1 \leq l \leq n_2} \sup_{u \in \mathcal{U}_l} \left(nb^d\right)^{-1} \sum_{i=1}^n \left|a^+_{i,u} - a^+_{i,u_l}\right| \\
&\leq C e_n \max_{1 \leq l \leq n_2} \left(nb^d\right)^{-1} \sum_{i=1}^n \left(K^*_{iu_l} + K_{iu_l}\right) = O_P\left(e_n\right) = o_P(\vartheta_n).
\end{aligned}$$

Thus we have proved that $\sup_{\tilde{\tau}, \tau \in \mathcal{T}_0^\epsilon, |\tilde{\tau} - \tau| \leq M v_b} \sup_{u \in \mathcal{U}_0} \left|W^+\left(\tilde{\tau}, \tau; u\right)\right| = o_P\left(\vartheta_n\right)$. ■

By Lemmas 10.1 and 10.3, with probability approaching 1 we have

$$\sup_{(a,x,z) \in \mathcal{A}_H \times \mathcal{X}_0 \times \mathcal{Z}_0} \sqrt{nb^{d_X}} \left|V_{p,b}\left(\hat{G}_{p,b}(a|x^*, z); x, z\right) - V_{p,b}\left(G(a|x^*, z); x, z\right)\right| = o_P\left(1\right).$$

Proof of Theorem 5.1 Letting $\hat{\tau}_z \equiv \hat{G}_{p,b}(a|x^*, z)$ and $\tau_z \equiv G(a|x^*, z)$, we have $\hat{m}_H\left(x, a\right) - m^*_H\left(x, a\right) = \int [G^{-1}(\hat{\tau}_z \mid x, z) - G^{-1}(\tau_z \mid x, z)] dH(z) + \int [\hat{G}^{-1}_{p,b}(\hat{\tau}_z \mid x, z) - G^{-1}(\hat{\tau}_z \mid x, z)] dH(z) \equiv M_{n1}\left(x, a\right) + M_{n2}\left(x, a\right)$, say. Note that

$$G^{-1}(\hat{\tau}_z|x, z) - G^{-1}(\tau_z|x, z) = \frac{\hat{\tau}_z - \tau_z}{g\left(G^{-1}\left(\tau_z|x, z\right)|x, z\right)} + \hat{r}\left(a; x^*, x, z\right),$$

where $\hat{r}\left(a; x^*, x, z\right) \equiv -\frac{g'\left(G^{-1}(\tau^*_z|x,z)|x,z\right)}{g(G^{-1}(\tau^*_z|x,z)|x,z)^3}(\hat{\tau}_z - \tau_z)^2$ and τ^*_z lies between $\hat{\tau}_z$ and τ_z. By the remark after Lemma 10.1 and Assumption C.6, $\hat{r}\left(a; x^*, x, z\right) = O_P\left(n^{-1} b^{-d} \log n + b^{2(p+1)}\right) = o_P\left(n^{-1/2} b^{-d_X/2}\right)$ uniformly in $(a, x, z) \in \mathcal{A}_H \times \mathcal{X}_0 \times \mathcal{Z}_0$. It follows that for all $(a, x) \in \mathcal{A}_H \times \mathcal{X}_0$

$$\begin{aligned}
\sqrt{nb^{d_X}} M_{n1}\left(x, a\right) &= \sqrt{nb^{d_X}} \int \frac{\hat{\tau}_z - \tau_z}{g\left(G^{-1}\left(\tau_z|x, z\right)|x, z\right)} dH\left(z\right) + o_P\left(1\right) \\
&= \sqrt{nb^{d_X}} \int \frac{e'_{1,p} \overline{\mathbf{S}}_{p,b}\left(x^*, z\right)^{-1} \overline{\mathbf{B}}_{p,b}\left(a; x^*, z\right)}{g\left(G^{-1}\left(\tau_z|x, z\right)|x, z\right)} dH\left(z\right) \\
&\quad + \sqrt{nb^{d_X}} \int \frac{e'_{1,p} \overline{\mathbf{S}}_{p,b}\left(x^*, z\right)^{-1} \mathbf{V}_{p,b}\left(a; x^*, z\right)}{g\left(G^{-1}\left(\tau_z|x, z\right)|x, z\right)} dH\left(z\right) + o_P\left(1\right) \\
&\equiv M_{n11}\left(x, a\right) + M_{n12}\left(x, a\right) + o_P\left(1\right), \text{ say,}
\end{aligned}$$

where the second line follows from Lemma 10.1(a). Noting that $\overline{\mathbf{B}}_{p,b}(a;u) = E[K_b(U_i-u)\mu_{p,b}(U_i-u)\Delta_{i,a}(u)] = b^{p+1}g(u)\mathbb{B}_p\mathbf{G}_{p+1}(a|u) + o\left(b^{p+1}\right)$ and $\overline{\mathbf{S}}_{p,b}(u) = \mathbb{S}_p g(u) + o(1)$ uniformly in $(a,u) \in \mathcal{A}_H \times \mathcal{U}_0$, we have

$$M_{n11}(x,a) = \sqrt{nb^{d_X}}b^{p+1}\int \frac{e_{1,p}'\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}_{p+1}(a|x^*,z)}{g(G^{-1}(\tau_z|x,z)|x,z)}dH(z)\{1+o_P(1)\} \tag{10.8}$$

and

$$\begin{aligned} M_{n12}(x,a) &= \sqrt{nb^{d_X}}\int \frac{e_{1,p}'\mathbb{S}_p^{-1}\mathbf{V}_n(a;x^*,z)}{g(x^*,z)\,g(G^{-1}(\tau_z|x,z)|x,z)}dH(z)\{1+o_P(1)\} \\ &= \bar{M}_{n12}(x,a)\{1+o_P(1)\} \xrightarrow{d} N(0,V_1), \end{aligned} \tag{10.9}$$

where $\bar{M}_{n12}(x,a) \equiv \sqrt{\frac{b^{d_X}}{n}}\sum_{i=1}^n\int \frac{e_{1,p}'\mathbb{S}_p^{-1}K_b(X_i-x^*,Z_i-z)\mu_{p,b}(X_i-x^*,Z_i-z)\bar{\mathbf{1}}_a(W_i)}{g(x^*,z)g(G^{-1}(\tau_z|x,z)|x,z)}dH(z)$, (10.8) holds true uniformly in $(a,x) \in \mathcal{A}_H \times \mathcal{X}_0$, $V_1 \equiv \kappa_{1p}\int \frac{\tau_z(1-\tau_z)h(z)^2}{g(x^*,z)g(G^{-1}(\tau_z|x,z)|x,z)^2}dz$, $\kappa_{1p} \equiv \int e_{1,p}'\mathbb{S}_p^{-1}\mu_p(\tilde{x},\tilde{z})\mu_p(\tilde{x},\tilde{z}-\bar{z})'\mathbb{S}_p^{-1}e_{1,p}K(\tilde{x},\tilde{z})\ K(\tilde{x},\tilde{z}-\bar{z})\,d(\tilde{x},\tilde{z},\bar{z})$, and (10.9) follows from straightforward moment calculations and Liapounov's central limit theorem.

For M_{n2}, noting that $\sqrt{nb^{d_X}}o(b^{p+1}+n^{-1/2}b^{-d_X/2}) = o(1)$ under Assumption C.6, by Lemma 10.2 we have that uniformly in $(a,x) \in \mathcal{A}_H \times \mathcal{X}_0$

$$\begin{aligned} \sqrt{nb^{d_X}}M_{n2}(x,a) &= \sqrt{nb^{d_X}}\int b^{p+1}e_{1,p}'\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}_{p+1}^{-1}(\hat{\tau}_z|x,z)\,dH(z) \\ &\quad + \sqrt{nb^{d_X}}\int e_{1,p}'S_p(\hat{\tau}_z;x,z)^{-1}V_n(\hat{\tau}_z;x,z)\,dH(z) + o_P(1), \\ &\equiv M_{n21}(x,a) + M_{n22}(x,a) + o_P(1), \text{ say,} \end{aligned}$$

where $S_p(\tau;u) \equiv \mathbb{S}_p g\left(G^{-1}(\tau|u)|u\right)g(u)$. By Lemmas 10.2 and 10.3, we have that uniformly in $(a,x) \in \mathcal{A}_H \times \mathcal{X}_0$

$$M_{n21}(x,a) = \sqrt{nb^{d_X}}b^{p+1}\int e_{1,p}'\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}_{p+1}^{-1}(\tau_z|x,z)\,dH(z) + o_P(1), \tag{10.10}$$

and $M_{n22}(x,a) = \bar{M}_{n22}(x,a) + o_P(1)$, where $\bar{M}_{n22}(x,a) \equiv \int e_{1,p}'S_p(\tau_z;x,z)^{-1}V_n(\tau_z;x,z)dH(z)$. Furthermore,

$$\begin{aligned} \bar{M}_{n22}(x,a) &= \sqrt{\frac{b^{d_X}}{n}}\sum_{i=1}^n\int \frac{e_{1,p}'\mathbb{S}_p^{-1}K_b(U_i-u)\,\mu_{p,b}(U_i-u)\psi_{\tau_z}\left(Y_i-G^{-1}(\tau_z|U_i)\right)}{g(x,z)\,g(G^{-1}(\tau_z|x,z)|x,z)}dH(z) \\ &\xrightarrow{d} N(0,V_2), \end{aligned}$$

where $V_2 \equiv \kappa_{1p}\int \frac{\tau_z(1-\tau_z)h(z)^2}{g(x,z)g(G^{-1}(\tau_z|x,z)|x,z)^2}dz$. The asymptotic normality result follows by the Cramér-Wold device and the fact that the asymptotic covariance of $\bar{M}_{n12}$ and $\bar{M}_{n22}$ zero. In sum, we have $\sqrt{nb^{d_X}}\{\hat{m}_H(x,a) - m_H^*(x,a) - B_m(x,a;x^*)\} \xrightarrow{d} N\left(0,\sigma_m^2(x,a;x^*)\right)$, where $B_m(x,a;x^*)$ and $\sigma_m^2(x,a;x^*)$ are defined in (5.6) and (5.7), respectively.

Next, it is standard to show that $\sup_{(x,a)\in\mathcal{X}_0\times\mathcal{A}_H}|M_{n12}(x,a)| = O_P(\sqrt{\log n})$ and $\sup_{(x,a)\in\mathcal{X}_0\times\mathcal{A}_H}\left|\bar{M}_{n22}(x,a)\right| = O_P(\sqrt{\log n})$. Then the uniform convergence result follows. ■

Lemma 10.4 *Suppose that Assumptions C.1-C.6 hold. Then for any* $\delta_n = O(\nu_b)$, *we have*

(a) $\hat{G}_{p,b}(a+\delta_n|u) - \hat{G}_{p,b}(a|u) = g(a|u)\delta_n + o_P\left(\delta_n + n^{-1/2}b^{-d_X/2}\right)$ *uniformly in* $u \in \mathcal{U}_0$,

(b) $\hat{m}_H(x, a+\delta_n) - \hat{m}_H(x,a) = D_H^*(x,a)\,\delta_n + o_P\left(\delta_n + n^{-1/2}b^{-d_X/2}\right)$,

where $D_H^*(x,a) \equiv \int \frac{g(a|x^*,z)}{g(G^{-1}(G(a|x^*,z)|x,z)|x,z)} dH(z)$.

Proof. By Lemma 10.1,

$$\begin{aligned}\hat{G}_{p,b}(a+\delta_n|u) - \hat{G}_{p,b}(a|u) &= [G(a+\delta_n|u) - G(a|u)] \\ &\quad + e_{1,p}'\overline{\mathbf{S}}_{p,b}(u)^{-1}\left[\overline{\mathbf{B}}_{p,b}(a+\delta_n;u) - \overline{\mathbf{B}}_{p,b}(a;u)\right] \\ &\quad + e_{1,p}'\overline{\mathbf{S}}_{p,b}(u)^{-1}\left[\mathbf{V}_{p,b}(a+\delta_n;u) - \mathbf{V}_{p,b}(a;u)\right] + O_P(\nu_b^2 + \nu_b b^{p+1}).\end{aligned}$$

Clearly, the first term on the right hand side of the last expression is $g(a|x,z)\delta_n + o(\delta_n)$; the second term is $o(b^{p+1}) = o_P(n^{-1/2}b^{-d_X/2})$ uniformly in $u \in \mathcal{U}_0$ by the fact that $\overline{\mathbf{B}}_{p,b}(a;u) = b^{p+1}\mathbb{B}_p\mathbf{G}_{p+1}(a|u)\,g(u) + o\left(b^{p+1}\right)$ uniformly in u and $\overline{\mathbf{S}}_{p,b}(u) = \mathbb{S}_p g(u) + o(1)$, and the continuity of $\mathbf{G}_{p+1}$. Analogously to the proof of Lemma 10.3, we can show that $\mathbf{V}_{p,b}(a+\delta_n;u) - \mathbf{V}_{p,b}(a;u) = o_P(n^{-1/2}b^{-d_X/2})$ uniformly in $u \in \mathcal{U}_0$. Thus (a) follows.

To show (b), decompose $\hat{m}_H(x, a+\delta_n) - \hat{m}_H(x,a) = D_{n1} + D_{n2}$, where

$$D_{n1} \equiv \int \left[G^{-1}(\hat{G}_{p,b}(a+\delta_n|x^*,z)|x,z)) - G^{-1}(\hat{G}_{p,b}(a|x^*,z)|x,z))\right] dH(z),$$

and

$$\begin{aligned}D_{n2} &\equiv \int \left[\hat{G}_{p,b}^{-1}(\hat{G}_{p,b}(a+\delta_n|x^*,z)|x,z)) - G^{-1}(\hat{G}_{p,b}(a+\delta_n|x^*,z)|x,z))\right] dH(z) \\ &\quad - \int \left[\hat{G}_{p,b}^{-1}(\hat{G}_{p,b}(a|x^*,z)|x,z)) - G^{-1}(\hat{G}_{p,b}(a|x^*,z)|x,z))\right] dH(z).\end{aligned}$$

For D_{n1}, we have

$$\begin{aligned}D_{n1} &= \int \frac{\hat{G}_{p,b}(a+\delta_n|x^*,z) - \hat{G}_{p,b}(a|x^*,z)}{g\left(G^{-1}\left(\hat{G}_{p,b}(a|x^*,z)|x,z\right)|x,z\right)} dH(z) + o_P\left(\delta_n + n^{-1/2}b^{-d_X/2}\right) \\ &= \int \frac{g(a|x^*,z)\delta_n}{g\left(G^{-1}\left(\hat{G}_{p,b}(a|x^*,z)|x,z\right)|x,z\right)} dH(z) + o_P\left(\delta_n + n^{-1/2}b^{-d_X/2}\right) \\ &= D_H^*(x,a)\,\delta_n + o_P\left(\delta_n + n^{-1/2}b^{-d_X/2}\right),\end{aligned}$$

where the first equality follows from the Taylor expansion, the second from (a), and the third from Lemma 10.1. By the proof of Theorem 5.1, we have

$$\begin{aligned}D_{n2} &\approx b^{p+1}\int e_{1,p}'\mathbb{S}_p^{-1}\mathbb{B}_p\left[\mathbf{G}_{p+1}^{-1}(G(a+\delta_n|x^*,z)|u) - \mathbf{G}_{p+1}^{-1}(G(a|x^*,z)|u)\right] dH(z) \\ &\quad + \int e_{1,p}'[S_p(G(a+\delta_n|x^*,z);x,z)^{-1}V_{p,b}(G(a+\delta_n|x^*,z);x,z) \\ &\qquad - S_p(G(a|x^*,z);x,z)^{-1}V_{p,b}(G(a|x^*,z);x,z)]dH(z) \\ &\equiv D_{n21} + D_{n22}, \text{ say.}\end{aligned}$$

It is easy to see that $D_{n21} = o(b^{p+1}) = o_P(n^{-1/2}b^{-d_X/2})$ by the continuity of G and $\mathbf{G}_{p+1}^{-1}$. Next, we write $D_{n22} = D_{n22,1} + D_{n22,2}$, where

$$\begin{aligned} D_{n22,1} &= \int e_{1,p}' S_p\left(G(a+\delta_n|x^*,z);x,z\right)^{-1}\left[V_{p,b}\left(G(a+\delta_n|x^*,z);x,z\right) - V_{p,b}\left(G(a|x^*,z);x,z\right)\right] dH(z), \\ D_{n22,2} &= \int e_{1,p}'[S_p\left(G(a+\delta_n|x^*,z);x,z\right)^{-1} - S_p\left(G(a|x^*,z);x,z\right)^{-1}]V_{p,b}\left(G(a|x^*,z);x,z\right) dH(z). \end{aligned}$$

One can readily show that $D_{n22,1} = o_P(n^{-1/2}b^{-d_X/2})$ and $D_{n22,2} = o_P(n^{-1/2}b^{-d_X/2})$ by standard moment calculations and the dominated convergence theorem, and (b) follows. ■

Proof of Theorem 5.3 By the strict monotonicity of $m_H^*(x,\cdot)$ for all x, its inverse function $m_H^{*-1}(x,\cdot)$ exists and is unique. This implies that for any fixed (x,a) with $y = m_H^*(x,a)$ (and thus $a = m_H^{*-1}(x,y)$), there is an $\epsilon = \epsilon(x) > 0$ such that

$$\delta = \delta(\epsilon) = \min\{m_H^*(x,a) - m_H^*(x,a-\epsilon),\ m_H^*(x,a+\epsilon) - m_H^*(x,a)\} > 0. \qquad (10.11)$$

It follows that for sufficiently large n,

$$\begin{aligned} & P\left\{\left|\hat{m}_H^{-1}(x,y) - m_H^{*-1}(x,y)\right| > \epsilon\right\} \\ = \;& P\left\{\hat{m}_H^{-1}(x,y) > m_H^{*-1}(x,y) + \epsilon \text{ or } \hat{m}_H^{-1}(x,y) < m_H^{*-1}(x,y) - \epsilon\right\} \\ = \;& P\left\{m_H^*\left(x,\hat{m}_H^{-1}(x,y)\right) > m_H^*\left(x,m_H^{*-1}(x,y)+\epsilon\right) \text{ or } m_H^*\left(x,\hat{m}_H^{-1}(x,y)\right) < m_H^*\left(x,m_H^{*-1}(x,y)\text{-}\epsilon\right)\right\} \\ \le \;& P\left\{\left|m_H^*\left(x,\hat{m}_H^{-1}(x,y)\right) - m_H^*\left(x,m_H^{*-1}(x,y)\right)\right| > \delta\right\} \\ = \;& P\left\{\left|m_H^*\left(x,\hat{m}_H^{-1}(x,y)\right) - y\right| > \delta\right\} \\ = \;& P\left\{\left|m_H^*\left(x,\hat{m}_H^{-1}(x,y)\right) - \hat{m}_H\left(x,\hat{m}_H^{-1}(x,y)\right)\right| > \delta\right\} \\ \le \;& P\left\{\sup_{a\in\mathcal{A}_H^\eta} \left|\hat{m}_H(x,a) - m_H^*(x,a)\right| > \delta\right\} \to 0, \end{aligned}$$

where the third line follows from the monotonicity of $m_H^*(x,\cdot)$, the fourth line holds by (10.11), the fifth and six lines follow from the fact $m_H^*\left(x,m_H^{*-1}(x,y)\right) = y = \hat{m}_H\left(x,\hat{m}_H^{-1}(x,y)\right)$, and $\mathcal{A}_H^\eta \equiv \{a : |a-c| \le \eta \text{ for some } c \in \mathcal{A}_H\}$ and $\eta > 0$.

Let $\Phi_n(v) \equiv P\left\{n^{1/2}b^{d_X/2}\sigma_{m^{-1}}^{-1}(x,y)\left[\hat{m}_H^{-1}(x,y) - m_H^{*-1}(x,y) - B_{m^{-1}}(x,y)\right] \le v\right\}$ for any $v \in \mathbb{R}$. Then

$$\begin{aligned} \Phi_n(v) &= P\left\{\hat{m}_H^{-1}(x,y) \le m_H^{*-1}(x,y) + \delta_n(v;x,y)\right\} \\ &= P\left\{\hat{m}_H\left(x,\hat{m}_H^{-1}(x,y)\right) \le \hat{m}_H\left(x,m_H^{*-1}(x,y) + \delta_n(v;x,y)\right)\right\} \\ &= P\left(\hat{m}_H\left(x,m_H^{*-1}(x,y) + \delta_n(v;x,y)\right) \ge y\right), \end{aligned}$$

where $\delta_n(v;x,y) \equiv B_{m^{-1}}(x,y) + n^{-1/2} b^{-d_X/2} \sigma_{m^{-1}}(y;x)\, v$. By Lemma 10.4(b) and Theorem 5.1,

$$\begin{aligned}
\Phi_n(v) &\approx P\left\{\hat{m}_H\left(x, m_H^{*-1}(x,y)\right) \geq -D_H^*\left(x, m_H^{*-1}(x,y)\right)\delta_n(v;x,y) + y\right\} \\
&= P\left\{\hat{m}_H\left(x, m_H^{*-1}(x,y)\right) - y + D_H^*\left(x, m_H^{*-1}(x,y)\right) B_{m^{-1}}(x,y) \geq \right. \\
&\qquad \left. -(n^{-1/2} b^{-d_X/2}) D_H^*\left(x, m_H^{*-1}(x,y)\right)\sigma_{m^{-1}}(y;x)\, v\right\} \\
&= P\left\{\sqrt{nb^{d_X}}\left[D_H^*\left(x, m_H^{*-1}(x,y)\right)\sigma_{m^{-1}}(y;x)\right]^{-1}\right. \\
&\qquad \left. \times \left[\hat{m}_H\left(x, m_H^{*-1}(x,y)\right) - y + D_H^*\left(x, m_H^{*-1}(x,y)\right) B_{m^{-1}}(x,y)\right] \geq -v\right\} \\
&\to 1 - \Phi(-v) = \Phi(v),
\end{aligned}$$

where Φ is the CDF for the standard normal distribution. ■

Proofs of Theorems 6.1 and 6.2 We prove Theorem 6.2, as Theorem 6.1 is a special case. Put $G = G_n$, and let $\bar{J}_n \equiv b^{d_X} \sum_{i=1}^n \{\int G^{-1}(\hat{G}_{p,b}(Y_i|X_i,z)\,|x^*,z) d\Delta(z)\}^2 \pi_i$, $\alpha_n(\tau|u) \equiv \hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u)$, $\alpha_{n1}(\tau;u) \equiv e_1' \bar{S}_{p,b}(\tau;u)^{-1} \bar{V}_{p,b}(\tau;u)$, and $\alpha_{n2}(\tau;u) \equiv \alpha_n(\tau|u) - \alpha_{n1}(\tau;u)$, $\tau_{iz} \equiv G(Y_i|X_i,z)$, and $\hat{\tau}_{iz} \equiv \hat{G}_{p,b}(Y_i|X_i,z)$. Noting that $a^2 - b^2 = (a-b)^2 + 2(a-b)b$, we have

$$\begin{aligned}
\hat{J}_n &= \bar{J}_n + (\hat{J}_n - \bar{J}_n) \\
&= \bar{J}_n + b^{d_X} \sum_{i=1}^n \left\{\int \left[\hat{G}_{p,b}^{-1}(\hat{\tau}_{iz}|x^*,z) - G^{-1}(\hat{\tau}_{iz}|x^*,z)\right] d\Delta(z)\right\}^2 \pi_i \\
&\quad + 2b^{d_X} \sum_{i=1}^n \int \left[\hat{G}_{p,b}^{-1}(\hat{\tau}_{iz}|x^*,z) - G^{-1}(\hat{\tau}_{iz}|x^*,z)\right] d\Delta(z) \int G^{-1}(\hat{\tau}_{iz}|x^*,z)\, d\Delta(z)\, \pi_i \\
&= \bar{J}_n + b^{d_X} \sum_{i=1}^n \left[\int \alpha_{n1}(\hat{\tau}_{iz};x^*,z)\, d\Delta(z)\right]^2 \pi_i + b^{d_X} \sum_{i=1}^n \left[\int \alpha_{n2}(\hat{\tau}_{iz};x^*,z)\, d\Delta(z)\right]^2 \pi_i \\
&\quad + 2b^{d_X} \sum_{i=1}^n \int \alpha_{n1}(\hat{\tau}_{iz};x^*,z)\, d\Delta(z) \int \alpha_{n2}(\hat{\tau}_{iz};x^*,z)\, d\Delta(z)\, \pi_i \\
&\quad + 2b^{d_X} \sum_{i=1}^n \int \alpha_{n1}(\hat{\tau}_{iz};x^*,z)\, d\Delta(z) \int G^{-1}(\hat{\tau}_{iz}|x^*,z)\, d\Delta(z)\, \pi_i \\
&\quad + 2b^{d_X} \sum_{i=1}^n \int \alpha_{n2}(\hat{\tau}_{iz};x^*,z)\, d\Delta(z) \int G^{-1}(\hat{\tau}_{iz}|x^*,z)\, d\Delta(z)\, \pi_i \\
&\equiv \bar{J}_n + \hat{J}_{n1} + \hat{J}_{n2} + 2\hat{J}_{n3} + 2\hat{J}_{n4} + 2\hat{J}_{n5}, \text{ say.}
\end{aligned} \tag{10.12}$$

We prove the theorem by first demonstrating that

$$\hat{J}_n = J_n + o_P(1) \tag{10.13}$$

and then showing that

$$J_n - B_n \xrightarrow{d} N\left(\mu_0, \sigma_J^2\right), \tag{10.14}$$

where

$$J_n \equiv b^{d_X} \sum_{i=1}^{n} \left[\int \frac{e_1' \bar{\mathbf{S}}_{p,b}(X_i,z)^{-1} \mathbf{V}_{p,b}(Y_i;X_i,z)}{g(G^{-1}(\tau_{iz}|x^*,z)|x^*,z)} + e_1' \bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1} V_{p,b}(\tau_{iz};x^*,z)\, d\Delta(z) \right]^2 \pi_i.$$

We prove (10.13) by showing that

$$\bar{J}_n = b^{d_X} \sum_{i=1}^{n} \left[\int \frac{e_1' \bar{\mathbf{S}}_{p,b}(X_i,z)^{-1} \mathbf{V}_{p,b}(Y_i;X_i,z)}{g(G^{-1}(\tau_{iz}|x^*,z)|x^*,z)} \right]^2 \pi_i + \mu_0 + o_P(1), \tag{10.15}$$

$$\hat{J}_{n1} = b^{d_X} \sum_{i=1}^{n} \left[\int e_1' \bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1} V_{p,b}(\tau_{iz};x^*,z)\, d\Delta(z) \right]^2 \pi_i + o_P(1), \tag{10.16}$$

$$\hat{J}_{n4} = \tilde{J}_{n4} + o_P(1), \quad \text{and} \tag{10.17}$$

$$\hat{J}_{ns} = o_P(1) \text{ for } s = 2,3,5, \tag{10.18}$$

where

$$\tilde{J}_{n4} \equiv b^{d_X} \sum_{i=1}^{n} \int \frac{e_1' \bar{\mathbf{S}}_{p,b}(X_i,z)^{-1} \mathbf{V}_{p,b}(Y_i;X_i,z)}{g(G^{-1}(\tau_{iz}|x^*,z)|x^*,z)} e_1' \bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1} V_{p,b}(\tau_{iz};x^*,z)\, d\Delta(z)\, \pi_i. \tag{10.19}$$

To show (10.15), write

$$\begin{aligned} \bar{J}_n &= b^{d_X} \sum_{i=1}^{n} \left[\int G^{-1}(\tau_{iz}|x^*,z) d\Delta(z) \right]^2 \pi_i \\ &\quad + b^{d_X} \sum_{i=1}^{n} \left\{ \int \left[G^{-1}(\hat{\tau}_{iz}|x^*,z) - G^{-1}(\tau_{iz}|x^*,z) \right] d\Delta(z) \right\}^2 \pi_i \\ &\quad + 2b^{d_X} \sum_{i=1}^{n} \int G^{-1}(\tau_{iz}|x^*,z) d\Delta(z) \int \left[G^{-1}(\hat{\tau}_{iz}|x^*,z) - G^{-1}(\tau_{iz}|x^*,z) \right] d\Delta(z)\, \pi_i \\ &\equiv \bar{J}_{n1} + \bar{J}_{n2} + 2\bar{J}_{n3}, \text{ say.} \end{aligned}$$

Under $\mathbb{H}_1(\gamma_n)$ with $\gamma_n = n^{-1/2} b^{-d_X/2}$, $\bar{J}_{n1} = n^{-1} \sum_{i=1}^{n} \delta_n(X_i,Y_i)^2 \pi_i \xrightarrow{P} \mu_0$. Noting that

$$G^{-1}(\hat{\tau}_{iz}|x^*,z) - G^{-1}(\tau_{iz}|x^*,z) = \frac{\hat{\tau}_{iz} - \tau_{iz}}{g(G^{-1}(\tau_{iz}|x^*,z)|x^*,z)} + \hat{r}_i(z), \tag{10.20}$$

where $\hat{r}_i(z) = -\frac{g'\left(G^{-1}\left(\tau_{iz}^*|x^*,z\right)|x^*,z\right)}{g\left(G^{-1}\left(\tau_{iz}^*|x^*,z\right)|x^*,z\right)^3} (\hat{\tau}_{iz} - \tau_{iz})^2$ and τ_{iz}^* lies between τ_{iz} and $\hat{\tau}_{iz}$, we have that under $\mathbb{H}_1(\gamma_n)$,

$$\begin{aligned} \bar{J}_{n3} &= n^{-1/2} b^{d_X/2} \sum_{i=1}^{n} \delta_n(X_i,Y_i) \int \left[G^{-1}(\hat{\tau}_{iz}|x^*,z) - G^{-1}(\tau_{iz}|x^*,z) \right] d\Delta(z)\, \pi_i \\ &= n^{-1/2} b^{d_X/2} \sum_{i=1}^{n} \delta_n(X_i,Y_i) \int \frac{\hat{\tau}_{iz} - \tau_{iz}}{g(G^{-1}(\tau_{iz}|x^*,z)|x^*,z)} d\Delta(z)\, \pi_i \\ &\quad + n^{-1/2} b^{d_X/2} \sum_{i=1}^{n} \delta_n(X_i,Y_i) \int \hat{r}_i(z)\, d\Delta(z)\, \pi_i \\ &\equiv \bar{J}_{n31} + \bar{J}_{n32}, \text{ say.} \end{aligned}$$

Observing that $\hat{r}_i(z) = O_P\left(n^{-1}b^{-d}\log n + b^{2(p+1)}\right)$ uniformly in z, we have $\bar{J}_{n32} = n^{1/2}b^{d_X/2} \times O_P\left(n^{-1}b^{-d}\log n + b^{2(p+1)}\right) = o_P(1)$. By Lemma 10.1 and Assumption C.6, we have

$$\bar{J}_{n31} = n^{-1/2}b^{d_X/2}\sum_{i=1}^{n}\delta_n(X_i, Y_i)\int \frac{\overline{\mathbf{S}}_{p,b}(X_i,z)^{-1}\mathbf{V}_{p,b}(Y_i;X_i,z)}{g\left(G^{-1}(\tau_{iz}|x^*,z)\,|x^*,z\right)}d\Delta(z)\,\pi_i + o_P(1).$$

Writing the dominant term in the last expression as a second order U-statistic plus a smaller order term ($O_P(n^{-1/2}b^{-d_X/2})$), it is easy to show that this dominant term is $O_P(b^{d_X/2}+n^{-1/2}b^{-d_X/2}) = o_P(1)$ by Chebyshev. Thus, $\overline{J}_{n3} = o_P(1)$ under $\mathbb{H}_1(\gamma_n)$. Using (10.12), we decompose $\bar{J}_{n2}$ as follows

$$\begin{aligned}\bar{J}_{n2} &= b^{d_X}\sum_{i=1}^{n}\left[\int\frac{\hat{\tau}_{iz}-\tau_{iz}}{g\left(G^{-1}(\tau_{iz}|x^*,z)\,|x^*,z\right)}d\Delta(z)\right]^2\pi_i + b^{d_X}\sum_{i=1}^{n}\left[\int\hat{r}_i(z)\,d\Delta(z)\right]^2\pi_i\\ &\quad +2b^{d_X}\sum_{i=1}^{n}\int\frac{\hat{\tau}_{iz}-\tau_{iz}}{g\left(G^{-1}(\tau_{iz}|x^*,z)\,|x^*,z\right)}d\Delta(z)\int\hat{r}_i(z)\,d\Delta(z)\,\pi_i\\ &\equiv \bar{J}_{n21}+\bar{J}_{n22}+2\bar{J}_{n23},\text{ say.}\end{aligned}$$

By Lemma 10.1 and Assumption C.11, we can readily show that

$$\bar{J}_{n21} = b^{d_X}\sum_{i=1}^{n}\left[\int\frac{\overline{\mathbf{S}}_{p,b}(X_i,z)^{-1}\mathbf{V}_{p,b}(Y_i;X_i,z)}{g\left(G^{-1}(\tau_{iz}|x^*,z)\,|x^*,z\right)}d\Delta(z)\right]^2\pi_i + o_P(1) = O_P(1),$$

and $\bar{J}_{n22} = nb^{d_X}O_P(n^{-2}b^{-2d}(\log n)^2 + b^{4(p+1)}) = O_P(n^{-1}b^{-3d/2}(\log n)^2 + nh^{4(p+1)+d_X}) = o_P(1)$. Then $\bar{J}_{n23} = o_P(1)$ by the Cauchy-Schwarz inequality. Consequently, (10.15) follows.

By Proposition 10.5 below, (10.16) holds. With (10.16), it is standard to show that $\hat{J}_{n1} = O_P(1)$. Using (10.20) we can decompose $\hat{J}_{n4}$ as

$$\begin{aligned}\hat{J}_{n4} &= b^{d_X}\sum_{i=1}^{n}\int\alpha_{n1}(\hat{\tau}_{iz};x^*,z)\,d\Delta(z)\int G^{-1}(\tau_{iz}|x^*,z)d\Delta(z)\,\pi_i\\ &\quad +b^{d_X}\sum_{i=1}^{n}\int\alpha_{n1}(\hat{\tau}_{iz};x^*,z)\,d\Delta(z)\int\frac{\hat{\tau}_{iz}-\tau_{iz}}{g\left(G^{-1}(\tau_{iz}|x^*,z)\,|x^*,z\right)}d\Delta(z)\,\pi_i\\ &\quad +b^{d_X}\sum_{i=1}^{n}\int\alpha_{n1}(\hat{\tau}_{iz};x^*,z)\,d\Delta(z)\int\hat{r}_i(z)\,d\Delta(z)\,\pi_i \equiv \sum_{s=1}^{3}\hat{J}_{n4s},\text{ say.}\end{aligned}$$

Analogously to the case of $\bar{J}_{n31}$, we can readily show that $\hat{J}_{n41} = 0$. $\hat{J}_{n43}$ is of smaller order. For $\hat{J}_{n42}$, we can apply Lemmas 10.1 and 10.2 to obtain $\hat{J}_{n42} = \tilde{J}_{n4} + o_P(1)$, where $\tilde{J}_{n4}$ is defined in (10.19).[12] Thus (10.17) follows.

We now show (10.18). By Lemma 10.2(a), $\hat{J}_{n2} = nb^{d_X}[O_P(\nu_b^4) + o_P\left(b^{2(p+1)} + n^{-1}b^{-d_X}\right)] = o_P(1)$. By the fact that $\hat{J}_{n1} = O_P(1)$ and the Cauchy-Schwarz inequality, $\hat{J}_{n3} = o_P(1)$. For $\hat{J}_{n5}$,

[12] Using the expressions for $V_{p,b}$ and $\mathbf{V}_{p,b}$, we can write $\tilde{J}_{n42}$ as a third order U-statistic. By straightforward moment conditions, we can verify that $E[(\tilde{J}_{n42})^2] = o(1)$. Despite the asymptotic negligibility of $\tilde{J}_{n42}$, we keep it in our asymptotic analysis, as it will simplify notation in other places.

we have $\hat{J}_{n5} = nb^{d_X}[O_P(\nu_b^2) + o_P\left(b^{p+1} + n^{-1/2}b^{-d_X/2}\right)]\ O_P(n^{-1/2}b^{-d/2}\ \sqrt{\log n} + b^{p+1}) = o_P(1)$. Consequently, (10.13) follows. To show (10.14), let

$$\begin{aligned}\eta_{1k}(\tau;x,z) &\equiv e_1'\bar{\mathbf{S}}_{p,b}(x,z)\,\mu_{p,b}(X_k-x,Z_k-z)\,K_b(X_k-x,Z_k-z)/g\left(G^{-1}(\tau|x^*,z)\,|x^*,z\right),\\ \eta_{2k}(\tau;x,z) &\equiv e_1'\bar{S}_{p,b}(\tau;x,z)\,\mu_{p,b}(X_k-x,Z_k-z)\,K_b(X_k-x,Z_k-z),\end{aligned}$$

and $\zeta_0(W_i,W_k;z) \equiv \eta_{1k}(\tau_{iz};X_i,z)\,\bar{\mathbf{1}}_{Y_i}(W_k) + \eta_{2k}(\tau_{iz};x^*,z)\,\psi_{\tau_{iz}}\left(Y_k - G^{-1}(\tau_{iz}|U_k)\right)$. Then

$$\frac{e_1'\bar{\mathbf{S}}_{p,b}(X_i,z)^{-1}\,\mathbf{V}_{p,b}(Y_i;X_i,z)}{g\left(G^{-1}(\tau_{iz}|x^*,z)\,|x^*,z\right)} + e_1'\bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1}\,V_{p,b}(\tau_{iz};x^*,z) = \frac{1}{n}\sum_{k=1}^{n}\zeta_0(W_i,W_k;z).$$

It follows that

$$J_n = b^{d_X}\sum_{i=1}^{n}\left[\int n^{-1}\sum_{k=1}^{n}\zeta_0(W_i,W_k;z)\,d\Delta(z)\right]^2\pi_i = n^{-2}b^{d_X}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\zeta(W_{i_1},W_{i_2},W_{i_3}),$$

where $\zeta(W_{i_1},W_{i_2},W_{i_3}) \equiv \int\int\zeta_0(W_{i_1},W_{i_2};z)\,\zeta_0(W_{i_1},W_{i_3};\bar{z})\,d\Delta(z)\,d\Delta(\bar{z})\,\pi_{i_1}$. Let $\varphi(w_{i_1},w_{i_2}) \equiv E[\zeta(W_1,w_{i_1},w_{i_2})]$, and $\bar{\zeta}(w_{i_1},w_{i_2},w_{i_3}) \equiv \zeta(w_{i_1},w_{i_2},w_{i_3}) - \varphi(w_{i_2},w_{i_3})$. Then we can decompose J_n as $J_n = J_{n1} + J_{n2}$, where

$$J_{n1} = n^{-1}b^{d_X}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\varphi(W_{i_1},W_{i_2}) \text{ and } J_{n2} = n^{-2}b^{d_X}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3}).$$

Consider J_{n2} first. Write $E\left[J_{n2}^2\right] = n^{-4}b^{2d_X}\sum_{i_1,\ldots,i_6}^{n}E\left[\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3})\,\bar{\zeta}(W_{i_4},W_{i_5},W_{i_6})\right]$. Noting that $E\left[\bar{\zeta}(W_{i_1},w_{i_2},w_{i_3})\right] = E\left[\bar{\zeta}(w_{i_1},W_{i_2},w_{i_3})\right] = E[\bar{\zeta}(w_{i_1},w_{i_2},W_{i_3})] = 0$, $E[\bar{\zeta}(W_{i_1},W_{i_2},W_{i_3})\bar{\zeta}(W_{i_4},W_{i_5},W_{i_6})] = 0$ if there are more than three distinct elements in $\{i_1,\ldots,i_6\}$. With this, it is easy to show that $E\left[J_{n2}^2\right] = O(n^{-1}b^{-2d_X} + n^{-2}b^{-3d_X} + n^{-3}b^{-4d_X}) = o(1)$. Hence $J_{n2} = o_P(1)$ by the Chebyshev inequality.

For J_{n1}, let $\varphi(W_i,W_j) = \int\int\int\zeta_0(\tilde{w},W_i;z)\,\zeta_0(\tilde{w},W_j;\bar{z})\,\pi(\tilde{x},\tilde{y})\,d\Delta(z)\,d\Delta(\bar{z})\,dG(\tilde{w})$. Then $J_{n1} = n^{-1}b^{d_X}\sum_{i=1}^{n}\varphi(W_i,W_i) + 2n^{-1}b^{d_X}\sum_{1\le i<j\le n}\varphi(W_i,W_j) \equiv \mathbb{B}_{J_n} + \mathbb{V}_{J_n}$, say, where $\mathbb{B}_{J_n}$ and $\mathbb{V}_{J_n}$ contribute to the asymptotic bias and variance of our test statistic, respectively. Note that as $\mathbb{V}_n$ is a second-order degenerate U-statistic, we can easily verify that all the conditions of Theorem 1 of Hall (1984) are satisfied and a central limit theorem applies to it: $\mathbb{V}_{J_n} \stackrel{d_X}{\to} N\left(0,\sigma_J^2\right)$, where $\sigma_J^2 = \lim_{n\to\infty}\sigma_{J_n}^2$ and $\sigma_{J_n}^2 = 2h^{2d_X}E[\varphi(W_1,W_2)]^2$. ■

Proposition 10.5 *$\hat{J}_{n1} = \bar{J}_{n1} + o_P(1)$, where $\bar{J}_{n1} = b^{d_X}\sum_{i=1}^{n}[\int e_1'\bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1}\ V_{p,b}(\tau_{iz};x^*,z)d\Delta(z)]^2\pi_i$.*

Proof. To prove the result, we define $\tilde{J}_{n1}$ analogously as $\bar{J}_{n1}$ with τ_{iz} replaced by $\hat{\tau}_{iz}$: $\tilde{J}_{n1} = b^{d_X}\sum_{i=1}^{n}\left[\int e_1'\bar{S}_{p,b}(\hat{\tau}_{iz};x^*,z)^{-1}V_{p,b}(\hat{\tau}_{iz};x^*,z)\,d\Delta(z)\right]^2\pi_i$. It suffices to show (i) $\hat{J}_{n1} = \tilde{J}_{n1} +$

$o_P(1)$, and (ii) $\tilde{J}_{n1} = \bar{J}_{n1} + o_P(1)$. To prove (i), let $d_{n1}(\tau; u) = e_1' \bar{S}_{p,b}(\tau; x, z)^{-1} [\bar{V}_{p,b}(\tau; x, z) - V_{p,b}(\tau; x, z)]$. Then we have $\hat{J}_{n1} - \tilde{J}_{n1} = D_{n1} + 2D_{n2}$, where

$$\begin{aligned} D_{n1} &\equiv b^{d_X} \sum_{i=1}^{n} \left[\int d_{n1}(\hat{\tau}_{iz}; x^*, z)\, d\Delta(z) \right]^2 \pi_i, \text{ and} \\ D_{n2} &\equiv b^{d_X} \sum_{i=1}^{n} \int d_{n1}(\hat{\tau}_{iz}; x^*, z)\, d\Delta(z) \int e_1' \bar{S}_{p,b}(\hat{\tau}_{iz}; x^*, z)^{-1} V_{p,b}(\hat{\tau}_{iz}; x^*, z)\, d\Delta(z)\, \pi_i. \end{aligned}$$

As $d_{n1}(\tau; u) = O_P\left(b^{p+1}\right)$ uniformly in $(\tau, u) \in \mathcal{T}_0 \times \mathcal{U}_0$, we have $D_{n1} = nb^{d_X} O_P(b^{2(p+1)}) = o_P(1)$. For D_{n2}, we get $D_{n2} = \bar{D}_{n2} + o_P(1)$ using Lemmas 10.1 and 10.3, where $\bar{D}_{n2} = b^{d_X} \sum_{i=1}^{n} \int d_{n1}(\tau_{iz}; x^*, z)\, d\Delta(z) \int e_1' \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz}; x^*, z)\, d\Delta(z)\, \pi_i$. Standard moment calculations give $\bar{D}_{n2} = nb^{d_X} O_P(b^{p+1}) O_P(n^{-1/2}\ b^{-d_X/2}) = o_P(1)$, so (i) holds.

Next, we show (ii). Let $d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) = e_1' \bar{S}_{p,b}(\hat{\tau}_{iz}; x^*, z)^{-1} V_n(\hat{\tau}_{iz}; x^*, z) - e_1' \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_n(\tau_{iz}; x^*, z)$, $\bar{d}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) = e_1' \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} [V_n(\hat{\tau}_{iz}; x^*, z) - V_n(\tau_{iz}; x^*, z)]$, and $\bar{r}_{n2} = d_{n2} - \bar{d}_{n2}$. Then uniformly in $z \in \mathcal{Z}_0$ and conditional on $\hat{\tau}_{iz} \in \mathcal{T}_0^{\epsilon}$,

$$\begin{aligned} \bar{r}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) &= e_1' \left[\bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} - \bar{S}_{p,b}(\hat{\tau}_{iz}; x^*, z)^{-1} \right] V_n(\hat{\tau}_{iz}; x^*, z) \qquad (10.21) \\ &= O_P(\hat{\tau}_{iz} - \tau_{iz})\, O_P(v_b) = O_P\left(v_b(v_b + b^{p+1})\right). \end{aligned}$$

Decompose $\tilde{J}_{n1} - \bar{J}_{n1}$ as

$$\begin{aligned} \tilde{J}_{n11} + 2\tilde{J}_{n12} \equiv{}& b^{d_X} \sum_{i=1}^{n} \left[\int d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)\, d\Delta(z) \right]^2 \pi_i \\ &+ 2b^{d_X} \sum_{i=1}^{n} \int [d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)]\, d\Delta(z) \int e_1' \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz}; x^*, z)\, d\Delta(z)\, \pi_i \end{aligned}$$

Further decompose $\tilde{J}_{n11}$ as$\tilde{J}_{n11} = \tilde{J}_{n11,a} + \tilde{J}_{n11,b} + 2\tilde{J}_{n11,c}$, say, with

$$\begin{aligned} \tilde{J}_{n11,a} + \tilde{J}_{n11,b} + 2\tilde{J}_{n11,c} \equiv{}& b^{d_X} \sum_{i=1}^{n} \left[\int \bar{d}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)\, d\Delta(z) \right]^2 \pi_i \\ &+ b^{d_X} \sum_{i=1}^{n} \left[\int \bar{r}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)\, d\Delta(z) \right]^2 \pi_i \\ &+ 2b^{d_X} \sum_{i=1}^{n} \sum_{j=1}^{n} \int \bar{d}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)\, d\Delta(z) \int \bar{r}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)\, d\Delta(z)\, \pi_i. \end{aligned}$$

Fix $\epsilon > 0$. By the uniform consistency of $\hat{\tau}_{iz}$ for τ_{iz}, there exists $M > 0$ such that $P(\sup_z \max_{1 \le i,j \le n} |\hat{\tau}_{iz} - \tau_{iz}| \ge M\nu_b) < \epsilon/2$ for sufficiently large n. It follows that

$$P\left(\left|\tilde{J}_{n11,a}\right| \ge \vartheta_n \epsilon\right) \le P\left(\left|\tilde{J}_{n11,a}\right| \ge \vartheta_n \epsilon, \sup_z \max_{1 \le i \le n} |\hat{\tau}_{iz} - \tau_{iz}| \le M\nu_b\right) + \epsilon/2,$$

and showing $\tilde{J}_{n11,a} = o_P(1)$ is equivalent to showing that the first term in the last expression is $o(1)$. Conditional on $\sup_{z\in\mathcal{Z}_0}\max_{1\le i\le n}|\hat{\tau}_{iz}-\tau_{iz}| \le M\nu_b$ and $\tau_{iz}\in\mathcal{T}_0\subset(0,1)$, by Lemma 10.3

$$\begin{aligned}\tilde{J}_{n11,a} &= b^{d_X}\sum_{i=1}^{n}\left[\int e_1'\bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1}\left[V_{p,b}(\hat{\tau}_{iz};x^*,z)-V_{p,b}(\tau_{iz};x^*,z)\right]d\Delta(z)\right]^2\pi_i \\ &\le Cnb^{d_X}\sup_{\underline{\tau}\le\tau\le\bar{\tau},\ |\tilde{\tau}-\tau|\le M\nu_n}\ \sup_{u\in\mathcal{U}_0}\|V_{p,b}(\tilde{\tau};u)-V_{p,b}(\tau;u)\|^2 = o_P(1),\end{aligned}$$

By (10.21), $\tilde{J}_{n11,b} = nb^{d/2}O_P((n^{-1}b^{-d}\log n + b^{2(p+1)})\ n^{-1}b^{-d}\log n) = O_P(n^{-1}b^{-3d/2}(\log n)^2 + b^{2(p+1)-d/2}\log n) = o_P(1)$. By Cauchy-Schwarz inequality, $\tilde{J}_{n11,c} = o_P(1)$, so $\tilde{J}_{n11} = o_P(1)$.

Analogously to the determination of the probability order of $\tilde{J}_{n11}$, we can show that $\tilde{J}_{n12} = \tilde{J}_{n12,a} + o_P(1)$, where $\tilde{J}_{n12,a} = b^{d_X}\sum_{i=1}^{n}\int\bar{d}_{n2}(\hat{\tau}_{iz},\tau_{iz};x^*,z)\,d\Delta(z)\int e_1'\bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1}V_{p,b}(\tau_{iz};x^*,z)d\Delta(z)\,\pi_i$. By the Cauchy-Schwarz inequality and Lemmas 10.1 and 10.3, $\tilde{J}_{n12,a} = o_P(1)$. Thus $\tilde{J}_{n12} = o_P(1)$. This completes the proof of (ii). ■

Proof of Theorem 6.3 The proof is much simpler than that of Theorems 6.1-6.2, so we only sketch the main steps. Under $\mathbb{H}_1$, we can apply Lemmas 10.2 and 10.1 in turn to obtain

$$\begin{aligned}n^{-1}b^{-d_X}\hat{J}_n &= n^{-1}\sum_{i=1}^{n}\left\{\int G^{-1}\left(\hat{G}_{p,b}(Y_i|X_i,z)\,|x^*,z\right)d\Delta(z)\right\}^2\pi_i + o_P(1) \\ &= n^{-1}\sum_{i=1}^{n}\left\{\int G^{-1}\left(G(Y_i|X_i,z)\,|x^*,z\right)d\Delta(z)\right\}^2\pi_i + o_P(1).\end{aligned}$$

The dominant term in the last equality tends to $\mu_A > 0$ in probability; the result follows. ■

The power of the test: discussion To gain insight into the power of the test, we examine the conditions under which $\mu_A \equiv E\{[\int G^{-1}(G(Y\mid X,z)\mid x^*,z))d\Delta(z)]^2\pi(X,Y)\} = 0$ or $\mu_A > 0$. Let $w\equiv(x',y')'$, and write

$$f(w,z)\equiv 1\{\pi(x,y)>0\}\ G^{-1}(G(y\mid x,z)\mid x^*,z)).$$

First, we note that if $f(w,z)$ does not depend on z, then $\mu_A = 0$. This can happen either because the "implicit null" holds, i.e., Theorem 3.3(iii) holds with singleton sets in Theorem 3.3(ii) (e.g., strict monotonicity); or because the implicit null fails, but π puts zero weight on the values of w that could reveal this failure. Although this latter case is possible, we consider this part of the price to be paid for constructing a tractable test.

Thus, suppose that $f(w,z)$ depends on both z and on $w\equiv(x',y')'$ (it must depend at least on y). We now show that if $f(w,z)$ has a unique Fourier series representation, then $\mu_A>0$, provided that H_1 and H_2 have no common harmonic moments, as defined below.

By Bartle (1966, lemma 4.10), we have $\mu_A > 0$ if and only if $[\int G^{-1}(G(Y\mid X,z)\mid x^*,z))d\Delta(z)]^2\pi(X,Y) > 0$ $a.s.$ As $\pi(x,y)\ge 0$, it is necessary and sufficient that

$$\int f(w,z)\,d\Delta(z) = 1\{\pi(x,y)>0\}\int G^{-1}(G(y\mid x,z)\mid x^*,z))\,d\Delta(z)\neq 0\quad a.e. \tag{10.22}$$

To ensure (10.22), it is equivalent that $\int f(w,z)\,dH_1(z) \neq \int f(w,z)\,dH_2(z)$ *a.e.*

Let $f(w,z)$ have the Fourier representation

$$f(w,z) = a_0 + \sum_{j=1}^{\infty} a_j \cos(w'\beta_j + z'\gamma_j) + b_j \sin(w'\beta_j + z'\gamma_j),$$

where $\{(a_j, b_j)\}$ are Fourier coefficients and $\{(\beta_j, \gamma_j)\}$ are the Fourier (multi-)frequencies. Recall the trigonometric identities $\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ and $\sin(\alpha+\beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)$. Applying these for $j = 1, 2, ...$, we have

$$\begin{aligned}
\cos(w'\beta_j + z'\gamma_j) &= \cos(w'\beta_j)\cos(z'\gamma_j) - \sin(w'\beta_j)\sin(z'\gamma_j) \\
\sin(w'\beta_j + z'\gamma_j) &= \sin(w'\beta_j)\cos(z'\gamma_j) + \cos(w'\beta_j)\sin(z'\gamma_j),
\end{aligned}$$

so that

$$\begin{aligned}
f(w,z) &= a_0 + \sum_{j=1}^{\infty} a_j[\cos(w'\beta_j)\cos(z'\gamma_j) - \sin(w'\beta_j)\sin(z'\gamma_j)] \\
&\quad + b_j[\sin(w'\beta_j)\cos(z'\gamma_j) + \cos(w'\beta_j)\sin(z'\gamma_j)] \\
&= a_0 + \sum_{j=1}^{\infty} (a_j \cos((z'\gamma_j) + b_j \sin(z'\gamma_j))\cos(w'\beta_j) \\
&\quad + (b_j \cos(z'\gamma_j) - a_j \sin(z'\gamma_j))\sin(w'\beta_j).
\end{aligned}$$

For $i = 1, 2$, integrating gives

$$\begin{aligned}
\int f(w,z) dH_i(z) &= a_0 + \sum_{j=1}^{\infty} [\int (a_j \cos(z'\gamma_j) + b_j \sin(z'\gamma_j)) dH_i(z)] \cos(w'\beta_j) \\
&\quad + \sum_{j=1}^{\infty} [\int (b_j \cos(z'\gamma_j) - a_j \sin(z'\gamma_j)) dH_i(z)] \sin(w'\beta_j) \\
&= a_0 + \sum_{j=1}^{\infty} \tilde{a}_{i,j} \cos(w'\beta_j) + \tilde{b}_{i,j} \sin(w'\beta_j),
\end{aligned}$$

where $\tilde{a}_{i,j} \equiv \int (a_j \cos(z'\gamma_j) + b_j \sin(z'\gamma_j)) dH_i(z)$ and $\tilde{b}_{i,j} \equiv \int (b_j \cos(z'\gamma_j) - a_j \sin(z'\gamma_j)) dH_i(z)$. By the uniqueness of the Fourier representation, it follows that $\mu_A > 0$ if and only if for some j we have $\tilde{a}_{1,j} \neq \tilde{a}_{2,j}$ or $\tilde{b}_{1.j} \neq \tilde{b}_{2,j}$. When $\tilde{a}_{1,j} = \tilde{a}_{2,j}$ and $\tilde{b}_{1.j} = \tilde{b}_{2,j}$, we have

$$\begin{aligned}
\int (a_j \cos(z'\gamma_j) + b_j \sin(z'\gamma_j)) dH_1(z) &= \int (a_j \cos(z'\gamma_j) + b_j \sin(z'\gamma_j)) dH_2(z) \\
\int (b_j \cos(z'\gamma_j) - a_j \sin(z'\gamma_j)) dH_1(z) &= \int (b_j \cos(z'\gamma_j) - a_j \sin(z'\gamma_j)) dH_2(z), \qquad \text{or}
\end{aligned}$$

$$\begin{bmatrix} a_j & b_j \\ b_j & -a_j \end{bmatrix} \begin{bmatrix} \mu_{11j} \\ \mu_{12j} \end{bmatrix} = \begin{bmatrix} a_j & b_j \\ b_j & -a_j \end{bmatrix} \begin{bmatrix} \mu_{21j} \\ \mu_{22j} \end{bmatrix},$$

where $\mu_{11j} \equiv \int \cos(z'\gamma_j))dH_1(z)$, $\mu_{12j} \equiv \int \sin(z'\gamma_j))dH_1(z)$, $\mu_{21j} \equiv \int \cos(z'\gamma_j))dH_2(z)$, and $\mu_{22j} \equiv \int \sin(z'\gamma_j))dH_2(z)$, $j = 1, 2, ...$, are the "harmonic moments" of H_1 and H_2, respectively.

Because $f(w, z)$ depends on both w and z, it follows that for some j, $a_j \neq 0$ or $b_j \neq 0$, ensuring that $a_j^2 + b_j^2 \neq 0$, so that the matrix inverse exists. Thus, for $\mu_A > 0$, it suffices that for this j, $\mu_{11j} \neq \mu_{21j}$ or $\mu_{12j} \neq \mu_{22j}$. For this it suffices that for *all* j, $\mu_{11j} \neq \mu_{21j}$ and $\mu_{12j} \neq \mu_{22j}$; that is, H_1 and H_2 have no common harmonic moments. If H_1 and H_2 differ but do have some common harmonic moments, then it is possible that $\mu_A = 0$, since it could happen that whenever $a_j \neq 0$ or $b_j \neq 0$ we have $\mu_{11j} = \mu_{21j}$ and $\mu_{12j} = \mu_{22j}$. Nevertheless, this is clearly a quite special case.

Use of the Fourier series representation here is convenient, but not necessary. Under suitable regularity conditions, one could use other series representations, such as polynomials, or transforms, such as the Fourier or Laplace transforms, and obtain similar results.

Proof sketch for (6.4). Let $\hat{G}_{p,b}^*$, $\hat{G}_{p,b}^{-1*}$ and $\hat{J}_n^*$ be defined as $\hat{G}_{p,b}$, $\hat{G}_{p,b}^{-1}$ and $\hat{J}_n$, with $\mathcal{W}_n^*$ replacing $\mathcal{W}_n$. Define $\mathcal{S}(C) \equiv \{\sup_{(y,u)\in\mathbb{R}\times\mathcal{U}_0} |\hat{G}_{p,b}(y|u) - G(y|u)| \leq Cn^{-1/2}b^{-d/2}(\log n)^{1/2} + b^{p+1}$, $\sup_{(\tau,u)\in\mathcal{T}\times\mathcal{U}_0} |\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u)| \leq Cn^{-1/2}b^{-d/2}(\log n)^{1/2} + b^{p+1}\}$, where $\mathcal{T} = [\epsilon_0, 1-\epsilon_0]$ for some small $\epsilon_0 \in (0, 1/2)$. Then by Lemmas 10.1 and 10.2, for any $\epsilon > 0$, there exists a sufficiently large constant C such that $P(\mathcal{S}^c(C)) \leq \epsilon$ for sufficiently large n where $\mathcal{S}^c(C)$ is the complement of $\mathcal{S}(C)$. Noting that

$$P(T_n^* \leq t|\mathcal{W}_n) = P(T_n^* \leq t|\mathcal{W}_n \cap \mathcal{S}(C))\, P(\mathcal{S}(C)) + P(T_n^* \leq t|\mathcal{W}_n \cap \mathcal{S}^c(C))\, P(\mathcal{S}^c(C))$$

and that the second term in the above expression can be made arbitrarily small for sufficiently large n, it suffices to prove (i) by showing that $P(T_n^* \leq t|\mathcal{W}_n \cap \mathcal{S}(C)) \to \Phi(t)$ for all $t \in \mathbb{R}$. Conditional on $\mathcal{W}_n \cap \mathcal{S}(C)$, $\hat{A}_i$ is well defined, and one can follow the proof of Theorem 6.2 and that of Theorem 4.1 in Su and White (2008) to show that

$$\begin{aligned} \hat{J}_n^* &= n^{-1}b^{d_X}\sum_{i=1}^{n}\varphi^*(W_i^*, W_i^*) + 2n^{-1}b^{d_X}\sum_{1\leq i<j\leq n}\varphi^*\left(W_i^*, W_j^*\right) + o_{P^*}(1) \\ &\equiv \mathbb{B}_{J_n^*} + \mathbb{V}_{J_n^*} + o_{P^*}(1), \end{aligned}$$

where P^* denotes probability conditional on $\mathcal{W}_n \cap \mathcal{S}(C)$, and φ^* is defined analogously to φ with E replaced by E^*, the expectation with respect to P^*. Noting that $\mathbb{V}_{J_n^*}$ is a second-order U-statistic based on the triangular process $\{W_i^*\}$ and that the W^*'s are IID conditional on $\mathcal{W}_n$, one can continue to apply the CLT of Hall (1984) to $\mathbb{V}_{J_n^*}$ to demonstrate that it is asymptotically $N\left(0, \sigma_J^{*2}\right)$ conditional on $\mathcal{W}_n$, where $\sigma_J^{*2} \equiv 2\text{plim}_{n\to\infty} E^*[\varphi^*(W_1^*, W_2^*)^2]$. The asymptotic bias and variance terms can be estimated analogously as $\hat{\mathbb{B}}_{J_n}$ and $\hat{\sigma}_{J_n}^2$ in the paper. The asymptotic normality of T_n^* conditional on $\mathcal{W}_n \cap \mathcal{S}(C)$ then follows.

For (ii), let $\bar{z}_\alpha^*$ denote the $1-\alpha$ conditional quantile of T_n^* given $\mathcal{W}_n$, i.e., $P(T_n^* \geq \bar{z}_\alpha^*|\mathcal{W}_n) = \alpha$. By choosing B sufficiently large, the approximation error of z_α^* to $\bar{z}_\alpha^*$ can be made arbitrarily small and negligible. By (i), $\bar{z}_\alpha^* \to z_\alpha$ in probability where z_α is the $1-\alpha$ quantile of the standard normal distribution. Then, in view of Theorem 6.1, T_n diverges to ∞ at the rate nb^{d_X}, implying that $\lim_{n\to\infty} P\left(T_n \geq z_\alpha^*\right) = \lim_{n\to\infty} P\left(T_n \geq z_\alpha\right) = 1$ under $\mathbb{H}_1$. ■

REFERENCES

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

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

BLUNDELL, R., X. CHEN, and D. KRISTENSEN (2007): "Semi-Nonparametric IV Estimation of Shape Invariant Engel Curves," *Econometrica,* 75, 1613-1669.

BLUNDELL, R., J. HOROWITZ, and PAREY (2009): "Measuring the Price Responsiveness of Gasoline Demand," CeMMAP Working Paper.

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

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

CARNEIRO, P., J. HECKMAN, and D. MASTEROV (2005): "Labor Market Discrimination and Racial Differences in Premarket Factors," *Journal of Law and Economics*, 48, 1-39.

CHALAK, K., and H. WHITE (2011a): "An Extended Class of Instrumental Variables for the Estimation of Causal Effects," *Canadian Journal of Economics*, 44, 1-51.

CHALAK, K., and H. WHITE (2011b): "Causality, Conditional Independence, and Graphical Separation in Settable Systems," *Discussion Paper*, Dept. of Economics, UCSD.

CHO, J.-S. and H. WHITE (2011): "Testing Correct Model Specification Using Extreme Learning Machines," *Neurocomputing*, 74, 2552–2565.

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

GIACOMINI, R., D. N. POLITIS, and H. WHITE (2007): "A Warp-Speed Method for Conducting Monte Carlo Experiments Involving Bootstrap Estimators," *Discussion Paper*, Dept. of Economics, UCSD.

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

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

HECKMAN, J. J. and E. VYTLACIL (1999) "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Science*, USA, Vol. 96, pp. 4730-4734.

HECKMAN, J. J. and E. VYTLACIL (2005) "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 73, 669-738.

HODERLEIN, S. (2011): "How Many Consumers are Rational?" *Journal of Econometrics*, forthcoming.

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

HODERLEIN, S. and E. MAMMEN (2009): "Identification and Estimation of Local Average Derivatives in Non-separable Models without Monotonicity," *Econometrics Journal*, 12, 1-25.

HUANG, M. and H. WHITE (2010): "A Flexible Test for Conditional Independence," *Discussion Paper*, Dept. of Economics, UCSD.

HURWICZ, L. (1950): "Generalization of the Concept of Identification," in T. C. Koopmans (eds.), *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph 10. New York: Wiley.

IMBENS, G. W. and J. D. ANGRIST (1994) "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62, 467-475.

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

KING, M., X. ZHANG, and M. AKRAM (2011): "A New Procedure for Multiple Testing of Econometric Models," *Working Paper* 07/11, Monash University .

LEWBEL, A. (1999): "Consumer Demand Systems and Household Expenditure," in H. Pesaran and M. Wickens (Eds.), *Handbook of Applied Econometrics*. Oxford: Blackwell.

LU, X. and H. WHITE (2011): "Testing for Separability in Structural Equations," *Discussion Paper*, Dept. of Economics, UCSD.

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

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

NEAL, D. and W. JOHNSON (1996): "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104, 869-895.

OLLEY G. S. , and A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297.

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: Wiley.

SU, L., HODERLEIN, S., and H. WHITE (2011): "Testing Monotonicity in Unobservables with Panel Data," *Discussion Paper*, Dept. of Economics, UCSD.

SU, L., and H. WHITE (2008): "A Nonparametric Hellinger Metric Test for Conditional Independence," *Econometric Theory,* 24, 829-864.

SU, L., and H. WHITE (2011): "Local Polynomial Quantile Regression for Dependent Heterogeneous Processes: Uniform Bahadur Representation with Applications to Testing Conditional Independence," *Discussion Paper*, Dept. of Economics, UCSD.

WHITE, H. and K. CHALAK (2010): "Testing a Conditional Form of Exogeneity," *Economics Letters*, 109, 88-90.

WHITE, H., and X. LU (2011): "Causal Diagrams for Treatment Effect Estimation with Application to Efficient Covariate Selection," *Review of Economics & Statistics,* forthcoming.