Testing for Monotonicity in Unobservables under Unconfoundedness*

Stefan Hoderlein[a], Liangjun Su[b], Halbert White[c], and Thomas Tao Yang[a]
[a] Department of Economics, Boston College
[b] School of Economics, Singapore Management University
[c] Department of Economics, University of California, San Diego

October 23, 2015

Abstract

Monotonicity in a scalar unobservable is a common assumption when modeling heterogeneity in structural models. 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 in some economic applications unlikely to hold, e.g., random coefficient models. Its failure can have substantive adverse consequences, in particular inconsistency of any estimator that is based on it. Having a test for this hypothesis is hence desirable. This paper provides such a test for cross-section data. We show how to exploit an exclusion restriction together with a conditional independence assumption, which in the binary treatment literature is commonly called unconfoundedness, to construct a test. Our statistic is asymptotically normal under local alternatives and consistent against global alternatives. 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.

JEL Classification: C12, C14, C21, C26

Keywords: Control variables, Conditional exogeneity, Endogenous variables, Monotonicity, Nonparametrics, Nonseparable, Specification test, Unobserved heterogeneity

1 Introduction

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

*The authors gratefully thank the co-editor Jianqing Fan, an associate editor, and two anonymous referees for their many constructive comments on the previous version of the paper. They also thank Steve Berry, Jim Heckman, Joel Horowitz, Lars Nesheim, and Maxwell Stinchcombe for helpful comments and discussion, and are indebted to Richard Blundell, Whitney Newey, and Derek Neal for providing us with their data and to Xun Lu for superlative computational assistance. The second author gratefully acknowledges the Singapore Ministry of Education for Academic Research Fund under grant number MOE2012-T2-2-021.

of such important structural features as average marginal effects or various average effects of treatment. Seminal examples are the local average treatment effects (LATE) of Imbens and Angrist (1994), the marginal treatment effects (MTE) of Heckman and Vytlacil (1999, 2005), or the control function model of Imbens and Newey (2009, IN hereafter), to name just 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 to be necessary to identify and estimate 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). It has received a lot of attention recently; see Altonji and Matzkin (2005, AM hereafter), IN, Torgovitsky (2011), and d'Haultfoeuille and Février (2012), among others.

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 affects wages 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, where bids are monotonic functions of a scalar unobserved private valuation.

Given the wide use of monotonicity in economics and econometrics, a test for monotonicity seems desirable, not least because it has been repeatedly criticized; see, e.g., Hoderlein and Mammen (2007) or Kasy (2011). Alternatives have been suggested in the case of triangular systems (Hoderlein et al. 2014), and in the treatment effects setup (Huber and Mellace 2014). Nevertheless, to the best of our knowledge, generally applicable specification tests for monotonicity in unobservables are lacking in econometrics and statistics despite the enormous literature on nonparametric specification tests. Most closely related are specification tests in the treatment effect framework, see in particular Kitagawa (2013), but these are for a binary endogenous variable. Less closely related are tests for monotonicity in observable determinants; see, e.g., Birke and Dette (2007) and Delgado and Escanciano (2012). These latter tests are very different in structure, and generally compare a monotonized estimator with an unrestricted one. In addition, there are also tests on the structure in unobservables: Hoderlein and Mammen (2009), Lu and White (2014) and Su et al. (2015) propose convenient nonparametric tests for structural separability, but they cannot handle monotonicity. Su, Hoderlein, and White (2014, SHW hereafter) do provide a test for scalar monotonicity under a strict exogeneity assumption for large dimensional panel data models, which allows for several structural errors, but its applicability is limited by the panel data requirement.[1] Thus, our main goal and contribution here is to provide a new generally applicable test designed specifically to detect the failure of scalar monotonicity in a scalar unobservable in cross section data.

Under the null hypothesis of monotonicity of a structural function in a scalar unobservable and a con-

[1] See also Ghanem (2014) for related work on identification of these models.

ditional exogeneity assumption, we derive a testable implication that is used to construct our test statistic. We derive the asymptotic distribution of our test statistic under a sequence of Pitman local alternatives and prove its global consistency. Simulations indicate that the empirical level of our test behaves reasonably well and it has good power against non-monotonicity. The conditional exogeneity assumption holds in many important examples, including control function treatments of exogeneity, unconfoundedness assumptions as in the treatment effects literature, and generalizations of the classical proxy assumption. Note that our test does not rely on the assumption of unconditional exogeneity, and hence also works in a situation where regressors are endogenous, as long as instruments are available.

To illustrate our test, we apply our 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), who include unobserved ability, A, as scalar monotonic factor, 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 to reject monotonicity is illustrated by an analysis of Engel curves, where a scalar monotone unobservable is implausible (see 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, motivate our testing approach, and discuss identification under monotonicity. Based on these results, we discuss the heuristics for our test in Section 3, turning to the formal asymptotics of our estimators and tests in Sections 4 and 5. A Monte Carlo study is given in Section 6, and in Section 7 we present our two applications. Section 8 concludes. The proofs of all results are relegated to the appendix. Further technical details are contained in supplementary material which can be found online.

2 Scalar monotonicity and test motivation

The appeal of monotonicity stems at least in part from the fact that it permits 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 function from the regression quantiles. When we talk about structural models, we mean that there are random vectors Y, X and Z, and scalar random variable A, with supports $\mathcal{Y}$, $\mathcal{X}$, $\mathcal{Z}$, and $\mathcal{A}$, and only the former three being directly observable, which admit a structural relationship in the sense that there exists a measurable function $m : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$ such that Y is structurally determined as

$$Y = m(X, A).$$

Note that 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. As in SHW, we are interested in testing the following null hypothesis

$$\mathbb{H}_0 : m(x, \cdot) \text{ is strictly monotone for each } x \in \mathcal{X}. \tag{2.1}$$

Without loss of generality, we further restrict our attention to the case where $m(x, \cdot)$ is strictly increasing for each $x \in \mathcal{X}$ under the null; otherwise, one can always consider $-m(x, \cdot)$ if $m(x, \cdot)$ is strictly decreasing.

As SHW note, Y always has a quantile representation given X. If X is independent of A, and m is monotone in A, it allows the recovery of m. Specifically, let $G(\cdot|x)$ and $G^{-1}(\tau|x)$ denotes the conditional cumulative distribution function (CDF) and conditional τ-th quantile of Y given $X = x$, respectively. Then, the strict monotonicity of $m(x, \cdot)$, combined with full independence of A and X (strict exogeneity of X) and a normalization, allows the recovery of m as $m(x, a) = G^{-1}(a|x)$ for all (a, x).

Apparently, scalar monotonicity for a structural function 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 individuals must be preserved under interventions to x. For example, under independence, if individual j attains the conditional median food consumption $G^{-1}(0.5|x_j)$, then he would remain at the conditional median for all other values of x.

The generic existence of the regression quantile representation, however, makes it impossible to test for monotonicity without further information. One source of such information is that provided by panel data, as exploited by SHW, who use the discrete variation provided by time and consider several unobservables in the structural model. Here, we follow a different strategy, using additional cross-section information. In particular, we assume there is random vector Z that is excluded from the structural function, and conditional on which X is independent of A (for this, we use the shorthand $X \perp A \mid Z$). There are many models in economics that admit such a representation, as we list in the following:

First, conditional independence is a very common notion in the treatment effect literature, and is often called "unconfoundedness" (e.g., Wooldridge, 2010), or "conditionally exogeneity" (e.g., AM) with respect to A given Z. For example, X could be randomly assigned given Z, i.e., the intensity of a treatment, say an income supplement or a dose of a treatment, is randomly assigned, conditional on having certain socioeconomic characteristics Z, but unconditionally there may exist correlation between X and the unobservables, because certain socioeconomic groups may be treated preferentially.

Another important classical example is when Z is a valid proxy for A. In this case, we have no omitted variable bias, because controlling for Z removes the correlation - whatever is left over of A is not correlated. This extends the classical notion of proxy to nonseparable models.

But this type of structure does not just arise in exogenous settings. A similar conditional exogeneity assumption is also the key to the control function approach. In particular, Z could be the unobservable in a first-stage equation that relates X to an instrument ξ. For instance, the first stage could be

$$X = \Psi(\xi, Z),$$

where $\xi \perp (A, Z)$ and Ψ is strictly monotonic in Z, with Z being 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 now state the assumptions we impose on the structural model introduced above in a formal sense. The structural model satisfies:

Assumption A.1 *For all $x \in \mathcal{X}$, $m(x, \cdot)$ is strictly increasing in its last argument.*

As mentioned above, we also require additional (excluded) variable Z, such that:

Assumption A.2 $X \perp A \mid Z$, where Z is not measurable with respect to the sigma-field generated by X.

By requiring Z not to be solely a function of X, we permit important flexibility for recovering objects of interest. See, for example, Hoderlein and Mammen (2007, 2009) and IN. White and Lu (2011) explicitly discuss structures ensuring A.2 where Z is not a function of X.

Finally, we also impose an invertibility condition on the conditional CDF of Y given (X, Z) :

Assumption A.3 Let $G(\cdot|x, z)$ denote the conditional CDF of Y given $(X, Z) = (x, z)$. For each (x, z) in $\mathcal{X} \times \mathcal{Z}$, $G(\cdot|x, z)$ is invertible.

With these assumptions and the normalization condition $a = m(x^*, a)$ $\forall a$ for some reference point x^* (see eq. (2.4) in Matzkin (2003)), A becomes comparable across observations. The main result that we build our test upon is then

Proposition 2.1 *Suppose that Assumptions A.1-A.3 hold. Then with the normalization $a = m(x^*, a)$ $\forall a$,*

$$m(x, a) = G^{-1}(G(a \mid x^*, z) \mid x, z) \quad \forall\ (a, x, z) \in \mathcal{A} \times \mathcal{X} \times \mathcal{Z}, \tag{2.2}$$

$$A = G^{-1}(G(Y \mid X, z) \mid x^*, z) \qquad \forall z \in \mathcal{Z}. \tag{2.3}$$

The main implication of this proposition is that $G^{-1}(G(Y|X, z)|x^*, z)$ has to be invariant with respect to the changes in z. In the following, we will make use of this fact and propose a test statistic for testing the null hypothesis in (2.1).

3 Heuristics of estimation and specification testing

3.1 Estimation through sample counterparts

Proposition 2.1 provides the basis for convenient estimators complementary to those proposed by AM. Because this result ensures that $m(x, a) = G^{-1}(G(a|x^*, z)|x, z)$ for given x^* and any z, one can estimate $m(x, a)$ as $\hat{m}_z(x, a) = \hat{G}^{-1}(\hat{G}(a|x^*, z)|x, z)$ 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 user-chosen distribution function supported on $\mathcal{Z}_0 \subseteq \mathcal{Z}$, e.g., the uniform distribution. 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.

Similarly, one can estimate A as $\hat{A}_z = \hat{G}^{-1}(\hat{G}(Y|X,z)|x^*,z))$ 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

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

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

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

where m_H^* denotes the probability limit of $\hat{m}_H$. Under the correct specification of both Assumptions A1 and A2, $A_H^* = A_H^\dagger = A$.

3.2 Specification testing

Proposition 2.1 motivates constructing specification tests by comparing various estimators of A, as there are multiple consistent estimators of A under correct specification. Under Assumptions A.1-A.3, the failure of these estimators to coincide asymptotically signals non-monotonicity. Below we will study the asymptotic properties of the test statistic

$$\begin{aligned} \hat{J}_n &\equiv b^{d_X} \sum_{i=1}^n (\hat{A}_{1,i} - \hat{A}_{2,i})^2 \pi(X_i, Y_i) \\ &= b^{d_X} \sum_{i=1}^n \left\{ \int \hat{G}^{-1}\left(\hat{G}(Y_i | X_i, z) | x^*, z\right) d\Delta(z) \right\}^2 \pi_i, \end{aligned} \tag{3.1}$$

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); $\pi_i \equiv \pi(X_i, Y_i)$ and $\pi(\cdot, \cdot)$ is a nonnegative weight function with support on a compact subset $\mathcal{X}_0 \times \mathcal{Y}_0$ of $\mathcal{X} \times \mathcal{Y}$. The weight π_i downweights observations for which $\hat{G}(Y_i \mid X_i, z)$ is close to either 0 or 1, so that G^{-1} can not be accurately estimated. For example, one can set $\pi(X_i, Y_i) = \mathbf{1}\{C_{\epsilon_0, X} \leq X_i \leq C_{1-\epsilon_0, X}\} \times$

$\mathbf{1}\{C_{\epsilon_0,Y} \leq Y_i \leq C_{1-\epsilon_0,Y}\}$ in case $d_X = 1$, where, $\mathbf{1}\{\cdot\}$ denotes the usual indicator function, and, e.g., $C_{\epsilon_0,X}$ denotes the ϵ_0th sample quantile of $\{X_i\}_{i=1}^n$. We take $\epsilon_0 = 0.0125$ in our simulations below.

Finally, $\Delta(z) \equiv H_1(z) - H_2(z)$ denotes the contrast between two CDFs. We require that H_1 and H_2 be distinct CDFs 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, and the most important thing is that we want the contrast Δ to be quite different from a zero function. Since they are chosen by the users, it is not restrictive to focus on the case of known H's. Different choices for Δ focus the power of the test in different directions, but as long as they put some positive weight on any fixed interval they will generally exhibit some power against global alternatives. But it is extremely challenging, if possible at all, to consider the optimal choice of the contrast Δ, see Remark 5.5 in Section 5.2. In the Monte Carlo simulations, we experiment with the uniform and re-scaled beta(2,2) distributions for H_1 and H_2 (see Section 6.1 for details), but we also discuss the results with different choices of H_1 and H_2 in footnote 6 in Section 6.2.

As we show below, T_n, a standardized version of $\hat{J}_n$, is asymptotically standard normal under the correct specification of both Assumptions A1 and A2. If T_n is incompatible with this distribution, we have evidence against the correct specification, which suggests the failure of the monotonicity hypothesis under the maintained conditional exogeneity assumption. 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 that violate the null of monotonicity. Despite having a standard normal asymptotic null distribution, T_n requires the use of bootstrap to compute useful critical values or P-values, which is standard practice in econometrics.

In a companion paper, SHW provide a test for scalar monotonicity under the assumption of *strict exogeneity* for large dimensional panel data models. The basic model they consider is

$$Y_{it} = m(X_{it}, A_i) + \varepsilon_{it},\ \ i = 1, ..., N,\ \ t = 1, ..., T,$$

where A_i denotes an individual's unobserved time-invariant attribute, ε_{it} is a time-varying idiosyncratic error term, and they assume that the regressor X_{it} is independent of A_i ($X_{it} \perp A_i$). Let $F_t(\cdot|x)$ denote the conditional CDF of $m_{it} \equiv m(X_{it}, A_i)$ given $X_{it} = x$. Exogeneity ($X_{it} \perp A_i$) and the time-invariance of A_i jointly ensure that F_t is time invariant and can be written as F. Under the assumption that m is strictly increasing in its second argument, we have $m_{it} = F^{-1}(A_i \mid X_{it})$, which yields the following implied null hypothesis

$$\mathbb{H}_0^{\text{SHW1}} : \ F(m_{it} \mid X_{it}) = A_i \ \text{ a.s. for all } t = 1, 2, ... \tag{3.2}$$

The above null hypothesis motivates SHW to consider a test statistic based on the comparison of $F_t(m_{it} \mid X_{it})$ and $F_s(m_{is} \mid X_{is})$ for all $t \neq s$. Nevertheless, a direct comparison is infeasible because m_{it} is not observable. Let $w_\tau(x)$, $\tau = 1, ..., \mathcal{T}$, be non-negative weight functions. Let $\tilde{Y}_{\tau,i} = \tilde{Y}_{\tau,it} \equiv E[Y_{it}w_\tau(X_{it}) \mid A_i]$ under the assumption that $(X_{it}, \varepsilon_{it})$ are identically distributed over t. Let $\tilde{F}_\tau(y) \equiv P[\tilde{Y}_{\tau,i} \leq y]$. Under the maintained exogeneity ($X_{it} \perp A_i$), SHW show that $\tilde{F}_\tau(\tilde{Y}_{\tau,i}) = A_i$ and the monotonicity of m in its second argument implies the following testable null hypothesis

$$\mathbb{H}_0^{\text{SHW2}}: \tilde{F}_\tau(\tilde{Y}_{\tau,i}) = \tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i}) \text{ a.s. for all } (\tau,\varsigma) \text{ with } \tau \neq \varsigma.$$

SHW construct a test statistic based on the squared distance between consistent estimates of $\tilde{F}_\tau(\tilde{Y}_{\tau,i})$ and $\tilde{F}_\varsigma(\tilde{Y}_{\varsigma,i})$ under the assumption that $T/N \to \infty$ as $N \to \infty$. Even though both SHW and we study the test of monotonicity, there are several major differences between the two papers. First, SHW's test only works for panel data and our test works for cross sectional data. Second, SHW assumes strict exogeneity ($X_{it} \perp A_i$) whereas we assume conditional exogeneity (see Assumption A.2 above). Third, the unobservable term A_i can be recovered through certain *unconditional* CDFs in SHW while it can be recovered only through certain *conditional* CDF and its inverse function (see eq. (2.3) in Proposition 2.1). As a result, SHW's test is based on the comparison of different estimates of unconditional CDFs under different weighting scheme whereas our test is based on the estimation of certain conditional CDFs and their inverse functions. Fourth, SHW's test statistic is not asymptotically pivotal and its asymptotic null distribution is a mixture χ^2, while our test statistic is asymptotically standard normal under the null.

Like many other nonparametric tests in the literature (e.g., Chiappori et al. 2015, SHW 2014, Lewbel et al. 2015), we can only test for some implied hypothesis under certain maintained conditions. Once we reject the null, we need to be careful about the interpretation of our test. The rejection may be due to either the failure of the null hypothesis or the violation of the maintained hypothesis.

4 Asymptotics for estimation and inference

4.1 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 (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 $\sum_{0 \le |\boldsymbol{j}| \le p} \equiv \sum_{k=0}^p \sum_{\substack{j_1=0 \\ j_1+\cdots+j_d=k}}^k \cdots \sum_{j_d=0}^k$.

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

$$\min_{\boldsymbol{\beta}} n^{-1} \sum_{i=1}^n \left[\mathbf{1}\{Y_i \le y\} - \sum_{0 \le |\boldsymbol{j}| \le p} \beta_{\boldsymbol{j}}' \left((U_i - u)/b \right)^{\boldsymbol{j}} \right]^2 K_b(U_i - u), \tag{4.1}$$

where $\boldsymbol{\beta}$ stacks the $\beta_{\boldsymbol{j}}$'s ($0 \le |\boldsymbol{j}| \le p$) 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

$$\hat{\boldsymbol{\beta}}(y|u) = [\mathbf{S}_{p,b}(u)]^{-1} n^{-1} \sum_{i=1}^{n} K_b(U_i - u)\, \mu_{p,b}(U_i - u)\ \mathbf{1}\{Y_i \leq y\}, \tag{4.2}$$

$$\mathbf{S}_{p,b}(u) \equiv n^{-1} \sum_{i=1}^{n} K_b(U_i - u)\, \mu_{p,b}(U_i - u)\, \mu_{p,b}(U_i - u)', \tag{4.3}$$

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 $\rho_\tau(u) \equiv u(\tau - 1\{u \leq 0\})$ be the "check" function. We obtain $\hat{G}_{p,b}^{-1}(\tau|u)$ as the minimizing intercept term in the weighted quantile estimation problem

$$\min_{\boldsymbol{\alpha}} n^{-1} \sum_{i=1}^{n} \rho_\tau\left(Y_i - \sum_{0 \leq |\boldsymbol{j}| \leq p} \alpha'_{\boldsymbol{j}}\left((U_i - u)/b\right)^{\boldsymbol{j}}\right) K_b(U_i - u), \tag{4.4}$$

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) and Li and Racine (2008). 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.

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

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 support of Y_i and $\mathcal{Y}_0 \equiv [\underline{y}, \bar{y}]$ for finite real numbers $\underline{y}$, $\bar{y}$. We use the following assumptions.

Assumption C.1 Let $W_i \equiv (X'_i, Y_i, Z'_i)'$, $i = 1, 2, ...$, be independently and identically distributed (IID) random variables that have the same distribution as $(X', Y, Z')'$.

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 $\underline{\tau} \leq \inf_{u \in \mathcal{U}_0} G(\underline{y}|u) \leq \sup_{u \in \mathcal{U}_0} G(\bar{y}|u) \leq \bar{\tau}$ 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: $\forall \epsilon > 0$, $\exists c > 0 : |y - \tilde{y}| < c \Rightarrow \sup_{u \in \mathcal{U}_0} |G(y|u) - G(\tilde{y}|u)| < \epsilon$. 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 c_K, $\bar{c}_1$, and $\bar{c}_2$, either $K(\cdot)$ is compactly supported such that $K(u) = 0$ for $\|u\| > c_K$, and $|\mathbf{K}_{\boldsymbol{j}}(u) - \mathbf{K}_{\boldsymbol{j}}(\tilde{u})| \leq \bar{c}_2 \|u - \tilde{u}\|$ 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, $\|\partial \mathbf{K}_{\boldsymbol{j}}(u) / \partial u\| \leq \bar{c}_1$, and for some $c > 1$, $|\partial \mathbf{K}_{\boldsymbol{j}}(u) / \partial u| \leq \bar{c}_1 \|u\|^{-c}$ for all $\|u\| > c_K$ 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 $nb^{2(p+1)+d_X} \to c_0 \in [0, \infty)$. 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. 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 is imposed by implicitly assuming that Z is continuously distributed, simplifying the analysis. Assumption C.6 appropriately restricts the choices of bandwidth sequence and the order of local polynomial regressions; see the remark after Assumption C.7 below.

To proceed, arrange the $N_{p,l}$ d-tuples as a sequence in lexicographical order, so that $\ell_l(1) \equiv (0, 0, ..., l)$ is the first element and $\ell_l(N_{p,l}) \equiv (l, 0, ..., 0)$ is the last, and let ℓ_l^{-1} be the mapping inverse to ℓ_l. 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

$$\mathbb{S}_p \equiv \begin{bmatrix} \mathbb{M}_{0,0} & \mathbb{M}_{0,1} & \cdots & \mathbb{M}_{0,p} \\ \mathbb{M}_{1,0} & \mathbb{M}_{1,1} & \cdots & \mathbb{M}_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{M}_{p,0} & \mathbb{M}_{p,1} & \cdots & \mathbb{M}_{p,p} \end{bmatrix} \text{ and } \mathbb{B}_p = \begin{bmatrix} \mathbb{M}_{0,p+1} \\ \mathbb{M}_{1,p+1} \\ \vdots \\ \mathbb{M}_{p,p+1} \end{bmatrix}, \tag{4.5}$$

where $\mathbb{M}_{i,j}$ are $N_{p,i} \times N_{p,j}$ matrices whose (l, s) element is $\mu_{\ell_i(l)+\ell_j(s)}$. 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 4.1 *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*

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

where

$$B_m(x, a; x^*) \equiv b^{p+1} e_{1,p}' \mathbb{S}_p^{-1} \mathbb{B}_p \int \left[\frac{\mathbf{G}_{p+1}(a|x^*, z)}{g\left(G^{-1}\left(G(a|x^*, z)|x, z\right)|x, z\right)} + \mathbf{G}_{p+1}^{-1}\left(G(a|x^*, z)|x, z\right) \right] dH(z), \tag{4.6}$$

$$\sigma_m^2(x, a; x^*) \equiv \kappa_{1p} \int \frac{G(a|x^*, z)\left[1 - G(a|x^*, z)\right] h(z)^2}{g\left(G^{-1}\left(G(a|x^*, z)|x, z\right)|x, z\right)^2} \left[\frac{1}{g(x^*, z)} + \frac{1}{g(x, z)} \right] dz, \tag{4.7}$$

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,*

$$\sup_{(x,a)\in\mathcal{X}_0\times\mathcal{A}_H} |\hat{m}_H(x,a) - m_H^*(x,a)| = O_P(n^{-1/2} b^{-d_X/2} \sqrt{\log n} + b^{p+1}). \tag{4.8}$$

Note that the result in Theorem 4.1 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 estimation, 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 estimation.

Define $A_{H,i}^*$ and $A_{H,i}^\dagger$ in the obvious manner. Theorem 4.1 implies the following result for $\hat{A}_{H,i}$.

Corollary 4.2 *Suppose Assumptions C.1-C.6 hold. Then conditional on* $(X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\sqrt{nb^{d_X}}[\hat{A}_{H,i} - A_{H,i}^* - B_m(x^*, Y_i; X_i)] \xrightarrow{d} \mathbb{N}(0, \sigma_m^2(x^*, Y_i; X_i))$. *Further, for i such that* $(X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\hat{A}_{H,i} - A_{H,i}^* = O_P(n^{-1/2} b^{-d_X/2} \sqrt{\log n} + b^{p+1})$ *uniformly in i.*

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

Theorem 4.3 *Suppose Assumptions C.1-C.6 hold. Then for any* $(x,y) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\hat{m}_H^{-1}(x,y) \xrightarrow{P} m_H^{*-1}(x,y)$ *and* $\sqrt{nb^{d_X}}\left(\hat{m}_H^{-1}(x,y) - m_H^{*-1}(x,y) - B_{m^{-1}}(x,y;x^*)\right) \xrightarrow{d} \mathbb{N}\left(0, \sigma_{m^{-1}}^2(x,y;x^*)\right)$, *where* $B_{m^{-1}}(x,y;x^*) \equiv -B_m\left(x, m_H^{*-1}(x,y); x^*\right) / \lambda_H^*\left(x, m_H^{*-1}(x,y)\right)$, $\sigma_{m^{-1}}^2(x,y;x^*) \equiv \sigma_m^2\left(x, m_H^{*-1}(x,y); x^*\right) / \left[\lambda_H^*\left(x, m_H^{*-1}(x,y)\right)\right]^2$, *and* $\lambda_H^*(x,a) \equiv \int \frac{g(a|x^*,z)}{g(G^{-1}(G(a|x^*,z)|x,z)|x,z)} dH(z)$.

When correct specification holds, we can show that $\lambda_H^*(x,a) = \int \frac{g(a|x^*,z)}{g(m(x,a)|x,z)} dH(z) = \partial m(x,a)/\partial a$ by Proposition 2.1, 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 4.3 implies that conditional on $(X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0$, $\sqrt{nb^{d_X}}(\tilde{A}_{H,i} - A_{H,i}^\dagger - B_{m^{-1}}(X_i, Y_i; x^*)) \xrightarrow{d} \mathbb{N}\left(0, \sigma_{m^{-1}}^2(X_i, Y_i; x^*)\right)$.

5 Asymptotics for specification testing

In this section, we study the asymptotic behavior of the test statistic in (3.1).

5.1 Asymptotic null distribution

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

$$\begin{aligned}
S_{p,b}(\tau;u) &\equiv n^{-1}\sum_{i=1}^n K_b(U_i-u)\, g(G^{-1}(\tau|U_i)|U_i)\mu_{p,b}(U_i-u)\,\mu_{p,b}(U_i-u)', \\
\eta_{1k}(\tau;u) &\equiv e_{1,p}'\bar{\mathbf{S}}_{p,b}(u)\,\mu_{p,b}(U_k-u)\,K_b(U_k-u)/g\left(G^{-1}(\tau|x^*,z)\,|x^*,z\right), \\
\eta_{2k}(\tau;u) &\equiv e_{1,p}'\bar{S}_{p,b}(\tau;u)\,\mu_{p,b}(U_k-u)\,K_b(U_k-u), \\
\zeta_0(W_i,W_k;z) &\equiv \eta_{1k}(\tau_{iz};X_i,z)\,\bar{\mathbf{1}}_{Y_i}(W_k) + \eta_{2k}(\tau_{iz};x^*,z)\,\psi_{\tau_{iz}}\left(Y_k - G^{-1}(\tau_{iz}|U_k)\right), \quad \text{and} \\
\varphi(w_i,w_j) &\equiv E\left[\int\int \zeta_0(W_1,w_i;z)\,\zeta_0(W_1,w_j;\bar{z})\,d\Delta(z)\,d\Delta(\bar{z})\,\pi_1\right],
\end{aligned} \tag{5.1}$$

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

$$\mathbb{B}_{J_n} \equiv n^{-2} b^{d_X} \sum_{i=1}^n \sum_{j=1}^n \left[\int \zeta_0(W_i, W_j; z)\, d\Delta(z) \right]^2 \pi_i, \text{ and } \sigma_{J_n}^2 = 2b^{2d_X} E[\varphi(W_1,W_2)^2]. \tag{5.2}$$

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., $p \ge 2$) 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 $2(p+1)+d_X > \max(2d_X, d+2d_Z, 3d/2)$. By choosing $b \propto n^{-1/[2(p+1)+d_X]}$, it suffices to set $p=1$ if $d_X \le 3$ and $d_Z = 1$.[2] $p \ge 2$ would be required as long as $d_Z \ge 2$. 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 5.1 *Suppose Assumptions A.2-A.3, C.1-C.4, C.6, and C.7 hold. Suppose that Assumption C.5 hold with H replaced by H_1 and H_2. Then under Assumption A.1, we have $\hat{J}_n - \mathbb{B}_{J_n} \xrightarrow{d} \mathbb{N}(0, \sigma_J^2)$, where $\sigma_J^2 \equiv \lim_{n\to\infty} \sigma_{J_n}^2$.*

Remark 5.1. 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.

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

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

where $\hat{g}_{iz} \equiv \hat{g}(\hat{G}_{p,b}^{-1}(\hat{\tau}_{zi}|x^*, z)\,|x^*, z)$ with $\hat{g}(y|x^*, z)$ being a consistent estimator of $g(y|x^*, z)$, and $\hat{\mathbf{1}}_{Y_i}(W_k) \equiv \mathbf{1}\{Y_k \le Y_i\} - \hat{G}_{p,b}(Y_i|U_k)$. We propose to estimate $\mathbb{B}_{J_n}$ and $\sigma_{J_n}^2$ respectively by

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

[2] In our two empirical applications, $d_X = 2$ and $d_Z = 1$. So we consider the local linear estimation ($p = 1$).

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

$$T_n \equiv \left(\hat{J}_n - \hat{\mathbb{B}}_{J_n}\right) / \sqrt{\hat{\sigma}^2_{J_n}} \tag{5.3}$$

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

5.2 Asymptotic local power and consistency

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

$$\mathbb{H}_1(c_n) : Y_{ni} = m(X_i, A_i) + c_n \gamma(X_i, A_i), \tag{5.4}$$

where c_n is nonnegative such that $c_n \to 0$ as $n \to \infty$, $m(x, \cdot)$ is strictly increasing for each $x \in \mathcal{X}$, but $m_n(x, \cdot) \equiv m(x, \cdot) + c_n \gamma(x, \cdot)$ is not strictly increasing for x in a nontrivial subset of $\mathcal{X}_0$. Apparently, $\{Y_{ni}, 1 \le i \le n\}$ becomes a triangular array process. Let $G_n(y|x,z)$ and $g_n(y|x,z)$ denote the conditional CDF and PDF of Y_{ni} given $(X_i, Z_i) = (x, z)$, respectively.[3] Let $G_n^{-1}(\cdot|x,z)$ denote the inverse function of $G_n(\cdot|x,z)$. For notational simplicity, we continue to use Y_i to denote $m(X_i, A_i)$ and $G(y|x,z)$ and $g(y|x,z)$ to denote the conditional CDF and PDF of Y_i given $(X_i, Z_i) = (x, z)$. Let $F_{A|Z}(\cdot|z)$ and $f_{A|Z}(\cdot|z)$ denote the conditional CDF and PDF if A_i given $Z_i = z$, respectively.

We add the following assumption.

Assumption A.4 $f_{A|Z}(\cdot|z)$ is a continuous function for each $z \in \mathcal{Z}$; $m(x, \cdot)$ is a continuously differentiable function for each $x \in \mathcal{X}$; $\gamma(x, \cdot)$ is a continuous function for each $x \in \mathcal{X}$.

The following theorem lays down the foundation for the asymptotic local power analysis of our test.

Theorem 5.2 *Suppose that Assumptions A.1-A.4 hold. Then*

$$G_n^{-1}(G_n(y|x,z)|x^*, z) = G^{-1}\left(G(y|x,z)|x^*, z\right) + c_n \Theta_n^\dagger(y; x, z)$$

for all $(y, x, z) \in \mathcal{Y}_0 \times \mathcal{X}_0 \times \mathcal{Z}_0$ *such that* $g\left(G^{-1}\left(G(y|x,z)|x^*, z\right) \ |x^*, z\right) > 0$, *where* $\Theta_n^\dagger(y; x, z)$ *is defined in (B.7).*

Remark 5.2. By Proposition 2.1, $A_i = G^{-1}(G(Y_i \mid X_i, z) \mid x^*, z)$ under the normalization $a = m(x^*, a)\ \forall a$. This fact, in conjunction with Theorem 5.2, indicates

$$\begin{aligned}
& \int G_n^{-1}(G_n(Y_{ni}|X_i,z)|x^*,z))d\Delta(z) \\
= & \int \left[G^{-1}\left(G(Y_{ni}|X_i,z)|x^*,z\right) - G^{-1}\left(G(Y_i|X_i,z)|x^*,z\right)\right] d\Delta(z) + c_n \int \Theta_n^\dagger(Y_{ni}; X_i, z)\, d\Delta(z) \\
= & \int c_n \gamma(X_i, A_i) \int_0^1 \frac{g\left(Y_i + t c_n \gamma(X_i, A_i)|X_i, z\right)}{g\left(G^{-1}\left(G\left(Y_i + t c_n \gamma(X_i, A_i)|X_i, z\right)|x^*, z\right)|x^*, z\right)} dt d\Delta(z) \\
& + c_n \int \Theta_n^\dagger\left(Y_i + c_n \gamma(X_i, A_i); X_i, z\right) d\Delta(z) \\
= & \; c_n \Theta_n(Y_i; X_i, A_i),
\end{aligned} \tag{5.5}$$

[3] With a little bit more complicated notation, one could also allow $\{(X_i, Z_i, A_i)\}$ to be a triangular array.

where the first equality follows from Theorem 5.2 and the fact that $\int d\Delta(z)=0$; the second equality follows from the chain rule, and the fact that $\alpha(y)-\alpha(y_0)=(y-y_0)\int_0^1\alpha'(y_0+t(y-y_0))\,dt$ (by Taylor expansion with an integral remainder, here prime denotes derivative) and $d\alpha^{-1}(y)/dy=1/\left[\alpha'\left(\alpha^{-1}(y)\right)\right]$ for any continuously differentiable function $a(y)$; and $\Theta_n(Y_i;X_i,A_i)\equiv\int[\gamma(X_i,A_i)\int_0^1\frac{g(Y_i+tc_n\gamma(X_i,A_i)|X_i,z)}{g(G^{-1}(G(Y_i+tc_n\gamma(X_i,A_i)|X_i,z)|x^*,z)|x^*,z)}dt$ $+\Theta_n^{\dagger}(Y_i+c_n\gamma(X_i,A_i);X_i,z)]d\Delta(z)$. Ignoring terms of smaller order, we have

$$\Theta_n(Y_i;X_i,A_i)\approx\int\left[\gamma(X_i,A_i)\frac{g(Y_i|X_i,z)}{g(A_i|x^*,z)}+\Theta_n^{\dagger}(Y_i;X_i,z)\right]d\Delta(z)=\int\Theta_n^{\dagger}(Y_i;X_i,z)d\Delta(z). \tag{5.6}$$

To see why the last equality holds, note that under Assumption A.2, $Y_i=m(X_i,A_i)$ has the conditional CDF satisfying $G(y|x,z)=P[m(x,A_i)\leq y|X_i=x,\ Z_i=z]=F_{A|Z}\left(m^{-1}(x,y)|z\right)$. It follows that $g(y|x,z)=f_{A|Z}\left(m^{-1}(x,y)|z\right)\frac{\partial m^{-1}(x,y)}{\partial y}$. Under the normalization rule $m(x^*,a)=a\ \forall a$, we have $m^{-1}(x^*,a)=a\ \forall a$ and

$$\frac{g(Y_i|X_i,z)}{g(A_i|x^*,z)}=\frac{f_{A|Z}\left(m^{-1}(X_i,Y_i)|z\right)\frac{\partial m^{-1}(X_i,Y_i)}{\partial y}}{f_{A|Z}\left(m^{-1}(x^*,A_i)|z\right)}=\frac{\partial m^{-1}(X_i,Y_i)}{\partial y} \tag{5.7}$$

which is free of z. Nevertheless, without knowing the functional form $m_n(\cdot,\cdot)$, it is difficult to derive the explicit formula for $\Theta_n^{\dagger}$ in general.

The following theorem studies the asymptotic local power property of our test.

Theorem 5.3 *Suppose that Assumptions A.1-A.4, C.1, C.4, C.6, and C.7 hold. Suppose that Assumptions C.2 and C.3 hold with $g(y|u)$ and $G(y|u)$ replaced by $g_n(y|u)$ and $G_n(y|u)$ and Assumption C.5 hold with H replaced by H_1 and H_2. Suppose $c_{\Theta_0}\equiv\lim_{n\to\infty}E[\Theta_n(Y_i;\ X_i,A_i)^2\pi(X_i,Y_i)]$ exists. Then under $\mathbb{H}_1(c_n)$ with $c_n=n^{-1/2}b^{-d_X/2}$, $T_n\xrightarrow{d}\mathbb{N}(c_{\Theta_0}/\sigma_J,1)$.*

Remark 5.3. Theorem 5.3 implies that the T_n test has non-trivial power against Pitman local alternatives that converge to the null at rate $n^{-1/2}b^{-d_X/2}$, provided $0<c_{\Theta_0}<\infty$. The asymptotic local power function of the test is given by $1-\Phi(z_\alpha-c_{\Theta_0}/\sigma_J)$, where Φ is the standard normal CDF. Unfortunately, we are unable to characterize primitive conditions under which the limit c_{Θ_0} is strictly positive, ensuring the non-trivial power of our test. In the supplementary Appendix E.2, we consider a simple example where

$$m(X_i,A_i)=\left(1+0.1X_i^2\right)A_i\ \text{ and }\gamma(X_i,A_i)=X_ie^{-A_i},$$

where the PDFs of X_i and A_i have supports $\mathcal{X}=[\underline{c}_x,\bar{c}_x]$ and $\mathcal{A}=(-\infty,\bar{c}_a]$, respectively, where $\underline{c}_x<0$, $\bar{c}_x>0$, and $\bar{c}_a>0$. It is possible that $\underline{c}_x=-\infty$, $\bar{c}_x=\infty$, and $\bar{c}_a=\infty$, in which case we have $\mathcal{X}=\mathcal{A}=\mathbb{R}$. Note that $m(x,\cdot)$ is strictly increasing for all $x\in\mathbb{R}$ and the normalization $m(x^*,a)=a\ \forall a$ is achieved at $x^*=0$. In addition, we assume that for sufficiently large positive number a,

$$P(A_i\leq -a|Z_i=z)=\phi\left(e^{-a},z\right), \tag{5.8}$$

where $\phi(\cdot,\cdot)$ is continuously differentiable with respect to its first argument (as a CDF), a non-constant function of z, and $\lim_{t\downarrow 0}\phi(t,z)\to 0$. Essentially, the condition in (5.8) requires that $F_{A|Z}(a|z)$ decay at the exponential rate as $a\to-\infty$. In this case we can show that

$$\Theta_n^{\dagger}(y;x,z)=\left[\frac{x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}}+\omega(y;x,z)+o(1)\right]\mathbf{1}\{x>0\}+\left[\frac{x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}}+o(1)\right]\mathbf{1}\{x\leq 0\}, \tag{5.9}$$

where $\upsilon(x) = 1 + 0.1x^2$, $\omega(y;x,z) = \frac{-\phi_1(0,z)\psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{x}{\upsilon(x)}}{f_{A|Z}\left(\frac{y}{\upsilon(x)}|z\right)}$, ϕ_1 (resp. ψ_2) is the derivative of $\phi(\cdot,\cdot)$ (resp. $\psi(\cdot,\cdot)$) with respect to its first (reps. second) argument, ϕ and ψ are some functions defined via the use of the implicit function theorem in Appendix F.2, and $\psi_2\left(\frac{y}{\upsilon(x)},0\right) > 0$. In this case, we have

$$c_{\Theta_0} = E\left\{\left[\int \omega(Y_i;X_i,z)d\Delta(z)\right]^2 \pi(X_i,Y_i)\right\}, \tag{5.10}$$

which is positive for properly chosen $\Delta(z)$ and $\pi(\cdot,\cdot)$ provided that $\phi_1(0,z)/f_{A|Z}\left(\frac{y}{\upsilon(x)}\middle| z\right)$ is not invariant with respect to z.

Remark 5.4. On the other hand, we do find that our test does not have power against some particular sequence of local alternatives. To see this, we consider a simple sequence of local alternatives where

$$m(X_i,A_i) = -1 + X_i + A_i,\ \gamma(X_i,A_i) = X_iA_i,$$

and we assume that the PDF of X_i has support $(-\infty,\infty)$. In this case, $m(x,a) = -1 + x + a$ is strictly increasing in a for all x and the normalization $m(x^*,a) = a\ \forall a$ is achieved at $x^* = 1$. In addition, $m_n(x,a) = -1 + x + a + c_nxa$ is strictly increasing in a for all $x > -1/c_n$ and strictly decreasing in a for all $x < -1/c_n$. But the region where $m_n(x,\cdot)$ is not strictly increasing keeps shrinking as $c_n \downarrow 0$ and eventually it becomes a strictly increasing function for all x on the real line as $n \to \infty$. It is easy to verify that under Assumption A.2,

$$\begin{aligned}
G(y|x,z) &= F_{A|Z}(y+1-x|z), \\
G^{-1}(\tau|x,z) &= x - 1 + F_{A|Z}^{-1}(\tau|z), \\
G_n(y|x,z) &= F_{A|Z}\left(\frac{y+1-x}{1+c_nx}|z\right)e_n(x) + \bar{F}_{A|Z}\left(\frac{y+1-x}{1+c_nx}|z\right)\bar{e}_n(x), \\
G_n^{-1}(\tau|x,z) &= x - 1 + (1+c_nx)\left[F_{A|Z}^{-1}(\tau|z)e_n(x) + F_{A|Z}^{-1}(1-\tau|z)\bar{e}_n(x)\right],
\end{aligned}$$

where $\bar{F}_{A|Z} = 1 - F_{A|Z}$, $e_n(x) = \mathbf{1}\{1+c_nx > 0\}$, and $\bar{e}_n(x) = \mathbf{1}\{1+c_nx < 0\}$. Noting that $e_n(x^*) = 1$ and $G^{-1}(G(Y_i|X_i,z)|x^*,z) = A_i$, by repeated applications of Taylor expansions we can show that

$$\begin{aligned}
G_n^{-1}(G_n(Y_{ni}|X_i,z)|x^*,z) &= x^* - 1 + (1+c_nx^*)F_{A|Z}^{-1}(G_n(Y_{ni}|X_i,z)|z) \\
&\approx A_i + c_nA_i + \frac{\left[1 - 2F_{A|Z}(A_i|z)\right]\bar{e}_n(X_i)}{f_{A|Z}(A_i|z)}
\end{aligned}$$

where we ignore terms that are $o_P(c_n)$. Noting that for any $\epsilon > 0$, $P(\bar{e}_n(X_i) > \epsilon c_n) \le P(X_i < -\frac{1}{c_n}) \to 0$, we can conclude that the last term in the last displayed expression is $o_P(c_n)$. Then $c_{\Theta_0} = E\{[\int A_i d\Delta(z)]^2\pi(X_i,Y_i)\} = 0$ in this case, and our test does not have power to detect local alternatives of the form $m_n(x,a) = -1 + x + a + c_nxa$ for $c_n \propto n^{-1/2}b^{-d_X/2}$. A close examination of the above derivation suggests that higher order Taylor expansions should be called upon and the dominant term in $G_n^{-1}(G_n(Y_{ni}|X_i,z)|x^*,z) - G^{-1}(G(Y_i|X_i,z)|x^*,z)$ that is not a constant function of z will be of probability order $O_P(c_n^2)$, implying that our test has power against such type of local alternative that converges to the null at the much slower rate: $n^{-1/4}b^{-d_X/4}$.

Remark 5.5. To maximize the asymptotic local power, one may consider choosing Δ to maximize c_{Θ_0}/σ_J, where both c_{Θ_0} and σ_J depend on Δ implicitly. Noting σ_J^2 is the limit of $\sigma_{J_n}^2 = 2b^{2d_X} E[\varphi(W_1, W_2)^2]$ with $\varphi(w_1, w_2) \equiv E\left[\int\int \zeta_0(W_1, w_1; z)\, \zeta_0(W_1, w_2; \bar{z})\, d\Delta(z)\, d\Delta(\bar{z})\, \pi(X_1, Y_1)\right]$, the Δ function enters σ_J^2 fourfold. Similarly, it enters c_{Θ_0} twofold. The local power function is distinct from what we have commonly seen in the literature (e.g., Tripathi and Kitamura, 2003; Su and White, 2014), and we are not aware of any variational analysis that can help us optimize the local asymptotic power over Δ. In addition, since the above local power function depends on the particular local alternative through Θ_n which is generally unobserved, one should consider a weighted average local power following the spirit of Andrews and Ploberger (1994) in the parametric literature. But this further complicates the issue to a great deal and goes beyond the scope of this paper.

On the other hand, as the associate editor kindly indicates, one can consider improving the power performance of our test by studying a sup-type statistic of the form $\sup_{\Delta\in\Xi} T_n(\Delta)$, where $\Xi = \{H_1 - H_2 :$ both H_1 and H_2 are CDFs defined on $\mathbb{R}^{d_Z}$ with continuous PDFs$\}$ and we have made the dependence of T_n on Δ explicit. Without any restriction on the size of Ξ, it seems impossible to consider the above optimization problem. Nevertheless, if Ξ contains only a finite number of elements, our asymptotic analysis can be extended to this case straightforwardly. We do not consider this option because our test based on a single choice of Δ is already computationally expensive. Instead, in the simulations we will consider the effect of different choices of Δ on the size and power of our test.

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

$$\mathbb{H}_1 : c_{H_1} \equiv E\left\{\left[\int G^{-1}(G(Y_1|X_1, z) \mid x^*, z))d\Delta(z)\right]^2 \pi(X_1, Y_1)\right\} > 0.$$

Theorem 5.4 *Suppose Assumptions C.1-C.4, C.6, and C.7 hold. Suppose that Assumption C.5 hold with H replaced by H_1 and H_2. Then under $\mathbb{H}_1$, $P(T_n > c_n) \to 1$ for any nonstochastic sequence $c_n = o(nb^{d_X})$.*

5.3 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, and Monte Carlo experiments show this to be true here as well. Thus, we suggest to use a bootstrap method to obtain bootstrap P-values.

Let $\mathbb{W}_n \equiv \{W_i\}_{i=1}^n$. 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 $\tilde{f}_Z(z) = n^{-1}\sum_{i=1}^n \phi_{b_z}(Z_i - z)$, where $\phi_b(z) = b^{-d_Z}\phi(z/b)$, $\phi(\cdot)$ 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 $b_z > 0$ is a bandwidth parameter.

3. For $i = 1, ..., n$, given Z_i^*, draw X_i^* and A_i^* independently from the smoothed conditional density $\tilde{f}_{X|Z}(x|Z_i^*) = \sum_{j=1}^n \phi_{b_x}(X_j - x)\phi_{b_z}(Z_j - Z_i^*) / \sum_{l=1}^n \phi_{b_z}(Z_l - Z_i^*)$ and $\tilde{f}_{A|Z}(a|Z_i^*) = \sum_{j=1}^n \phi_{b_a}(\hat{A}_j - a)\phi_{b_z}(Z_j - Z_i^*) / \sum_{l=1}^n \phi_{b_z}(Z_l - Z_i^*)$, respectively, where b_z, b_x, and b_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 $\mathbb{W}_n^* \equiv \{W_i^* = (X_i^{*\prime}, Y_i^*, Z_i^{*\prime})'\}_{i=1}^n$ replacing $\mathbb{W}_n$.

6. Repeat Steps 2-5 N_B times to obtain bootstrap test statistics $\{T_{nj}^*\}_{j=1}^{N_B}$. Calculate the bootstrap P-values $P^* \equiv N_B^{-1} \sum_{j=1}^{N_B} 1\{T_{nj}^* \geq T_n\}$ 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 supplemental appendix sketches the main ideas needed to show that this bootstrap method is asymptotically valid under suitable conditions, that is,

$$(i)\ \ P(T_n^* \leq t|\mathbb{W}_n) \to \Phi(t) \text{ for all } t \in \mathbb{R}, \text{ and } (ii)\ \ P(T_n > z_\alpha^*) \to 1 \text{ under } \mathbb{H}_1, \tag{5.11}$$

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

5.4 Asymptotics with nonparametrically generated regressors

In the control function approach literature, the control variable Z is often not directly observed. Let Q denote the endogenous regressor and ϖ the instrument. Following Newey et al. (1999), a seminal reference in the control function literature, we suppose that Q, ϖ, and Z satisfy

$$Q = r_0(\varpi) + Z, \tag{5.12}$$

where (Q, ϖ) takes value in $\mathcal{Q} \times \Omega_{d_\varpi}$, Z is a $d_Z \times 1$ random vector such that $Z = Q - E(Q|\varpi)$ and $E\|Z\|^2 < \infty$. This implies that $r_0(\varpi) = E(Q|\varpi) \equiv (r_{0,1}(\varpi), ..., r_{0,d_Z}(\varpi))'$. In the application below, we generalize the specification to $Q = \Psi(\varpi, Z)$, and Ψ strictly monotonic in Z, but for brevity of exposition we focus on the additive Z case here.

We use d_ϖ to denote the dimension of ϖ and assume that $r_{0,\ell} : \mathbb{R}^{d_\varpi} \to \mathbb{R}$, $\ell = 1, ..., d_Z$, are measurable functions. Moreover, we consider the $\tilde{p}$-th order local polynomial estimator of $r_{0,\ell}$:

$$\hat{r}_\ell(\varpi) = e_{1,\tilde{p}}' \mathbf{S}_{\tilde{p},\tilde{b}}(\varpi)^{-1} n^{-1} \sum_{i=1}^n \tilde{K}_{\tilde{b}}(\varpi_i - \varpi)\mu_{\tilde{p},\tilde{b}}(\varpi_i - \varpi) Q_{i,\ell}, \tag{5.13}$$

where $e_{1,\tilde{p}}$, $\mathbf{S}_{\tilde{p},\tilde{b}}(\varpi)$, $\mu_{\tilde{p},\tilde{b}}(\varpi_i - \varpi)$ are defined similarly as before, $\tilde{K}_{\tilde{b}}(\cdot) \equiv \tilde{K}(\cdot/\tilde{b})/\tilde{b}$, $\tilde{K}(\cdot)$ is a PDF on $\mathbb{R}^{d_\varpi}$, and $Q_i = (Q_{i,1}, ..., Q_{i,d_Z})'$ and ϖ_i, $i = 1, ..., n$, are IID copy of Q and ϖ. Let $\hat{r}(\varpi) = (\hat{r}_1(\varpi), ..., \hat{r}_{d_Z}(\varpi))'$. For notational simplicity, we use the same bandwidth $\tilde{b}$ for each element in ϖ. We now consider the asymptotics of our test with the generated Z: $\hat{Z} \equiv Q - \hat{r}(\varpi)$.

Let $\mathfrak{R}$ denote a class of functions from $\mathbb{R}^{d_\varpi} \to \mathbb{R}^{d_Z}$, and $r_0(\cdot) \in \mathfrak{R}$. Let $\|\cdot\|_{\mathfrak{R}}$ be a generic pseudo-norm on $\mathfrak{R}$. Given two vector functions r_l, r_u, a bracket $[r_l, r_u]$ is set of vector functions $r \in \mathfrak{R}$ such that $r_{l,\ell} \leq r_\ell \leq r_{u,\ell}$ for all $1 \leq \ell \leq d_Z$. An ϵ-bracket with respect to $\|\cdot\|_{\mathfrak{R}}$ is a bracket $[r_l, r_u]$ with $\|r_l - r_u\|_{\mathfrak{R}} \leq \epsilon$, $\|r_l\|_{\mathfrak{R}} < \infty$, $\|r_u\|_{\mathfrak{R}} < \infty$. The covering number with bracketing $N_{[\cdot]}(\epsilon, \mathfrak{R}, \|\cdot\|_{\mathfrak{R}})$ is the minimal number of ϵ-bracket with respect to $\|\cdot\|_{\mathfrak{R}}$ needed to cover $\mathfrak{R}$. We let $\|\cdot\|_2$ and $\|\cdot\|_\infty$ denote the L_2 and sup-norms, respectively: $\|r\|_2 \equiv \{\int \sum_{\ell=1}^{d_Z} r_\ell^2(\varpi)\, dP(\varpi)\}^{1/2}$ and $\|r\|_\infty \equiv \sup_{\varpi \in \Omega_{d_\varpi}} \max_{1 \leq \ell \leq d_Z} |r_\ell(\varpi)|$, where P denotes the probability measure associated with ϖ. We typically require that $\mathfrak{R}$ should not be too complex, e.g., $\log N_{[\cdot]}(\epsilon, \mathfrak{R}, \|\cdot\|_\infty) < C\epsilon^{-c_r}$ for some $c_r < 2$ and for all $\epsilon > 0$. See van der Vaart and Wellner (1996) for examples of classes of functions satisfying such a requirement. The functional space $C_M^\varrho(\Omega_{d_\varpi})$ given in the following definition is one of them with $c_r = d_\varpi/\varrho$.

Definition. Let $\varrho > 0$, $M > 0$, and Ω_{d_ϖ} be a finite union of convex and bounded subsets of $\mathbb{R}^{d_\varpi}$. Let $C_M^\varrho(\Omega_{d_\varpi})$ denote a class of smooth functions such that for any smooth function $f : \Omega_{d_\varpi} \to \mathbb{R}$ in it, we have $\|f\|_{\infty,\varrho} \leq M$, where $\|f\|_{\infty,\varrho} \equiv \max_{|\boldsymbol{j}| \leq \underline{\varrho}} \sup_{\varpi \in \Omega_{d_\varpi}} \left| D^{\boldsymbol{j}} f(\varpi) \right| + \max_{|\boldsymbol{j}| = \underline{\varrho}} \sup_{\varpi \neq \varpi'} \frac{\left| D^{\boldsymbol{j}} f(\varpi) - D^{\boldsymbol{j}} f(\varpi') \right|}{\|\varpi - \varpi'\|^{\varrho - \underline{\varrho}}}$, and $\underline{\varrho}$ be the largest integer smaller than ϱ.

Assumption C.8 (i) Ω_{d_ϖ} is a finite union of convex and bounded subsets of $\mathbb{R}^{d_\varpi}$ with non-empty interior. $r_\ell \in C_M^\varrho(\Omega_{d_\varpi})$ for $\ell = 1, ..., d_Z$, $\varrho/d_\varpi > 1/2$, and $\varrho \geq \tilde{p} + 1$.

(ii) The conditional distribution of X given ϖ exhibits a conditional density $f_{X|\varpi}(x|\varpi)$ that is continuously differentiable in x and continuous in ϖ. $f_{X|\varpi}(\cdot|\varpi)$ is bounded almost surely on $\mathcal{X}_0$.

In addition, we add the following technical assumption:

Assumption C.9 As $n \to \infty$, $b \to 0$, $\|\hat{r} - r_0\|_\infty = o_P(n^{-1/4} b^{1/2 - d_X/4} \wedge b^{\frac{2+d_Z}{2-c_r}})$, $nb^{d_X} \tilde{b}^{2(\tilde{p}+1)} \to 0$, and $\tilde{b}/b \to 0$.

The rate restrictions on $\hat{r}$ in Assumption C.9 are essentially the same as those in Mammen et al. (2012, MRS hereafter); see the discussion in Appendix F.1. The condition on $\|\hat{r} - r_0\|_\infty$ requires that $\hat{r}$ converge faster than $n^{-1/4} b^{1/2 - d_X/4}$ for two reasons: one is that $\hat{r}$ appears in the kernel function $K_b(\cdot)$ whose Taylor expansion results in the appearance of b^{-1}, and the other is that we are aiming at controlling the remainder term in the difference between $\tilde{G}^{-1}(\tau|u)$ and $\hat{G}^{-1}(\tau|u)$ (resp. $\tilde{G}_{p,b}(y|u)$ and $\hat{G}_{p,b}(y|u)$) at the rate $o_P((nb^{d_X})^{-1/2})$ (not $o_P(n^{-1/2})$ as for parametric estimation in the second stage). It also requires that $\|\hat{r} - r_0\|_\infty^{1 - \frac{1}{2}c_r} = o_P(b^{1 + d_Z/2})$, which could be restrictive with respect to the smoothness of r. In view of the fact that $c_r = d_\varpi/\varrho$ for the functional space we are interested in, the smaller d_ϖ is or the larger ϱ is, the more likely for this rate restriction on $\hat{r} - r$ to hold. In general, we need a smoother class of functions (larger ϱ) for larger value of d_ϖ in order for the above condition to hold, reflecting the curse of dimensionality issue (c.f. Ai and Chen 2003). $nb^{d_X} \tilde{b}^{2(\tilde{p}+1)} \to 0$ is imposed to ensure the bias term associated with the estimation of r is asymptotically negligible. $\tilde{b} = o(b)$ is imposed to obtain the asymptotic linear expansion as in the first case of Proposition 1 in MRS and to simplify the proofs in various places. Consider the case where $d_X = 2$, $d_Z = 1$, and $d_\varpi = 2$, which resembles our second application below. By choosing $p = 1$, $\tilde{p} = 2$, $b \propto n^{-1/[2(p+1)+d_X]} = n^{-1/6}$, and $\tilde{b} \propto b/\log(n)$, we have the standard convergence result $\|\hat{r} - r_0\|_\infty = O_P(\tilde{b}^3 + n^{-1/2}\tilde{b}^{-1}(\log n)^{1/2})$. Then the rate condition on $\|\hat{r} - r_0\|_\infty$ in Assumption C.9 would be met provided $c_r < 1/2$ or equivalently $\varrho > 4$. Alternatively, if we choose $p = \tilde{p} = 2$, $b \propto n^{-1/[2(p+1)+d_X]} = n^{-1/8}$, and $\tilde{b} \propto b/\log(n)$, we can also check that $\|\hat{r} - r_0\|_\infty = O_P(\tilde{b}^3 + n^{-1/2}\tilde{b}^{-1}(\log n)^{1/2}) = o_P(n^{-1/4}b^{1/2-d_X/4} \wedge b^{\frac{2+d_Z}{2-c_r}})$ provided $c_r < 1$ or equivalently $\varrho > 2$. One can readily check the other conditions in Assumptions C.6 and C.9 are also met in either case.

To simplify notation, let $\boldsymbol{\theta}$ be the vector that stacks $\sqrt{nb^d}(\alpha_{\boldsymbol{j}} - \frac{b^{\boldsymbol{j}}}{\boldsymbol{j}!}D^{\boldsymbol{j}}G^{-1}(\tau|u))$ for $0 \leq |\boldsymbol{j}| \leq p$ in lexicographic order. Let $U_i(r) \equiv \left(X_i', (Q_i - r(\varpi_i))'\right)'$, $\mu_{p,b,i}(r;u) \equiv \mu_{p,b}(U_i(r) - u)$, $\check{K}_b(\cdot) \equiv K(\cdot/b)$, $\check{K}_{b,i}(r) \equiv K((U_i(r) - u)/b)$, and $Y_i^*(r) = Y_i - \sum_{0\leq|\boldsymbol{j}|\leq p} \frac{1}{\boldsymbol{j}!}D^{\boldsymbol{j}}G^{-1}(\tau|u)(U_i(r) - u)^{\boldsymbol{j}}$. Note that $Y_i^*(r_0) \equiv Y_i - \boldsymbol{\beta}(\tau;u)$, where $\boldsymbol{\beta}(\tau;u)$ is the pth-order Taylor expansion of $G^{-1}(\tau|U_i(r_0))$ around u. Note that we have suppressed the dependence of $\check{K}_{b,i}(r)$ and $Y_i^*(r)$ on u for notational simplicity.

In terms of $\boldsymbol{\theta}$, the minimization problem in (4.4) can now be rewritten as

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left[\rho_\tau\left(Y_i^*(r_0) - \mu_{p,b,i}'(r_0;u)\frac{\boldsymbol{\theta}}{\sqrt{nb^d}}\right) - \rho_\tau(Y_i^*(r_0))\right]\check{K}_{b,i}(r_0), \tag{5.14}$$

as $\rho_\tau(Y_i^*(r_0))$ is not a function of $\boldsymbol{\theta}$ and $\check{K}_{b,i}(r_0) = b^d K_b(U_i(r_0) - u)$. The problem with generated regressors can be similarly rewritten as

$$\min_{\boldsymbol{\theta}} \sum_{i=1}^{n} \left[\rho_\tau\left(Y_i^*(\hat{r}) - \mu_{p,b,i}'(\hat{r};u)\frac{\boldsymbol{\theta}}{\sqrt{nb^d}}\right) - \rho_\tau(Y_i^*(\hat{r}))\right]\check{K}_{b,i}(\hat{r}). \tag{5.15}$$

Denote the minimizer in (5.15) as $\tilde{\boldsymbol{\theta}}$. Then we can recover the estimate for $G^{-1}(\tau|u)$ by $\tilde{G}_{p,b}^{-1}(\tau|u)$ via: $\tilde{G}_{p,b}^{-1}(\tau|u) \equiv G^{-1}(\tau|u) + \frac{1}{\sqrt{nb^d}}e_{1,p}'\tilde{\boldsymbol{\theta}}$.

Proposition 5.5 *Suppose Assumptions C.1-C.4, C.6, and C.8-C.9 hold. Then* $\tilde{G}_{p,b}^{-1}(\tau|u) - \hat{G}_{p,b}^{-1}(\tau|u) = \Upsilon_{1,p,b}(\tau,u)\{1 + o_P(1)\} + o_P(n^{-1/2}b^{-d_X/2})$ *uniformly over* $(\tau,u) \in \mathcal{T} \times \mathcal{U}_0$, *where*

$$\Upsilon_{1,p,b}(\tau,u) = -n^{-1}\sum_{i=1}^{n} e_{1,p}'\bar{S}_{p,b}(\tau;u)^{-1} b^{-d_Z}\overline{\mu\check{K}}_{p,b}(z, r_0(\varpi_i))\Psi_{n1}(\tau,u,\varpi_i)Z_i,$$

$\overline{\mu\check{K}}_{p,b}(z, r_0(\varpi_i)) \equiv \int_{\mathbb{R}^{d_X}} \mu_p((\mathbf{t}_{d_X}, z_{ib}')')\ K((\mathbf{t}_{d_X}, z_{ib}')')\,d\mathbf{t}_{d_X}$, $\mathbf{t}_{d_X} \equiv (t_1, ..., t_{d_X})$, $z_{ib} = (\bar{Q}(x,\varpi_i) - r_0(\varpi_i) - z)/b$, *and* $\bar{Q}$ *and* Ψ_{n1} *are defined in equations (D.14) and (D.17) respectively.*

Consider the local polynomial regression estimate of $G(y|u)$ with the generated $\hat{r}$:

$$\min_{\boldsymbol{\beta}} \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{1}\{Y_i \leq y\} - \sum_{0\leq|\boldsymbol{j}|\leq p} \boldsymbol{\beta}_{\boldsymbol{j}}'((U_i(\hat{r}) - u)/b)^{\boldsymbol{j}}\right]^2 K_b(U_i(\hat{r}) - u). \tag{5.16}$$

Denote the solution as $\tilde{\boldsymbol{\beta}}(y|u)$. Let $\tilde{G}_{p,b}(y|u) \equiv e_{1,p}'\tilde{\boldsymbol{\beta}}(y|u)$. Recall that the corresponding solution to the original problem (4.1) with observed U_i is denoted as $\hat{\boldsymbol{\beta}}(y|u)$ and that $\hat{G}_{p,b}(y|u) = e_{1,p}'\hat{\boldsymbol{\beta}}(y|u)$.

Proposition 5.6 *Suppose Assumptions C.1-C.4, C.6, and C.8-C.9 hold. Then* $\tilde{G}_{p,b}(y|u) - \hat{G}_{p,b}(y|u) = \Upsilon_{2,p,b}(y,u)\{1+o_P(1)\} + o_P(n^{-1/2}b^{-d_X/2})$ *uniformly over* $(y,u) \in \mathbb{R} \times \mathcal{U}_0$,

$$\Upsilon_{2,p,b}(y,u) = -n^{-1}\sum_{i=1}^{n} \frac{dG(y|u)}{du'} \begin{pmatrix} \mathbf{0}_{d_X} \\ Z_i \end{pmatrix} [e'_{1,p}\overline{\mathbf{S}}_{p,b}(u)^{-1} b^{-d_Z} \overline{\mu\check{K}}_{p,b}(z, r_0(\varpi_i))]\Psi_{n2}(x,\varpi_i),$$

$\overline{\mu\check{K}}_{p,b}(r_0(\varpi_i)) \equiv \int_{\mathbb{R}^{d_X}} \mu_p((\mathbf{t}_{d_X}, z'_{ib})') K((\mathbf{t}_{d_X}, z'_{ib})') d\mathbf{t}_{d_X}$, $\mathbf{0}_{d_X}$ *is a* $d_X \times 1$ *vector of zeros, and* Ψ_{n2} *is defined in equation (D.18).*

Remark 5.6. It is trivial to show that the dominant term in $\tilde{G}^{-1}_{p,b}(\tau|u) - \hat{G}^{-1}_{p,b}(\tau|u)$ or $\tilde{G}_{p,b}(y|u) - \hat{G}_{p,b}(y|u)$ is of the order $O_P\left(n^{-1/2}b^{-d_Z/2}\right)$ if $d_Z \geq d_X$. From the proof of Theorem 5.3, we see that the asymptotic distribution of the test statistic with observed Z under the local alternative depends on the interactions of the dominating terms in the linear representations of $\tilde{G}^{-1}(\tau|u)) - G^{-1}(\tau|u)$ and $\tilde{G}_{p,b}(y|u) - G(y|u)$. When Z is nonparametrically generated as above, a careful examination of the proof of Theorem 5.3 suggests that the interaction between $\Upsilon_{1,p,b}(\tau,u)$ and $\Upsilon_{2,p,b}(y,u)$ and the interactions between these two terms and the dominant terms in the aforementioned linear representations are asymptotically negligible for the asymptotic distribution of our test statistic due to the smooth integrating operator over $\Delta(z)$ in the definition of the test statistic. As a result, the asymptotic distribution of our test statistic under the null or local alternative with nonparametrically generated Z will be the same as the case with observed Z.

6 Estimation and specification testing in finite samples

In this section, we conduct simulations to evaluate the finite-sample performance of our estimators and tests. We first consider the estimation of the response and then examine the behavior of the T_n test.

6.1 Estimation of the response

We begin by considering the following two DGPs:

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

DGP 2: $Y_i = A_i/\left(1 + 0.1X_i^2\right)$,

where $i = 1, ..., n$, $A_i = 0.5Z_i + \varepsilon_{1i}$, $X_i = 0.25 + Z_i - 0.25Z_i^2 + \varepsilon_{2i}$, and ε_{1i}, ε_{2i}, and Z_i are each IID $\mathbb{N}(0,1)$ and mutually independent. Clearly, $m(x,a) = (0.5 + 0.1x^2)a$ in DGP 1 and $= a/\left(1+0.1x^2\right)$ 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.[4]

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, similarly, we let $a^* = a/c_2$, $c_2 = 1/\left(1 + 0.1x_{med}^2\right) = 0.9987$. For notational simplicity, we continue to use $m(x,a)$ and A_i to denote $m^*(x,a)$ and A_i^*, respectively.

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

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 $\hat{G}_{p,b}^{-1}(\hat{G}_{p,b}(a|x^*, z)|x, z)$, 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: $b = (1.06\sigma_X n^{-1/6}, 1.06\sigma_Z n^{-1/6})$, where, e.g., σ_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 $[C_{\epsilon_0}, C_{1-\epsilon_0}]$, and H_2 is a scaled beta$(2,2)$ distribution on $[C_{\epsilon_0}, C_{1-\epsilon_0}]$, where C_{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 $[-1.433,\ 1.433]$, where -1.433 and 1.433 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 $m(x,a)$ at these 225 points, and calculate the corresponding mean absolute deviations (MADs) and root mean squared errors (RMSEs):

$$MAD_{H_l}^{(i)} = \frac{1}{225} \sum_{j=1}^{225} \left| m(x_j, a_j) - \hat{m}_{H_l}^{(i)}(x_j, a_j) \right|, \tag{6.1}$$

$$RMSE_{H_l}^{(i)} = \left\{ \frac{1}{225} \sum_{j=1}^{225} \left[m(x_j, a_j) - \hat{m}_{H_l}^{(i)}(x_j, a_j) \right]^2 \right\}^{1/2}, \tag{6.2}$$

where, for $i = 1, ..., 500$, $\hat{m}_{H_l}^{(i)}(x_j, a_j)$ is the estimate of $m(x_j, a_j)$ in the ith 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 $MAD_{H_l}^{(i)}$ and $RMSE_{H_l}^{(i)}$ for the estimates of $m(x,a)$, together with their means obtained by averaging over 500 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. Second, as n quadruples, both the MADs and RMSEs tend to improve, as expected. Third, also as expected, the MADs and RMSEs improve at a rate slower than the parametric rate $n^{-1/2}$.

To examine the small sample properties of the our estimates, Figures 1 and 2 displays the Q-Q plots of standardized $\hat{m}_{H_1}^{(i)}(x,a)$ $(i = 1, ..., 500)$ at four different points, namely, $(x,a) = (-0.85, -0.31), (-0.85, 0.21)$, $(0.45, -0.31)$ and $(0.45, 0.21)$, against the standard normal distribution for $n = 100$ and 400 in DGPs 1 and 2, respectively. In both figures we first center $\hat{m}_{H_1}^{(i)}(x,a)$ around its sample mean and then divide by its sample standard deviation over 500 simulations to obtain the normalized estimate. We see that the points

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	0.192	0.269	0.405	0.278	0.246	0.348	0.510	0.361
		Beta	0.158	0.222	0.323	0.227	0.208	0.296	0.429	0.304
	400	Uniform	0.126	0.175	0.249	0.181	0.164	0.234	0.332	0.240
		Beta	0.093	0.132	0.181	0.135	0.126	0.182	0.258	0.185
2	100	Uniform	0.262	0.374	0.574	0.391	0.326	0.468	0.706	0.484
		Beta	0.208	0.295	0.424	0.303	0.259	0.368	0.536	0.383
	400	Uniform	0.169	0.239	0.352	0.249	0.217	0.298	0.435	0.310
		Beta	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.222

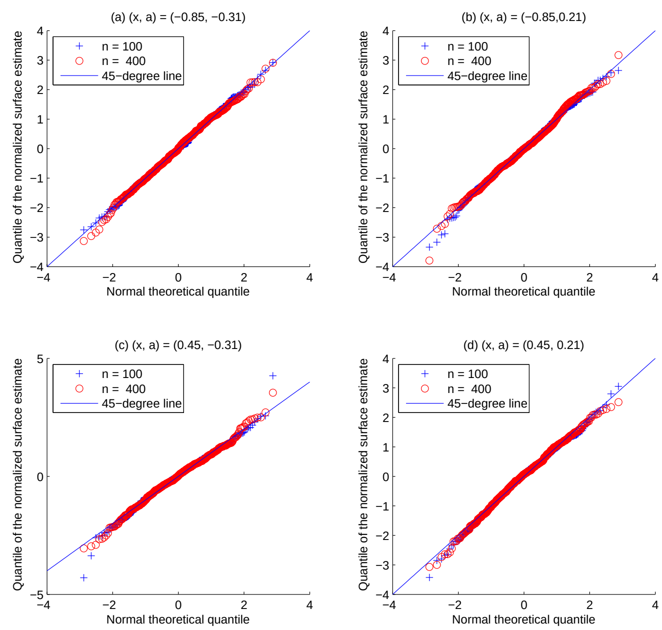

Figure 1: Q-Q plot of $\hat{m}_{H_1}^{(i)}(x, a)$, $i = 1, ..., 500$, at various points versus standard normal (DGP 1, $n = 100, 400$)

in the Q-Q plots follow closely along the 45-degree line at all four points for both sample sizes. We find that the result is similar when we change the evaluation points (x, a) or when we use H_2 instead of H_1.

6.2 Specification testing

To examine the finite-sample properties of the specification test, we consider 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: $Y_i = A_i/\left(1 + 0.1X_i^2\right) + 2\delta_0 X_i e^{0.1A_i}$,

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.

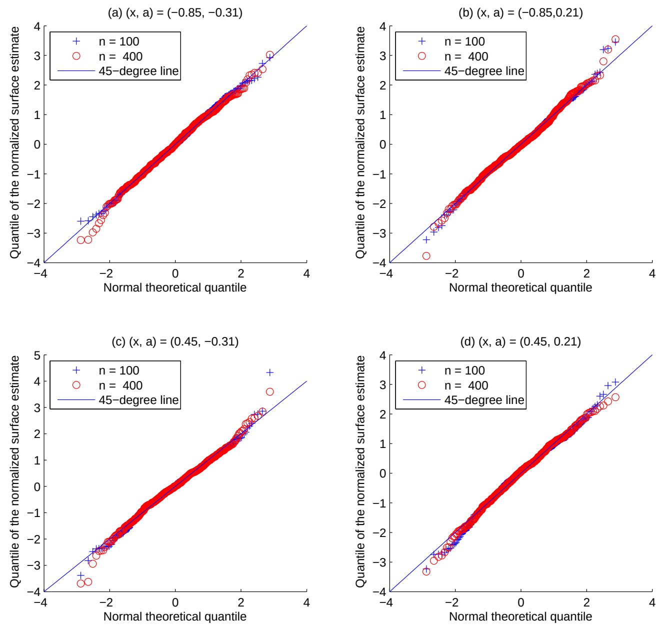

Figure 2: Q-Q plot of $\hat{m}^{(i)}_{H_1}(x,a)$, $i = 1, ..., 500$, at various points versus standard normal (DGP 2, $n = 100, 400$)

To construct the raw test statistic $\hat{J}_n$, we first obtain $\hat{G}_{p,b}(Y_i|X_i,z)$ and $\hat{G}^{-1}_{p,b}(\hat{G}_{p,b}(Y_i|X_i,z) \mid x^*, z)$ by choosing the order (p) of the local polynomial regression, the kernel function K, and the bandwidth b. As in the estimation of 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 $b = (c_2\sigma_X n^{-1/5},\ c_2\sigma_Z n^{-1/5})$, where c_2 is a positive scalar that we use to check the sensitivity of our test to the choice of bandwidth. Next, we choose H_1 and H_2 as above and set $\pi(X_i, Y_i) = \mathbf{1}\{C_{\epsilon_0,X} \leq X_i \leq C_{1-\epsilon_0,X}\} \times \mathbf{1}\{C_{\epsilon_0,Y} \leq Y_i \leq C_{1-\epsilon_0,Y}\}$, where, e.g., $C_{\epsilon_0,X}$ is the ϵ_0th sample quantile of $\{X_i\}_{i=1}^n$ and $\epsilon_0 = 0.0125$.[5] 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 $\hat{G}^{-1}_{p,b}$ in the tails.

In the bootstrap, we set $b_z = s_Z n^{-1/6}$, $b_x = s_X n^{-1/6}$, and $b_a = s_A n^{-1/6}$, where, e.g., s_Z denotes the sample standard deviation of Z_i. For computational feasibility, we consider two sample sizes ($n = 100$, 200) in our simulation study; for each sample size, our "full" bootstrap experiments use 500 replications and $N_B = 99$ bootstrap resamples in each replication. For the reason to choose 99 instead of 100 bootstrap resamples, see Racine and MacKinnon (2007). We also evaluate our test statistic using the warp-speed bootstrap of Giacomini et al. (2013). In this procedure, only one bootstrap resample is drawn in each replication. We use 499 replications for this study. We study the sensitivity of our test to the bandwidth b by setting $b = (c_2 s_X n^{-1/5},\ c_2 s_Z n^{-1/5})$ for $c_2 = 0.9, 1.1, 1.3, 1.5$.

Table 2 reports the empirical rejection frequencies for our test at various nominal levels for DGPs 3-4

[5] We experimented several different weight functions, like scaled beta(1,1) (uniform), beta(2,2), beta(3,3), beta(4,2) etc. for H_1 and H_2 and found our results are not very sensitive to the choice of H_1 and H_2. To get reasonable power, however, one may not choose some H_1 and H_2 that are too similar, i.e., beta(7,7) and beta(8,8). In practice, we recommend to trim off some extreme evaluation points and use some density functions with bounded support for H_1 and H_2, like what we did here.

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%
			$b = \left(0.9 s_X n^{-1/5}, 0.9 s_Z n^{-1/5}\right)$					
3	100	0	0.004	0.064	0.128	0.024	0.080	0.142
		1	0.218	0.604	0.746	0.440	0.678	0.766
	200	0	0.020	0.048	0.112	0.018	0.070	0.108
		1	0.346	0.688	0.878	0.558	0.798	0.886
4	100	0	0.006	0.050	0.100	0.012	0.066	0.124
		1	0.330	0.764	0.846	0.572	0.768	0.854
	200	0	0.008	0.032	0.072	0.016	0.052	0.094
		1	0.326	0.818	0.912	0.598	0.840	0.904
			$b = \left(1.1 s_X n^{-1/5}, 1.1 s_Z n^{-1/5}\right)$					
3	100	0	0.020	0.042	0.096	0.016	0.072	0.146
		1	0.386	0.562	0.622	0.406	0.524	0.565
	200	0	0.036	0.092	0.158	0.009	0.061	0.141
		1	0.544	0.704	0.738	0.523	0.613	0.656
4	100	0	0.018	0.034	0.100	0.012	0.056	0.130
		1	0.566	0.734	0.798	0.456	0.580	0.642
	200	0	0.002	0.034	0.080	0.006	0.056	0.106
		1	0.538	0.818	0.866	0.618	0.728	0.772
			$b = \left(1.3 s_X n^{-1/5}, 1.3 s_Z n^{-1/5}\right)$					
3	100	0	0.018	0.036	0.106	0.024	0.079	0.116
		1	0.376	0.468	0.532	0.368	0.460	0.496
	200	0	0.028	0.082	0.130	0.008	0.048	0.106
		1	0.504	0.570	0.622	0.477	0.550	0.585
4	100	0	0.012	0.042	0.104	0.010	0.060	0.132
		1	0.542	0.676	0.718	0.392	0.498	0.546
	200	0	0.002	0.044	0.080	0.004	0.064	0.116
		1	0.570	0.778	0.826	0.538	0.634	0.672
			$b = \left(1.5 s_X n^{-1/5}, 1.5 s_Z n^{-1/5}\right)$					
3	100	0	0.012	0.038	0.084	0.014	0.058	0.099
		1	0.316	0.412	0.442	0.338	0.414	0.462
	200	0	0.016	0.054	0.097	0.012	0.042	0.093
		1	0.440	0.505	0.521	0.424	0.503	0.531
4	100	0	0.000	0.054	0.091	0.010	0.054	0.111
		1	0.476	0.595	0.712	0.458	0.696	0.786
	200	0	0.016	0.044	0.082	0.008	0.050	0.107
		1	0.569	0.722	0.762	0.673	0.740	0.762

based on both the warp-speed and full bootstrap. 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. First, the level of our test is reasonably 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. Third, our test is not very sensitive to the choice of b.

7 Empirical applications

This section illustrates the usefulness of our test with two examples. To show their broad applicability, we consider two very different applications. The first application analyzes the determinants of the Black-White earnings gap. The second application comes from classical consumer demand using Engel curves.

7.1 The black-white earnings gap: just ability?

7.1.1 Economic background

The quest for the sources of the apparent differences in economic circumstances between the three major races in the United States, 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 hereafter), 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. See Carneiro, Heckman, and Masterov (2005, CHM hereafter) for an overview of this literature.

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 (2011) discuss 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 decreases, 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 $X = (X_1, X_2)'$ 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.

7.1.2 The data

Our data come from NJ's original study, 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 we are otherwise using exactly the same data as NJ, we refer to their paper for summary statistics and other data details.

7.1.3 Implementation detail

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 cross validation as in Li and Racine (2004). We perform 199 bootstrap replications.

Table 3: The monotonicity test results for the Black-White earnings data

	Male	Female
p-value	0.965	0.652

7.1.4 Empirical results

Table 3 report the bootstrap p-values for our test. Since the p-values are so large, we believe that it is 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, Assumption A.2. Without this, the test is a joint test for Assumptions A.1 and A.2. Under this interpretation, we have no evidence against either A.1 or A.2. To illustrate the use of a multiple test for misspecification, we also report the results of a pure test of A.2. By White and Chalak (2010, Prop.2), we can test A.2 by testing $X \perp \xi \mid Z$, where $\xi = q(Z, A, V)$, with $X \perp V \mid (A, Z)$. A plausible candidate for ξ is another proxy for A, viewing V as a measurement error. Here, a natural choice for ξ 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 et al. (2013). Table 4 reports the results. We fail to reject the null of conditional exogeneity for both males and females, consistent with our monotonicity test findings.

Table 4: The conditional exogeneity test results for the Black-White earnings data

	Male	Female
p-value	0.54	0.51

7.2 Engel Curves in a Heterogeneous Population

7.2.1 Economic Background

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

$$Y = m(X_1, X_2, A), \tag{7.1}$$

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. It is commonly thought that log total expenditure is endogenous[6] and is hence instrumented for, typically by labor income, say S. This is justified by the same intertemporal separability assumption.

The model is different from the model considered in Blundell et al. (2014), which also features scalar heterogeneity, for two reasons: First, they consider price effects (even though they do not estimate them, and merely use bounds stemming from revealed preference analysis). Second, and more importantly, in their main specification they do not control for endogeneity, nor do they invoke our exact conditional independence assumption. As such, our results are not directly applicable. However, if Blundell et al. (2014) were to extend their analysis to control for endogeneity using a control function approach, our results would indeed have

[6] Nevertheless, the evidence is not strong; see Blundell, Horowitz, and Parey (2012) or Hoderlein (2011).

implications for their analysis. If, in a demand setting, monotonicity in the outcome equation is rejected, approaches like the one put forward in Hoderlein (2011), Dette et al. (2015), or Hoderlein and Stoye (2014) have to be pursued.

To allow for endogeneity, we follow IN and write the X_1 structural equation as

$$X_1 = \Psi(S, X_2, Z), \tag{7.2}$$

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 is $\mathcal{U}[0,1]$. Because $(S, X_2) \perp (A, Z)$, $Z|(S, X_2)$ is also $\mathcal{U}[0,1]$ and Z is identified as

$$Z = F(X_1 \mid S, X_2), \tag{7.3}$$

where F denotes the conditional CDF of X_1 given (S, X_2). 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 (7.1) 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 et al. (2007), who assume $Y = m(X_1, X_2) + A$. Nevertheless, to test this specification, equation (7.2) must also be specified as above, while the IV approach pursued in Blundell et al. (2007) does not require this to be the case. Hence, our test is only valid, if this part of the model is also correctly specified. It is conceivable (though perhaps not extremely likely) that the rejection we find below stems from this part of the model.

7.2.2 The Data

For our test, we use the British Family Expenditure Survey (FES) data in exactly the form employed in IN. The FES reports a yearly cross section of labor income, expenditures, demographic composition, and other characteristics of about 7,000 households in every year. 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 2 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.

7.2.3 Details of Implementation

To apply the test, we let Y be the budget shares of Food, Catering, Transportation, or Leisure in (7.1).[7] 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 cross validation the same as in the previous application. Again, we performed 199 bootstrap replications.

The major difference is that the instrument $Z = F(X_1|S, X_2)$ must be estimated from the data in the first stage and our theory for nonparametrically generated regressors in Section 5.4 can be extended to this application with appropriate modifications. First note that using the fitted value of the first step estimation only changes the signs of the influence terms in Propositions 5.5 and 5.6, because when we plug in the first step estimator in Lemmas C.10 and C.12 using the fitted value, e.g. equation (D.11), $U_i(\hat{r}) - U_i(r_0) = \begin{pmatrix} X_i \\ \hat{r}(\varpi_i) \end{pmatrix} - \begin{pmatrix} X_i \\ r_0(\varpi_i) \end{pmatrix} = \begin{pmatrix} 0 \\ \hat{r}(\varpi_i) - r_0(\varpi_i) \end{pmatrix}$ is of the opposite sign than using residuals as $\hat{Z}$, and the analysis followed is the same as before. Thus, we can continue using the results in Section 5.4. Second, define $Q(x_1) = \mathbf{1}\{X_1 \le x_1\}$ and let $\varpi = (S, Z_2)'$. Let $F(x_1|S, X_2) = E[Q(x_1)|S, X_2]$. Observe then that we have

$$Q(x_1) = E[Q(x_1)|S, Z_2] + \varepsilon(x_1) = F(x_1|S, Z_2) + \varepsilon(x_1),$$

where $E[\varepsilon(x_1)|S, X_2] = 0$. Apparently, $Z = F(X_1|S, Z_2)$ now denotes the cdf (i.e., a mean regression involving the new dependent variable $Q(X_1)$) instead of the residual as in section 5.4. To implement our test, we need to replace $Z_i = F(X_{1i}|S_i, Z_{2i})$ by its $\tilde{p}$-th order local polynomial estimate $\hat{F}(X_{1i}|S_i, Z_{2i})$. Other than the sign change discussed above, we also have to content with the fact that this quantity depends on X_{1i}. However, note that the convergence rate of $\hat{F}(x_1|s, z_2)$ to $F(x_1|s, z_2)$ only depends on the dimension of the conditioning vector $(S_i, Z_{2i})'$ due to the monotonicity of the conditional CDF function $F(\cdot|S, Z_2)$. In fact, under standard conditions for local polynomial regression (e.g., Boente and Fraiman 1991), we can establish the following uniform result

$$\sup_{x_1 \in \mathbb{R}} \sup_{(s,z_2) \in \mathcal{S}} \left\| \hat{F}(x_1|s, z_2) - F(x_1|s, z_2) \right\| = O_P\left(n^{-1/2} \tilde{b}^{-1} (\log n)^{1/2} + \tilde{b}^{\tilde{p}+1} \right),$$

where $\mathcal{S}$ is a compact set in the domain of (S, Z_2). Hence, assumption C.9 continues to hold with $\hat{r}$ and r replaced by $\hat{F}$ and F, respectively. From the discussion of Assumption C.9 in Section 5.4, $n^{-1/2}\tilde{b}^{-1}(\log n)^{1/2} + \tilde{b}^{\tilde{p}+1} = o(n^{-1/4} b^{1/2 - d_X/4} \wedge b^{\frac{2+d_Z}{2-c_r}})$, if we choose p, $\tilde{p}$, b and $\tilde{b}$ as suggested there (and provided c_r is small

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

enough). Indeed, we set $p = 1$ and $\tilde{p} = 2$, choose the two bandwidth sequences as suggested, and employ a Gaussian kernel throughout.

7.2.4 Empirical Results

Table 5 summarizes our test results. 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.

Table 5: The monotonicity test results for the British FES data

	Food	Catering	Transportation	Leisure
Test statistic	1.659	0.752	1.235	1.634
p-value	≤ 0.005	0.010	≤ 0.005	≤ 0.005

8 Conclusion

This paper provides a test for monotonicity in unobservables for cross-section data. We show how to exploit the power of an exclusion restriction together with a conditional independence assumption 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.

Appendix

A Important notations

In this appendix we summarize some important notations in the main body of the paper.

$\mathbf{1}\{\cdot\}$: indicator function;

$\bar{\mathbf{1}}_{Y_i}(W_k) \equiv \mathbf{1}\{Y_k \le Y_i\} - G(Y_i|U_k)$,

$\mathbf{j} \equiv (j_1, ..., j_d)'$, $u^{\mathbf{j}} \equiv \Pi_{i=1}^{d} u_i^{j_i}$, $|\mathbf{j}| = \sum_{i=1}^{d} j_i$, $\sum_{0 \le |\boldsymbol{j}| \le p} \equiv \sum_{k=0}^{p} \sum_{\substack{j_1=0 \\ j_1+\cdots+j_d=k}}^{k} \cdots \sum_{j_d=0}^{k}$;

A : the unobservable defined in the structural function m, with realization a and support $\mathcal{A}$;

$\mathcal{A}_H \equiv \{a : m_H^*(x, a) = y, x \in \mathcal{X}_0, y \in \mathcal{Y}_0\}$;

$\hat{A}_z, \hat{A}_H, \tilde{A}_H, A_H^*, A_H^\dagger$: various terms defined in Section 3.1;

$b, \tilde{b}$: bandwidths used in nonparametric estimation;

$B_m, B_{m^{-1}}, \mathbb{B}_p, \mathbb{B}_{J_n},$: bias terms defined in Theorems 4.1 and 4.3, and equations (4.5) and (5.2);

$\mathbf{B}_{p,b}$: bias term defined before the proof of Theorem 4.1, $\overline{\mathbf{B}}_{p,b} \equiv E[\mathbf{B}_{p,b}]$;

$e_{1,p}$: $N_p \times 1$ vector with 1 in the first position and zeros elsewhere;

E, E_n : expectation and expectation in empirical processes, respectively;

$G(y|x,z), g(y|x,z)$: conditional CDF and PDF of Y_i on $(X_i, Z_i) = (x, z)$, similarly for $G_n(y|x,z), g_n(y|x,z)$;

$\hat{G}_{p,b} \equiv e_{1,p}'\hat{\boldsymbol{\alpha}}$, $\hat{G}_{p,b}^{-1} \equiv e_{1,p}'\hat{\boldsymbol{\beta}}$: the estimators from equations (4.4) and (4.1), respectively;

$\tilde{G}_{p,b}, \tilde{G}_{p,b}^{-1}$: analogue of $\hat{G}_{p,b}$ and $\hat{G}_{p,b}^{-1}$ in the case of generated regressors, defined in (5.16) and (5.15);

$\mathbf{G}_{p+1}$: a vector that stacks the $(p+1)$-th order derivatives of G, defined after equation (4.5);

H, h : CDF and PDF with support on $\mathcal{Z}$;

$\hat{J}_n$: statistic defined in equation (3.1) with standardized version T_n;

$K, \tilde{K}$: kernel functions; $\mathbf{K}_{\mathbf{j}}(u) \equiv u^{\mathbf{j}} K(u)$, $K_b(\cdot) \equiv 1/b^d K(\cdot/b)$, $\tilde{K}_{\tilde{b}}(\cdot) \equiv 1/\tilde{b}^d \tilde{K}\left(\cdot/\tilde{b}\right)$, $\check{K}_b(\cdot) \equiv K(\cdot/b)$;

$m(x, a)$: structural function;

$\hat{m}_z(x, a) \equiv \hat{G}^{-1}(\hat{G}(a|x^*, z)|x, z)$, $\hat{m}_H(x, a) \equiv \int \hat{G}^{-1}(\hat{G}(a|x^*, z)|x, z) dH(z)$ with probability limit $m_H^*(x, a)$;

$\hat{m}_H^{-1}(X, Y) \equiv \inf\{a : \hat{m}_H(X, a) \ge Y\}$, $m_H^{*-1}(X, Y) \equiv \inf\{a : m_H^*(X, a) \ge Y\}$;

$N_{p,l} \equiv \frac{(l+d-1)!}{l!(d-1)!}$, $N_p \equiv \sum_{l=0}^{p} N_{p,l}$;

$N_{[\cdot]}$: the covering number in empirical processes;

$p, \tilde{p}$: the orders of the kernel;

Q : dependent variable defined in Section 5.4 with support $\mathcal{Q}$;

r : structural function between Q and ϖ, defined in the functional space $\Re$;

$\mathbf{S}_{p,b}, \mathbb{S}_{p,b}, S_{p,b}$: technical terms defined in equation (4.3), (4.5), and (5.1) respectively;

$\overline{\mathbf{S}}_{p,b} \equiv E(\mathbf{S}_{p,b})$, $\overline{S}_{p,b} \equiv E(S_{p,b})$;

$U \equiv (X', Z')'$, $u \equiv (x', z')'$, $\mathcal{U} \equiv \mathcal{X} \times \mathcal{Z}$, $\hat{U} \equiv (X', \hat{r}(\varpi)')'$;

$\overline{V}_{p,b}, V_{p,b}, \mathbf{V}_{p,b}$: the decomposition of the variance term defined before the proof of Theorem 4.1;

$W \equiv (X', Y, Z')'$, $\mathbb{W}_n \equiv \{W_i\}_{i=1}^{n}$;

$Y_i^*(r) \equiv Y_i - \sum_{0 \le |\boldsymbol{j}| \le p} \frac{1}{\boldsymbol{j}!} D^{\boldsymbol{j}} G^{-1}(\tau | u)(U_i(r) - u)^{\boldsymbol{j}}$;

$\boldsymbol{\alpha}$, $\boldsymbol{\beta}$: the parameters in equations (4.4) and (4.1), respectively;

$\boldsymbol{\theta}$: the parameter vector that stacks $\sqrt{nb^d}(\alpha_{\boldsymbol{j}} - \frac{b^{\boldsymbol{j}}}{\boldsymbol{j}!} D^{\boldsymbol{j}} G^{-1}(\tau|u))$ for $0 \le |\mathbf{j}| \le p$;

$\eta_{1k}, \eta_{2k}, \varsigma_0, \varphi$: technical terms defined after equation (5.1);

μ : some stacking function defined right before equation (4.2), $\mu_{p,b}(\cdot) \equiv \mu_p(\cdot/b)$;

$\pi_i = \pi(X_i, Y_i)$: some weight function; one defined after equation (3.1);

ϖ : independent variable defined in Section 5.4 with support Ω_{d_ϖ};

$\rho_\tau(u) \equiv u(\tau - \mathbf{1}\{u \le 0\})$;

$\sigma_m^2, \sigma_{m^{-1}}^2, \sigma_{J_n}^2$: variance terms defined in Theorem 4.1 and 4.3, and equation (5.2);

ϕ, Φ : PDF and CDF of the standard normal;

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

$\Delta(z) \equiv H_1(z) - H_2(z)$.

B Proof of the main results

In this appendix we prove all the main results but Theorem 5.3 in the paper. The proof of Theorem 5.3 is lengthy and is relegated to the online supplemental material.

Proof of Proposition 2.1. Assumption A.2 ensures that $G(y|x,z) = P[m(X,A) \le y | X = x, Z = z] = P[m(x,A) \le y | Z = z]$ $\forall (y,x,z)$. By Assumptions A.1-A.3 and setting $y = m(x,a)$, we have that for all $(a, x, \tilde{x}, z)$

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

Now, setting $\tilde{x} = x^*$ and using the normalization $a = m(x^*, a)$ gives $m(x,a) = G^{-1}(G(a \mid x^*, z) \mid x, z)$, ensuring (2.2). Successively inverting $Y = G^{-1}(G(A \mid x^*, z) \mid X, z)$ for any z gives (2.3). ■

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 (4.3) and (5.1), 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 $\psi_\tau(u) \equiv \tau - \mathbf{1}\{u \le 0\}$, $\Delta_{i,y}(u) \equiv G(y|U_i) - G(y|u) - \sum_{1 \le |\mathbf{j}| \le p} \frac{1}{\mathbf{j}!} G^{(\mathbf{j})}(y|u)(U_i - u)^{\mathbf{j}}$, and $\bar{\mathbf{1}}_y(W_i) \equiv \mathbf{1}\{Y_i \le y\} - G(y|U_i)$. Let U_{is} and u_s denote the sth elements of U_i and u.

To prove Theorem 4.1, we will need lemmas C.1-C.3 in Appendix C. Let $\overline{\mathbf{S}}_{p,b}(u) \equiv E[\mathbf{S}_{p,b}(u)]$ and $\overline{\mathbf{B}}_{p,b}(y;u) \equiv E[\mathbf{B}_{p,b}(y;u)]$. In particular, Lemmas C.1 and C.2 establish uniform consistency of $\hat{\boldsymbol{\beta}}(y|u)$ and $\hat{G}_{p,b}^{-1}(\tau|u)$, respectively.

Proof of Theorem 4.1. Letting $\hat{\tau}_z \equiv \hat{G}_{p,b}(a|x^*,z)$ and $\tau_z \equiv G(a|x^*,z)$, we have $\hat{m}_H(x,a) - m_H^*(x,a) = \int [G^{-1}(\hat{\tau}_z \mid x,z) - G^{-1}(\tau_z \mid x,z)]dH(z) + \int [\hat{G}_{p,b}^{-1}(\hat{\tau}_z \mid x,z) - G^{-1}(\hat{\tau}_z \mid x,z)]dH(z) \equiv M_{n1}(x,a) + M_{n2}(x,a)$, say. Note that

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

where $\hat{R}(a;x^*,x,z) \equiv -\frac{g'\left(G^{-1}(\tau_z^*|x,z)|x,z\right)}{g(G^{-1}(\tau_z^*|x,z)|x,z)^3}(\hat{\tau}_z - \tau_z)^2$ and τ_z^* lies between $\hat{\tau}_z$ and τ_z. Noting that $\hat{\tau}_z - \tau_z$ is the first element of $\hat{\boldsymbol{\beta}}(y|u) - \boldsymbol{\beta}(y|u)$ with $u = (x', z')'$, we have that by Lemma C.1(b) and Assumption C.6, $\hat{R}(a;x^*,x,z) = O_P\left(n^{-1}b^{-d}\log n + b^{2(p+1)}\right) = o_P\left(n^{-1/2}b^{-d_X/2}\right)$ uniformly in $(a,x,z) \in \mathcal{A}_H \times \mathcal{X}_0 \times \mathcal{Z}_0$. It follows that for all $(a,x) \in \mathcal{A}_H \times \mathcal{X}_0$

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

where the second line follows from Lemma C.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

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

and

$$\begin{aligned}
M_{n12}(x,a) &= \sqrt{nb^{d_X}} \int \frac{e_{1,p}' \mathbb{S}_p^{-1} \mathbf{V}_{p,b}(a;x^*,z)}{g(x^*,z)\, g(G^{-1}(\tau_z|x,z)|x,z)} dH(z) \{1 + o_P(1)\} \\
&= \bar{M}_{n12}(x,a) \{1 + o_P(1)\} \xrightarrow{d} \mathbb{N}(0, V_1), \qquad \text{(B.2)}
\end{aligned}$$

where $\bar{M}_{n12}(x,a) \equiv \sqrt{\frac{b^{d_X}}{n}} \sum_{i=1}^n \int \frac{e_{1,p}' \mathbb{S}_p^{-1} K_b(X_i - x^*, Z_i - z) \mu_{p,b}(X_i - x^*, Z_i - z) \bar{\mathbf{1}}_a(W_i)}{g(x^*,z) g(G^{-1}(\tau_z|x,z)|x,z)} dH(z)$, (B.1) holds true uniformly in $(a,x) \in \mathcal{A}_H \times \mathcal{X}_0$, $V_1 \equiv \kappa_{1p} \int \frac{\tau_z(1-\tau_z)h(z)^2}{g(x^*,z) g(G^{-1}(\tau_z|x,z)|x,z)^2} dz$, $\kappa_{1p} \equiv \int e_{1,p}' \mathbb{S}_p^{-1} \mu_p(\tilde{x},\tilde{z})\, \mu_p(\tilde{x}, \tilde{z} - \bar{z})' \mathbb{S}_p^{-1} e_{1,p} \times K(\tilde{x},\tilde{z}) K(\tilde{x}, \tilde{z} - \bar{z})\, d(\tilde{x},\tilde{z},\bar{z})$, and (B.2) 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 C.2(c) we have that uniformly in $(a,x) \in \mathcal{A}_H \times \mathcal{X}_0$

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

where $S_p(\tau; u) \equiv \mathbb{S}_p g\left(G^{-1}(\tau|u)|u\right) g(u)$. Using Lemmas C.2 and C.3, we can show that uniformly in $(a, x) \in \mathcal{A}_H \times \mathcal{X}_0$

$$M_{n21}(x,a) = \sqrt{nb^{d_X}} b^{p+1} \int e_{1,p}' \mathbb{S}_p^{-1} \mathbb{B}_p \mathbf{G}_{p+1}^{-1}(\tau_z|x,z)\, dH(z) + o_P(1), \tag{B.3}$$

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

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

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 $\sqrt{nb^{d_X}}[\hat{m}_H(x,a) - m_H^*(x,a) - B_m(x,a;x^*)] \xrightarrow{d} \mathbb{N}\left(0, \sigma_m^2(x,a;x^*)\right)$, where $B_m(x,a;x^*)$ and $\sigma_m^2(x,a;x^*)$ are defined in (4.6) and (4.7), 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. ■

Proof of Theorem 4.3. 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

$$\delta = \delta(\epsilon) = \min\{m_H^*(x,a) - m_H^*(x,a-\epsilon),\ m_H^*(x,a+\epsilon) - m_H^*(x,a)\} > 0. \tag{B.4}$$

It follows that for sufficiently large n,

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

where the third line follows from the monotonicity of $m_H^*(x,\cdot)$, the fourth line holds by (B.4), 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 $\mathcal{A}_H^c \equiv \{a : |a - c_A| \le c$ for some $c_A \in \mathcal{A}_H\}$ and $c > 0$.

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 C.4(b) and Theorem 4.1,

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

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

Proof of Theorem 5.1. The proof is a special case of that of Theorem 5.3 and omitted. ■

Proof of Theorem 5.2. By Assumption A2

$$\begin{aligned}
G_n(y|x,z) &= P\left[m(X,A) + c_n\gamma(X,A) \leq y | X = x, Z = z\right] \\
&= P\left[m(x,A) \leq y - c_n\gamma(x,A)\,|Z = z\right] \\
&= E\left[\mathbf{1}\{m(x,A) \leq y\}\,|Z = z\right] + R_n(y;x,z) \\
&= E\left[\mathbf{1}\{m(X,A) \leq y\}\,|X = x, Z = z\right] + R_n(y;x,z) = G(y|x,z) + R_n(y;x,z),
\end{aligned}$$

where $R_n(y;x,z) = E\left\{\left[\mathbf{1}\{m(x,A) \leq y - c_n\gamma(x,A)\} - \mathbf{1}\{m(x,A) \leq y\}\right]|Z = z\right\}$. Let $\underline{\gamma}(x) \equiv \inf_{a\in\mathcal{A}}\gamma(x,a)$ and $\bar{\gamma}(x) \equiv \sup_{a\in\mathcal{A}}\gamma(x,a)$. By the monotonicity of $m(x,\cdot)$,

$$\begin{aligned}
R_n(y;x,z) &\leq E\left\{\left[\mathbf{1}\left\{m(x,A) \leq y - c_n\underline{\gamma}(x)\right\} - \mathbf{1}\{m(x,A) \leq y\}\right]|Z = z\right\} \\
&= E\left\{\left[\mathbf{1}\left\{A \leq m^{-1}\left(x, y - c_n\underline{\gamma}(x)\right)\right\} - \mathbf{1}\left\{A \leq m^{-1}(x,y)\right\}\right]|Z = z\right\} \\
&= F_{A|Z}\left(m^{-1}\left(x, y - c_n\underline{\gamma}(x)\right)|z\right) - F_{A|Z}\left(m^{-1}(x,y)\,|z\right) \\
&= -c_n\underline{\gamma}(x)\int_0^1 \varsigma\left(y - tc_n\underline{\gamma}(x); x, z\right)dt,
\end{aligned}$$

where the last equality holds by Taylor expansion of $F_{A|Z}\left(m^{-1}(x,\cdot)\,|z\right)$ around y with an integral remainder, $\varsigma(y;x,z) \equiv \partial F_{A|Z}\left(m^{-1}(x,y)\,|z\right)/\partial y = f_{A|Z}\left(m^{-1}(x,y)\,|z\right)\frac{1}{m_2(x,m^{-1}(x,y))}$ and $m_2(x,a) = \partial m(x,a)/\partial a$. Similarly, $R_n(y;x,z) \geq E\{[\mathbf{1}\{m(x,A) \leq y - c_n\bar{\gamma}(x)\} - \mathbf{1}\{m(x,A) \leq y\}]|Z = z\} = -c_n\bar{\gamma}(x)\int_0^1 \varsigma(y - tc_n\bar{\gamma}(x); x, z)dt$. It follows that there exists a continuous function $\gamma_n^\dagger(y;x,z)$ such that

$$\underline{\gamma}(x)\int_0^1 \varsigma\left(y - tc_n\underline{\gamma}(x); x, z\right)dt \leq \gamma_n^\dagger(y;x,z) \leq \bar{\gamma}(x)\int_0^1 \varsigma\left(y - tc_n\bar{\gamma}(x); x, z\right)dt,$$

and

$$G_n(y|x,z) = G(y|x,z) - c_n\gamma_n^\dagger(y;x,z). \tag{B.5}$$

Let $\bar{\gamma}_n^\dagger(x,z) = \sup_y \gamma_n^\dagger(y;x,z)$ and $\underline{\gamma}_n^\dagger(x,z) = \inf_y \gamma_n^\dagger(y;x,z)$. For any given $(x,z) \in \mathcal{X}\times\mathcal{Z}$ and $\tau \in (0,1)$, the inverse function of $G_n(\cdot|x,z)$ satisfies

$$\begin{aligned}
G_n^{-1}(\tau|x,z) &= \inf\left\{y : G(y|x,z) - c_n\gamma_n^\dagger(y;x,z) \geq \tau\right\} \\
&\geq \inf\left\{y : G(y|x,z) \geq \tau + c_n\underline{\gamma}_n^\dagger(x,z)\right\} = G^{-1}\left(\tau + c_n\underline{\gamma}_n^\dagger(x,z)\,|x,z\right);
\end{aligned}$$

and similarly $G_n^{-1}(\tau|x,z) \leq \inf\{y : G(y|x,z) \geq \tau + c_n\bar{\gamma}_n^{\dagger}(x,z)\} = G^{-1}\left(\tau + c_n\bar{\gamma}_n^{\dagger}(x,z)|x,z\right)$. So there exists a continuous function $\gamma_n^{\perp}(\tau;x,z)$ such that $\underline{\gamma}_n^{\dagger}(x,z) \leq \gamma_n^{\perp}(\tau;x,z) \leq \bar{\gamma}_n^{\dagger}(x,z)$ for all $\tau \in (0,1)$ and

$$\begin{aligned} G_n^{-1}(\tau|x,z) &= G^{-1}\left(\tau + c_n\gamma_n^{\perp}(\tau;x,z)|x,z\right) \\ &= G^{-1}(\tau|x,z) + c_n\gamma_n^{\perp}(\tau;x,z)\int_0^1 \frac{1}{g\left(G^{-1}\left(\tau + tc_n\gamma_n^{\perp}(\tau;x,z)|x,z\right)\ |x,z\right)}dt, \end{aligned} \quad \text{(B.6)}$$

where the second equality holds by Taylor expansion with an integral remainder (see the explanation after (5.5)).

Combining (B.5) with (B.6) and using the Taylor expansion, we have

$$\begin{aligned} G_n^{-1}(G_n(y|x,z)|x^*,z) &= G^{-1}\left(G_n(y|x,z)|x^*,z\right) + c_n\Theta_{1n}^{\dagger}(y;x,z) \\ &= G^{-1}\left(G(y|x,z) - c_n\gamma_n^{\dagger}(y;x,z)|x^*,z\right) + c_n\Theta_{1n}^{\dagger}(y;x,z) \\ &= G^{-1}\left(G(y|x,z)|x^*,z\right) + c_n\Theta_n^{\dagger}(y;x,z) \end{aligned}$$

where $\Theta_{1n}^{\dagger}(y;x,z) = \gamma_n^{\perp}(G_n(y|x,z);x^*,z)\int_0^1 \frac{1}{g(G^{-1}(G_n(y|x,z)+tc_n\gamma_n^{\perp}(G_n(y|x,z);x^*,z)|x^*,z)\ |x^*,z)}dt$, and

$$\begin{aligned} \Theta_n^{\dagger}(y;x,z) &= \Theta_{1n}^{\dagger}(y;x,z) - \gamma_n^{\dagger}(y;x,z)\int_0^1 \frac{1}{g\left(G^{-1}\left(G(y|x,z) - tc_n\gamma_n^{\dagger}(y;x,z)|x^*,z\right)|x^*,z\right)}dt \\ &= \left[\gamma_n^{\perp}(G_n(y|x,z);x^*,z) - \gamma_n^{\dagger}(y;x,z)\right]\frac{1}{g\left(G^{-1}\left(G(y|x,z)|x^*,z\right)|x^*,z\right)} + o(1). \blacksquare \end{aligned} \quad \text{(B.7)}$$

Proof of Theorem 5.4. The proof is much simpler than that of Theorem 5.3, so we only sketch the main steps. Under $\mathbb{H}_1$, we can apply Lemmas C.2 and C.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 $c_{H_1} > 0$ in probability; the result follows. ■

Proof of Proposition 5.5. Define the following empirical process

$$\begin{aligned} \mathcal{V}_n(\tau,u,U(r);\boldsymbol{\theta}) &= n^{-1/2}b^{-d/2}\sum_{i=1}^n \psi_\tau\left(Y_i^*(r) - n^{-1/2}b^{-d/2}\boldsymbol{\theta}'\mu_{p,b,i}(r;u)\right)\mu_{p,b,i}(r;u)\check{K}_{b,i}(r) \\ &= n^{-1/2}b^{-d/2}\sum_{i=1}^n \psi_\tau(Y_i^*(r),u,\boldsymbol{\theta})\mu_{p,b,i}(r;u)\check{K}_{b,i}(r), \end{aligned}$$

where $\psi_\tau(Y_i^*(r),u,\boldsymbol{\theta}) \equiv \psi_\tau\left(Y_i^*(r) - n^{-1/2}b^{-d/2}\boldsymbol{\theta}'\mu_{p,b,i}(r;u)\right)$. Then $\mathcal{V}_n(\tau,u,U(r);0) = n^{-1/2}b^{-d/2} \times \sum_{i=1}^n \psi_\tau(Y_i^*(r))\mu_{p,b,i}(r;u)\check{K}_{b,i}(r)$. Let $\mathcal{V}_n(\tau,u,U(\hat{r});\boldsymbol{\theta})$ and $\mathcal{V}_n(\tau,u,U(\hat{r});0)$ denote the corresponding terms with generated regressors.

In Lemma C.7, we show that $\mathcal{V}_n(\tau,u,U(\hat{r});\boldsymbol{\theta}) - \mathcal{V}_n(\tau,u,U(\hat{r});0) - [\mathcal{V}_n(\tau,u,U(r_0);\boldsymbol{\theta}) - \mathcal{V}_n(\tau,u,U(r_0);0)] = o_P(b^{d_Z/2})$ uniformly over $(\tau,u) \in \mathcal{T}\times\mathcal{U}_0$ and $\|\boldsymbol{\theta}\| \leq L$ for any fixed positive constant L. With this in hand, we have by Lemma C.8,

$$\tilde{\boldsymbol{\theta}} = -\, S_{p,b}\left(\tau;u\right)^{-1}\mathcal{V}_n\left(\tau,u,U\left(\hat{r}\right);0\right)+o_P(b^{d_Z/2})$$

uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$. Then uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$,

$$\begin{aligned}
\tilde{\boldsymbol{\theta}} &= -S_{p,b}\left(\tau;u\right)^{-1}\{\mathcal{V}_n\left(\tau,u,U\left(r_0\right);0\right)+E_n\left[\mathcal{V}_n\left(\tau,u,U\left(\hat{r}\right);0\right)-\mathcal{V}_n\left(\tau,u,U\left(r_0\right);0\right)\right]\\
&\quad+\{\mathcal{V}_n\left(\tau,u,U\left(\hat{r}\right);0\right)-\mathcal{V}_n\left(\tau,u,U\left(r_0\right);0\right)-E_n\left[\mathcal{V}_n\left(\tau,u,U\left(\hat{r}\right);0\right)-\mathcal{V}_n\left(\tau,u,U\left(r_0\right);0\right)\right]\}\}\\
&\quad+o_P(b^{d_Z/2})\\
&= -S_{p,b}\left(\tau;u\right)^{-1}n^{-1/2}b^{d/2}\sum_{i=1}^{n}\psi_\tau\left(Y_i^*\left(r_0\right)\right)\mu_{p,b,i}\left(r_0;u\right)b^{-d}\check{K}_{b,i}\left(r_0\right)\\
&\quad-\left\{S_{p,b}\left(\tau;u\right)^{-1}n^{-1/2}b^{d/2}\sum_{i=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_i\right)\right)\Psi_{n1}\left(\tau,u,\varpi_i\right)Z_i\right\}\{1+o_P\left(1\right)\}+o_P(b^{d_Z/2}),
\end{aligned}$$

where the second equality holds by Lemmas C.9 and C.10. $E_n\left(\cdot\right)$ is the expectation in empirical processes and it is the same as the usual $E\left(\cdot\right)$ except that inside $E_n\left(\cdot\right)$ $\hat{r}$ is taken as a constant function. We provide a specific example in Appendix F.1 to help understand $E_n\left(\cdot\right)$.

In view of the fact that the influence term of $\sqrt{nb^d}\left[\hat{G}^{-1}\left(\tau|u\right)-G^{-1}\left(\tau|u\right)\right]$ is given by the first component in the last expression premultiplied by $e_{1,p}'$, and $e_{1,p}'\tilde{\boldsymbol{\theta}}=\sqrt{nb^d}\left[\tilde{G}_{p,b}^{-1}\left(\tau|u\right)-G^{-1}\left(\tau|u\right)\right]$, we have

$$\begin{aligned}
&\tilde{G}_{p,b}^{-1}\left(\tau|u\right)-\hat{G}_{p,b}^{-1}\left(\tau|u\right)\\
&= -\left\{e_{1,p}'S_{p,b}\left(\tau;u\right)^{-1}n^{-1}\sum_{i=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_i\right)\right)\Psi_{n1}\left(\tau,u,\varpi_i\right)Z_i\right\}\{1+o_P\left(1\right)\}+o_P(n^{-1/2}b^{-d_X/2})\\
&= -\left\{e_{1,p}'\bar{S}_{p,b}\left(\tau;u\right)^{-1}n^{-1}\sum_{i=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_i\right)\right)\Psi_{n1}\left(\tau,u,\varpi_i\right)Z_i\right\}\{1+o_P\left(1\right)\}+o_P(n^{-1/2}b^{-d_X/2})
\end{aligned}$$

uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$, where the last equality holds by the fact that $S_{p,b}\left(\tau;u\right)=\bar{S}_{p,b}\left(\tau;u\right)+O_P(n^{-1/2}b^{-d/2}\ (\log n)^{1/2})$ and $n^{-1}\sum_{i=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_i\right)\right)\Psi_{n1}\left(\tau,u,\varpi_i\right)Z_i=O_P(n^{-1/2}b^{-d_Z/2}(\log n)^{1/2})$ uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$. ■

Proof of Proposition 5.6. This is a direct result from MRS (Theorem 1). We discuss in Appendix F.1 how those conditions for Theorem 1 in MRS are met here. The remainder term for the linear expansion is $o_P\left((nb^{d_X})^{-1/2}\right)$, as discussed in Appendix F.1. Then by Theorem 1 in MRS, we have $\tilde{G}_{p,b}\left(y|u\right)-\hat{G}_{p,b}\left(y|u\right)=\frac{\partial G(y|u)}{\partial u'}\hat{\Lambda}\left(u\right)+o_P\left((nb^{d_X})^{-1/2}\right)$ uniformly over $u\in\mathcal{U}_0$, where

$$\hat{\Lambda}\left(u\right)\equiv e_{1,p}'\overline{\mathbf{S}}_{p,b}\left(u\right)^{-1}E_n\left\{b^{-d}\check{K}_{b,i}\left(r_0\right)\mu_{p,b,i}\left(r_0;u\right)\left[U_i\left(\hat{r}\right)-U_i\left(r_0\right)\right]\right\}. \tag{B.8}$$

The uniform convergence over $y\in\mathbb{R}$ similarly follows from Boente and Fraiman (1991). The conclusion then follows from Lemma C.12 and the finiteness of $\frac{\partial G(y|u)}{\partial u}$.■

C Some technical lemmas

Lemma C.1 *Suppose that Assumptions C.1-C.4 and C.6 hold. Let $\boldsymbol{\beta}\left(y|u\right)$ be a vector that stacks $\frac{1}{\mathbf{j}!}D^{\mathbf{j}}G\left(y|u\right)$, $0\le|\mathbf{j}|\le p$, in lexicographic order. 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) $\hat{\boldsymbol\beta}(y|u) - \boldsymbol\beta(y|u) = O_P(\nu_b + b^{p+1})$.

Lemma C.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) $\hat{G}^{-1}_{p,b}(\tau|u) - G^{-1}(\tau|u) = e'_{1,p}\bar{S}_{p,b}(\tau;u)^{-1}\bar{V}_{p,b}(\tau;u) + O_P(\nu_b^2 + \nu_b b^{p+1}) + o_P\left(n^{-1/2}b^{-d_X/2}\right)$,
(b) $\hat{G}^{-1}_{p,b}(\tau|u) - G^{-1}(\tau|u) = e'_{1,p}\bar{S}_{p,b}(\tau;u)^{-1}V_{p,b}(\tau;u) + b^{p+1}e'_{1,p}\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}^{-1}_{p+1}(\tau|u) + O_P(\nu_b^2) + o_P(b^{p+1} + n^{-1/2}b^{-d_X/2})$,
(c) $\hat{G}^{-1}_{p,b}(\tau|u) - G^{-1}(\tau|u) = e'_{1,p}S_p(\tau;u)^{-1}V_{p,b}(\tau;u)[1 + o_P(1)] + b^{p+1}e'_{1,p}\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}^{-1}_{p+1}(\tau|u) + o_P(b^{p+1} + n^{-1/2}b^{-d_X/2})$,
where $S_p(\tau;u) \equiv \mathbb{S}_p g\left(G^{-1}(\tau|u)|u\right)g(u)$ *is the limit of* $\bar{S}_{p,b}(\tau;u) \equiv E[S_{p,b}(\tau;u)]$.

Lemma C.3 *Suppose that Assumptions C.1-C.4 and C.6 hold. Then*
$\sup_{\tilde\tau,\tau\in\mathcal{T}_0^\epsilon, |\tilde\tau-\tau|\le M\nu_b}\sup_{u\in\mathcal{U}_0}\sqrt{nb^{d_X}}\,\|V_{p,b}(\tilde\tau;u) - V_{p,b}(\tau;u)\| = o_P(1)$.

By Lemmas C.1 and C.3, with probability approaching 1 we have $\sup_{(a,x,z)\in\mathcal{A}_H\times\mathcal{X}_0\times\mathcal{Z}_0}\sqrt{nb^{d_X}}|V_{p,b}(\hat{G}_{p,b}(a|x^*,z);x,z) - V_{p,b}(G(a|x^*,z);x,z)| = o_P(1)$.

Lemma C.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) $\hat{m}_H(x, a+\delta_n) - \hat{m}_H(x,a) = \lambda_H^*(x,a)\delta_n + o_P\left(\delta_n + n^{-1/2}b^{-d_X/2}\right)$,
where $\lambda_H^*(x,a) \equiv \int \frac{g(a|x^*,z)}{g(G^{-1}(G(a|x^*,z)|x,z)|x,z)}dH(z)$.

Lemma C.5 *Suppose the conditions in Theorem 5.3 hold. Then* $\hat{J}_{n1} = \bar{J}_{n1} + o_P(1)$, *where* $\bar{J}_{n1} = b^{d_X}\sum_{i=1}^n [\int e'_{1,p}\bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1}\, V_{p,b}(\tau_{iz};\,x^*,z)d\Delta(z)]^2\pi_i$. *We assume the assumptions in Proposition 5.5, 5.6 hold for Lemma C.6–C.12.*

We assume that the conditions in Proposition 5.5-5.6 hold for Lemmas C.6−C.12 below.

Lemma C.6 $P(\hat{r} \in C_M^{\varrho}(\Omega_{d_\varpi})) \to 1$.

Lemma C.7 $\mathcal{V}_n(\tau,u,U(\hat r);\boldsymbol\theta) - \mathcal{V}_n(\tau,u,U(\hat r);0) - [\mathcal{V}_n(\tau,u,U(r_0);\boldsymbol\theta) - \mathcal{V}_n(\tau,u,U(r_0);0)] = o_P(b^{d_Z/2})$ *uniformly over* $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ *and* $\|\boldsymbol\theta\| \le L$.

Lemma C.8 $\tilde{\boldsymbol\theta} = S_{p,b}(\tau;u)^{-1}\mathcal{V}_n(\tau,u,U(\hat r);0) + o_P(b^{d_Z/2})$ *uniformly over* $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$.

Lemma C.9 $\|\mathcal{V}_n(\tau,u,U(\hat r);\boldsymbol\theta) - \mathcal{V}_n(\tau,u,U(r_0);\boldsymbol\theta) - E_n[\mathcal{V}_n(\tau,u,U(\hat r);\boldsymbol\theta) - \mathcal{V}_n(\tau,u,U(r_0);\boldsymbol\theta)]\| = o_P(b^{d_Z/2})$ *uniformly over* $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ *and* $\|\boldsymbol\theta\|\le L$.

Lemma C.10 $E_n[\mathcal{V}_n(\tau,u,U(\hat r);0) - \mathcal{V}_n(\tau,u,U(r_0);0)] = \left\{n^{-1/2}b^{d/2}\sum_{i=1}^n b^{-d_Z}\overline{\mu\breve{K}}_{p,b}(z,r_0(\varpi_i))\Psi_{n1}(\tau,u,\varpi_i)Z_i\right\} \times\{1+o_P(1)\} + o_P(b^{d_Z/2})$ *uniformly over* $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$.

Lemma C.11 $E_n[\mathcal{V}_n(\tau,u,U(\hat r);\boldsymbol\theta) - \mathcal{V}_n(\tau,u,U(r_0);\boldsymbol\theta)] - E_n[\mathcal{V}_n(\tau,u,U(\hat r);0) - \mathcal{V}_n(\tau,u,U(r_0);0)] = o_P(b^{d_Z/2})$ *uniformly over* $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ *and* $\|\boldsymbol\theta\|\le L$.

Lemma C.12 $\hat\Lambda(u) = -\left\{n^{-1}\sum_{i=1}^n \begin{pmatrix}\mathbf{0}_{d_X}\\ Z_i\end{pmatrix} e'_{1,p}\overline{\mathbf{S}}_{p,b}(u)^{-1}b^{-d_Z}\overline{\mu\breve{K}}_{p,b}(z,r_0(\varpi_i))\Psi_{n2}(x,\varpi_i)\right\}\times\{1+o_P(1)\} + o_P(n^{-1/2}b^{-d_X/2})$ *uniformly over* $u\in\mathcal{U}_0$.

REFERENCES

Ai, C., Chen, X., 2003. Efficient estimation of models with conditional moment restrictions containing unknown functions. Econometrica 71, 1795-1843.

Altonji, J.G., Matzkin, R.L., 2005. Cross section and panel data estimators for nonseparable models with endogenous regressors. Econometrica 73, 1053-1102.

Andrews, D.W.K., Ploberger, W., 1994. Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62, 1383-1414.

Birke, M., Dette, H., 2007. Testing strict monotonicity in nonparametric regression. Mathematical Methods of Statistics 16, 110-123.

Blundell, R., Chen, X., Kristensen, D., 2007. Semi-nonparametric IV estimation of shape invariant Engel curves. Econometrica 75, 1613-1669.

Blundell, R., Kristensen, D., Matzkin, R., 2014. Bounding quantile demand functions using revealed preference inequalities. Journal of Econometrics 179, 112-127

Blundell, R., Horowitz, J.L., Parey, M., 2012. Measuring the price responsiveness of gasoline demand: economic shape restrictions and nonparametric demand estimation. Quantitative Economics 3, 39-51.

Boente, G., Fraiman, R., 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., Heckman, J., Masterov, D., 2005. Labor market discrimination and racial differences in premarket factors. Journal of Law and Economics 48, 1-39.

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

Chiappori, P-A., Komunjer, I., Kristensen, D., 2015. Nonparametric identification and estimation of transformation models. Journal of Econometrics 188, 22-39.

Delgado, M., Escanciano, J. C., 2012. Distribution-free tests of stochastic monotonicity. Journal of Econometrics 170, 68-75.

Dette, H., Hoderlein, S., Neumayer, N., 2015. Testing multivariate economic restrictions using quantiles: the example of Slutsky negative semidefiniteness. Cemmap working paper.

d'Haultfoeuille, X., Février, P., 2012. Identification of nonseparable models with endogeneity and discrete instruments. Working paper, CREST.

Fan, J., Hu, T., Truong, Y., 1994. Robust non-parametric function estimation. Scandinavian Journal of Statistics 21, 433-446.

Ghanem, D., 2014. Average partial effects in nonseparable panel data models: identification and testing. Working Paper, UC Davis.

Giacomini, R., Politis, D. N., White, H., 2013. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. Econometric Theory 29, 567-589.

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

Heckman, J.J., Vytlacil, E., 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., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73, 669-738.

Hoderlein, S., 2011. How many consumers are rational? Journal of Econometrics 164(2), 294-309.

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

Hoderlein, S., Mammen, E., 2009. Identification and estimation of local average derivatives in non-separable models without monotonicity. Econometrics Journal 12, 1-25.

Hoderlein, S., Stoye, J., 2014. Revealed preferences in a heterogeneous population. Review of Economics and Statistics 96, 197-213.

Hoderlein, S., Holzmann, H., Meister, A., 2015. The triangular model with random coefficients. Cemmep WP.

Huber, M. , Mellace, G., 2014. Relaxing monotonicity in the identification of local average treatment effects, Working paper, St. Gallen.

Huang, M., Sun, Y., White, H., 2013. A flexible nonparametric test for conditional independence. Discussion Paper, Dept. of Economics, UCSD.

Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62, 467-475.

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

Kasy, M., 2011. Identification in triangular systems using control functions. Econometric Theory 27.

Kitagawa, T., 2013. Bootstrap test for instrument validity in the heterogeneous treatment effect models, Cemmap Working papers.

Lewbel, A., 1999. Consumer demand systems and household expenditure. H. Pesaran and M. Wickens (Eds.), Handbook of Applied Econometrics. Oxford: Blackwell.

Lewbel, A., Lu, X., Su, L., 2015. Specification testing for transformation models with applications to generalized accelerated failure-time models. Journal of Econometrics 184, 81-96.

Li, Q., Racine, J.S., 2004. Cross-validated local linear nonparametric regression. Statistica Sinica 14, 485-512.

Li, Q., Racine, J.S., 2008. Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. Journal of Business and Economic Statistics 26, 423-434.

Lu, X., White, H., 2014. Testing for separability in structural equations. Journal of Econometrics 182, 14-26.

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

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

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

Newey, W. K., Powell, J. L., Vella, F., 1999. Nonparametric estimation of triangular simultaneous equations models. Econometrica, 67, 3, 565-603.

Neal, D., Johnson, W., 1996. The role of premarket factors in black-white wage differences. Journal of Political Economy 104, 869-895.

Olley, G.S., Pakes, A., 1996. The dynamics of productivity in the telecommunications equipment industry. Econometrica 64, 1263-1297.

Pollard, D., 1991. Asymptotics for the least absolute deviation regression estimators. Econometric Theory 7, 186-199.

Racine, J.S., MacKinnon, J., 2007. Simulation-based tests that can use any number of simulations. Communications in Statistics: Simulation and Computation 36, 357-365.

Roehrig, C. S., 1988. Conditions for identification in nonparametric and parametric models. Econometrica 56, 433-447.

Su, L., Hoderlein, S., White, H., 2014. Testing monotonicity in unobservables with panel data. Discussion Paper, Dept. of Economics, Boston College.

Su, L., Tu, Y., Ullah, A., 2015. Testing additive separability of error term in nonparametric structural models. Econometric Reviews 34, 1056-1087.

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

Su, L., White, H., 2012. Local polynomial quantile regression for dependent heterogeneous processes: uniform Bahadur representation with applications to testing conditional independence. Advances in Econometrics 29, 355-434.

Su, L., White, H., 2014. Testing conditional independence via empirical likelihood. Journal of Econometrics 182, 27-44.

Torgovitsky, A., 2011. Identification of nonseparable models with general instruments. Working paper, Northwestern University.

Tripathi, G., Kitamura, Y., 2003. Testing conditional moment restrictions. Annals of Statistics 31, 2059-2095.

White, H., Chalak, K., 2010. Testing a conditional form of exogeneity. Economics Letters 109, 88-90.

White, H., Lu, X., 2011. Causal diagrams for treatment effect estimation with application to efficient covariate selection. Review of Economics & Statistics 93, 1453-1459.

Wooldridge, J. M., 2010. Econometric Analysis of Cross Section and Panel Data. Second edition, MIT Press, London.

van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes with Applications to Statistics. Springer-Verlag, New York.

Supplementary Appendix to
"Testing for Monotonicity in Unobservables under Unconfoundedness"
(Not for publication)

Stefan Hoderlein[a], Liangjun Su[b], Halbert White[c], and Thomas Tao Yang[a]

[a] Department of Economics, Boston College
[b] School of Economics, Singapore Management University
[c] Department of Economics, University of California, San Diego

In this appendix, we first prove the technical lemmas in Appendix C, and then provide proofs of some claims in the main text. Finally, we make further discussion on the local power property of the proposed test in the above paper.

D Proofs of the technical lemmas

Proof of Lemma C.1. Since $[\mathbf{S}_{p,b}(u)]^{-1}\mathbf{S}_{p,b}(u) = I_{N_p}$ where I_{N_p} is an $N_p \times N_p$ identity matrix, by (4.2) we obtain the following standard bias and variance decomposition:

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

By Theorems 2 and 4 in Masry (1996) with some modification to account for the non-compact support of the kernel function,[8]

$$\mathbf{S}_{p,b}(u) = \overline{\mathbf{S}}_{p,b}(u) + O_P(\nu_b), \mathbf{V}_{p,b}(y;u) = O_P(\nu_b), \ \mathbf{B}_{p,b}(y;u) - \overline{\mathbf{B}}_{p,b}(y;u) = O_P(\nu_p b^{p+1}), \tag{D.2}$$

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 (D.2) also hold uniformly in $y \in \mathbb{R}$ under Assumption C.3. In addition, by the Slutsky lemma,

$$\mathbf{S}_{p,b}(u)^{-1} = \left\{\overline{\mathbf{S}}_{p,b}(u) + \left[\mathbf{S}_{p,b}(u) - \overline{\mathbf{S}}_{p,b}(u)\right]\right\}^{-1} = [\overline{\mathbf{S}}_{p,b}(u)]^{-1} + O_P(\nu_b). \tag{D.3}$$

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

Recall that $\hat{G}_{p,b}(y|x,z) = e_{1,p}'\hat{\boldsymbol{\beta}}(y|u)$ where $e_{1,p}$ is defined after (4.3). 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 $\hat{G}_{p,b}(y|u) - G(y|u) = b^{p+1}e_{1,p}'\mathbb{S}_p^{-1}\mathbb{B}_p\mathbf{G}_{p+1}(y|u) + g(u)^{-1}e_{1,p}'\mathbb{S}_p^{-1}\mathbf{V}_{p,b}(y;u) + O_P(b\bar{\nu}_b)$.

Proof of Lemma C.2. 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 (2012). To prove (b), write $\bar{V}_{p,b}(\tau;u) = V_{p,b}(\tau;u) + R_{p,b}(\tau;u)$, where $R_{p,b}(\tau;u) \equiv \frac{1}{n}\sum_{i=1}^n\{\mathbf{1}\left(Y_i \le G^{-1}(\tau|U_i)\right) - \mathbf{1}(Y_i \le \boldsymbol{\beta}_b(\tau;u)'\mu_{iu})\}K_b(U_i-u)\mu_{iu}$ 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

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

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

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 C.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 C.2 holds uniformly in $(\tau, u) \in \mathcal{T}_0^\epsilon \times \mathcal{U}_0$.

Proof of Lemma C.3. 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

$$\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon, |\tilde{\tau}-\tau|\le Mv_b} \sup_{u\in\mathcal{U}_0} |W(\tilde{\tau}, \tau; u)| = o_P(\vartheta_n) \text{ with } \vartheta_n = n^{-1/2}b^{-d_X/2}. \tag{D.4}$$

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 $W(\tilde{\tau}, \tau; u) = (nb^d)^{-1}\sum_{i=1}^n a_{i,u}\left[\tilde{\tau} - 1\{Y_i \le G^{-1}(\tilde{\tau}|U_i)\} - \tau + 1\{Y_i \le G^{-1}(\tau|U_i)\}\right]$, 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, (D.4) will hold if $\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\le Mv_b} \sup_{u\in\mathcal{U}_0} |W^+(\tilde{\tau}, \tau; u)| = o_P(\vartheta_n)$ and $\sup_{\tilde{\tau},\tau\in\mathcal{T}_0^\epsilon,|\tilde{\tau}-\tau|\le Mv_b} \sup_{u\in\mathcal{U}_0} |W^-(\tilde{\tau}, \tau; u)| = o_P(\vartheta_n)$. We will only show the first part as the other case is similar.

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

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

Furthermore,

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

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

$$\begin{aligned}
P(W_{n11} > \vartheta_n\epsilon_0) &\le C_1 n_1 n_2 \nu_b n^{1/2} \max_{1\le l\le n_2} \max_{1\le j,k\le n_1:|\tau_j-\tau_k|\le M\nu_b} P\left(W^+(\tau_j, \tau_k; u_l) > \vartheta_n\epsilon_0\right) \\
&\le 2C_1 n_1 n_2 \nu_b n^{1/2} \exp\left(-\frac{n^2 b^{2d}\vartheta_n^2\epsilon_0^2}{2C_2 nb^d\nu_b + \frac{2}{3}C_3 nb^d\vartheta_n\epsilon_0}\right) \\
&= O\left(n_1 n_2 \nu_b n^{1/2}\right) \exp\left(-\frac{b^{d_Z}\epsilon_0^2}{C_4\left(n^{-1/2}b^{-d/2}\sqrt{\log n} + n^{-1/2}b^{-d_X/2}\epsilon_0\right)}\right) = o(1),
\end{aligned}$$

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

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

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

$$|K(u_2)-K(u_1)|\le\delta K^{*}(u_1), \tag{D.5}$$

where $K^{*}(u)=\overline{C}\mathbf{1}(\|u\|\le 2dc_K)$ for some constant $\overline{C}$ that depends on $\overline{c}_1$ and $\overline{c}_2$ in the assumption. For any $u\in\mathcal{U}_l$, $\|u-u_l\|/b\le e_n$. It follows from (D.5) that $|K_{iu}-K_{iu_l}|\le e_nK_{iu_l}^{*}$ where $K_{iu}\equiv K((U_i-u_l)/b)$ and $K_{iu_l}^{*}\equiv K^{*}((U_i-u_l)/b)$, and

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

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

$$\left|a_{i,u}^{+}-a_{i,u_l}^{+}\right|=\left|K_{iu}\omega'\mu_{p,b}(U_i-u)-K_{iu_l}\omega'\mu_{p,b}(U_i-u_l)\right|\le Ce_n(K_{iu_l}^{*}+K_{iu_l}).$$

It follows that

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

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

Proof of Lemma C.4. By Lemma C.1,

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

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 C.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

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

For D_{n1}, we have

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

where the first equality follows from the Taylor expansion, the second from (a), and the third from Lemma C.1. By the proof of Theorem 4.1, we have

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

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 $\mathbf{G}^{-1}_{p+1}$. Next, we write $D_{n22} = D_{n22,1} + D_{n22,2}$, where

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

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 Lemma C.5. To prove the result, we define $\tilde{J}_{n1}$ analogously as $\bar{J}_{n1}$ with τ_{iz} replaced by $\hat{\tau}_{iz}$: $\tilde{J}_{n1} = b^{d_X}\sum_{i=1}^n \left[\int e'_{1,p}\bar{S}_{p,b}(\hat{\tau}_{iz};x^*,z)^{-1}V_{p,b}(\hat{\tau}_{iz};x^*,z)\, d\Delta(z)\right]^2 \pi_i$. 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 $d_{n1}(\tau; u) = e'_{1,p}\bar{S}_{p,b}(\tau; x, z)^{-1}[\bar{V}_{p,b}(\tau; x, z) - V_{p,b}(\tau; x, z)]$. Then we have $\hat{J}_{n1} - \tilde{J}_{n1} = D_{n1} + 2D_{n2}$, where

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

As $d_{n1}(\tau; u) = O_P(b^{p+1})$ 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 C.1 and C.3, where $\bar{D}_{n2} = b^{d_X}\sum_{i=1}^{n}\int d_{n1}(\tau_{iz}; x^*, z)\, d\Delta(z)\int e'_{1,p} \times \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz}; x^*, z)\, d\Delta(z)\, \pi_i$. 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 $d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) = e'_{1,p}\bar{S}_{p,b}(\hat{\tau}_{iz}; x^*, z)^{-1} V_{p,b}(\hat{\tau}_{iz}; x^*, z) - e'_{1,p}\bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} \times V_{p,b}(\tau_{iz}; x^*, z)$, $\bar{d}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) = e'_{1,p}\bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1}[V_{p,b}(\hat{\tau}_{iz}; x^*, z) - V_{p,b}(\tau_{iz}; x^*, z)]$, and $\overline{R}_{n2} = d_{n2} - \bar{d}_{n2}$. Then uniformly in $z \in \mathcal{Z}_0$ and conditional on $\hat{\tau}_{iz} \in \mathcal{T}_0^{\epsilon}$,

$$\begin{aligned} \overline{R}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) &= e'_{1,p}\left[\bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} - \bar{S}_{p,b}(\hat{\tau}_{iz}; x^*, z)^{-1}\right] V_{p,b}(\hat{\tau}_{iz}; x^*, z) \qquad \text{(D.6)} \\ &= O_P(\hat{\tau}_{iz} - \tau_{iz})\, O_P(v_b) = O_P\left(v_b(v_b + b^{p+1})\right). \end{aligned}$$

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

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

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

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

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 C.3

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

By (D.6), $\tilde{J}_{n11,b} = nb^{d/2}O_P((n^{-1}b^{-d}\log n + b^{2(p+1)})\, n^{-1}b^{-d}\log n) = O_P(n^{-1}b^{-3d/2}(\log n)^2 + b^{2(p+1)-d/2}\log n) = o_P(1)$. By Cauchy-Schwarz (CS hereafter) 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 $\tilde{J}_{n12,a} = b^{d_X}\sum_{i=1}^{n}\int \bar{d}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)\, d\Delta(z) \int e_{1,p}'\bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz};\, x^*, z) d\Delta(z)\,\pi_i$. By the CS inequality and Lemmas C.1 and C.3, $\tilde{J}_{n12,a} = o_P(1)$. Thus $\tilde{J}_{n12} = o_P(1)$. This completes the proof of (ii). ■

Proof of Lemma C.6. The derivation follows from similar arguments as used in Escanciano et al. (2014, Appendix C). ■

Proof of Lemma C.7. To show the conclusion, we note that by Lemma C.9

$$\mathcal{V}_n(\tau, u, U(\hat{r}); \boldsymbol{\theta}) - \mathcal{V}_n(\tau, u, U(r_0); \boldsymbol{\theta}) - E_n[\mathcal{V}_n(\tau, u, U(\hat{r}); \boldsymbol{\theta}) - \mathcal{V}_n(\tau, u, U(r_0); \boldsymbol{\theta})] = o_P(b^{d_Z/2}) \quad \text{(D.7)}$$
$$\mathcal{V}_n(\tau, u, U(\hat{r}); 0) - \mathcal{V}_n(\tau, u, U(r_0); 0) - E_n[\mathcal{V}_n(\tau, u, U(\hat{r}); 0) - \mathcal{V}_n(\tau, u, U(r_0); 0)] = o_P(b^{d_Z/2}) \quad \text{(D.8)}$$

uniformly over $(\tau, u) \in \mathcal{T} \times \mathcal{U}_0$ and $\|\boldsymbol{\theta}\| \le L$. By Lemma C.11 and (D.7)-(D.8), we have

$$\mathcal{V}_n(\tau, u, U(\hat{r}); \boldsymbol{\theta}) - \mathcal{V}_n(\tau, u, U(\hat{r}); 0) - [\mathcal{V}_n(\tau, u, U(r_0); \boldsymbol{\theta}) - \mathcal{V}_n(\tau, u, U(r_0); 0)] = o_P(b^{d_Z/2}).$$

uniformly over $(\tau, u) \in \mathcal{T} \times \mathcal{U}_0$ and $\|\boldsymbol{\theta}\| \le L$. ■

Proof of Lemma C.8. The proof is similar to that of Theorem 1 in Su and White (2012, SW hereafter) which is built on their Lemmas A.1-A.5. Because we assume IID data, the mixing conditions in that theorem are automatically satisfied. The bandwidth condition (Assumption A.6 in SW) is essentially our Assumption C.6. The other conditions of SW can also be verified, and we can set $\kappa_n = b^{d_Z/2}$ in SW. The object of interest in SW corresponds to our $\mathcal{V}_n(\tau, u, U(r_0); \boldsymbol{\theta})$. We now argue that a similar result holds for $\mathcal{V}_n(\tau, u, U(\hat{r}); \boldsymbol{\theta})$.

Lemma A.2 in SW shows that $\mathcal{V}_n(\tau, u, U(r_0); 0) = O_P(1)$ uniformly over $(\tau, u) \in \mathcal{T} \times \mathcal{U}_0$ in our notation. With this, by equation (D.8) and Lemma C.10, we know $\mathcal{V}_n(\tau, u, U(\hat{r}); 0) = O_P(1)$ uniformly over $(\tau, u) \in \mathcal{T} \times \mathcal{U}_0$. By Lemmas A.3-A.4 with $\kappa_n = b^{d_Z/2}$ in SW and Lemma C.7 above, we have uniformly over $(\tau, u) \in \mathcal{T} \times \mathcal{U}_0$ and $\|\boldsymbol{\theta}\| \le L$,

$$\|\mathcal{V}_n(\tau, u, U(\hat{r}); \boldsymbol{\theta}) - \mathcal{V}_n(\tau, u, U(\hat{r}); 0) + S_{p,b}(\tau; u)\,\boldsymbol{\theta}\| = o_P(b^{d_Z/2}). \quad \text{(D.9)}$$

By the same argument as used in the proof of Lemma A.5 in SW and the fact that $\tilde{\boldsymbol{\theta}}$ is the solution to problem (5.15), we have $\left\|\mathcal{V}_n(\tau, u, U(\hat{r}); \tilde{\boldsymbol{\theta}})\right\| = o_P\left(b^{d_Z/2}\right)$ uniformly over $(\tau, u) \in \mathcal{T} \times \mathcal{U}_0$. Note that $\boldsymbol{\theta}'\mathcal{V}_n(\tau, u, U(\hat{r}); \lambda\boldsymbol{\theta})$ is non-decreasing in λ for any $\boldsymbol{\theta}$ because ψ_τ is a non-decreasing function and

$$\boldsymbol{\theta}'\mathcal{V}_n(\tau, U(\hat{r}); \lambda\boldsymbol{\theta}) = n^{-1/2}b^{-d/2}\sum_{i=1}^{n}\psi_\tau\left(Y_i^*(r) - n^{-1/2}b^{-d/2}\lambda\boldsymbol{\theta}'\mu_{p,b,i}(r; u)\right)\boldsymbol{\theta}'\mu_{p,b,i}(r; u)\,\check{K}_{b,i}(r).$$

$S_{p,b}(\tau; u)$ is positive definite as n tends to infinity. These results, in conjunction with Lemma C.7, imply that all conditions in Lemma A.1 in SW are verified, and by (D.9) we have

$$\begin{aligned}
\tilde{\boldsymbol{\theta}} &= S_{p,b}(\tau; u)^{-1}\mathcal{V}_n(\tau, u, U(\hat{r}); 0) + o_P(b^{d_Z/2}), \\
&= S_{p,b}(\tau; u)^{-1} n^{-1/2}b^{-d/2}\sum_{i=1}^{n}\psi_\tau(Y_i^*(\hat{r}))\,\mu_{p,b,i}(\hat{r}; u)\,\check{K}_{b,i}(\hat{r}) + o_P(b^{d_Z/2}),
\end{aligned}$$

uniformly over $(\tau, u) \in \mathcal{T} \times \mathcal{U}_0$. ■

Proof of Lemma C.9. Let $\tilde{\mathcal{V}}_n(\tau,u,U(r);\boldsymbol{\theta}) \equiv n^{-1}\sum_{i=1}^{n}\psi_\tau(Y_i^*(r),u,\boldsymbol{\theta})\mu'_{p,b,i}(r;u)\check{K}_{b,i}(r)/b^d$. Then $\mathcal{V}_n(\tau,u,U(r);\boldsymbol{\theta}) = n^{1/2}b^{d/2}\tilde{\mathcal{V}}(\tau,u,U(r);\boldsymbol{\theta})$. Noting that $\tilde{\mathcal{V}}_n(\tau,u,U(r);\boldsymbol{\theta})$ has the same structure as $\mu(\cdot,\cdot)$ in Lemma 4 in MRS (where $(\tau,u,\boldsymbol{\theta})$ plays the role of x in MRS), we can apply their result to conclude that uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ and $\|\boldsymbol{\theta}\|\le L$,

$$\left\|\tilde{\mathcal{V}}_n(\tau,u,U(\hat{r});\boldsymbol{\theta}) - \tilde{\mathcal{V}}_n(\tau,u,U(r_0);\boldsymbol{\theta}) - E_n\left[\tilde{\mathcal{V}}_n(\tau,u,U(\hat{r});\boldsymbol{\theta}) - \tilde{\mathcal{V}}_n(\tau,u,U(r_0);\boldsymbol{\theta})\right]\right\| \equiv O_P(\Xi_{n,5}),$$

where by the discussion in Appendix F.1 and Assumption C.9

$$\begin{aligned}\Xi_{n,5} &= O_P\left((nb^d)^{-1/2}\frac{\|\hat{r}-r_0\|_\infty}{b}\|\hat{r}-r_0\|_\infty^{-\frac{1}{2}c_r}\right) = \|\hat{r}-r_0\|_\infty^{1-\frac{1}{2}c_r}O_P\left(b^{-1}(nb^d)^{-1/2}\right)\\ &= o_P(b^{1+d_Z/2})O_P\left(b^{-1}(nb^d)^{-1/2}\right) = o_P\left((nb^{d_X})^{-1/2}\right).\end{aligned}$$

It follows that

$$\begin{aligned}&\|\mathcal{V}_n(\tau,u,U(\hat{r});\boldsymbol{\theta}) - \mathcal{V}_n(\tau,u,U(r_0);\boldsymbol{\theta}) - E_n[\mathcal{V}_n(\tau,u,U(\hat{r});\boldsymbol{\theta}) - \mathcal{V}_n(\tau,u,U(r_0);\boldsymbol{\theta})]\|\\ =\;& n^{1/2}b^{d/2}o_P\left((nb^{d_X})^{-1/2}\right) = o_P(b^{d_Z/2})\end{aligned}$$

uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ and $\|\boldsymbol{\theta}\|\le L$. ■

Proof of Lemma C.10. For notational simplicity, let $\psi_{\tau,i} = \psi_\tau(Y_i^*(r_0))$, $\mu_i = \mu_{p,b,i}(r_0;u)$, and $K_i = \check{K}_{b,i}(r_0)$. Define $\hat{\psi}_{\tau,i}$, $\hat{\mu}_i$, and $\hat{K}_i$ analogously with r_0 replaced by $\hat{r}$. We make the following decomposition for $E_n[\mathcal{V}_n(\tau,u,U(\hat{r});0)] - E_n[\mathcal{V}_n(\tau,u,U(r_0);0)]$:

$$\begin{aligned}&E_n[\mathcal{V}_n(\tau,u,U(\hat{r});0)] - E_n[\mathcal{V}_n(\tau,u,U(r_0);0)]\\ =\;& E_n\left\{n^{-1/2}b^{-d/2}\sum_{i=1}^{n}\left[\hat{\psi}_{\tau,i}\hat{\mu}_i\hat{K}_i - \psi_{\tau,i}\mu_iK_i\right]\right\}\\ =\;& E_n\left\{n^{-1/2}b^{-d/2}\sum_{i=1}^{n}\left[\left(\hat{\psi}_{\tau,i}-\psi_{\tau,i}\right)\mu_iK_i + \psi_{\tau,i}\left(\hat{\mu}_i\hat{K}_i-\mu_iK_i\right) + \left(\hat{\psi}_{\tau,i}-\psi_{\tau,i}\right)\left(\hat{\mu}_i\hat{K}_i-\mu_iK_i\right)\right]\right\}\\ =\;& E_n\left[n^{1/2}b^{-d/2}\left(\hat{\psi}_{\tau,i}-\psi_{\tau,i}\right)\mu_iK_i\right] + E_n\left[n^{1/2}b^{-d/2}\psi_{\tau,i}\left(\hat{\mu}_i\hat{K}_i-\mu_iK_i\right)\right]\\ &+E_n\left[n^{1/2}b^{-d/2}\left(\hat{\psi}_{\tau,i}-\psi_{\tau,i}\right)\left(\hat{\mu}_i\hat{K}_i-\mu_iK_i\right)\right]\\ \equiv\;& \Xi_{n,0}(\tau,u) + \Xi_{n,1}(\tau,u) + \Xi_{n,2}(\tau,u),\ \text{say},\end{aligned}$$

where the third equality holds by IID assumption.

First, we study $\Xi_{n,2}$. Noting that $Y_i^*(r)$ does not contain b, we can readily show that $E_n[\hat{\psi}_{\tau,i} - \psi_{\tau,i} | U_i(r_0)] = O_P(\|\hat{r}-r\|_\infty)$ uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ by Assumptions C.2-C.3. With this, by the Taylor expansion of $\hat{\mu}_i\hat{K}_i - \mu_iK_i$ at $U_i(r_0)$ and taking expectation, we can readily show that $\Xi_{n,2}(\tau,u) = O_P(\sqrt{nb^d}\|\hat{r}-r\|_\infty^2/b) = o_P\left(b^{d_Z/2}\right)$ uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ by Assumptions C.4, C.6 and C.9.

Second, we study $\Xi_{n,1}$. By the same reasoning as used in the proof of Lemma 2 in Fan et al. (1994), we have $E[\psi_\tau(Y_i^*(r_0))\mu_iK_i] = O\left(b^{p+1}\right)$. With this and following the analysis of $\Xi_{n,2}$, we can show that uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$

$$\begin{aligned}\Xi_{n,1}(\tau,u) &= O_P\left(b^{p+1}\sqrt{nb^d}\|\hat{r}-r\|_\infty/b\right) = O_P\left(b^{p+1}\sqrt{nb^{d_X}}b^{d_Z/2}\|\hat{r}-r\|_\infty/b\right)\\ &= O_P\left(b^{d_Z/2}\|\hat{r}-r\|_\infty/b\right) = o_P\left(b^{d_Z/2}\right)\end{aligned}$$

by Assumption C.6 and C.9.

To study $\Xi_{n,0}$, without loss of generality, we assume for now that $d_Z = 1$, i.e., r is a scalar function. The generalization to multivariate case is straightforward. By the IID assumption we have

$$\Xi_{n,0}(\tau,u) = E_n\left[n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)E_n\left[\psi_\tau(Y_i^*(\hat{r})) - \psi_\tau(Y_i^*(r_0))|\,U_i(r_0)\right]\right]. \tag{D.10}$$

Note that

$$\begin{aligned}
&E_n\left(\hat{\psi}_{\tau,i} - \psi_{\tau,i}|\,U_i(r_0)\right) = E_n\left[-\mathbf{1}\{Y_i^*(\hat{r}) \le 0\} + \mathbf{1}\{Y_i^*(r_0) \le 0\}|\,U_i(r_0)\right] \qquad \text{(D.11)}\\
=\;& -G\left(\sum_{0\le|\boldsymbol{j}|\le p}\frac{1}{\boldsymbol{j}!}D^{\boldsymbol{j}}G^{-1}(\tau|\,u)(U_i(\hat{r}) - u)^{\boldsymbol{j}}\right|U_i(r_0)\Bigg) + G(\boldsymbol{\beta}(\tau;u)|\,U_i(r_0))\\
=\;& -g(\boldsymbol{\beta}(\tau;u)|\,U_i(r_0))\left\{\sum_{0\le|\boldsymbol{j}|\le p}\frac{1}{\boldsymbol{j}!}D^{\boldsymbol{j}}G^{-1}(\tau|\,u)\left.\frac{\partial U^{\boldsymbol{j}}}{\partial U'}\right|_{U=U_i(r_0)-u}\right\}(U_i(\hat{r}) - U_i(r_0)) + \Xi_{n,3}(\tau,u)\\
=\;& g(\boldsymbol{\beta}(\tau;u)|\,U_i(r_0))\left\{\sum_{0\le|\boldsymbol{j}|\le p}\frac{1}{\boldsymbol{j}!}D^{\boldsymbol{j}}G^{-1}(\tau|\,u)\left.\frac{\partial U^{\boldsymbol{j}}}{\partial Z'}\right|_{U=U_i(r_0)-u}\right\}(\hat{r}(\varpi_i) - r_0(\varpi_i)) + \Xi_{n,3}(\tau,u)\\
\equiv\;& \chi(\tau,u,r_0(\varpi_i))(\hat{r}(\varpi_i) - r_0(\varpi_i)) + \Xi_{n,3}(\tau,u),
\end{aligned}$$

where $\Xi_{n,3}$ is the residual term from the Taylor expansion which is of the same order as $\|\hat{r} - r\|_\infty^2$, $\frac{\partial U^{\boldsymbol{j}}}{\partial Z'}$ denotes the first order derivative of $U^{\boldsymbol{j}}$ with respect to Z (so it is a scalar by assuming $d_Z = 1$) and the fourth equality holds by the fact that $U_i(\hat{r}) - U_i(r_0) = \begin{pmatrix} X_i \\ Q_i - \hat{r}(\varpi_i)\end{pmatrix} - \begin{pmatrix} X_i \\ Q_i - r_0(\varpi_i)\end{pmatrix} = \begin{pmatrix} 0 \\ r_0(\varpi_i) - \hat{r}(\varpi_i)\end{pmatrix}$, χ is defined as

$$\chi(\tau,u,r_0(\varpi_i)) \equiv g(\boldsymbol{\beta}(\tau;u)|\,U_i(r_0))\left\{\sum_{0\le|\boldsymbol{j}|\le p}\frac{1}{\boldsymbol{j}!}D^{\boldsymbol{j}}G^{-1}(\tau|\,u)\left.\frac{\partial U^{\boldsymbol{j}}}{\partial Z'}\right|_{U=U_i(r_0)-u}\right\}. \tag{D.12}$$

Substituting (D.11) into (D.10) yields

$$\begin{aligned}
\Xi_{n,0}(\tau,u) &= E_n\left\{n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)\left[\chi(\tau,u,r_0(\varpi_i))(\hat{r}(\varpi_i) - r_0(\varpi_i)) + \Xi_{n,3}(\tau,u)\right]\right\}\\
&= E_n\left[n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)\chi(\tau,u,r_0(\varpi_i))(\hat{r}(\varpi_i) - r_0(\varpi_i))\right] + o_P\left(b^{d_Z/2}\right)\\
&= \int_{\Omega_{d_\varpi}} n^{1/2}b^{-d_Z/2+d_X/2}E\left[\bar{\mu}_{p,b}(r_0(\varpi);u)b^{-d_X}\check{K}_b(r_0(\varpi))\chi(\tau,u,r_0(\varpi))\big|\,\varpi\right]\\
&\quad\times(\hat{r}(\varpi) - r_0(\varpi))F_\varpi(d\varpi) + o_P(b^{d_Z/2})\\
&\equiv \int_{\Omega_{d_\varpi}} n^{1/2}b^{-d_Z/2+d_X/2}\iota(\tau,u,r_0(\varpi))(\hat{r}(\varpi) - r_0(\varpi))F_\varpi(d\varpi) + o_P(b^{d_Z/2}),
\end{aligned} \tag{D.13}$$

uniformly over $(\tau,u) \in \mathcal{T}\times\mathcal{U}_0$, where $\bar{\mu}_{p,b}(r(\varpi);u) = \mu_{p,b}(U(r(\varpi)) - u)$, $U(r(\varpi)) = (X', (Q - r(\varpi))')'$, the second equality holds by the fact that $\Xi_{n,3}(\tau,u) = O(\|\hat{r} - r\|_\infty^2)$ and $E_n\left[\left|n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)\right|\right]$ $= O\left(n^{1/2}b^{d/2}\right)$ and uniformly over $(\tau,u) \in \mathcal{T}\times\mathcal{U}_0$ and that $n^{1/2}b^{d/2}\|\hat{r} - r\|_\infty^2 = o_P\left(b^{d_Z/2}\right)$ by Assumption C.9, the third equality holds by taking the expectation over ϖ_i, and

$$\iota(\tau,u,r_0(\varpi)) \equiv E\left[\bar{\mu}_{p,b}(r_0(\varpi);u)\frac{\check{K}_b(r_0(\varpi))}{b^{d_X}}\chi(\tau,u,r_0(\varpi))\Big|\,\varpi\right].$$

The dominant term in $\Xi_{n,0}$ could be analyzed analogously as in the proof of Proposition 1 in MRS. We decompose $\hat{r}(\varpi) - r_0(\varpi)$ in the same way as in equation (A.24) in MRS. Following the same lines of proof as theirs and noting that $\tilde{b}^{\tilde{p}+1} = o_P(n^{-1/2}b^{-d_X/2})$ by Assumption C.9, the bias term is asymptotically negligible uniformly over $u \in \mathcal{U}_0$ and we can obtain a similar expression as equation (A.26) in MRS:

$$\begin{aligned} \Xi_{n,0}(\tau,u) &= n^{-1/2}b^{-d_Z/2+d_X/2}\sum_{i=1}^{n}\int_{\Omega_{d_\varpi}} \iota(\tau,u,r_0(\varpi))\, e_{1,\tilde{p}}'\overline{\mathbf{S}}_{\tilde{p},\tilde{b}}(\varpi)^{-1}\tilde{K}_{\tilde{b}}(\varpi_i-\varpi)\,\mu_{\tilde{p},\tilde{b},i}(\varpi)\,F_\varpi(d\varpi)\,Z_i \\ &\quad + o_P(b^{d_Z/2}), \\ &\equiv \Xi_{n,4}(\tau,u) + o_P(b^{d_Z/2}). \end{aligned}$$

Let $t = (\varpi_i - \varpi)/\tilde{b}$. Using $\tilde{b} = o(b)$, we have $\iota(\tau,u,r_0(\varpi_i + t\tilde{b})) = \iota(\tau,u,r_0(\varpi_i))\{1 + O(\tilde{b}/b)\}$, similarly for $\overline{\mathbf{S}}_{p,b}(\varpi)$, and $\overline{\mathbf{S}}_{\tilde{p},\tilde{b}}(\varpi_i) = \mathbb{S}_{\tilde{p}} f_\varpi(\varpi_i) + o(1)$.

By the smoothness of $\chi(\tau,u,r_0(\varpi))$ and $f_{X|\varpi}$

$$\begin{aligned} &\iota(\tau,u,r_0(\varpi_i)) \qquad \text{(D.14)} \\ \equiv\;& E\left[\bar{\mu}_{p,b}(r_0(\varpi);u)\frac{\check{K}_b(r_0(\varpi))}{b^{d_X}}\chi(\tau,u,r_0(\varpi))\middle|\,\varpi=\varpi_i\right] \\ =\;& \int_{\mathbb{R}^{d_X}}\left[\int_{\mathcal{Q}}\mu_{p,b}\left((X',(Q-r_0(\varpi_i))')'\right)\frac{K\left(\left(\left(\frac{X-x}{b}\right)',\left(\frac{Q-r_0(\varpi_i)-z}{b}\right)'\right)'\right)}{b^{d_X}}F_{Q|X,\varpi}(dQ,X,\varpi_i)\right] \\ & \chi(\tau,u,r_0(\varpi))\,F_{X|\varpi}(dX,\varpi_i) \\ =\;& \int_{\mathbb{R}^{d_X}}\mu_{p,b}\left((X',(\bar{Q}(X,\varpi_i)-r_0(\varpi_i))')'\right)\frac{K\left(\left(\left(\frac{X-x}{b}\right)',\left(\frac{\bar{Q}(X,\varpi_i)-r_0(\varpi_i)-z}{b}\right)'\right)'\right)}{b^{d_X}}\chi(\tau,u,r_0(\varpi))\,F_{X|\varpi}(dX,\varpi_i) \\ =\;& \overline{\mu\check{K}}_{p,b}(z,\varpi_i)\,\chi(\tau,u,r_0(\varpi_i))\,f_{X|\varpi}(x|\varpi_i)\,(1+o_P(1)), \end{aligned}$$

for some $\bar{Q}(X,\varpi_i)$, where $\bar{Q}(X,\varpi_i)$ is obtained by the Second Mean Value Theorem for Definite Integrals and $\int_{\mathcal{Q}} F_{Q|X,\varpi}(dQ,X,\varpi_i) = 1$,[9]

$$\overline{\mu\check{K}}_{p,b}(z,r_0(\varpi_i)) \equiv \int_{\mathbb{R}^{d_X}}\mu_p\left((\mathbf{t}_{d_X},z_{ib}')'\right)K\left((\mathbf{t}_{d_X},z_{ib}')'\right)d\mathbf{t}_{d_X}, \qquad \text{(D.15)}$$

$\mathbf{t}_{d_X} \equiv (t_1,...,t_{d_X})$, $z_{ib} = (\bar{Q}(x,\varpi_i) - r_0(\varpi_i) - z)/b$, and $\overline{\mu\check{K}}_{p,b}(r_0(\varpi_i))$ is obtained by the change of variable: $\mathbf{t}_{d_X} = (X_i - x)/b$ and the continuity of $\bar{Q}(X,\varpi_i)$ in X. It follows that

$$\begin{aligned} \Xi_{n,4}(\tau,u) &\approx n^{-1/2}b^{-d_Z/2+d_X/2}\sum_{i=1}^{n}\overline{\mu\check{K}}_{p,b}(z,r_0(\varpi_i))\,\chi(\tau,u,r_0(\varpi_i))\,f_{X|\varpi}(x|\varpi_i)\,e_{1,\tilde{p}}'\mathbb{S}_{\tilde{p}}^{-1}\int_{\mathbb{R}^{d_\varpi}}\tilde{K}(t)\,\mu_{\tilde{p}}(t)\,dtZ_i \\ &\approx n^{-1/2}b^{d/2}\sum_{i=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}(z,r_0(\varpi_i))\,\chi(\tau,u,r_0(\varpi_i))\,f_{X|\varpi}(x|\varpi_i)\,e_{1,\tilde{p}}'\mathbb{S}_{\tilde{p}}^{-1}\int_{\mathbb{R}^{d_\varpi}}\tilde{K}(t)\,\mu_{\tilde{p}}(t)\,dt\;Z_i \\ &\approx n^{-1/2}b^{d/2}\sum_{i=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}(z,r_0(\varpi_i))\,\Psi_{n1}(\tau,u,\varpi_i)\,Z_i, \qquad \text{(D.16)} \end{aligned}$$

[9] The Second Mean Value Theorem for Definite Integrals does not hold for vector functions in general, but it holds here because each element of $\mu_{p,b}\left((X',(Q-r(\varpi_i))')'\right)$ contains at most one element of Q and each element of Q appears only once in $\mu_{p,b}\left((X',(Q-r(\varpi_i))')'\right)$.

where $A(\tau,u) \approx B(\tau,u)$ denotes that $A(\tau,u) = B(\tau,u)\{1+o_P(1)\}$ uniformly over $(\tau,u) \in \mathcal{T} \times \mathcal{U}_0$ and

$$\Psi_{n1}(\tau,u,\varpi_i) \equiv \chi(\tau,u,r_0(\varpi_i)) f_{X|\varpi}(x|\varpi_i) e'_{1,\tilde{p}} \mathbb{S}_{\tilde{p}}^{-1} \int_{\mathbb{R}^{d_\varpi}} \tilde{K}(t) \mu_{\tilde{p}}(t) dt. \tag{D.17}$$

In sum, we have uniformly over $(\tau,u) \in \mathcal{T} \times \mathcal{U}_0$

$$\begin{aligned} & E_n[\mathcal{V}_n(\tau,u,U(\hat{r});0) - \mathcal{V}_n(\tau,u,U(r_0);0)] \\ = & \left\{n^{-1/2}b^{d/2}\sum_{i=1}^n b^{-d_Z}\overline{\mu\check{K}}_{p,b}(z,r_0(\varpi_i))\Psi_{n1}(\tau,u,\varpi_i)Z_i\right\}\{1+o_P(1)\} + o_P(b^{d_Z/2}). \end{aligned}$$

This also holds for the multivariate case of r. ■

Proof of Lemma C.11. As in the proof of Lemma C.10, we make the following decomposition

$$\begin{aligned} & E_n[\mathcal{V}_n(\tau,u,U(\hat{r});\boldsymbol{\theta})] - E_n[\mathcal{V}_n(\tau,u,U(r_0);\boldsymbol{\theta})] \\ = & E_n\left[n^{1/2}b^{-d/2}(\psi_\tau(Y_i^*(\hat{r}),u,\boldsymbol{\theta}) - \psi_\tau(Y_i^*(r_0),u,\boldsymbol{\theta}))\mu_i K_i\right] \\ & + E_n\left[n^{1/2}b^{-d/2}\psi_\tau(Y_i^*(r_0),u,\boldsymbol{\theta})(\hat{\mu}_i\hat{K}_i - \mu_i K_i)\right] \\ & + E_n\left[n^{1/2}b^{-d/2}(\psi_\tau(Y_i^*(\hat{r}),u,\boldsymbol{\theta}) - \psi_\tau(Y_i^*(r_0),u,\boldsymbol{\theta}))(\hat{\mu}_i\hat{K}_i - \mu_i K_i)\right] \\ \equiv & \Xi_{n,0}(\tau,u;\boldsymbol{\theta}) + \Xi_{n,1}(\tau,u;\boldsymbol{\theta}) + \Xi_{n,2}(\tau,u;\boldsymbol{\theta}), \text{ say.} \end{aligned}$$

Apparently, $\Xi_{n,j}(\tau,u) = \Xi_{n,j}(\tau,u;0)$ for $j=0,1,2$.

For $\Xi_{n,2}(\tau,u;\boldsymbol{\theta})$, the analysis is almost the same as $\Xi_{n,2}(\tau,u)$. The main difference is that now have $E_n[\psi_\tau(Y_i^*(\hat{r}),u,\boldsymbol{\theta}) - \psi_\tau(Y_i^*(r_0),u,\boldsymbol{\theta})|U_i(r_0)] = O_P\left(\|\hat{r}-r\|_\infty + \|\hat{r}-r\|_\infty n^{-1/2}b^{-d/2}\|\boldsymbol{\theta}\|/b\right)$ uniformly over $(\tau,u) \in \mathcal{T}\times\mathcal{U}_0$, because $\boldsymbol{\theta}'\mu_{p,b,i}(\hat{r};u)$ contains b in the denominator associated with $\hat{r}$. So $\Xi_{n,2}(\tau,u,\boldsymbol{\theta}) = o_P\left(b^{d_Z/2}\right)$ uniformly over $(\tau,u) \in \mathcal{T} \times \mathcal{U}_0$ and $\|\boldsymbol{\theta}\| \leq L$ as $n^{-1/2}b^{-d/2}/b = O(1)$ by Assumption C.6.

For $\Xi_{n,1}(\tau,u;\boldsymbol{\theta})$, following the analysis of $\Xi_{n,1}(\tau,u)$ in the proof of Lemma C.10, we can show that $E[\psi_\tau(Y_i^*(r_0),u,\boldsymbol{\theta})\mu_i K_i] = O_P\left(b^{p+1} + n^{-1/2}b^{-d/2}\|\boldsymbol{\theta}\|\right)$ uniformly over $(\tau,u) \in \mathcal{T}\times\mathcal{U}_0$ and $\|\boldsymbol{\theta}\| \leq L$. Then

$$\Xi_{n,1}(\tau,u;\boldsymbol{\theta}) = o_P(b^{d_Z/2}) + O_P\left(n^{-1/2}b^{-d/2}\|\boldsymbol{\theta}\|\sqrt{nb^d}\|\hat{r}-r\|_\infty/b\right) = o_P(b^{d_Z/2}) + O_P(\|\hat{r}-r\|_\infty/b) = o_P(b^{d_Z/2}),$$

uniformly over $(\tau,u) \in \mathcal{T} \times \mathcal{U}_0$ and $\|\boldsymbol{\theta}\| \leq L$, where the last equality holds by Assumption C.9.

For $\Xi_{n,0}(\tau,u;\boldsymbol{\theta})$, we have

$$\Xi_{n,0}(\tau,u;\boldsymbol{\theta}) = E_n\left\{n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)E_n[\psi_\tau(Y_i^*(\hat{r}),\boldsymbol{\theta}) - \psi_\tau(Y_i^*(r_0),\boldsymbol{\theta})|U_i(r_0)]\right\}.$$

Following the derivation of (D.11), we have

$$\begin{aligned}
& E_n\left[\psi_\tau\left(Y_i^*(\hat{r}), u, \boldsymbol{\theta}\right)-\psi_\tau\left(Y_i^*\left(r_0\right), u, \boldsymbol{\theta}\right) \mid U_i\left(r_0\right)\right] \\
= & -G\left(\left.\sum_{0 \leq|\boldsymbol{j}| \leq p} \frac{1}{\boldsymbol{j}!} D^{\boldsymbol{j}} G^{-1}(\tau \mid u)\left(U_i(\hat{r})-u\right)^{\boldsymbol{j}}-n^{-1/2} b^{-d/2} \boldsymbol{\theta}^{\prime} \mu_{p, b, i}(\hat{r} ; u)\right| U_i\left(r_0\right)\right) \\
& +G\left(\boldsymbol{\beta}(\tau ; u)-n^{-1/2} b^{-d/2} \boldsymbol{\theta}^{\prime} \mu_{p, b, i}\left(r_0 ; u\right) \Big| U_i\left(r_0\right)\right) \\
= & -G\left(\left.\sum_{0 \leq|\boldsymbol{j}| \leq p} \frac{1}{\boldsymbol{j}!}\left(b^{|\boldsymbol{j}|} D^{\boldsymbol{j}} G^{-1}(\tau \mid u)-n^{-1/2} b^{-d/2} \boldsymbol{\theta}_{\boldsymbol{j}}\right)\left(\frac{U_i(\hat{r})-u}{b}\right)^{\boldsymbol{j}}\right| U_i\left(r_0\right)\right) \\
& +G\left(\left.\sum_{0 \leq|\boldsymbol{j}| \leq p} \frac{1}{\boldsymbol{j}!}\left(b^{|\boldsymbol{j}|} D^{\boldsymbol{j}} G^{-1}(\tau \mid u)-n^{-1/2} b^{-d/2} \boldsymbol{\theta}_{\boldsymbol{j}}\right)\left(\frac{U_i\left(r_0\right)-u}{b}\right)^{\boldsymbol{j}}\right| U_i\left(r_0\right)\right) \\
= & \; g\left(\left.\sum_{0 \leq|\boldsymbol{j}| \leq p} \frac{1}{\boldsymbol{j}!}\left(b^{|\boldsymbol{j}|} D^{\boldsymbol{j}} G^{-1}(\tau \mid u)-n^{-1/2} b^{-d/2} \boldsymbol{\theta}_{\boldsymbol{j}}\right)\left(\frac{U_i\left(r_0\right)-u}{b}\right)^{\boldsymbol{j}}\right| U_i\left(r_0\right)\right) \\
& \left\{\sum_{0 \leq|\boldsymbol{j}| \leq p} \frac{1}{\boldsymbol{j}!}\left(b^{|\boldsymbol{j}|-1} D^{\boldsymbol{j}} G^{-1}(\tau \mid u)-n^{-1/2} b^{-d/2} b^{-1} \boldsymbol{\theta}_{\boldsymbol{j}}\right) \left.\frac{\partial U^{\boldsymbol{j}}}{\partial Z^{\prime}}\right|_{U=\frac{U_i\left(r_0\right)-u}{b}}\right\}\left(\hat{r}\left(\varpi_i\right)-r_0\left(\varpi_i\right)\right)+\Xi_{n, 3}(\tau, u ; \boldsymbol{\theta}), \\
\equiv & \; \chi\left(\tau, u, r_0\left(\varpi_i\right) ; \boldsymbol{\theta}\right)\left(\hat{r}\left(\varpi_i\right)-r_0\left(\varpi_i\right)\right)+\Xi_{n, 3}(\tau, u ; \boldsymbol{\theta}),
\end{aligned}$$

where $\boldsymbol{\theta}_{\boldsymbol{j}}$ denotes the $\boldsymbol{j}$-th element of $\boldsymbol{\theta}$ and $\Xi_{n,3}(\tau,u;\boldsymbol{\theta})$ denotes the remainder term in the Taylor expansion. Then uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ and $\|\boldsymbol{\theta}\|\leq L$,

$$\begin{aligned}
\Xi_{n,0}(\tau,u;\boldsymbol{\theta}) &= E_n\left[n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\left(\check{K}_{b,i}(r_0)\chi(\tau,u,r_0(\varpi_i);\boldsymbol{\theta})(\hat{r}(\varpi_i)-r_0(\varpi_i))+\Xi_{n,3}(\tau,u;\boldsymbol{\theta})\right)\right] \\
&= E_n\left[n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)\chi(\tau,u,r_0(\varpi_i);\boldsymbol{\theta})(\hat{r}(\varpi_i)-r_0(\varpi_i))\right]+o_P(b^{d_Z/2}),
\end{aligned}$$

where

$$E_n\left[\left|n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)\Xi_{n,3}(\tau,u;\boldsymbol{\theta})\right|\right]=O_P\left(n^{1/2}b^{d/2}\|\hat{r}-r\|_\infty^2(1+n^{-1/2}b^{-d/2}b^{-2}\|\boldsymbol{\theta}\|)\right)=o_P(b^{d_Z/2})$$

and b^{-2} in front of $\|\boldsymbol{\theta}\|$ arises from the second order derivative. This, in conjunction with (D.13), implies that

$$\begin{aligned}
& \Xi_{n,0}(\tau,u;\boldsymbol{\theta})-\Xi_{n,0}(\tau,u) \\
= & \; E_n\left[n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)\left[\chi(\tau,u,r_0(\varpi_i);\boldsymbol{\theta})-\chi(\tau,u,r_0(\varpi_i))\right](\hat{r}(\varpi_i)-r_0(\varpi_i))\right]+o_P(b^{d_Z/2}) \\
= & \; O_P\left(E_n\left[\left|n^{1/2}b^{-d/2}\mu_{p,b,i}(r_0;u)\check{K}_{b,i}(r_0)\left[\chi(\tau,u,r_0(\varpi_i);\boldsymbol{\theta})-\chi(\tau,u,r_0(\varpi_i))\right]\right|\right]\|\hat{r}-r\|_\infty\right)+o_P(b^{d_Z/2}) \\
= & \; O_P\left(n^{1/2}b^{d/2}\left(n^{-1/2}b^{-d/2}b^{-1}\|\boldsymbol{\theta}\|\right)\|\hat{r}-r\|_\infty\right)+o_P(b^{d_Z/2})=o_P(b^{d_Z/2}),
\end{aligned}$$

uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ and $\|\boldsymbol{\theta}\|\leq L$.

To summarize, we have

$$E_n\left[\mathcal{V}_n(\tau,u,U(\hat{r});\boldsymbol{\theta})-\mathcal{V}_n(\tau,u,U(r_0);\boldsymbol{\theta})\right]-E_n\left[\mathcal{V}_n(\tau,u,U(\hat{r});0)-\mathcal{V}_n(\tau,u,U(r_0);0)\right]=o_P(b^{d_Z/2}),$$

which holds uniformly over $(\tau,u)\in\mathcal{T}\times\mathcal{U}_0$ and $\|\boldsymbol{\theta}\|\leq L$. ■

Proof of Lemma C.12. Without loss of generality, we assume that r is a scalar function. Following the same lines of analysis of $\Xi_{n,0}$ in the proof of Lemma C.10 (see the derivations of (D.13) to (D.16) in particular), we have

$$\begin{aligned}
& E_n\left\{b^{-d}\check{K}_{b,i}\left(r_0\right)\mu_{p,b,i}\left(r_0;u\right)\left[\hat{r}\left(\varpi_i\right)-r_0\left(\varpi_i\right)\right]\right\} \\
= & n^{-1}\sum_{j=1}^{n}\int_{\Omega_{d_\varpi}} b^{-d_Z} E\left[b^{-d_X}\check{K}_{b,i}\left(r_0\right)\mu_{p,b,i}\left(r_0;u\right)\right|\varpi_i=\varpi\left.\right] e_{1,\tilde{p}}^{\prime}\overline{\mathbf{S}}_{\tilde{p},\tilde{b}}\left(\varpi_j\right)^{-1}\tilde{K}_{\tilde{b}}\left(\varpi_j-\varpi\right)\mu_{\tilde{p},\tilde{b}}\left(\varpi_j-\varpi\right)F_{\varpi}\left(d\varpi\right)Z_j \\
& +o_P(n^{-1/2}b^{-d_X/2}) \\
= & \left(n^{-1}\sum_{j=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_j\right)\right)f_{X|\varpi}\left(x|\varpi_j\right)e_{1,\tilde{p}}^{\prime}\mathbb{S}_{\tilde{p}}^{-1}\int_{\mathbb{R}^{d_\varpi}}\tilde{K}\left(t\right)\mu_{\tilde{p}}\left(t\right)dtZ_j\right)\left\{1+o_P\left(1\right)\right\}+o_P(n^{-1/2}b^{-d_X/2}) \\
= & \left(n^{-1}\sum_{i=1}^{n}b^{-d_Z}\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_i\right)\right)\Psi_{n2}\left(x,\varpi_i\right)Z_i\right)\left\{1+o_P\left(1\right)\right\}+o_P(n^{-1/2}b^{-d_X/2})
\end{aligned}$$

uniformly over $u\in\mathcal{U}_0$, where $\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_i\right)\right)$ is defined in equation (D.15), and

$$\Psi_{n2}\left(x,\varpi_i\right)\equiv f_{X|\varpi}\left(x|\varpi_i\right)e_{1,\tilde{p}}^{\prime}\mathbb{S}_{\tilde{p}}^{-1}\int_{\mathbb{R}^{d_\varpi}}\tilde{K}\left(t\right)\mu_{\tilde{p}}\left(t\right)dt. \tag{D.18}$$

Noting that $U_i\left(\hat{r}\right)-U_i\left(r_0\right)=\begin{pmatrix} X_i \\ Q_i-\hat{r}\left(\varpi_i\right)\end{pmatrix}-\begin{pmatrix} X_i \\ Q_i-r_0\left(\varpi_i\right)\end{pmatrix}=\begin{pmatrix} \mathbf{0}_{d_X} \\ r_0\left(\varpi_i\right)-\hat{r}\left(\varpi_i\right)\end{pmatrix}$, we have

$$\begin{aligned}
\hat{\Lambda}\left(u\right) & \equiv e_{1,p}^{\prime}\overline{\mathbf{S}}_{p,b}\left(u\right)^{-1}E_n\left\{b^{-d}\check{K}_{b,i}\left(r_0\right)\mu_{p,b,i}\left(r_0;u\right)\left[U_i\left(\hat{r}\right)-U_i\left(r_0\right)\right]\right\} \\
& = -e_{1,p}^{\prime}\overline{\mathbf{S}}_{p,b}\left(u\right)^{-1}E_n\left\{b^{-d}\check{K}_{b,i}\left(r_0\right)\mu_{p,b,i}\left(r_0;u\right)\begin{pmatrix}\mathbf{0}_{d_X} \\ \hat{r}\left(\varpi_i\right)-r_0\left(\varpi_i\right)\end{pmatrix}\right\} \\
& = -\left\{n^{-1}\sum_{i=1}^{n}\begin{pmatrix}\mathbf{0}_{d_X} \\ Z_i\end{pmatrix}\left[e_{1,p}^{\prime}\overline{\mathbf{S}}_{p,b}\left(u\right)^{-1}b^{-d_Z}\overline{\mu\check{K}}_{p,b}\left(z,r_0\left(\varpi_i\right)\right)\right]\Psi_{n2}\left(x,\varpi_i\right)\right\}\left\{1+o_P\left(1\right)\right\} \\
& \quad +o_P(n^{-1/2}b^{-d_X/2})
\end{aligned}$$

uniformly over $u\in\mathcal{U}_0$. It is easy to show that the dominant term in the last expression is $O_P(n^{-1/2}b^{-d_Z/2})$. ■

E Proofs of some results in the main text

Proof of Theorem 5.3. Note that the results in Lemmas C.1 and C.2 continue to hold when $\{Y_i\}$ is replaced by $\{Y_{ni}\}$ in the estimation. In this case, both the conditional CDF and PDF of Y_{ni} given $(X_i,Z_i)=(x,z)$ become n-dependent. Let

$$\bar{J}_n\equiv b^{d_X}\sum_{i=1}^{n}\{\int G_n^{-1}(\hat{G}_{p,b}\left(Y_i|X_i,z\right)|x^*,z)d\Delta\left(z\right)\}^2\pi_i,\ \alpha_n\left(\tau|u\right)\equiv\hat{G}_{p,b}^{-1}\left(\tau|u\right)-G_n^{-1}\left(\tau|u\right),$$

$$\alpha_{n1}\left(\tau;u\right)\equiv e_{1,p}^{\prime}\bar{S}_{p,b}\left(\tau;u\right)^{-1}\bar{V}_{p,b}\left(\tau;u\right),\text{ and }\alpha_{n2}\left(\tau;u\right)\equiv\alpha_n\left(\tau|u\right)-\alpha_{n1}\left(\tau;u\right).$$

Let $\tau_{iz} \equiv G_n(Y_i|X_i,z)$, 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

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

Since it is straightforward to show that $\hat{\mathbb{B}}_{J_n} - \mathbb{B}_{J_n} = o_P(1)$ and $\hat{\sigma}_{J_n}^2 - \sigma_{J_n}^2 = o_P(1)$, it suffices to prove the theorem by showing that

$$\hat{J}_n = J_n + o_P(1), \tag{E.2}$$

and

$$J_n - \mathbb{B}_{J_n} \xrightarrow{d} \mathbb{N}\left(c_{\Theta_0}, \sigma_J^2\right), \tag{E.3}$$

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

We prove (E.2) by showing that

$$\begin{aligned}
\bar{J}_n &= b^{d_X}\sum_{i=1}^n \left[\int \frac{e_{1,p}'\bar{\mathbf{S}}_{p,b}(X_i,z)^{-1}\mathbf{V}_{p,b}(Y_i;X_i,z)}{g_n\left(G_n^{-1}(\tau_{iz}|x^*,z)|x^*,z\right)}\right]^2 \pi_i + c_{\Theta_0} + o_P(1), && \text{(E.4)} \\
\hat{J}_{n1} &= b^{d_X}\sum_{i=1}^n \left[\int e_{1,p}'\bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1} V_{p,b}(\tau_{iz};x^*,z)\, d\Delta(z)\right]^2 \pi_i + o_P(1), && \text{(E.5)} \\
\hat{J}_{n4} &= \tilde{J}_{n4} + o_P(1), \quad \text{and} && \text{(E.6)} \\
\hat{J}_{ns} &= o_P(1) \text{ for } s = 2,3,5, && \text{(E.7)}
\end{aligned}$$

where $\tilde{J}_{n4} \equiv b^{d_X}\sum_{i=1}^n \int \frac{e_{1,p}'\bar{\mathbf{S}}_{p,b}(X_i,z)^{-1}\mathbf{V}_{p,b}(Y_i;X_i,z)}{g_n\left(G_n^{-1}(\tau_{iz}|x^*,z)|x^*,z\right)} e_{1,p}'\bar{S}_{p,b}(\tau_{iz};x^*,z)^{-1} V_{p,b}(\tau_{iz};x^*,z)\, d\Delta(z)\, \pi_i$. To show (E.4), write

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

Under $\mathbb{H}_1(c_n)$ with $c_n = n^{-1/2}b^{-d_X/2}$, $\bar{J}_{n1} = n^{-1}\sum_{i=1}^n \Theta_n(Y_i; X_i, A_i)^2 \pi_i \xrightarrow{P} c_{\Theta_0}$ by Remark 5.2. Noting that

$$G_n^{-1}(\hat{\tau}_{iz}|x^*, z) - G_n^{-1}(\tau_{iz}|x^*, z) = \frac{\hat{\tau}_{iz} - \tau_{iz}}{g_n\left(G_n^{-1}(\tau_{iz}|x^*, z)\,|x^*, z\right)} + \hat{R}_i(z), \tag{E.8}$$

where $\hat{R}_i(z) = -\frac{g_n'\left(G_n^{-1}(\tau_{iz}^*|x^*,z)|x^*,z\right)}{g_n\left(G_n^{-1}(\tau_{iz}^*|x^*,z)|x^*,z\right)^3}(\hat{\tau}_{iz} - \tau_{iz})^2$, $g_n'(\cdot|x^*, z)$ denotes the derivative of $g_n(\cdot|, x^*, z)$ with respect to $\cdot$, and τ_{iz}^* lies between τ_{iz} and $\hat{\tau}_{iz}$, we have that under $\mathbb{H}_1(c_n)$,

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

Observing that $\hat{R}_i(z) = O_P\left(n^{-1}b^{-d}\log n + b^{2(p+1)}\right)$ uniformly in z by Lemma C.1(b), we have $\bar{J}_{n32} = n^{1/2}b^{d_X/2}O_P\left(n^{-1}b^{-d}\log n + b^{2(p+1)}\right) = o_P(1)$. By Lemma C.1(a), the fact that $\overline{\mathbf{B}}_{p,b}(y; x, z) = O_P(b^{p+1})$ uniformly in $(y, x, z) \in \mathbb{R} \times \mathcal{X}_0 \times \mathcal{Z}_0$, and Assumption C.6, we have $\bar{J}_{n31} = n^{-1/2}b^{d_X/2}\sum_{i=1}^n \Theta_n(Y_i; X_i, A_i)$ $\times \int \frac{e_{1,p}'\overline{\mathbf{S}}_{p,b}(X_i,z)^{-1}\mathbf{V}_{p,b}(Y_i;X_i,z)}{g_n\left(G_n^{-1}(\tau_{iz}|x^*,z)|x^*,z\right)} d\Delta(z)\,\pi_i + o_P(1)$. 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 inequality. Thus, $\overline{J}_{n3} = o_P(1)$ under $\mathbb{H}_1(c_n)$. Using (E.8), we decompose $\bar{J}_{n2}$ as follows

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

By Lemmas C.1(a)-(b) and Assumption C.11, we can readily show that

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

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

By Lemma C.5, (E.5) holds. With (E.5), it is standard to show that $\hat{J}_{n1} = O_P(1)$. Using (E.8) we can decompose $\hat{J}_{n4}$ as

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

Using arguments as used in the analysis of $\bar{J}_{n31}$ and $\bar{J}_{n32}$, we can readily show that $\hat{J}_{n4s} = o_P(1)$ for $s = 1, 3$. For $\hat{J}_{n42}$, we can apply Lemmas C.1 and C.2 to obtain $\hat{J}_{n42} = \tilde{J}_{n4} + o_P(1)$, where $\tilde{J}_{n4}$ is defined after (E.7).[10] Thus (E.6) follows.

We now show (E.7). By Lemma C.2(a), $\hat{J}_{n2} = nb^{d_X}[O_P(\nu_b^4) + o_P(b^{2(p+1)} + n^{-1}b^{-d_X})] = o_P(1)$. By the fact that $\hat{J}_{n1} = O_P(1)$ and Cauchy-Schwarz inequality, $\hat{J}_{n3} = o_P(1)$. For $\hat{J}_{n5}$, we have $\hat{J}_{n5} = nb^{d_X}[O_P(\nu_b^2) + o_P(b^{p+1} + n^{-1/2}b^{-d_X/2})]\, O_P(n^{-1/2}b^{-d/2}\sqrt{\log n} + b^{p+1}) = o_P(1)$. Consequently, (E.2) follows.

To show (E.3), let

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

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

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

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 Chebyshev inequality.

For J_{n1}, let $\varphi(W_i, W_j) = \int\int\int\zeta_0(\tilde{w}, W_i; z)\,\zeta_0(\tilde{w}, W_j; \bar{z})\,\pi(\tilde{x}, \tilde{y})\,d\Delta(z)\,d\Delta(\bar{z})\,dG_n(\tilde{w})$. 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: $\mathbb{V}_{J_n} \xrightarrow{d} \mathbb{N}\left(0, \sigma_J^2\right)$, where $\sigma_J^2 = \lim_{n\to\infty}\sigma_{J_n}^2$ and $\sigma_{J_n}^2 = 2b^{2d_X}E\left[\varphi(W_1, W_2)\right]^2$. ■

Sketch of the Proof of Equation (5.11). 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 $\mathbb{W}_n^*$ replacing $\mathbb{W}_n$. Define $\mathcal{S}(C) \equiv \{\sup_{(y,u)\in\mathbb{R}\times\mathcal{U}_0}|\hat{G}_{p,b}(y|u) - G(y|u)| \le Cn^{-1/2}b^{-d/2}(\log n)^{1/2} + b^{p+1}$, $\sup_{(\tau,u)\in\mathcal{T}\times\mathcal{U}_0}|\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u)| \le Cn^{-1/2}b^{-d/2}(\log n)^{1/2} + b^{p+1}\}$, where $\mathcal{T} = [\epsilon_0, 1-\epsilon_0]$ for some small $\epsilon_0 \in (0, 1/2)$. Then by Lemmas C.1 and C.2, for any $\epsilon > 0$, there exists a sufficiently large constant C such that $P(\mathcal{S}^c(C)) \le \epsilon$ for sufficiently large n where $\mathcal{S}^c(C)$ is the complement of $\mathcal{S}(C)$. Noting that

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

[10] 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.

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 $P(T_n^* \leq t|\mathbb{W}_n \cap \mathcal{S}(C)) \to \Phi(t)$ for all $t \in \mathbb{R}$. Conditional on $\mathbb{W}_n \cap \mathcal{S}(C)$, $\hat{A}_i$ is well defined, and one can follow the proof of Theorem 5.2 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 $\mathbb{W}_n \cap \mathcal{S}(C)$, 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 $\mathbb{W}_n$, one can continue to apply the CLT of Hall (1984) to $\mathbb{V}_{J_n^*}$ to demonstrate that it is asymptotically $\mathbb{N}\left(0, \sigma_J^{*2}\right)$ conditional on $\mathbb{W}_n$, 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 $\mathbb{W}_n \cap \mathcal{S}(C)$ then follows.

For (ii), let $\bar{z}_\alpha^*$ denote the $1-\alpha$ conditional quantile of T_n^* given $\mathbb{W}_n$, i.e., $P(T_n^* \geq \bar{z}_\alpha^*|\mathbb{W}_n) = \alpha$. By choosing N_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 5.1, T_n diverges to ∞ at the rate nb^{d_X}, implying that $\lim_{n\to\infty} P(T_n \geq z_\alpha^*) = \lim_{n\to\infty} P(T_n \geq z_\alpha) = 1$ under $\mathbb{H}_1$. ■

F Some other details

F.1 More details on the generated regressors problem

F.1.1 Discussion on the assumptions on the generated regressors problem

We apply the result of Theorem 1 in MRS to study the asymptotic property of our test with some generated regressors. In this section, we discuss how the conditions required for this theorem are met and how the result in this theorem works in our paper.

As discussed in MRS, four conditions are required for this theorem, namely, the "Regularity", "Accuracy", "Complexity", and "Continuity" conditions. We discuss how these conditions are met in this paper in order.

- The "Regularity" condition is about some standard requirements for the kernel, the bandwidth, and the smoothness of the objective function which are trivially satisfied here. Other than these, it requires a moment condition that $E[\exp(l|\varepsilon|)|S] \leq C$ a.s. for some $l > 0$ and $C < \infty$, where S and ε are the covariate and error term in MRS. This means that the error term ε needs to have a thinner tail than exponential. Our error term takes value between 0 and 1 when we estimate a CDF function. So this moment condition is trivially satisfied here.

- The "Accuracy" condition corresponds to our Assumption C.9, which we will discuss in detail later.

- The "Complexity" condition corresponds to our Assumption C.8, which is about the functional space of r.

- The "Continuity" condition is irrelevant in our paper, because this condition is about model misspecification and we assume we do not have this problem here.

Now we discuss Assumption C.9 and how we apply the result in MRS in our case. This Assumption is needed to obtain $o_P\left(n^{-1/2}b^{-d_X/2}\right)$ for the small order term resulted from the generated regressors. Specifically, we need the $O_P\left(n^{-\kappa}\right)$ term in MRS's Theorem 1 to be $o_P\left(n^{-1/2}b^{-d_X/2}\right)$ in our case.

Note that $\kappa = \min\{\kappa_1, \kappa_2, \kappa_3\}$ in MRS, where

$$\begin{aligned} \kappa_1 &\leq \frac{1}{2}\left(1-\eta_+\right) + (\delta-\eta)_{\min} - \frac{1}{2}\max_{1\leq j\leq d}\left(\delta_j\alpha_j + \xi_j\right), \\ \kappa_2 &\leq 2\eta_{\min} + (\delta-\eta)_{\min}, \\ \kappa_3 &\leq \delta_{\min} + (\delta-\eta)_{\min}, \end{aligned} \tag{F.1}$$

where all notations are as defined in MRS. Strictly speaking, MRS specifies "$<$" instead of "$\leq$" in the above three relationships. But a careful check of equation (A.2)−(A.6) in their proofs suggests that we can make the above replacement. In the following six steps, we explain and translate those notations in equation (F.1) to ours.

1. We use the same bandwidth for each covariate, and have the same convergence rate for each component of r, so $\eta_+ = d\eta$ ($d = d_X + d_Z$ in our case) and we can drop out the "min" and "max" operators in (F.1);

2. Since we assume the all derivatives of r up to $\underline{\varrho}$-th are bounded, $\xi_j = 0$ in our case;

3. $\alpha_j = \alpha$ corresponds to our c_r;

4. η denotes the convergence rate of the second stage bandwidth: $b \propto n^{-\eta}$;

5. δ denotes the convergence rate of the first-stage nonparametric estimate: $\|\hat{r}-r\|_\infty \propto n^{-\delta}$;

6. In MRS they have $2\eta_{\min}$ ($n^{-2\eta_{\min}} = b^2$) in κ_2, because they use local linear estimator and have $O(n^{-2\eta_{\min}}) = O(b^2)$ order of bias from their first stage estimation (for details, see their equation A.4, A.19, A.21 and the proof of Lemma 3 in their appendix). Here we use pth order local polynomial regression and thus have $O\left(b^{p+1}\right)$ bias term and the condition on κ_2 in our case becomes

$$\kappa_2 \leq (p+1)\,\eta + (\delta-\eta)\,. \tag{F.2}$$

By the analyses in Points 1-6, if we let κ_1 and κ_3 take the largest possible value in equation (F.1) (i.e., equality holds) and κ_2 take the largest value in equation (F.2) (i.e., equality holds), then we can translate $n^{-\kappa_1}$, $n^{-\kappa_2}$, and $n^{-\kappa_3}$ into our notation as:

$$\begin{aligned} n^{-\kappa_1} &= n^{-\frac{1}{2}\left(1-\eta_+\right)}n^{-(\delta-\eta)}n^{\frac{1}{2}\delta\alpha} \propto n^{-\frac{1}{2}}b^{-\frac{d}{2}}\frac{\|\hat{r}-r\|_\infty}{b}\|\hat{r}-r\|_\infty^{-\frac{1}{2}c_r} = o_P\left(n^{-1/2}b^{-d_X/2}\right), \\ n^{-\kappa_2} &= n^{-(p+1)\eta}n^{-(\delta-\eta)} \propto b^{p+1}\frac{\|\hat{r}-r\|_\infty}{b} = o_P\left(n^{-1/2}b^{-d_X/2}\right), \\ n^{-\kappa_3} &= n^{-\delta}n^{-(\delta-\eta)} \propto \|\hat{r}-r\|_\infty\frac{\|\hat{r}-r\|_\infty}{b} = o_P\left(n^{-1/2}b^{-d_X/2}\right), \end{aligned}$$

where the equality in the first equation holds by the condition $\|\hat{r}-r\|_\infty^{1-\frac{1}{2}c_r} = o_P\left(b^{1+d_Z/2}\right)$ in Assumption C.9, the equality in the second equation holds by the condition $b^{p+1} = O(n^{-1/2}b^{-d_X/2})$ in Assumption C.6 and that $\|\hat{r}-r\|_\infty = o_P(b^{\frac{2+d_Z}{2-c_r}}) = o_P(b)$ in Assumption C.9, and the last equality holds by the condition $\|\hat{r}-r\|_\infty = o_P\left(n^{-1/4}b^{1/2-d_X/4}\right)$ in Assumption C.9. Therefore,

$$n^{-\kappa} = n^{-\min\{\kappa_1,\kappa_2,\kappa_3\}} = \max\left\{n^{-\kappa_1}, n^{-\kappa_2}, n^{-\kappa_3}\right\} = o\left(n^{-1/2}b^{-d_X/2}\right).$$

By Theorem 1 in MRS, the small order term resulted from the generated regressors is $o\left(n^{-1/2}b^{-d_X/2}\right)$ in our case.

F.1.2 An explanation on E_n

E_n in empirical processes is the same as the E in the usual sense, except that it treats every r in the functional space we are interested in as a constant function. Below is an illustration of how we calculate E_n. Suppose we observe an IID random sample $\{(Y_i, X_i), i = 1, ..., n\}$, where Y_i is a scalar and X_i is $d_X \times 1$ vector. Suppose that $\hat{r}(x) = n^{-1}\sum_{j=1}^{n}\mathcal{K}_n(X_j - x)$, where $\mathcal{K}_n$ is some measurable function that might depend on n. Then we calculate $E_n[Y_i\hat{r}(X_i)]$ by treating $\hat{r}$ as a constant function as follows:

$$
\begin{aligned}
E_n[Y_i\hat{r}(X_i)] &= E_n[E(Y_i|X_i)\hat{r}(X_i)] \\
&= \int_{\mathbb{R}^{d_X}} E(Y|X=x)\hat{r}(x)F_X(dx) \\
&= \int_{\mathbb{R}^{d_X}} E(Y|X=x)\left(n^{-1}\sum_{j=1}^{n}\mathcal{K}_n(X_j - x)\right)F_X(dx) \\
&= n^{-1}\sum_{j=1}^{n}\int_{\mathbb{R}^{d_X}} E(Y|X=x)\mathcal{K}_n(X_j - x)F_X(dx),
\end{aligned}
$$

where the second equality is obtained by taking expectation over X_i, and F_X denotes the CDF of X.

F.2 Discussion on the local power property of the proposed test

In this appendix, we give details in deriving the results in Remark 5.3. Here, $Y_{ni} = m_n(X_i, A_i) = Y_i + c_n\gamma(X_i, A_i)$, where $Y_i = m(X_i, A_i) = \left(1 + 0.1X_i^2\right)A_i$ and $\gamma(X_i, A_i) = X_i e^{-A_i}$.

It is easy to verify that under Assumption A.2,

$$
\begin{aligned}
G(y|x,z) &= P\left(\left(1+0.1X^2\right)A_i \le y|X_i = x, Z_i = z\right) = F_{A|Z}\left(\frac{y}{\upsilon(x)}|z\right), \\
G^{-1}(\tau|x,z) &= \upsilon(x)F_{A|Z}^{-1}(\tau|z), \\
G^{-1}(G(y|x,z)|x^*,z) &= F_{A|Z}^{-1}\left(F_{A|Z}\left(\frac{y}{\upsilon(x)}|z\right)|z\right) = \frac{y}{\upsilon(x)},
\end{aligned}
$$

where $\upsilon(x) = 1 + 0.1x^2$.

Now, we derive the $G_n(y|x,z)$ and $G_n^{-1}(\tau|x^*,z)$. Noting that $G_n(y|x^*,z) = P(A_i \le y|X_i = x^*, Z_i = z) = F_{A|Z}(y|z)$, we have

$$
G_n^{-1}(\tau|x^*,z) = F_{A|Z}^{-1}(\tau|z).
$$

We consider two cases: (1) $x > 0$, and (2) $x \le 0$. Note that $m_n(x,a) = \upsilon(x)a + c_n x e^{-a}$.

First, we consider case (1). Noting that

$$
\frac{\partial m_n(x,a)}{\partial a} = \begin{cases} \upsilon(x) - c_n x e^{-a} > 0 & \text{if } a > \ln\left(\frac{c_n x}{\upsilon(x)}\right) \\ \upsilon(x) - c_n x e^{-a} < 0 & \text{if } a < \ln\left(\frac{c_n x}{\upsilon(x)}\right) \end{cases},
$$

$m_n(x,a)$ is strictly increasing in a when $a > \ln\left(\frac{c_n x}{\upsilon(x)}\right)$ and strictly decreasing in a when $a < \ln\left(\frac{c_n x}{\upsilon(x)}\right)$. In addition, when $a = \ln\left(\frac{c_n x}{\upsilon(x)}\right)$, we have

$$
m_n(x,a) = \upsilon(x)\ln\left(\frac{c_n x}{\upsilon(x)}\right) + c_n x e^{-\ln\left(\frac{c_n x}{\upsilon(x)}\right)} = \upsilon(x)\left[\ln\left(\frac{c_n x}{\upsilon(x)}\right) + 1\right] < 0 \text{ for sufficiently small } c_n.
$$

With these, we can easily argue that for any $y > m_n\left(x, \ln\left(\frac{c_n x}{\upsilon(x)}\right)\right)$ and $x > 0$, the equation $m_n(x, a) = y$ has exactly two solutions, $a_{1n}(x, y)$ and $a_{2n}(x, y)$, such that $a_{2n}(x, y) < a_{1n}(x, y)$, $a_{2n}(x, y) < \ln\left(\frac{c_n x}{\upsilon(x)}\right) < 0$ and $a_{1n}(x, y) > \ln\left(\frac{c_n x}{\upsilon(x)}\right)$, and obtain that

$$\begin{aligned} G_n(y|x, z) &= P(m_n(x, A) \leq y | Z = z) = P(a_{2n}(x, y) \leq A \leq a_{1n}(x, y) | Z = z) \\ &= F_{A|Z}(a_{1n}(x, y) | z) - F_{A|Z}(a_{2n}(x, y) | z). \end{aligned} \tag{F.3}$$

In addition, it is easy to argue that $a_{1n}(x, y)$ is bounded above from infinity for any fixed $x > 0$ and $y > m_n\left(x, \ln\left(\frac{c_n x}{\upsilon(x)}\right)\right)$ and $a_{2n}(x, y) \to -\infty$ as $n \to \infty$.

We now take a close look at the forms of both solutions. Rewrite the equation $\upsilon(x) a + c_n x e^{-a} = y$ as

$$a + \frac{c_n x}{\upsilon(x)} e^{-a} = \frac{y}{\upsilon(x)}. \tag{F.4}$$

By the implicit function theorem, we can write the solution $a_{1n}(x, y)$ in terms of $\frac{c_n x}{\upsilon(x)}$ and $\frac{y}{\upsilon(x)}$:

$$a_{1n}(x, y) = \varphi\left(\frac{y}{\upsilon(x)}, \frac{c_n x}{\upsilon(x)}\right),$$

where φ is a function that is continuously differentiable with respect to its second argument and does not depend on n. By the Taylor expansion, we have

$$a_{1n}(x, y) = \varphi\left(\frac{y}{\upsilon(x)}, 0\right) + \varphi_2\left(\frac{y}{\upsilon(x)}, 0\right) \frac{c_n x}{\upsilon(x)} + o(c_n) \tag{F.5}$$

where $\varphi_2(\cdot, \cdot)$ denotes the partial derivative of φ with respect to its second argument. In addition, (F.4) implies

$$\begin{aligned} a_{1n}(x, y) &= \frac{y}{\upsilon(x)} - \frac{c_n x}{\upsilon(x)} e^{-a_{1n}(x, y)} \\ &= \frac{y}{\upsilon(x)} - \frac{c_n x}{\upsilon(x)} e^{-\left[\varphi\left(\frac{y}{\upsilon(x)}, 0\right) + \varphi_2\left(\frac{y}{\upsilon(x)}, 0\right)\frac{c_n x}{\upsilon(x)} + o(c_n)\right]} \\ &= \frac{y}{\upsilon(x)} - \frac{c_n x}{\upsilon(x)} e^{-\varphi\left(\frac{y}{\upsilon(x)}, 0\right)} + o(c_n) \end{aligned} \tag{F.6}$$

Comparing (F.5) with (F.6), we can conclude that $\varphi\left(\frac{y}{\upsilon(x)}, 0\right) = \frac{y}{\upsilon(x)}$, implying that

$$a_{1n}(x, y) = \frac{y}{\upsilon(x)} - \frac{c_n x}{\upsilon(x)} e^{-\frac{y}{\upsilon(x)}} + o(c_n). \tag{F.7}$$

For the negative solution $a_{2n}(x, y)$, the above argument does not work because $a_{2n}(x, y) \to -\infty$ when $n \to \infty$. In this case, we consider the change of variables to solve for (F.4) by setting $\alpha = e^a$ and rewriting (F.4) in terms of α :

$$\ln(\alpha)\alpha + \frac{c_n x}{\upsilon(x)} = \frac{y}{\upsilon(x)}\alpha. \tag{F.8}$$

The negative solution $a_{2n}(x, y)$ to (F.4) corresponds to a positive solution $\alpha_{2n}(x, y)$ to (F.8). But $\alpha_{2n}(x, y)$ is now well behaved and converges from the right to 0 as $n \to \infty$. By the implicit function theorem, we have

$$\alpha_{2n}(x, y) = \psi\left(\frac{y}{\upsilon(x)}, \frac{c_n x}{\upsilon(x)}\right) \tag{F.9}$$

where ψ is a function that is continuously differentiable with respect to its second argument and does not depend on n. By the Taylor expansion,

$$\alpha_{2n}(x,y) = \psi\left(\frac{y}{\upsilon(x)},0\right) + \psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{c_n x}{\upsilon(x)} + o(c_n) \tag{F.10}$$

where $\psi_2(\cdot,\cdot)$ denotes the partial derivative of ψ with respect to its second argument. Noting that $a_{2n}(x,y) = \ln(\alpha_{2n}(x,y)) \to -\infty$ as $n \to \infty$ implies that $\psi\left(\frac{y}{\upsilon(x)},0\right) = 0$ and $\psi_2\left(\frac{y}{\upsilon(x)},0\right) > 0$ in (F.10). Consequently, we have

$$a_{2n}(x,y) = \ln\left(\psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{c_n x}{\upsilon(x)} + o(c_n)\right) \tag{F.11}$$

where $\psi_2\left(\frac{y}{\upsilon(x)},0\right) > 0$.

Combining (F.3), (F.7) and (F.11), we have

$$\begin{aligned} G_n(y|x,z) &= F_{A|Z}\left(\frac{y}{\upsilon(x)} - \frac{c_n x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}} + o(c_n)\middle| z\right) - F_{A|Z}\left(\ln\left(\psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{c_n x}{\upsilon(x)} + o(c_n)\right)\middle| z\right) \\ &= F_{A|Z}\left(\frac{y}{\upsilon(x)} - \frac{c_n x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}} + o(c_n)\middle| z\right) - \phi\left(\psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{c_n x}{\upsilon(x)} + o(c_n), z\right) + o(c_n) \\ &= F_{A|Z}\left(\frac{y}{\upsilon(x)}\middle| z\right) - f_{A|Z}\left(\frac{y}{\upsilon(x)}\middle| z\right)\frac{c_n x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}} - \phi_1(0,z)\psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{c_n x}{\upsilon(x)} + o(c_n), \end{aligned}$$

where the second equality follows from (5.8) and the last equality follows from Taylor expansions, and $\phi_1(\cdot,\cdot)$ is the partial derivative of ϕ with respect to its first element. Therefore, we have for $y > 0$ and $x > 0$.

$$\begin{aligned} & G_n^{-1}(G_n(y|x,z)|x^*,z) - G^{-1}(G(y|x,z)|x^*,z) \\ = \; & F_{A|Z}^{-1}(G_n(y|x,z)|z) - \frac{y}{\upsilon(x)} \\ = \; & -c_n\left[\frac{x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}} + \frac{\phi_1(0,z)\psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{x}{\upsilon(x)}}{f_{A|Z}\left(\frac{y}{\upsilon(x)}\middle| z\right)} + o(1)\right], \end{aligned} \tag{F.12}$$

where the second term in the above square bracket is a non-constant function of z and it will contribute to the asymptotic local power of our test.

Now, we consider case (2). When $x \leq 0$, $\frac{\partial m_n(x,a)}{\partial a} = \upsilon(x) - c_n x e^{-a} > 0 \; \forall a \in \mathcal{A}$, and the equation

$$\upsilon(x)a + c_n x e^{-a} = y$$

has a unique positive solution, which we continue to denote as $a_{1n}(x,y)$. As in case (1), the result in (F.7) continues to hold, and we have

$$\begin{aligned} G_n(y|x,z) &= P(m_n(x,A) \leq y|Z=z) = P(-\infty \leq A \leq a_{1n}(x,y)|Z=z) \\ &= F_{A|Z}(a_{1n}(x,y)|z). \end{aligned} \tag{F.13}$$

Then for any $x \leq 0$, we have

$$G_n^{-1}(G_n(y|x,z)|x^*,z) - G^{-1}(G(y|x,z)|x^*,z) = c_n\left[\frac{x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}} + o(1)\right]. \tag{F.14}$$

It follows that for any $y > m_n\left(x, \ln\left(\frac{c_n x}{\upsilon(x)}\right)\right)$, we have

$$\Theta_n^{\dagger}(y;x,z) = \left[-\frac{x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}} + \omega(y;x,z) + o(1)\right]\mathbf{1}\{x>0\} + \left[\frac{x}{\upsilon(x)}e^{-\frac{y}{\upsilon(x)}} + o(1)\right]\mathbf{1}\{x \leq 0\}. \tag{F.15}$$

where

$$\omega(y;x,z) = \frac{-\phi_1(0,z)\,\psi_2\left(\frac{y}{\upsilon(x)},0\right)\frac{x}{\upsilon(x)}}{f_{A|Z}\left(\frac{y}{\upsilon(x)}\middle| z\right)}. \tag{F.16}$$

As $c_n \to 0$, $m_n\left(x, \ln\left(\frac{c_n x}{\upsilon(x)}\right)\right) \to -\infty$, $\Theta_n^{\dagger}(y;x,z)$ in (F.15) is defined for all $y \in \mathbb{R}$ asymptotically.

REFERENCES

Andrews, D.W.K., 1995. Nonparametric kernel estimation for semiparametric models. Econometric Theory 11, 560-596.

Boente, G., Fraiman, R., 1991. Strong uniform convergence rates for some robust equivariant nonparametric regression estimates for mixing processes. International Statistical Review 59, 355-372.

Escanciano, J., Jacho-Chavez, D., Lewbel, A., 2014. Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. Journal of Econometrics 178, 426-443.

Fan, J., Hu, T., Truong, Y., 1994. Robust non-parametric function estimation. Scandinavian Journal of Statistics 21, 433-446.

Hall, P., 1984. Central limit theorem for integrated square error properties of multivariate nonparametric density estimators. Journal of Multivariate Analysis 14, 1-16.

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

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

Pollard, D., 1991. Asymptotics for the least absolute deviation regression estimators. Econometric Theory 7, 186-199.

Serfling, R. J., 1980. Approximation theorems of mathematical statistics. New York: Wiley.

Su, L., White, H., 2012. Local polynomial quantile regression for dependent heterogeneous processes: uniform Bahadur representation with applications to testing conditional independence. Advances in Econometrics 29, 355-434.

van der Vaart, A.W., Wellner, J.A., 1996. Weak Convergence and Empirical Processes with Applications to Statistics. Springer-Verlag, New York.