Testing for Monotonicity in Unobservables under Unconfoundedness∗

Stefan Hoderlein\textsuperscript{a}, Liangjun Su\textsuperscript{b}, Halbert White\textsuperscript{c}, and Thomas Tao Yang\textsuperscript{a}
\textsuperscript{a} Department of Economics, Boston College
\textsuperscript{b} School of Economics, Singapore Management University
\textsuperscript{c} Department of Economics, University of California, San Diego
March 12, 2015

Abstract

Monotonicity in a scalar unobservable is a now 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; 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 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. Moreover, the authors 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.
the dependent variable, possibly conditioned on other observables). These assumptions permit identification of such important structural features as average marginal effects or various average effects of treatment. Seminal examples are the 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 been criticized, see e.g., Hoderlein and Mammen (2007) or Kasy (2011). Alternatives have been suggested in the case of triangular systems, most notably Kasy (2014), and in the treatment effects setup, see 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 effects 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 monotized 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. 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.

---

1See also Ghanem (2014) for related work on identification of these models.
Under the null hypothesis of monotonicity of a structural function in a scalar unobservable and a conditional 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 deviations from monotonicity. The conditional exogeneity assumption holds in many important examples, including control function treatments of exogeneity, unconfoundedness assumptions as in the treatment effects literature, or generalizations of the classical proxy assumption, and we will discuss it in great detail below. 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, $\alpha$, 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.

## 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 $Y$, $X$, $Z$, and $A$, 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

$$
H_0 : m(x, \cdot) \text{ is strictly monotone for each } x \in \mathcal{X}.
$$

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 $G^{-1}(\tau|x)$ denote the conditional cumulative distribution function (CDF) and conditional $\tau$-th quantile of $Y$ given $X = x$, respectively. Then, the strict monotonocity 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, the 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 $\mathcal{Y}$ 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.

First, this is a very common notion in the treatment effect literature, and is often called “unconfoundedness” (e.g., Wooldridge, 2010), or “conditionally exogeneity” (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 $S$. For instance, the first stage could be

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

4
where $\xi \perp (A, Z)$ and $\Psi$ 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 develop the asymptotic distribution theory and propose a bootstrap method suitable for testing whether the differences between multiple estimators of $A$ accord with the null of scalar monotonicity or whether this null must be rejected.

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 $x \in X$, $m(x, \cdot)$ 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 $\Gamma(\cdot \mid x, z)$ denote the conditional CDF of $Y$ given $(X, Z) = (x, z)$. For each $(x, z)$ in $X \times Z$, $\Gamma(\cdot \mid x, z)$ 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 $a = m(x^*, a) \forall a$,

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

\[
A = G^{-1}(G(Y \mid X, z) \mid x^*, z) \quad \forall z \in Z.
\]

The main implication of this proposition is that for every $Z = z$, we can use $G^{-1}(G(Y \mid X, z) \mid x^*, z))$ to determine $A$. Put, reversely, $G^{-1}(G(Y \mid X, z) \mid x^*, z)$ has to be invariant with respect to the changes in $z$. In the following, we will show how to make use of this fact and propose a test statistic for testing the null hypothesis in (2.1).

---

2 In the strictly exogenous case where covariates $Z$ are absent, the unconditional distribution of $A$ is typically normalized to be standard uniform, while this normalization makes more sense here. For example, $x^*$ can be the vector of medians of $X$; this normalization is actually innocuous, as we have showed in an earlier version of the paper.
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 \( m(x, a) = G^{-1}(G(a \mid x^*, z) \mid x, z) \) for given \( x^* \) and any \( z \), one can estimate \( m(x, a) \) as

\[
\hat{m}_x(x, a) = \hat{G}^{-1}(\hat{G}(a \mid x^*, z) \mid x, z)
\]

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

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

where \( H \) is a user-chosen distribution function supported on \( Z_0 \subseteq Z \), 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

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

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

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

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

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

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

\[
A^*_H \equiv \int G^{-1}(G(Y \mid X, z) \mid x^*, z)) dH(z),
\]

\[
A^\dagger_H \equiv m_H^{-1}(X, Y) \equiv \inf \{a : m_H^*(X, a) \geq Y\},
\]

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

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

\[
\hat{J}_n = b^d \sum_{i=1}^{n} (\hat{A}_{1,i} - \hat{A}_{2,i})^2 \pi(X_i, Y_i)
\]

\[
= b^d \sum_{i=1}^{n} \left\{ \int \hat{G}^{-1} \left( \hat{G}(Y_i | X_i, z) | x^*, z \right) d\Delta(z) \right\}^2 \pi_i,
\]

(3.1)

where \(b \equiv b_n\) is a suitable bandwidth; \(d_X\) is the dimension of \(X\); \(\hat{A}_{j,i} \equiv \int \hat{G}^{-1}(\hat{G}(Y_i | X_i, z) | x^*, z)) dH_j(z), j = 1, 2; \hat{G} and \hat{G}^{-1}\) 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 \(X_0 \times Y_0\) of \(X \times Y\). The weight \(\pi_i\) downweights observations for which \(\hat{G}(Y_i | X_i, z)\) is close to either 0 or 1, so that \(\hat{G}^{-1}\) cannot be accurately estimated. For example, one can set \(\pi(X_i, Y_i) = 1\{C_{\alpha,Y} \leq Y_i \leq C_{1-\alpha,Y}\} \times 1\{C_{\alpha,X} \leq X_i \leq C_{1-\alpha,X}\}\) in case \(d_X = 1\), where, 1\{\cdot\} denotes the usual indicator function, and, e.g., \(C_{\alpha,X}\) denotes the \(\alpha\)th sample quantile of \(\{X_i\}_{i=1}^{n}\). We take \(\alpha = 0.0125\) 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 \(Z_1\) and \(Z_2\) respectively, each a subset of a compact subset \(Z_0\) of \(Z\). These supports can be disjoint and the most important thing is that we want the contrast \(\Delta\) to be quite different from a zero function. Since they are chosen by the users, we think that it is not restrictive to focus on the case of known \(H\)'s. Different choices for \(\Delta\) 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. Our asymptotic power analysis below suggests ways of ensuring that \(\hat{J}_n\) can consistently detect violations of monotonicity under Assumptions A.2 and A.3. But it is extremely challenging, if possible at all, to consider the optimal choice of the contrast \(\Delta\). See Remark 5.5 in Section 5.2. We experiment the uniform and re-scaled beta(2,2) distributions for \(H_1\) and \(H_2\) in the Monte Carlo simulations (see Section 6.1 for details). We also discuss the results with different choices of \(H_1\) and \(H_2\) in footnote 8 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. Because the paper is already quite lengthy, we desist from a detailed analysis that shows that the bootstrap is consistent. Nevertheless, we will provide a sketch of the main arguments below. Moreover, since bootstrap is computationally feasible this statistic is relevant in practise.

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

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

where \( A_i \) denotes an individual’s unobserved time-invariant attribute, \( \varepsilon_{it} \) 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 \( F_1(\cdot|x) \) 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 \( F_1 \) 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

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

The above null hypothesis motivates SHW to consider a test statistic based on the comparison of \( F_1(m_{it} \mid X_{it}) \) and \( F_2(m_{is} \mid X_{is}) \) for all \( t \neq s \). Nevertheless, a direct comparison is infeasible because \( m_{it} \) is not observable. Let \( w_\tau(x), \tau = 1, ..., T \), be non-negative weight functions. Let \( \tilde{Y}_{\tau,i} = \tilde{Y}_{\tau,i} \equiv E[Y_{it}w_\tau(X_{it}) \mid A_i] \) 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 \( (X_{it} \perp A_i) \), HSW 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

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

SHW construct a test statistic based on the squared distance between consistent estimates of \( \tilde{F}_\tau(\tilde{Y}_{\tau,i}) \) and \( \tilde{F}_\zeta(\tilde{Y}_{\zeta,i}) \) under the assumption that \( T/N \rightarrow \infty \) as \( N \rightarrow \infty \). Apparently, 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 between estimates of certain unconditional CDFs corresponding to 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 \( \chi^2 \), 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 $u \equiv (x', z') = (u_1, \ldots, u_d)'$ be a $d \times 1$ vector, $d = d_X + d_Z$, where $x$ is $d_X \times 1$ and $z$ is $d_Z \times 1$. Let $j \equiv (j_1, \ldots, j_d)'$ be a $d \times 1$ vector of non-negative integers. Following Masry (1996), we adopt the notation: $w^j \equiv \Pi_{i=1}^d u_i^{j_i}$, $j! \equiv \Pi_{i=1}^d j_i!$, $|j| \equiv \sum_{i=1}^d j_i$, and $\sum_{0 \leq |j| \leq p} \equiv \sum_{j_1=0}^p \cdots \sum_{j_d=0}^p$.

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 $\{(Y_i, U_i), i = 1, \ldots, n\}$, $\hat{G}_{p,b}(y|x,z)$ can be obtained as the minimizing intercept term in the following minimization problem

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

where $\beta$ stacks the $\beta_j^{(i)}$’s ($0 \leq |j| \leq p$) in lexicographic order (with $\beta_0$, indexed by $0 \equiv (0, \ldots, 0)$, in the first position, the element with index $(0, 0, \ldots, 1)$ 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 $N_{p,t} \equiv (l + d - 1)/(l!(d-1)!)$ be the number of distinct $d$-tuples $j$ with $|j| = l$. In the above estimation problem, this denotes the number of distinct $l$th order partial derivatives of $G(y|u)$ with respect to $u$. Let $N_p \equiv \sum_{t=0}^p N_{p,t}$. 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 $((U_i - u)/b)^j$, $0 \leq |j| \leq p$, in lexicographic order (e.g., $\mu_p(u) = (1, u')'$ when $p = 1$). Let $\mu_{p,b}(u) \equiv \mu_p(u/b)$. Then $\hat{G}_{p,b}(y|u) = \epsilon_{1,p}\hat{\beta}(y|u)$ where

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

$$
S_{p,b}(u) \equiv n^{-1} \sum_{i=1}^n K_b(U_i - u) \mu_{p,b}(U_i - u) \mu_{p,b}(U_i - u)',
$$

and $\epsilon_{1,p} \equiv (1, 0, \ldots, 0)'$ 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 $\tau$th 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

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

where $\alpha$ stacks the $\alpha_j^{(i)}$’s ($0 \leq |j| \leq p$) 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 $\hat{\Delta}_H$ 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 $\tilde{A}_H$

We use the following assumptions.

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

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 $U \equiv \mathcal{X} \times \mathcal{Z}$ and $U_0 \equiv \mathcal{X}_0 \times \mathcal{Z}_0$. Let $\mathcal{Y}$ denote the support of $Y_i$, and let $\mathcal{Y}_0 \equiv [\underline{y}, \bar{y}]$ for finite real numbers $\underline{y}, \bar{y}$.

**Assumption C.2** (i) $g(u)$ is continuous in $u \in U$, and $g(y|u)$ is continuous in $(y, u) \in \mathcal{Y} \times U$.
(ii) There exist $C_1, C_2 \in (0, \infty)$ such that $C_1 \leq \inf_{u \in U_0} g(u) \leq \sup_{u \in U_0} g(u) \leq C_2$ and $C_1 \leq \inf_{(y, u) \in \mathcal{Y}_0 \times U_0} g(y|u) \leq \sup_{(y, u) \in \mathcal{Y}_0 \times U_0} g(y|u) \leq C_2$.

**Assumption C.3** (i) There exist $\underline{\tau}, \bar{\tau} \in (0, 1)$ such that $\underline{\tau} \leq \inf_{u \in U_0} G(y|u) \leq \sup_{u \in U_0} G(y|u) \leq \bar{\tau}$ and $\underline{\tau} \leq \inf_{z \in \mathcal{Z}_0} G(y|z) \leq \sup_{z \in \mathcal{Z}_0} G(y|z) \leq \bar{\tau}$.
(ii) $G(\cdot|u)$ is equicontinuous: $\forall \epsilon > 0$, $\exists \delta > 0$ such that $\|y - \tilde{y}\| < \delta \Rightarrow \sup_{u \in U_0} |G(y|u) - G(\tilde{y}|u)| < \epsilon$. For each $y \in \mathcal{Y}_0$, $G(y|\cdot)$ is Lipschitz continuous on $U_0$ and has all partial derivatives up to order $p + 1$, $p \in \mathbb{N}$.
(iii) Let $DG(y|u) \equiv \partial^j G(y|u) / \partial^i U_1 \ldots \partial^i U_d$. For each $y \in \mathcal{Y}_0$, $D^i DG(y|\cdot)$ with $|j| = p + 1$ is uniformly bounded and Lipschitz continuous on $U_0$; for all $u, \tilde{u} \in U_0$, $|D^i DG(y|u) - D^i DG(y|\tilde{u})| \leq C_3 ||u - \tilde{u}||$ for some $C_3 \in (0, \infty)$ where $\|\cdot\|$ is the Euclidean norm.
(iv) For each $u \in U_0$ and for all $y, \tilde{y} \in \mathcal{Y}_0$, $|D^i DG(y|u) - D^i DG(\tilde{y}|u)| \leq C_4 |y - \tilde{y}|$ for some $C_4 \in (0, \infty)$ where $|j| = p + 1$.

**Assumption C.4** (i) The kernel $K : \mathbb{R}^d \rightarrow \mathbb{R}^+$ is a continuous, bounded, and symmetric PDF.
(ii) $u \rightarrow \|u\|^{2p+1} K(\cdot)$ is integrable on $\mathbb{R}^d$ with respect to the Lebesgue measure.
(iii) Let $K_j(u) \equiv u^j K(u)$ for all $j$ with $0 \leq |j| \leq 2p + 1$. For some finite constants $c_K, \underline{\tau}_1, \bar{\tau}_2$, either $K(\cdot)$ is compactly supported such that $K(u) = 0$ for $\|u\| > c_K$, and $\|K_j(u) - K_j(\tilde{u})\| \leq \bar{\tau}_2 ||u - \tilde{u}||$ for any $u, \tilde{u} \in \mathbb{R}^d$ and for all $j$ with $0 \leq |j| \leq 2p + 1$; or $K(\cdot)$ is differentiable, $\|\partial K_j(u) / \partial u\| \leq \tau_1$, and for some $c > 1$, $|\partial K_j(u) / \partial u| \leq \tau_1 \|u\|^{-c}$ for all $\|u\| > c K$ and for all $j$ with $0 \leq |j| \leq 2p + 1$.

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

**Assumption C.6** As $n \rightarrow \infty$, $b \rightarrow 0$, $b^{p+1-d+2} \rightarrow 0$, and $\log^2(1 + d) \rightarrow \infty$.

The IID requirement of Assumption C.1 is standard in cross-section studies. Nevertheless, the asymptotic theory developed here can be readily extended to weakly dependent time series. To keep the results uncluttered, we leave the time-series case for study elsewhere. Assumption C.2 is standard for nonparametric local polynomial estimation of conditional mean and density. If $U_i$ has compact support $\mathcal{U}$ and $g(u)$ is bounded away from zero on $\mathcal{U}$, it is possible to choose $U_0 = \mathcal{U}$. 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 $N_{p,t}$ $d$-tuples as a sequence in lexicographical order, so that $\ell_t(1) \equiv (0, 0, ..., l)$ is the first element and $\ell_t(N_{p,t}) \equiv (l, 0, ..., 0)$ is the last, and let $\ell_t^{-1}$ be the mapping inverse to $\ell_t$. Define the $N_p \times N_p$ matrix $S_p$ and the $N_p \times N_{p+1}$ matrix $B_p$ respectively by

$$S_p \equiv \begin{pmatrix} M_{0,0} & M_{0,1} & \ldots & M_{0,p} \\ M_{1,0} & M_{1,1} & \ldots & M_{1,p} \\ \vdots & \vdots & \ddots & \vdots \\ M_{p,0} & M_{p,1} & \ldots & M_{p,p} \end{pmatrix} \quad \text{and} \quad B_p = \begin{pmatrix} M_{0,p+1} \\ M_{1,p+1} \\ \vdots \\ M_{p,p+1} \end{pmatrix},$$

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

Let $A_H = \{ a : m_H^*(x,a) = y, x \in X_0, y \in Y_0 \}$. The asymptotic behavior of $\hat{m}_H(x,a)$ follows:

**Theorem 4.1** Suppose Assumptions C.1-C.6 hold. Let $x^* \in X_0$ and $(x,a) \in X_0 \times A_H$. Then

$$\sqrt{n b^{1/2}} (\hat{m}_H(x,a) - m^*_H(x,a) - B_m(x,a;x^*)) \overset{d}{\rightarrow} \mathcal{N}(0, \sigma_m^2(x,a;x^*)), \quad \text{where}$$

$$B_m(x,a;x^*) \equiv b^{p+1} e_1' p S_p^{-1} B_p \int \left[ \frac{G_{p+1}(a|x^*,z)}{g(G^{-1}(G(a|x^*,z))|x,z)} + G_{p+1}^{-1}(G(a|x^*,z)|x,z) \right] dH(z),$$

$$\sigma_m^2(x,a;x^*) \equiv \kappa_1 \int G(a|x^*,z) [1 - G(a|x^*,z)] h(z)^2 \left[ \frac{1}{g(x^*,z)} + \frac{1}{g(x,z)} \right] dz,$$

and $\kappa_1 \equiv \int e_1' p S_p^{-1} \mu_p(\bar{x},\bar{z}) \mu_p(\hat{x},\hat{z} - \hat{z}) S_p^{-1} e_1, K(\hat{x},\hat{z}) K(\hat{x},\hat{z} - \hat{z}) d(\bar{x},\bar{z},\hat{z})$. In addition,

$$\sup_{(x,a) \in X_0 \times A_H} |\hat{m}_H(x,a) - m_H^*(x,a)| = O_P(n^{-1/2}b^{-d/2} \sqrt{\log n} + b^{p+1}).$$

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 $\frac{G_{p+1}(a|x^*,z)}{g(G^{-1}(G(a|x^*,z))|x,z)}$ in $B_m(x,a;x^*)$ and $\frac{1}{g(x,z)}$ in $\sigma_m^2(x,a;x^*)$ are due to the first stage estimation, whereas those involving $G_{p+1}^{-1}(G(a|x^*,z)|x,z)$ 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.

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

To state the next results, we define \( A_{H,i}^* \) (and \( A_{H,i}^T \) below) in the obvious manner. Theorem 4.1 implies
the following asymptotic properties for \( \hat{A}_{H,i} \).

**Corollary 4.2** Suppose Assumptions C.1-C.6 hold. Then conditional on \((X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0, \sqrt{n} \mathcal{A}_{H,i} - A_{H,i}^*\) follows a distribution \( \mathcal{N}(0, \sigma_{H,i}^2) \). Further, for \( i \) such that \((X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0, \mathcal{A}_{H,i} - A_{H,i}^* = O_P(n^{-1/2}b^{-d_x/2}\sqrt{\log n + b_i}) \) uniformly in \( i \).

The asymptotic properties of \( \hat{A}_{H,i} \) follow from the next theorem.

**Theorem 4.3** Suppose Assumptions C.1-C.6 hold. Then for any \((x, y) \in \mathcal{X}_0 \times \mathcal{Y}_0, \hat{A}_{H,i}^* - \lambda_H^*(x, y) \) and
\( \sqrt{n} \mathcal{A}_{H,i} - \mathcal{A}_{H,i}^* \) follow from the next proposition.

When correct specification holds, we can show that \( \lambda_H^* (x, a) = \int g(a|x, z) \mathcal{D}(z) = \partial m (x, a) / \partial a \) 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 \( \int \mathcal{D}(z) = 1 \). Further, Theorem 4.3 implies that conditional on \((X_i, Y_i) \in \mathcal{X}_0 \times \mathcal{Y}_0, \sqrt{n} \mathcal{A}_{H,i} - \mathcal{A}_{H,i}^* = -\mathcal{B}_{m,i}(X_i, Y_i; x^*)\)
follows a distribution \( \mathcal{N}(0, \sigma_{m,i}^2) \).

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 \( w_i = (x_i', y_i', z_i') \), and we introduce the following notation:

\[
S_{p,b}(\tau; u) = n^{-1} \sum_{i=1}^{n} K_b(U_i - u) g(G^{-1}(\tau|U_i)) \mu_{p,b}(U_i - u) \mu_{p,b}(U_i - u)', \quad (5.1)
\]

\[
\eta_{1k}(\tau; u) = e_{1,\mathcal{P}} S_{p,b}(u) \mu_{p,b}(U_k - u) K_b(U_k - u) g(G^{-1}(\tau|x^*, z)|x^*, z),
\]

\[
\eta_{2k}(\tau; u) = e_{1,\mathcal{P}} S_{p,b}(u) \mu_{p,b}(U_k - u) K_b(U_k - u) K_b(U_k - u),
\]

\[
\zeta_0(W_i, W_k, z) = \eta_{1k}(\tau_{iz}; X_i, z) Y_i(W_k) + \eta_{2k}(\tau_{iz}; x^*, z) \psi_{x^*}(Y_k - G^{-1}(\tau_{iz}|U_k))
\]

and

\[
\varphi(w_i, w_j) = E \left[ \int \zeta_0(W_i, w_i; z) \zeta_0(W_i, w_j; z) \mathcal{D}(z) \right],
\]

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

\[
\mathbb{E}_{\mathcal{J}_n} \equiv n^{-2} b^{d_x} \sum_{j=1}^{n} \int \zeta_0(W_i, W_j; z) \mathcal{D}(z) \pi_i, \quad \text{and} \quad \sigma_{2n}^2 = 2b^{d_x} E[\varphi(W_i, W_j)^2]. \quad (5.2)
\]
To establish the asymptotic properties of $\hat{J}_n$, we add the following condition on the bandwidth.

**Assumption C.7** As $n \to \infty$, $n b^{4d_X} \to \infty$, and $n b^{3d/2}/(\log n)^2 \to \infty$.

Assumptions C.6 and C.7 imply that a higher order local polynomial (i.e., $p \geq 2$) may be required in the case where $d_X$ or $d_Z$ is large in order to ensure that $p + 1 - d_Z/2 > 0$ and $2(p + 1) + d_X > \max(2d_X, d + 2d_Z, 3d/2)$. By choosing $b \propto n^{-1/2(p+1)+d_X}$, it suffices to set $p = 1$ if $d_X \leq 3$ and $d_Z = 1$. A 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 $\hat{J}_n - \mathbb{B}_{J_n} \overset{d}{\to} N(0, \sigma^2_{J})$, where $\sigma^2_{J} = \lim_{n \to \infty} \sigma^2_{J_n}$.

**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, $\eta_{1k}(\tau_{iiz} | X_i, z) \hat{I}_Y(W_k)$, in the definition of $\zeta_0$ reflects the influence of the first-stage estimator $\hat{G}_{p,b}(Y_i | X_i, z)$, whereas the second term $\eta_{2k}(\tau_{iiz}; x^*, z) \psi_{\tau_{iiz}}(Y_k - G^{-1}(\tau_{iiz} | U_k))$ embodies the effect of the second-stage estimator $\hat{G}_{p,b}^{-1}(\tau; x^*, z)$ evaluated at $\tau = \hat{G}_{p,b}(Y_i | X_i, z)$. 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 $\sigma^2_{J_n}$ shows that they contribute asymmetrically to the asymptotic variance: the asymptotic variance of $\hat{J}_n$ is mainly determined by the second-stage estimator, whereas the role played by the first-stage estimator is asymptotically negligible. The main reason for this is that the term $X_i$ in $\eta_{1k}(\tau_{iiz}; X_i, z)$ (but not the term $x^*$ appearing in $\eta_{2k}(\tau_{iiz}; x^*, z)$) is subject to a (smooth) expectation operator in the definition of $\varphi(w_i, w_j)$, which helps reduce the variation of the first-stage estimator. For the same reason, we need $b^{4d_X}$ instead of the usual term $b^{4/2}d_X$ as the normalization constant in the definition of $\hat{J}_n$, which avoidably reduces the size of the class of local alternatives that this test has power to detect.

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

$$\hat{\zeta}_0(W_i, W_k; z) = \frac{1}{\hat{g}_{iz}} \left[ e_1s_{p,b}(X_i, z) - 1 \mu_b(X_k - X_i, Z_k - z) K_b(X_k - X_i, Z_k - z) \hat{I}_Y(W_k) + e_1s_{p,b}(x^*, z) - 1 \mu_b(X_k - x^*, Z_k - z) K_b(X_k - x^*, Z_k - z) \psi_{\tau_{iiz}}(Y_k - \hat{G}_{p,b}^{-1}(\tau_{iiz} | U_k)) \right],$$

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

$$\hat{\mathbb{B}}_{J_n} = n^{-2} b^{4d_X} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[ \int \hat{\zeta}_0(W_i, W_j; z) d\Delta(z) \right]^2 \pi_i,$$

and

$$\hat{\sigma}^2_{J_n} = \frac{2b^{4d_X} b^{3d/2}}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} \sum_{k=1}^{n} \left[ \int \hat{\zeta}_0(W_i, W_j; z) d\Delta(z) \int \hat{\zeta}_0(W_i, W_j; z) d\Delta(z) \pi_i \pi_j \pi_k \pi_l \right]^2.$$

3In 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 \( \hat{B}_{J_n} - B_{J_n} = o_P(1) \) and \( \hat{\sigma}^2_{J_n} - \sigma^2_{J_n} = o_P(1) \). Then we can compare

\[
T_n \equiv \left( \hat{J}_n - \hat{B}_{J_n} \right) / \sqrt{\hat{\sigma}^2_{J_n}}
\]

(5.3)
to the critical value \( z_{\alpha} \), the upper \( \alpha \) percentile from the \( \mathcal{N}(0,1) \) 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

\[
\mathbb{H}_1 (c_n) : Y_{ni} = m(X_i, A_i) + c_n \gamma(X_i, A_i),
\]

(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, \( \{Y_{ni}, 1 \leq i \leq n\} \) 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.\(^4\) Let \( G_n^{-1}(|x, z) \) denote the inverse function of \( G_n (|x, z) \). 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 \( F_{A|Z} (|z) \) and \( f_{A|Z} (|z) \) denote the conditional CDF and PDF if \( A_i \) given \( Z_i = z \), respectively.

We add the following assumption.

**Assumption A.4** \( f_{A|Z} (|z) \) 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

\[
G_n^{-1}(G_n(y|x, z)|x^*, z) = G^{-1}(G(y|x, z)|x^*, z) + c_n \Theta_n^1(y|x, z)
\]

for all \( (y, x, z) \in \mathcal{Y}_0 \times \mathcal{X}_0 \times \mathcal{Z}_0 \) such that \( g(G^{-1}(G(y|x, z)|x^*, z)|x^*, z) > 0 \), where \( \Theta_n^1(y|x, z) \) 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

\[
\int G_n^{-1}(G_n(Y_{ni}|X_i, z)|x^*, z))d\Delta(z) = \int G^{-1}(G(Y_{ni}|X_i, z)|x^*, z) d\Delta(z) + c_n \int \Theta_n^1(Y_{ni}|X_i, z) d\Delta(z)
\]

\[
= \int G^{-1}(G(Y_{ni}|X_i, z)|x^*, z) d\Delta(z) + c_n \Theta_n(Y_i; X_i, A_i)
\]

\[
= c_n \Theta_n(Y_i; X_i, A_i),
\]

\(^4\)With a little bit more complicated notation, one could also allow \( \{(X_i, Z_i, A_i)\} \) to be a triangular array.
where

\[
\Theta_n (Y_i; X_i, A_i) = \int \left[ \gamma (X_i, A_i) \frac{g (Y_i + tc_\gamma (X_i, A_i) | X_i, z)}{g (G^{-1} (G (Y_i + tc_\gamma (X_i, A_i) | X_i, z) | x^*, z) | x^*, z, z)} dt + \Theta_n^\dagger (Y_i + c_\alpha \gamma (X_i, A_i); X_i, z) \right] d\Delta (z)
\]

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

where \( \approx \) holds by ignoring terms that are of smaller order term. 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 \( G (y | x, z) = P [m (x, A_i) \leq y | X_i = x, Z_i = z] = F_{A_i | Z} (m^{-1} (x, y) | z) \). It follows that \( g (y | x, z) = f_{A_i | Z} (m^{-1} (x, y) | z) \frac{\partial m^{-1} (x, y)}{\partial y} \). Under the normalization rule \( m (x^*, a) = a \forall a \), we have \( m^{-1} (x^*, a) = a \forall a \) and

\[
\frac{g (Y_i | X_i, z)}{g (A_i | x^*, z)} = \frac{f_{A_i | Z} (m^{-1} (X_i, Y_i) | z)}{f_{A_i | Z} (m^{-1} (x^*, A_i) | z)} = \frac{\partial m^{-1} (X_i, Y_i)}{\partial y},
\]

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 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 holds with \( H \) replaced by \( H_1 \) and \( H_2 \). Suppose \( c_{0n} = \lim_{n \to \infty} E [\Theta_n (Y_i; X_i, A_i)^2 \pi (X_i, Y_i)] \) exists. Then under \( \mathbb{H}_1 (c_n) \) with \( c_n = n^{-1/2}b^{-d/2}, T_n \overset{d}{\to} N (c_{0n} / \sigma_f, 1) \).

**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 \( n^{-1/2}b^{-d/2} \), provided \( 0 < c_{0n} < \infty \). The asymptotic local power function of the test is given by \( 1 - \Phi (z_\alpha - c_{0n} / \sigma_f) \), where \( \Phi \) is the standard normal CDF. Unfortunately, we are unable to characterize conditions under which the limit \( c_{0n} \) is strictly positive, ensuring the non-trivial power of our test. In the supplementary appendix (Appendix E), we consider a simple example where

\[
m (X_i, A_i) = (1 + 0.1X_i^2) A_i \quad \text{and} \quad \gamma (X_i, A_i) = X_i e^{-A_i},
\]

where the PDFs of \( X_i \) and \( A_i \) have supports \( \mathcal{X} = \left[ \bar{c}_\gamma, c_\gamma \right] \) and \( \mathcal{A} = (-\infty, \bar{c}_\alpha] \), respectively, where \( \bar{c}_\gamma < 0 \), \( \bar{c}_\gamma > 0 \), and \( \bar{c}_\alpha > 0 \). It is possible that \( \bar{c}_\gamma = -\infty, \bar{c}_\gamma = \infty \), and \( \bar{c}_\alpha = \infty \), 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 \( a \),

\[
P (A_i \leq -a | Z_i = z) = \phi (e^{-a}, z),
\]

where \( \phi (\cdot, \cdot) \) is continuously differentiable with respect to its first argument (as a CDF), a non-constant function of \( z \), and \( \lim_{a \to 0} \phi (t, z) \to 0 \). Essentially, the condition in (5.5) requires that \( F_{A_i | Z} (a | z) \) decay at the exponential rate as \( a \to -\infty \). In this case we can show that for any \( y > 0 \),

\[
\Theta_n^\dagger (y; x, z) = \left[ \frac{x}{v (x)} e^{-
abla x} + \omega (y; x, z) + o (1) \right] 1 \{ x > 0 \} + \left[ \frac{x}{v (x)} e^{-
abla x} + o (1) \right] 1 \{ x \leq 0 \}.
\]

\footnote{Observe that \( P (Y_{n,i} \leq 0) = P \left[ (1 + 0.1X_i^2) A_i + c_\alpha X_i e^{-A_i} \leq 0 \right] \to 0 \) as \( c_\alpha \to 0 \).}
where \( v(x) = 1 + 0.1x^2 \) and \( \omega(y; x, z) \) is a function that is nonconstant in \( z \) and is defined in (E.14). In this case, we have \( c_{\Theta_0} = E \left\{ \left[ \int \omega(Y_i; X_i, z) d\Delta z \right]^2 \pi (X_i, Y_i) \right\} \), which is generally positive.

**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

\[
m(X_i, A_i) = -1 + X_i + A_i, \quad \gamma(X_i, A_i) = X_i A_i,
\]

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,

\[
G(y|x, z) = F_{A|Z} \left( y + 1 - x | z \right),
\]

\[
G^{-1}(\tau|x, z) = x - 1 + F_{A|Z}^{-1}(\tau|z),
\]

\[
G_n(y|x, z) = F_{A|Z} \left( \frac{y + 1 - x}{1 + c_n x} | z \right) e_n(x) + \bar{F}_{A|Z} \left( \frac{y + 1 - x}{1 + c_n x} | z \right) \hat{e}_n(x),
\]

\[
G_n^{-1}(\tau|x, z) = x - 1 + (1 + c_n x) \left[ F_{A|Z}^{-1}(\tau|z) e_n(x) + F_{A|Z}^{-1}(1 - \tau|z) \hat{e}_n(x) \right],
\]

where \( \bar{F}_{A|Z} = 1 - F_{A|Z} \), \( e_n(x) = 1 \{1 + c_n x > 0\} \), and \( \hat{e}_n(x) = 1 \{1 + c_n x < 0\} \). 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

\[
g_n^{-1}(G_n(Y_{ni}|X_i, z)|x^*, z) = x^* - 1 + (1 + c_n x^*) F_{A|Z}^{-1}(A_i|z) e_n(X_i) \approx A_i + c_n A_i + \frac{1 - 2F_{A|Z}(A_i|z)}{f_{A|Z}(A_i|z)} \hat{e}_n(X_i),
\]

where we ignore terms that are \( o_P(c_n) \). Noting that for any \( \epsilon > 0 \), \( P(\hat{e}_n(X_i) > c_n) \leq P(X_i < -\frac{1}{c_n}) \to 0 \), 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 \( m_n(x, a) = -1 + x + a + c_n x a \) 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 \( G_n^{-1}(G_n(Y_{ni}|X_i, z)|x^*, z) - G^{-1}(G(Y_i|X_i, z)|x^*, z) \) 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 \( \Delta \) to maximize \( c_{\Theta_0}/\sigma_J \), where both \( c_{\Theta_0} \) and \( \sigma_J \) depend on \( \Delta \) implicitly. Noting \( \sigma_J^2 \) is the limit of \( \sigma_J^2_n = 2b^{d x} E[\varphi(W_1, W_2)^2] \) with \( \varphi(w_1, w_2) \equiv E \left[ \int \int \zeta_0(W_1, w_1; z) \zeta_0(W_1, w_2; z) d\Delta(z) d\Delta(z) \pi(X_i, Y_i) \right] \), the \( \Delta \) function enters \( \sigma_J^2 \) fourfold. Similarly, it enters \( c_{\Theta_0} \) 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 $\Delta$. In addition, since the above local power function depends on the particular local alternative through $\Theta_\mu$ 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 thus we leave it for future research.

On the other hand, as the associate editor kindly indicates, one can consider improving the power performance of our test by studying the following optimization problem

$$\sup_{\Delta \in \Xi} T_n(\Delta)$$

where $\Xi = \{H_1 - H_2 : \text{both } H_1 \text{ and } H_2 \text{ are CDFs defined on } \mathbb{R}^d \text{ with continuous PDFs}\}$ and we have made the dependence of $T_n$ on $\Delta$ explicit. Without any restriction on the size of $\Xi$, it seems impossible to consider the above optimization problem. Nevertheless, if $\Xi$ 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 $\Delta$ is already computationally expensive. Instead, in the simulations we will consider the effect of different choices of $\Delta$ on the size and power of our test.

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

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

**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 $c_n = o(n^{d_x})$.

The Appendix in the early version of the paper discusses how proper choice of $H_1$ and $H_2$ assures $c_{H_1} > 0$ when the conditional quantile representation (CQR) of the structural function in Proposition 2.1 fails. It sheds light on the primitive alternatives that our test has power to detect.

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

2. Draw a bootstrap sample $\{Z_i^*\}_{i=1}^n$ from the smoothed kernel density $\tilde{f}_Z(z) = n^{-1} \sum_{i=1}^n \phi_{\xi_i} (Z_i - z)$, where $\phi_b(z) = b^{-d_x} \phi(z/b)$, $\phi(\cdot)$ is the standard normal PDF in the case where $Z_i$ is scalar valued.
and becomes the product of univariate standard normal PDF otherwise, and \( b_2 > 0 \) is a bandwidth parameter.

3. For \( i = 1, \ldots, n \), given \( Z_i^* \), draw \( X_i^* \) and \( A_i^* \) independently from the smoothed conditional density 

\[
\hat{f}_{X|Z}(x|Z_i^*) = \sum_{j=1}^{n} \phi_{b_z}(Z_j - x) \phi_{b_z}(Z_j - Z_i^*) / \sum_{j=1}^{n} \phi_{b_z}(Z_i - Z_i^*) \text{ and } \hat{f}_{A|Z}(a|Z_i^*) = \sum_{j=1}^{n} \phi_{b_z}(A_j - a) \phi_{b_z}(Z_j - Z_i^*) / \sum_{j=1}^{n} \phi_{b_z}(Z_i - Z_i^*)
\]

respectively, where \( b_z, b_x, \) and \( b_\alpha \) are given bandwidths.

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

5. Compute a bootstrap statistic \( T_{n}^{*} \) in the same way as \( T_{n} \), with \( W_{i}^{*} = \{W_{i}^{*} = (X_{i}^{*}, Y_{i}^{*}, Z_{i}^{*})\}_{i=1}^{n} \) replacing \( W_{n} \).

6. Repeat Steps 2-5 \( N_B \) times to obtain bootstrap test statistics \( \{T_{n,j}^{*}\}_{j=1}^{N_B} \). Calculate the bootstrap 

\[
P^* \equiv N_B^{-1} \sum_{j=1}^{N_B} 1\{T_{n,j}^{*} \geq T_{n} \} \text{ and reject the null hypothesis if } P^* \text{ is smaller than the}
\]

prescribed nominal level of significance.

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

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

\[
(i) \ P(T_{n}^{*} \leq t|W_{n}) \to \Phi(t) \text{ for all } t \in \mathbb{R}, \text{ and } (ii) \ P(T_{n} > z_{n}^{*}) \to 1 \text{ under } H_1, \tag{5.7}
\]

where \( z_{n}^{*} \) is the \( \alpha \)-level bootstrap critical value based on \( N_B \) bootstrap resamples, i.e., \( z_{n}^{*} \) is the \( 1 - \alpha \) quantile of the empirical distribution of \( \{T_{n,j}^{*}\}_{j=1}^{N_B} \).

### 5.4 Asymptotics with generated regressors

In some applications, e.g., the control function approach, the instrument \( Z \) is not directly observed. To get a proxy for \( Z \), one may assume that we can recover \( Z \) via the following relation between \( Q, \varpi, \) and \( Z \),

\[
Q = Z + \varepsilon = r_0(\varpi) + \varepsilon, \tag{5.8}
\]

where \( Q \times \varpi \) takes value in \( \mathcal{Q} \times \Omega_{\varpi} \), \( \varepsilon \) is a \( d_Z \times 1 \) random component such that \( E(\varepsilon|\varpi) = 0 \) and \( E(\|\varepsilon\|^2_2) < \infty \). Apparently, \( Z = E(Q|\varpi) = r_0(\varpi) \equiv (r_{0,1}(\varpi), \ldots, r_{0,d_Z}(\varpi))^\top \). We use \( d_{\varpi} \) to denote the dimension of \( \varpi \) and assume that \( r_{0,\ell}: \mathbb{R}^{d_{\varpi}} \to \mathbb{R}, \ell = 1, \ldots, d_Z, \) are measurable functions. Here we consider the \( \hat{p} \)-th order local polynomial estimator of \( r_{0,\ell} \):

\[
\hat{r}_{\ell}(\varpi) = e_{1,\hat{p}}S_{\hat{p},\hat{b}}(\varpi)^{-1}n^{-1} \sum_{i=1}^{n} \hat{K}_{\hat{b}}(\varpi_i - \varpi) \mu_{\hat{p},\hat{b}}(\varpi_i - \varpi) Q_{i,\ell}, \tag{5.9}
\]

where \( e_{1,\hat{p}}, S_{\hat{p},\hat{b}}(\varpi), \mu_{\hat{p},\hat{b}}(\varpi_i - \varpi) \) are defined similarly as before, \( \hat{K}_{\hat{b}}(\cdot) \equiv \hat{K}(\cdot/\hat{b})/\hat{b} \), \( \hat{K}(\cdot) \) is a PDF on \( \mathbb{R}^{d_{\varpi}} \), and \( Q_i = (Q_{i,1}, \ldots, Q_{i,d_Z})^\top, i = 1, \ldots, N, \) are IID copy of \( Q \). Let \( \hat{r}(\varpi) = (\hat{r}_1(\varpi), \ldots, \hat{r}_{d_Z}(\varpi))^\top \). For
the convenience of exposition, we use the same bandwidth \( \tilde{b} \) for each element in \( \omega \). In this subsection, we consider the asymptotics of our test with generated regressors \( \hat{r} \).

In many cases, it is very tedious or practically impossible to obtain the pathwise derivatives of estimators with generated regressors as in Newey (1994), due to the potentially complicated structure of the estimators. An emerging literature bypasses this difficulty by using empirical process theory; see, e.g., Ichimura and Lee (2010), Mammen et al. (2012), and Escanciano et al. (2014). The asymptotics of our estimator and statistics with generated regressors are derived by adopting this technique. To this end, we introduce the following definitions, notations, and assumptions.

Let \( \mathcal{R} \) denote a class of functions from \( \mathbb{R}^d_{\omega} \rightarrow \mathbb{R}^d_{\omega} \), and \( r_0(\cdot) \in \mathcal{R} \). Let \( \|\cdot\|_{\mathcal{R}} \) be a generic pseudo-norm on \( \mathcal{R} \). Given two vector functions \( r_t, r_u \), a bracket \( [r_t, r_u] \) is set of vector functions \( r \in \mathcal{R} \) such that \( r_{t,\ell} \leq r_{u,\ell} \) for all \( 1 \leq \ell \leq d_Z \), where, e.g., \( r_{t,\ell} \) denotes the \( \ell \)th element of \( r \). An \( \epsilon \)-bracket with respect to \( \|\cdot\|_{\mathcal{R}} \) is a bracket \([r_t, r_u]\) with \( \|r_t - r_u\|_{\mathcal{R}} \leq \epsilon \), \( r_{t}\|_{\mathcal{R}} \leq \infty \), \( r_{u}\|_{\mathcal{R}} \leq \infty \). The covering number with bracketing \( \mathcal{N}_\epsilon(r) \) is the minimal number of \( \epsilon \)-bracket with respect to \( \|\cdot\|_{\mathcal{R}} \) needed to cover \( \mathcal{R} \). For the norms used in this section, we let \( \|\cdot\|_2 \) and \( \|\cdot\|_\infty \) denote the \( L_2 \) and sup-norms, respectively: \( \|r\|_2 = \{\sum_{t=1}^{d_Z} r_t^2(\omega) dP(\omega)\}^{1/2} \) and \( \|r\|_\infty = \sup_{\omega \in \Omega_{\omega}} \max_{1 \leq \ell \leq d_Z} |r_{\ell}(\omega)| \), where \( P \) denotes the probability measure associated with \( \omega \).

We typically require that \( \mathcal{R} \) should not be too complex, e.g., \( \log \mathcal{N}_\epsilon(\epsilon, \mathcal{R}, \|\cdot\|_\infty) < C_\epsilon^{-c_\epsilon} \) for some \( c_\epsilon < 2 \) and for all \( \epsilon > 0 \). van der Vaart and Wellner (1996) provide examples of classes of functions satisfying the above requirement. The functional space \( C^0_M(\Omega_{\omega}) \) given in the following definition is one of them with \( c_\epsilon = d_{\omega}/\bar{\varrho} \).

Definition. Let \( \varrho > 0 \), \( M > 0 \), and \( \Omega_{\omega} \) be a finite union of convex and bounded subsets of \( \mathbb{R}^d_{\omega} \). Let \( C^0_M(\Omega_{\omega}) \) denote a class of smooth functions such that for any smooth function \( f : \Omega_{\omega} \rightarrow \mathbb{R} \) in it, we have \( \|f\|_{\infty,\varrho} \leq M \), where \( \|f\|_{\infty,\varrho} = \max_{D^j f = g} \sup_{\omega \in \Omega_{\omega}} \left| D^j f(\omega) \right| + \max_{D^j f = g} \sup_{\omega \neq \omega'} \frac{|D^j f(\omega) - D^j f(\omega')|}{\|\omega - \omega'\|^2} \), and \( \varrho \) be the largest integer smaller than \( \varrho \).

Assumption C.8 \( \Omega_{\omega} \) is a finite union of convex and bounded subsets of \( \mathbb{R}^d_{\omega} \) with non-empty interior. \( r_\ell \in C^0_M(\Omega_{\omega}) \) for \( \ell = 1, ..., d_Z \), \( \varrho/d_{\omega} > 1/2 \), and \( \bar{\varrho} \geq \bar{\varrho} + 1 \).

In addition, we add the following technical assumption:

Assumption C.9 As \( n \rightarrow \infty \), \( b \rightarrow 0 \), \( \|\hat{r} - r\|_\infty = O_P(n^{-1/4} b) \), \( \|\hat{r} - r\|_\infty^{-1/2} c_r = O_P(b^{1+d_Z/2}) \), \( n b^{d_Z} \tilde{b}^{2(\bar{\varrho}+1)} \rightarrow 0 \), and \( \tilde{b}/b \rightarrow 0 \).

The rate restrictions on \( \hat{r} \) in Assumption C.9 are essentially the same as those in Mammen et al. (2012) (see the rate \( n^{-\kappa} \) in their paper). We require that \( \hat{r} \) converge faster than the usual rate \( n^{-1/4} \) because \( \hat{r} \) appears in the kernel function \( K_b(\cdot) \) and Taylor expansions will result in the appearance of \( b^{-1} \). As a referee kindly points out, the requirement that \( \|\hat{r} - r\|_\infty^{1-2/c_r} = O_P(b^{1+d_Z/2}) \) could be very restrictive with respect to the smoothness of \( r \). In view of the fact that \( c_r = d_{\omega}/\varrho \) for the functional space we are interested in, the smaller \( d_{\omega} \) is or the larger \( \varrho \) is, the more likely for this rate restriction on \( \hat{r} - r \) to hold. Indeed,
our asymptotic analysis requires that
\[ ||\hat{r} - r||_\infty = o_P\left( b^{2+\frac{d}{2}} \right) = o_P\left( b^{2+\frac{d}{2}+\gamma} \right). \]

\( b^{2+\frac{d}{2}+\gamma} \) can be smaller than \( o(b^2) \) if \( d_\infty/\varrho > 1 \). In general, we need a smoother class of functions (larger \( \varrho \)) for larger value of \( d_\infty \) in order for the above condition to hold, reflecting the curse of dimensionality issue. See Ai and Chen (2003) for a similar requirement. \( n\beta^d\hat{d}^{p(\hat{p}+1)} \to 0 \) is imposed to ensure the bias term associated with the estimation of \( r \) does not contribute to our asymptotic analysis. \( \hat{b} = o(b) \) is imposed to obtain the asymptotic linear expansion as in the first case of Proposition 1 in Mammen et al. (2012); it is not necessary for the final results but helps to simplify the proofs in various places. In the case when \( b \propto n^{-1/2(p+1)+d_x} \) (Assumption C.6), and \( \hat{r} \) attains the minimum requirement of convergence rate in Assumption C.9, e.g., \( ||\hat{r} - r||_\infty = n^{-1/4+1/[2(p+1)+d_x]} / \log(n) \), then we need \( c_\gamma < \frac{-2-4d_\infty/[2(p+1)+d_x]}{1+4/[2(p+1)+d_x]} \), \( \hat{b} \propto n^{-1/2(p+1)+d_x} / \log(n) \), and \( \hat{p} > p - d_\infty/2 \). In our applications, we have \( d_x = 2 \) and \( d_\infty = 1 \), but only the second application has nonparametrically generated \( Z \)'s with \( d_\infty = 2 \). Over there, by choosing \( p = 1 \) and \( \hat{p} = 2 \), one can set \( b \propto n^{-1/6}, \hat{b} \propto n^{-1/6}/\log(n) \), and we require \( c_\gamma < 4/5 \) or equivalently \( \varrho > 5/2 \) when \( d_\infty = 2 \). This condition would be met for a class of functions that are three times continuously differentiable.

To simplify notation, let \( \theta \) be the vector that stacks \( \sqrt{n}\theta^T(\alpha_j - \frac{\beta_j}{\hat{b}}D^jG^{-1}(\tau|u)) \) for \( 0 \leq |j| \leq p \) in lexicographic order. Let \( U_i(r) \equiv (X_i,r(\varpi_i))^T, \mu_{p,b,i}(r) \equiv \mu_{p,b}(U_i(r) - u), \hat{K}_b(\cdot) \equiv K(\cdot/b), \hat{K}_{b,i}(r) \equiv K((U_i(r) - u)/b), \) and \( Y_i^\ast(r) = Y_i - \sum_{0 \leq |j| \leq p} \frac{1}{\hat{b}}D^jG^{-1}(\tau|u)(U_i(r) - u)^j \). Note that \( Y_i^\ast(r_0) \equiv Y_i - \beta_b(\tau;u), \) where \( \beta_b(\tau;u) \) is the expansion of \( G^{-1}(\tau|U_i(r_0)) \) around \( u \).

Rewrite the original problem in terms of \( \theta \) with one additional term \( \rho_\tau(Y_i^\ast(r_0)) \), and to control the magnitude of the objective function, we use \( \hat{K}_{b,i}(\cdot) \) instead of \( K_{b,i}(\cdot) \) to obtain:

\[
\min_{\theta} \sum_{i=1}^{n} \left[ \rho_\tau \left( Y_i^\ast(r_0) - \mu_{p,b,i}(r_0) \frac{\theta}{\sqrt{n}\hat{b}} \right) - \rho_\tau \left( Y_i^\ast(r_0) \right) \right] \hat{K}_{b,i}(r_0). \tag{5.10}
\]

Since \( \rho_\tau(Y_i^\ast) \) is not a function of \( \theta \) and \( \hat{K}_{b,i}(r_0) = b^dK_b(U_i-u) \), this problem is equivalent to the original problem in (4.4). The problem with generated regressors could be similarly rewritten as

\[
\min_{\theta} \sum_{i=1}^{n} \left[ \rho_\tau \left( Y_i^\ast(\hat{r}) - \mu_{p,b,i}(\hat{r}) \frac{\theta}{\sqrt{n}\hat{b}} \right) - \rho_\tau \left( Y_i^\ast(\hat{r}) \right) \right] \hat{K}_{b,i}(\hat{r}). \tag{5.11}
\]

Denote the minimizers from problem (5.11) as \( \hat{\theta} \). Then we can recover the estimate for \( G^{-1}(\tau|u) \) by \( \hat{G}^{-1}(\tau|u) \) via:

\[
\hat{G}^{-1}(\tau|u) \equiv G^{-1}(\tau|u) + \frac{1}{\sqrt{n}\min|x_p|} \hat{\theta}.
\]

**Proposition 5.5** Suppose Assumptions C.1-C.4, C.8, and C.9 hold, then \( \hat{G}^{-1}(\tau|u) - G^{-1}(\tau|u) = o_P((nb\hat{d}_x)^{-1/2}). \)

Let's turn to the following nonparametric regression with generated regressors using the local polynomial regression technique. Let \( \hat{U}_i \equiv (X_i^T,\hat{r}(\varpi_i))^T \). Consider the following minimization problem:

\[
\min_{\beta} \frac{1}{n} \sum_{i=1}^{n} \left( 1\{Y_i \leq y\} - \sum_{0 \leq |j| \leq p} \beta_j \left( (\hat{U}_i - u)/b \right)^j \right)^2 K_b \left( \hat{U}_i - u \right). \tag{5.12}
\]
Denote the solution to the above problem as $\tilde{\beta}(y|u)$. Let $\tilde{G}_{p,b}(y|u) \equiv e_1^\dagger \mu \tilde{\beta}(y|u)$. Recall that the corresponding solution to the original problem (4.1) with observed $U_i$ is denoted as $\hat{\beta}(y|u)$ and that $\hat{G}_{p,b}(y|u) = e_1^\dagger \hat{\beta}(y|u)$.

**Proposition 5.6** Suppose Assumptions C.1-C.4, C.8, and C.9 hold. Then $\tilde{G}_{p,b}(y|u) = \tilde{G}_{p,b}(y|u) + o_P((nb^{d_s})^{-1/2})$.

**Remark 5.6.** By Proposition 5.5, we know that $\tilde{G}^{-1}(\tau|u)$ satisfies Lemma C.2. By Proposition 5.6, we know that $\tilde{G}_{p,b}(y|u)$ satisfies Lemma C.1(b) and Lemma C.1(a) with a small order remainder term $o_P((nb^{d_s})^{-1/2})$. From Theorems 5.1 and 5.3, the asymptotic distribution of the test statistic depends on the interactions of the linear representations of $\tilde{G}^{-1}(\tau|u) - G^{-1}(\tau|u)$ and $\tilde{G}_{p,b}(y|u) - G_{p,b}(y|u)$. We can readily show that all the asymptotic results for the statistic with the generated $Z$ remain the same as the case of observed $Z$ by using the results in Proposition 5.5 and 5.6. Note that our results do rely on the stronger than usual restriction on the convergence rate of $\hat{\rho}$ (Assumption C.9) which can be verified. Either assumption is standard in the literature on empirical process theory. The key steps to derive the final results are given by equations (D.7) and (B.20) in the appendix where $X$ in $K_b(U)$ are integrated out in the empirical process expectation calculation for the additional term resulted from generated regressors, giving a faster rate of convergence. This is in line with the asymptotics in IN, where only a subset of the regressors are estimated nonparametrically. This is different from the results in Mammen et al. (2012, Proposition 1) and Escanciano et al. (2014) where the asymptotics contain an additional term. This discrepancy comes from the fact that in their cases all regressors are not directly observed and generated via nonparametric regressions. The model setup in (5.8) includes the dependent variable $Q$ being a function of a subvector of $X$ as a special case, and all results in this session continue to go through with some modification. See Section 7.2 for details.

### 6 Estimation and specification testing in finite samples

In this section, we conduct a series of Monte Carlo experiments to evaluate the finite-sample performance of our estimators and tests. We first consider the estimation of the response under correct specification 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:** $Y_i = (0.5 + 0.1X_i^2)A_i$.

---

6 The regressors here are $X$ and $Z$ where $X$ is observed but $Z$ is not.
DGP 2: $Y_i = A_i/(1 + 0.1X_i^2)$, where $i = 1, \ldots, n$, $A_i = 0.5Z_i + \varepsilon_{1i}$, $X_i = 0.25 + Z_i - 0.25Z_i^2 + \varepsilon_{2i}$, and $\varepsilon_{1i}$, $\varepsilon_{2i}$, and $Z_i$ are each IID $N(0, 1)$ and mutually independent. Clearly, $m(x, a) = (0.5 + 0.1x^2)a$ in DGP 1 and $= a/(1 + 0.1x^2)$ 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 $a \in A$, where $x_{med} = 0.116$ is the population median of $X_i$.

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 $a^* \in A^*$, where $A^*$ is the support of $A_i/c_1$, we can solve for $c_1$ to obtain $c_1 = 1/(0.5 + 0.1x_{med}^2) = 1.9946$. For DGP 2, similarly, we let $a^* = a/c_2$, $c_2 = 1/(1 + 0.1x_{med}^2) = 0.9987$. For notational simplicity, we continue to use $m(x, a)$ and $A_i$ to denote $m^*(x, a)$ and $A^*$, respectively.

To estimate the response $m(x, a)$, we need to choose the local polynomial order $p$, the kernel function $K$, the bandwidth $b$, and the weight function $H$. Since $d_X = d_Z = 1$, it suffices to choose $p = 1$ to obtain the local linear estimates $\hat{G}_{p,b}(a|x^*, z)$ and $\hat{G}_{p,b}^{-1}(\hat{G}_{p,b}(a|x^*, z)|x, z)$, 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: $b = (1.06\sigma_X n^{-1/5}, 1.06\sigma_X n^{-1/6})$, where, e.g., $\sigma_X$ 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 $[C_{q_0}, C_{1-q_0}]$, and $H_2$ is a scaled beta$(2, 2)$ distribution on $[C_{q_0}, C_{1-q_0}]$, where $C_{q_0}$ is the $q_0$th sample quantile of $\{Z_i\}_{i=1}^n$ and $q_0 = 0.05$. 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 $(x_j, a_j), j = 1, \ldots, 225$, denote these values. We obtain the estimates $\hat{m}_{H_1}(x, a), l = 1, 2$, of $m(x, a)$ at these 225 points, and calculate the corresponding mean absolute deviations (MADs) and root mean squared errors (RMSEs):

$$MAD_{H_1}^{(i)} = \frac{1}{225} \sum_{j=1}^{225} \left| m(x_j, a_j) - \hat{m}_{H_1}^{(i)}(x_j, a_j) \right|, \quad (6.1)$$

$$RMSE_{H_1}^{(i)} = \left\{ \frac{1}{225} \sum_{j=1}^{225} \left[ m(x_j, a_j) - \hat{m}_{H_1}^{(i)}(x_j, a_j) \right]^2 \right\}^{1/2}, \quad (6.2)$$

where, for $i = 1, \ldots, 500$, $\hat{m}_{H_1}^{(i)}(x_j, a_j)$ is the estimate of $m(x_j, a_j)$ in the $i$th replication with weight function $H_1$, $l = 1, 2$. For feasibility in computation, we consider two sample sizes in our simulation study, namely, $n = 100$ and 400.

---

7For DGP 1, $m(x, a) = a$ for all $a$ at $x = x^* = \pm \sqrt{\pi}$. 22
Table 1: Finite sample performance of the estimates of the response

<table>
<thead>
<tr>
<th>DGP</th>
<th>n</th>
<th>H</th>
<th>MAD_H</th>
<th>RMSE_H</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>5th</td>
<td>50th</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>Uniform</td>
<td>0.192</td>
<td>0.269</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta</td>
<td>0.158</td>
<td>0.222</td>
</tr>
<tr>
<td>400</td>
<td>Uniform</td>
<td>0.126</td>
<td>0.175</td>
<td>0.249</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta</td>
<td>0.093</td>
<td>0.132</td>
</tr>
<tr>
<td>2</td>
<td>100</td>
<td>Uniform</td>
<td>0.262</td>
<td>0.374</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta</td>
<td>0.208</td>
<td>0.295</td>
</tr>
<tr>
<td>400</td>
<td>Uniform</td>
<td>0.169</td>
<td>0.239</td>
<td>0.352</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Beta</td>
<td>0.120</td>
<td>0.169</td>
</tr>
</tbody>
</table>

Table 1 reports the 5th, 50th, and 95th percentiles of \( \text{MAD}_H^{(i)} \) and \( \text{RMSE}_H^{(i)} \) 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 \( \hat{m}_{H_1}^{(i)}(x,a) \) 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 DGP 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. In theory, our normalized estimators converge to the standard normal and the quantiles of our normalized estimates should coincide with those of the normal asymptotically. We see that the points 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**: \( Y_i = (0.5 + 0.1X_i^2)A_i + 2\delta_0X_i/(0.1 + e^{A_i^2/2}) \),
- **DGP 4**: \( Y_i = A_i/(1 + 0.1X_i^2) + 2\delta_0X_ie^{0.1A_i} \),

where \( i = 1,\ldots,n \), 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 \( \delta_0 \), \( m(x,a) \) as defined in DGP 3 or 4 is not strictly monotonic in \( a \), permitting study of the test’s power against non-monotone alternatives.

To construct the raw test statistic \( \hat{J}_n \), we first obtain \( \hat{G}_{p,b}(Y_i|X_i,z) \) and \( \hat{G}_{p,b}^{-1}(\hat{G}_{p,b}(Y_i|X_i,z)|z^*,z) \).
Figure 1: Q-Q plot of $\hat{m}_{H_i}(x,a)$, $i = 1, ..., 500$, at various points versus standard normal (DGP 1, $n = 100, 400$)

Figure 2: Q-Q plot of $\hat{m}_{H_i}(x,a)$, $i = 1, ..., 500$, at various points versus standard normal (DGP 2, $n = 100, 400$)
by choosing the order 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 need to choose weight functions \( H_1 \), \( H_2 \), and \( \pi \). 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 \( \epsilon_0 \)th sample quantile of \[ \{X_i\}_{i=1}^n \] and \( \epsilon_0 = 0.0125 \). These sample quantiles converge to their population analogs at the parametric \( \sqrt{n} \)-rate, so they can be replaced by the latter in deriving the asymptotic theory. By construction, we trim \( \hat{G}_{p,b} \) and \( \hat{G}_{p,b}^{-1} \) in the tails.

In the bootstrap, we set \( b_x = s_Z n^{-1/6}, b_x = s_X n^{-1/6}, \) and \( b_a = s_A n^{-1/6} \), where, e.g., \( s_Z \) denotes the sample standard deviation of \( Z \). 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 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 an important question for policy analysis, namely, the determinants of the Black-White earnings gap. The second application comes from a traditional area of economics: classical consumer demand using Engel curves. We discuss the economic background, provide details of the data and the test implementation, and discuss our findings.

---

*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

<table>
<thead>
<tr>
<th>DGP</th>
<th>n</th>
<th>( \delta_0 )</th>
<th>Warp-speed bootstrap</th>
<th>Full bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1%</td>
<td>5%</td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>0</td>
<td>0.004</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.218</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.020</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.346</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0</td>
<td>0.006</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.330</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.008</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.326</td>
</tr>
<tr>
<td></td>
<td>0.10 n ( \chi^2 ) ( 1 ), 0.9 n ( Z ) ( 1/5 ), 0.9 n ( Z ) ( 1/5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>0</td>
<td>0.020</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.386</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.036</td>
<td>0.092</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.544</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0</td>
<td>0.018</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.566</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.002</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.538</td>
</tr>
<tr>
<td></td>
<td>0.1 n ( \chi^2 ) ( 1 ), 0.3 n ( Z ) ( 1/5 ), 0.3 n ( Z ) ( 1/5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>0</td>
<td>0.018</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.376</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.028</td>
<td>0.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.504</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0</td>
<td>0.012</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.542</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.002</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.570</td>
</tr>
<tr>
<td></td>
<td>0.1 n ( \chi^2 ) ( 1 ), 0.5 n ( Z ) ( 1/5 ), 0.5 n ( Z ) ( 1/5 )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>100</td>
<td>0</td>
<td>0.012</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.316</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.016</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.440</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>0</td>
<td>0.000</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.476</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0</td>
<td>0.016</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>0.569</td>
</tr>
</tbody>
</table>
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. Carneiro, Heckman, and Masterov (2005, CHM hereafter) give an overview of this literature, emphasizing the points important to our application.

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

As argued convincingly in CHM, however, schooling is no longer a plausible channel of discrimination, given the extent of affirmative action. Indeed, CHM show that when conditioned just on schooling, the wage gap increases, rather than 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 \( E[m(X_1, 1, A)|X_1, Z] - E[m(X_1, 0, A)|X_1, Z] = 0 \). 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 details

The details of the testing procedure we implement are largely identical to those for the simulation study of Section 6.2. The kernel is the product of univariate standard normal PDFs; the order of the local polynomial is 1. The bandwidth is chosen by 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

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0.965</td>
<td>0.652</td>
</tr>
</tbody>
</table>

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 B.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 $\xi$ is another proxy for $A$, viewing $V$ as a measurement error. Here, a natural choice for $\xi$ 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

<table>
<thead>
<tr>
<th>p-value</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.54</td>
<td>0.51</td>
</tr>
</tbody>
</table>

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

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

(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. However, it is commonly thought that log total expenditure is endogenous\(^9\) and is hence instrumented for, typically by labor income, say $\xi$. 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 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

$$X_1 = \Psi(\xi, X_2, Z),$$

(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 $\xi$ is exogenous, so we take $(\xi, X_2) \perp (A, Z)$, implying $(X_1, X_2) \perp A \mid Z$. With the usual

\(^9\)Nevertheless, the evidence is not strong; see Blundell, Horowitz, and Parey (2012) or Hoderlein (2011).
normalization, $Z \mid (\xi, X_2)$ is $U(0,1)$, and $Z$ is identified as

$$Z = F(X_1 \mid \xi, X_2),$$

(7.3)

where $F$ denotes the conditional CDF of $X_1$ given $(\xi, X_2)$. 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 $\alpha$-quantile regression of $Y$ on $X$ and $Z$, the parameter $\alpha$ 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 our rejection comes 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.\textsuperscript{10} The FES reports a yearly cross section of labor income, expenditures, demographic composition, and other characteristics of about 7,000 households in every year. Since we are considering Engel curves, we use only the cross section for 1995. We focus on households with two adults, where the adults are married or cohabiting, at least one is working, and the household head is aged between 20 and 55. We also exclude households with 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

\textsuperscript{10}We are grateful to Whitney Newey and Richard Blundell for providing us with the data.
Average Income study (HBAI). Roughly, this is net labor income to the household head after taxes, but including state transfers.

Table 5 gives some summary statistics for the FES data.

Table 5: Summary statistics for British FES data

<table>
<thead>
<tr>
<th>Variable</th>
<th>Food</th>
<th>Catering</th>
<th>Alcohol</th>
<th>Transport</th>
<th>Leisure</th>
<th>LogExp</th>
<th>LogWage</th>
<th>nKids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.2074</td>
<td>0.0805</td>
<td>0.0578</td>
<td>0.2204</td>
<td>0.1297</td>
<td>5.4215</td>
<td>5.8581</td>
<td>0.6205</td>
</tr>
<tr>
<td>Std</td>
<td>0.0971</td>
<td>0.0569</td>
<td>0.0624</td>
<td>0.1020</td>
<td>0.1343</td>
<td>0.4494</td>
<td>0.5381</td>
<td>0.4854</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.5754</td>
<td>1.1194</td>
<td>1.6135</td>
<td>0.7732</td>
<td>1.8381</td>
<td>0.2857</td>
<td>-0.4439</td>
<td>-0.4968</td>
</tr>
<tr>
<td>Minimum</td>
<td>0.0014</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>3.6090</td>
<td>2.1972</td>
<td>0.0000</td>
</tr>
<tr>
<td>25% Quantile</td>
<td>0.1393</td>
<td>0.0380</td>
<td>0.0112</td>
<td>0.1440</td>
<td>0.0372</td>
<td>5.1168</td>
<td>5.5520</td>
<td>0.0000</td>
</tr>
<tr>
<td>50% Quantile</td>
<td>0.1959</td>
<td>0.0700</td>
<td>0.0377</td>
<td>0.2094</td>
<td>0.0822</td>
<td>5.4019</td>
<td>5.8568</td>
<td>1.0000</td>
</tr>
<tr>
<td>75% Quantile</td>
<td>0.2716</td>
<td>0.1088</td>
<td>0.0852</td>
<td>0.2843</td>
<td>0.1744</td>
<td>5.6987</td>
<td>6.1746</td>
<td>1.0000</td>
</tr>
<tr>
<td>Maximum</td>
<td>0.6867</td>
<td>0.3595</td>
<td>0.3849</td>
<td>0.7003</td>
<td>0.8872</td>
<td>7.4287</td>
<td>8.0893</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

7.2.3 Details of Implementation

To apply the test, we let \( \Phi \) be the budget shares of Food, Catering, Transportation, or Leisure in (7.1). In each case, we specify \( \Phi_1 \) as the logarithm of total expenditure and \( \Phi_2 \) 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 \( \Phi = \Phi_1(\Phi_1, \Phi_2) \) must be estimated from the data in a first stage, and our theory for nonparametrically generated regressors in Section 5.3 is in effect with some slight modification. To see this, let \( \Phi_1(x_1) = 1 \{x_1 \leq x_1\} \) and \( F(x_1|S, X_2) = E[Q(x_1)|S, X_2] \). Then we have

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

where \( E[\varepsilon(x_1)|S, X_2] = 0 \). Apparently, \( Z \) in equation (5.8) corresponds to \( F(X_1|S, Z_2) \). To implement our test, we need to replace \( Z_i = F(X_1i|S_i, Z_2i) \) by its \( \hat{p} \)-th order local polynomial estimate \( \hat{F}(X_1i|S_i, Z_2i) \).

It is well known 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 variable \( Z_i = (S_i, Z_2i)' \). In fact, under some standard conditions for local polynomial regression (e.g., Hansen, 2008), we can establish the following uniform result

\[
\sup_{x_1} \sup_{z_2} \left\| \hat{F}(x_1|s, z_2) - F(x_1|s, z_2) \right\| = O_P \left( (n^{-1}b^{-2} \log n)^{1/2} + \hat{b}^{\hat{p}+1} \right).
\]

Assumption C.9 needs to hold with \( \hat{r} \) and \( r \) replaced by \( \hat{F} \) and \( F \), respectively. The major difference is that now only the dimension of \( z = (s, z_2)' \) and the smoothness of the conditional CDF \( F(x_1|s, z_2) \) (with respect to the conditioning variable \( z = (s, z_2)' \)) matter. To mitigate the bias from the first stage estimation of

---

\( ^{11} \)We did not consider the budget share for alcohol because there are too many 0 observations (258 out of 1,665) in the data.
we use local quadratic regression (i.e., \( \bar{p} = 2 \)), employing a Gaussian kernel and the bandwidth from cross validation. Using estimated \( Z \) also has implications for the bootstrap. Ideally, one would prefer a bootstrap method alternative to that applied here, based entirely on observables, while imposing the null. Developing and justifying this is a substantive undertaking, deserving of a paper in itself. We therefore leave this as a topic for future research. Another interesting topic would be to perform a test of the conditional independence assumption, which in this case would amount to testing the validity of the control function structure. Unfortunately, unlike in the first application where we had access to an additional proxy (i.e., a second measurement, AQFT 89 in addition to AQFT 80), we lack such a variable in our demand dataset. It would be good to perform such an analysis with other data, however, and we encourage such a future application of our method.

7.2.4 Empirical Results

Table 6 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 6: 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    | \( \leq 0.005 \) | 0.010           | \( \leq 0.005 \) | \( \leq 0.005 \) |

8 Concluding remarks

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

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

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

A Important notations

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

1 \{\cdot\} : indicator function;
\mathbb{1}_Y (W_k) \equiv \mathbb{1} \{Y_k \leq Y_i\} - G (Y_i | U_k),
\jmath \equiv (j_1, ..., j_d)', \ u^d \equiv \sum_{i=1}^d u_i^d, \ |\jmath| = \sum_{i=1}^d j_i, \ \sum_{0 \leq |\jmath| \leq p} \equiv \sum_{k=0}^p \sum_{j_1=0}^k \cdots \sum_{j_d=0}^k = 0;
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, \hat{A}_H^*, \hat{A}_H^1 : various terms defined in Section 3.1;
b, \hat{b} : bandwidths used in nonparametric estimation;
B_n, B_{n-1}, B_p, 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, \mathbf{B}_{p,b} = E [\mathbf{B}_{p,b}];
c_{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_t on (X_t, Z_t) = (x, z), similarly for \hat{G}, \hat{g};
\hat{G} \equiv \hat{G}_{p,b} \equiv c_{1,p} \hat{\alpha} : estimator from equation (4.4);
\hat{G}^{-1} \equiv \hat{G}_{-1}^{-1} \equiv c_{1,p} \hat{\beta} : estimator from equation (4.1);
\hat{G}_{p,b}, \hat{G}^{-1}_{p,b} : analogue of \hat{G}_{p,b} and \hat{G}^{-1}_{p,b} in the case of generated regressors, defined in (5.12) and (5.11);
\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}_u : statistic defined in equation (3.1) with standardized version \hat{T}_n;
K, \hat{K} : kernel functions; \mathbf{K}_j (u) \equiv u^d K (u), \ K_b (\cdot) \equiv 1/b^d K (\cdot | b), \ \hat{K}_b (\cdot) \equiv 1/b^d \hat{K} (\cdot | b), \ \hat{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^{*1} (x, a);
\hat{m}_H^{*1} (X, Y) \equiv \inf \{a : \hat{m}_H (X, a) \geq Y\}, \ m_H^{*1} (X, Y) \equiv \inf \{a : m_H (X, a) \geq Y\};
N_{p,t} \equiv \frac{(n_t - b)}{(n_t - 1)}, \ N_p \equiv \sum_{t=0}^p N_{p,t};
N_{|\cdot|} : the covering number in empirical processes;
p, \hat{p} : the orders of the kernel;
Q : dependent variable defined in Section 5.4 with support \mathcal{Q};
r : structural function between Q and \varphi, defined in the functional space \mathcal{R};
\mathbf{S}_{p,b}, \mathbf{S}_{p,b}, \mathbf{S}_{p,b} : technical terms defined in equation (4.3), (4.5), and (5.1) respectively;
\mathbf{F}_{p,b} \equiv E (S_{p,b}), \ \mathbf{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', r (x^*))';
\nabla_{p,b}, \nabla_{p,b}, \nabla_{p,b} : the decomposition of the variance term defined before the proof of Theorem 4.1;
$W \equiv (X', Y, Z')'$, $\mathcal{W}_n \equiv \{ W_i \}_{i=1}^n$.

$Y_i^t (r) \equiv Y_i - \sum_{0 \leq |j| \leq p} \frac{1}{P} D^j G^{-1} (\tau | u) (U_i - u)^j$;

$\alpha$ : the parameter in equation (4.4);

$\beta$ : the parameter in equation (4.1);

$\theta$ : the parameter vector that stacks $\sqrt{mb^j (\alpha_j - \frac{v_j}{P}) D^j G^{-1} (\tau | u)}$ for $0 \leq |j| \leq p$;

$\eta_{ik}, \eta_{2k}, s_0, \varphi$ : technical terms defined after equation (5.1);

$\mu$ : 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);

$\varpi$ : independent variable defined in Section 5.4 with support $\Omega_{\varpi}$;

$\rho_r (u) \equiv u (\tau - 1 \{ u \leq 0 \})$;

$\sigma_m^2, \sigma_m^2, \sigma_m^2, \sigma_m^2$ : variance terms defined in Theorem 4.1 and 4.3, and equation (5.2);

$\phi, \Phi$ : PDF and CDF of the standard normal;

$\psi_r (u) \equiv \tau - 1 \{ u \leq 0 \}$;

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

$\Theta_t^{\alpha} (y; x, z) : a$ function defined in equation (B.7).

## B Proof of the Main Results

### Proof of Proposition 2.1

Assumption A.2 ensures that for all $(y, x, z)$,

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

By Assumptions A.1-A.3 and setting $y = m(x, a)$, we have that for all $(a, x, \tilde{x}, z)$

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

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

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

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

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

For the next results, recall that $U_0 \equiv X_0 \times Z_0$, $U_i \equiv (X_i', Z_i')'$, $u \equiv (x', z')'$, $K_b (u) \equiv b^{-d} K (u/b)$, and $\mu_{p,b} (u) \equiv \mu_p (u/b)$. Let $W_i \equiv \langle Y_i, U_i' \rangle'$ and $w \equiv \langle y, u' \rangle'$. Let $S_{p,b} (u)$ and $S_{p,b} (u)$ be as defined in (4.3) and
(5.1), respectively. Define

\[ V_{p,b}(\tau; u) = \frac{1}{n} \sum_{i=1}^{n} K_b(U_i - u) \mu_{p,b}(U_i - u) \psi_r(Y_i - \beta_{b}(\tau; u)), \]

\[ V_{p,b}(\tau; u) = \frac{1}{n} \sum_{i=1}^{n} K_b(U_i - u) \mu_{p,b}(U_i - u) \psi_r(Y_i - G^{-1}(\tau|U_i)), \]

\[ B_{p,b}(y; u) = \frac{1}{n} \sum_{i=1}^{n} K_b(U_i - u) \mu_{p,b}(U_i - u) \Delta_{i,y}(u), \]

\[ V_{p,b}(y; u) = \frac{1}{n} \sum_{i=1}^{n} K_b(U_i - u) \mu_{p,b}(U_i - u) \tilde{I}_y(W_i), \]

where \( \psi_r(u) \equiv \rho - 1 \{ u \leq 0 \}, \Delta_{i,y}(u) \equiv G(y|U_i) - G(y|u) - \sum_{1 \leq |j| \leq p} \frac{1}{2} G(j)(y|u)(U_i - u)^j, \) and \( \tilde{I}_y(W_i) \equiv 1 \{ Y_i \leq y \} - G(y|U_i). \) Let \( U_{is} \) and \( u_s \) denote the \( s \)th elements of \( U_i \) and \( u \).

To prove Theorem 4.1, we will need lemmas C.1-C.3 in Appendix C. Let \( \hat{S}_{p,b}(u) \equiv E[S_{p,b}(u)] \) and \( \hat{B}_{p,b}(y; u) \equiv E[B_{p,b}(y; u)]. \) In particular, Lemmas C.1 and C.2 establish uniform consistency of \( \hat{\beta}(y|u) \) 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 \( \hat{m}_H(x, a) - m_H(x, a) = \int [G^{-1}(\hat{\tau}_z | x, z) - G^{-1}(\tau_z | x, z)] dH(z) + \int \hat{G}_{p,b}^{-1}(\hat{\tau}_z | x, z) - G^{-1}(\hat{\tau}_z | x, z)] dH(z) \equiv M_{n1}(x, a) + M_{n2}(x, a), \) 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 \( \hat{R}(a; x^*, x, z) \equiv \frac{g'(G^{-1}(\tau_z|x, z)|x, z)}{g(G^{-1}(\tau_z|x, z)|x, z)} (\hat{\tau}_z - \tau_z)^2 \) and \( \tau_z \) lies between \( \hat{\tau}_z \) and \( \tau_z. \) Noting that \( \hat{\tau}_z - \tau_z \) is the first element of \( \hat{\beta}(y|u) - \beta(y|u) \) with \( u = (x', z')', \) we have that by Lemma C.1(b) and Assumption C.6, \( \hat{R}(a; x^*, x, z) = O_P(\log n + b^{2(p+1)}) = O_P(n^{-1/2}b^{-d}a^2) \) uniformly in \( (a, x, z) \in A_H \times X_0 \times Z_0. \) It follows that for all \( (a, x) \in A_H \times X_0 \)

\[ \sqrt{nb^{d/2}} M_{n1}(x, a) = \sqrt{nb^{d/2}} \int \frac{\hat{\tau}_z - \tau_z}{g(G^{-1}(\tau_z|x, z)|x, z)} dH(z) + o_P(1) \]

\[ = \sqrt{nb^{d/2}} \int \frac{e_{1,p} S_{p,b}(x^*, z)^{-1} B_{p,b}(a; x^*, z)}{g(G^{-1}(\tau_z|x, z)|x, z)} dH(z) \]

\[ + \sqrt{nb^{d/2}} \int \frac{e_{1,p} S_{p,b}(x^*, z)^{-1} V_{p,b}(a; x^*, z)}{g(G^{-1}(\tau_z|x, z)|x, z)} dH(z) + o_P(1) \]

\[ \equiv M_{n1}(x, a) + M_{n2}(x, a) + o_P(1), \] say,

where the second line follows from Lemma C.1(a). Noting that \( \hat{B}_{p,b}(a; u) = E[K_b(U_i - u) \mu_{p,b}(U_i - u) \Delta_{i,a}(u)] = b^{p+1}g(u) \hat{B}_{p,G_{p+1}}(a|u) + o(b^{p+1}) \) and \( \hat{S}_{p,b}(u) = S_{p,g}(u) + o(1) \) uniformly in \( (a, u) \in A_H \times X_0, \) we have

\[ M_{n1}(x, a) = \sqrt{nb^{d/2}} b^{p+1} \int \frac{e_{1,p} S_{p,b}^{-1} B_{p,G_{p+1}}(a|x^*, z)}{g(G^{-1}(\tau_z|x, z)|x, z)} dH(z) \{ 1 + o_P(1) \} \] (B.1)
and
\[
M_{n12}(x, a) = \sqrt{nb^{d/2}} \int \frac{e_{1,p}^{-1} V_{p,b}(a; x^*, z)}{g(x^*, z) g(G^{-1}(\tau_z|x,z))} dH(z) \{1 + o_P(1)\}
\]
\[
= \tilde{M}_{n12}(x, a) \{1 + o_P(1)\} \distr N(0, V_1), \tag{B.2}
\]
where \(\tilde{M}_{n12}(x, a) \equiv \sqrt{\frac{b^{d/2}}{n}} \sum_{i=1}^{n} e_{1,p}^{-1} K_i(x_i-x^*, x_i-x^*) \mu_{p,b}(x_i-x^*, x_i-x^*) \tau_i(x_i) d\tau_i,\) \(\tilde{M}_{n12}(x, a) \equiv \sqrt{\frac{b^{d/2}}{n}} \sum_{i=1}^{n} e_{1,p}^{-1} K_i(x_i-x^*, x_i-x^*) \mu_{p,b}(x_i-x^*, x_i-x^*) \tau_i(x_i) d\tau_i,\) \(\tilde{M}_{n12}(x, a) \equiv \sqrt{\frac{b^{d/2}}{n}} \sum_{i=1}^{n} e_{1,p}^{-1} K_i(x_i-x^*, x_i-x^*) \mu_{p,b}(x_i-x^*, x_i-x^*) \tau_i(x_i) d\tau_i,\)

where \(M_{n12}(x, a) \equiv \sqrt{\frac{b^{d/2}}{n}} \sum_{i=1}^{n} e_{1,p}^{-1} K_i(x_i-x^*, x_i-x^*) \mu_{p,b}(x_i-x^*, x_i-x^*) \tau_i(x_i) d\tau_i,\)

Next, it is standard to show that uniformly in \((a, x) \in \mathcal{A}_H \times \mathcal{X}_0, V_1 \equiv \kappa_{1p} \int e_{1,p}^{-1} \mu_p(\tilde{z}, \tilde{z}) \mu_p(\tilde{z}, \tilde{z}) \tau_i(x_i) d\tau_i, \kappa_{1p} \equiv \sqrt{\frac{b^{d/2}}{n}} \sum_{i=1}^{n} e_{1,p}^{-1} K_i(x_i-x^*, x_i-x^*) \mu_{p,b}(x_i-x^*, x_i-x^*) \tau_i(x_i) d\tau_i, \) and \((B.2)\) follows from straightforward moment calculations and Liapounov’s central limit theorem.

For \(M_{n22},\) noting that \(\sqrt{nb^{d/2}} o(b^{d/2} + n^{-1/2} b^{-d/2}) = o(1)\) under Assumption C.6, by Lemma C.2(c) we have that uniformly in \((a, x) \in \mathcal{A}_H \times \mathcal{X}_0\)
\[
\sqrt{nb^{d/2}} M_{n22}(x, a) = \sqrt{nb^{d/2}} b^{d/2} \int e_{1,p}^{-1} \mu_p G_{p+1}^{-1}(\tau_z|x,z) dH(z) + o_P(1),
\]
\[
\equiv M_{n21}(x, a) + M_{n22}(x, a) + o_P(1), \tag{B.3}
\]
where \(S_p(\tau; u) \equiv \mu_p(\tau u) g(\tau u) g(u).\) Using Lemmas C.2 and C.3, we can show that uniformly in \((a, x) \in \mathcal{A}_H \times \mathcal{X}_0\)
\[
M_{n21}(x, a) = \sqrt{nb^{d/2}} b^{d/2} \int e_{1,p}^{-1} \mu_p G_{p+1}^{-1}(\tau_z|x,z) dH(z) + o_P(1),
\]
and \(M_{n22}(x, a) = M_{n22}(x, a) + o_P(1),\) where \(M_{n22}(x, a) \equiv \int e_{1,p} S_p(\tau_z|x,z) V_{p,b}(\tau_z|x,z) dH(z).\) Furthermore,
\[
\bar{M}_{n22}(x, a) = \sqrt{\frac{b^{d/2}}{n}} \sum_{i=1}^{n} \int e_{1,p}^{-1} K_i(U_i - u) \mu_{p,b}(U_i - u) \psi_{\tau_i} \mu_p(\tau_i^{-1} U_i) dH(z) \{1 + o_P(1)\}
\]
\[
\distr N(0, V_2),
\]
where \(V_2 \equiv \kappa_{1p} \int e_{1,p}^{-1} \mu_p(\tau_i^{-1} U_i) dH(z) \{1 + o_P(1)\},\) the asymptotic normality result follows by the Cranmer-Wold device and the fact that the asymptotic covariance of \(M_{n12}\) and \(M_{n22} \) are zero. In sum, we have \(\sqrt{nb^{d/2}} \{\bar{M}_{n12}(x, a) - m_{n12}(x, a)\} \distr N(0, \sigma^2_{m12}(x, a; x^*)) ,\) where \(B_m(x, a; x^*)\) and \(\sigma^2_{m12}(x, a; x^*)\) 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 \(\sup_{(x, a) \in \mathcal{X}_0 \times \mathcal{A}_H} |M_{n22}(x, a)| = O_P(\sqrt{\log n})\). Then the uniform convergence result follows. \(\blacksquare\)

**Proof of Theorem 4.3.** By the strict monotonicity of \(m_H^*(x, \cdot)\) for all \(x, a,\) 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
\[
\delta = \delta(\epsilon) = \min\{m_H^*(x, a) - m_H^*(x, a - \epsilon), m_H^*(x, a + \epsilon) - m_H^*(x, a)\} > 0.
\]

37
It follows that for sufficiently large \( n \),

\[
P \left\{ |\hat{m}_H^{-1}(x, y) - m_H^{-1}(x, y)| > \epsilon \right\} = P \left\{ |\hat{m}_H(x, y) - m_H(x, y)| + \epsilon > |\hat{m}_H(x, y) - m_H(x, y)| - \epsilon \right\} = P \left\{ m_H(x, \hat{m}_H^{-1}(x, y)) > m_H^{*}(x, y) + \epsilon \right\} \text{ or } m_H(x, \hat{m}_H^{-1}(x, y)) < m_H^{*}(x, y) - \epsilon \right\} \leq P \left\{ |m_H(x, \hat{m}_H^{-1}(x, y)) - m_H^{*}(x, y)| > \delta \right\} = P \left\{ \sup_{a \in \mathcal{A}_H} |\hat{m}_H(x, a) - m_H^{*}(x, a)| > \delta \right\} \rightarrow 0,
\]

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 \( m_H^{*}(x, y) = y = m_H(x, \hat{m}_H^{-1}(x, y)) \), and \( \mathcal{A}_H = \{ a : \langle a, c \rangle \leq c \} \) for some \( c \in \mathcal{A}_H \) and \( c > 0 \).

Let \( \Phi_n(v) = P \left\{ n^{1/2} d_{\mathbb{R}}^{1/2} \sigma_{m^{-1}}(x, y) \left[ \hat{m}_H^{-1}(x, y) - m_H^{-1}(x, y) - B_{m^{-1}}(x, y) \right] \leq v \right\} \) for any \( v \in \mathbb{R} \). Then

\[
\Phi_n(v) = P \left\{ \hat{m}_H(x, \hat{m}_H^{-1}(x, y)) \leq m_H^{*}(x, y) + \alpha_n(v; x, y) \right\} = P \left\{ \hat{m}_H(x, \hat{m}_H^{-1}(x, y)) \leq m_H^{*}(x, y) + \alpha_n(v; x, y) \right\} = P \left\{ \hat{m}_H(x, \hat{m}_H^{-1}(x, y) + \alpha_n(v; x, y)) \geq y \right\},
\]

where \( \alpha_n(v; x, y) \equiv B_{m^{-1}}(x, y) + n^{-1/2} d_{\mathbb{R}}^{1/2} \sigma_{m^{-1}}(y; x) v \). By Lemma C.4(b) and Theorem 4.1,

\[
\Phi_n(v) = P \left\{ \hat{m}_H(x, m_H^{-1}(x, y)) \geq -\lambda_H^{*}(x, m_H^{-1}(x, y)) \alpha_n(v; x, y) + y \right\} = P \left\{ \hat{m}_H(x, m_H^{-1}(x, y)) - y + \lambda_H^{*}(x, m_H^{-1}(x, y)) B_{m^{-1}}(x, y) \geq -(n^{-1/2} d_{\mathbb{R}}^{1/2} \lambda_H^{*}(x, m_H^{-1}(x, y)) \sigma_{m^{-1}}(y; x) v) \right\} = P \left\{ \sqrt{n} d_{\mathbb{R}}^{1/2} \lambda_H^{*}(x, m_H^{-1}(x, y)) \sigma_{m^{-1}}(y; x) \right\}^{-1} \times \hat{m}_H(x, m_H^{-1}(x, y)) - y + \lambda_H^{*}(x, m_H^{-1}(x, y)) B_{m^{-1}}(x, y) \geq -v \right\} \rightarrow 1 - \Phi(-v) = \Phi(v),
\]

where \( \Phi \) 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

\[
G_n(y|x, z) = P \{ m(X, A) + c_n \gamma(X, A) \leq y | X = x, Z = z \} = P \{ m(x, A) \leq y - c_n \gamma(x, A) | Z = z \} = E \{ 1 \{ m(x, A) \leq y \} | Z = z \} + R_n(y; x, z) = E \{ 1 \{ m(x, A) \leq y \} | X = x, Z = z \} + R_n(y; x, z) = G(y|x, z) + R_n(y; x, z),
\]

38
where $R_n(y; x, z) = E \{ [1 \{ m(x, A) \leq y - c_n \gamma(x, A) \} - 1 \{ m(x, A) \leq y \}] | Z = z \}$. Let $\gamma(x) = \inf_{a \in A} \gamma(x, a)$ and $\bar{\gamma}(x) = \sup_{a \in A} \gamma(x, a)$. By the monotonicity of $m(x, \cdot)$ and the Taylor expansion of $F_{A|Z}(m^{-1}(x, y) | z)$ around $y$ with an integral remainder,\textsuperscript{12}

\begin{equation}
R_n(y; x, z) \leq E \{ [1 \{ m(x, A) \leq y - c_n \gamma(x) \} - 1 \{ m(x, A) \leq y \}] | Z = z \}
\end{equation}

\begin{equation}
= E \{ [1 \{ A \leq m^{-1}(x, y - c_n \gamma(x)) \} - 1 \{ A \leq m^{-1}(x, y) \}] | Z = z \}
\end{equation}

\begin{equation}
= F_{A|Z}(m^{-1}(x, y - c_n \gamma(x)) | z) - F_{A|Z}(m^{-1}(x, y) | z)
\end{equation}

\begin{equation}
= -c_n \gamma(x) \int_0^1 \varsigma(y - tc_n \gamma(x); x, z) \, dt,
\end{equation}

where $\varsigma(y; x, z) = \partial F_{A|Z}(m^{-1}(x, y) | z) / \partial y = f_{A|Z}(m^{-1}(x, y) | z) \frac{1}{m_2(x, a)(x, y))}$ and $m_2(x, a) = \partial m(x, a) / \partial a$. Similarly, $R_n(y; x, z) \geq E \{ [1 \{ m(x, A) \leq y - c_n \bar{\gamma}(x) \} - 1 \{ m(x, A) \leq y \}] | Z = z \} = -c_n \bar{\gamma}(x) \int_0^1 \varsigma(y - tc_n \bar{\gamma}(x); x, z) \, dt$. It follows that there exists a continuous function $\gamma_n^1(y; x, z)$ such that

\begin{equation}
\gamma_n(x) \int_0^1 \varsigma(y - tc_n \gamma_n(x); x, z) \, d \leq \gamma_n^1(y; x, z) \leq \gamma(x) \int_0^1 \varsigma(y - tc_n \gamma(x); x, z) \, dt,
\end{equation}

and

\begin{equation}
G_n(y|x, z) = \bar{G}(y|x, z) - c_n \gamma_n^1(y; x, z).
\end{equation}

(B.5)

Let $\gamma_n^1(x, z) = \sup_y \gamma_n^1(y; x, z)$ and $\gamma_n^1(x, z) = \inf_y \gamma_n^1(y; x, z)$. 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{equation}
G_n^{-1}(\tau|x, z) = \inf \{ y : G(y|x, z) - c_n \gamma_n^1(y; x, z) \geq \tau \}
\end{equation}

\begin{equation}
\geq \inf \{ y : G(y|x, z) \geq \tau + c_n \gamma_n^1(y; x, z) \} = G_n^{-1}(\tau + c_n \gamma_n^1(y; x, z) | x, z),
\end{equation}

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

\begin{equation}
G_n^{-1}(\tau|x, z) = G_n^{-1}(\tau + c_n \gamma_n^1(\tau; x, z) | x, z)
\end{equation}

\begin{equation}
= G_n^{-1}(\tau + c_n \gamma_n^1(\tau; x, z) | x, z) \int_0^1 \frac{1}{g(G_n^{-1}(\tau + c_n \gamma_n^1(\tau; x, z) | x, z))} \, dt.
\end{equation}

(B.6)

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

\begin{equation}
G_n^{-1}(G_n(y|x, z)|x^*, z) = G_n^{-1}(G_n(y|x, z)|x^*, z) + c_n \Theta_1^1(y; x, z)
\end{equation}

\begin{equation}
= G_n^{-1}(G_n(y|x, z)|x^*, z) + c_n \Theta_1^1(y; x, z)
\end{equation}

\begin{equation}
= G_n^{-1}(G_n(y|x, z)|x^*, z) + c_n \Theta_1^1(y; x, z)
\end{equation}

\begin{equation}
\Theta_1^1(y; x, z) = \gamma_n^1(G_n(y|x, z)|x^*, z) \int_0^1 \frac{1}{g(G_n^{-1}(G_n(y|x, z)|x^*, z)|x^*, z)} \, dt,
\end{equation}

where $\Theta_1^1(y; x, z) = \gamma_n^1(G_n(y|x, z)|x^*, z) \int_0^1 \frac{1}{g(G_n^{-1}(G_n(y|x, z)|x^*, z)|x^*, z)} \, dt$, and

\begin{equation}
\Theta_1^1(y; x, z) = \gamma_n^1(G_n(y|x, z)|x^*, z) \int_0^1 \frac{1}{g(G_n^{-1}(G_n(y|x, z)|x^*, z)|x^*, z)} \, dt
\end{equation}

\begin{equation}
= \gamma_n^1(G_n(y|x, z)|x^*, z) \int_0^1 \frac{1}{g(G_n^{-1}(G_n(y|x, z)|x^*, z)|x^*, z)} \, dt + o(1).
\end{equation}

(B.7)

\textsuperscript{12}For any function $\alpha(t)$ that is continuously differentiable, we have $\alpha(y) - \alpha(x) = \alpha'(x)(y - x) + (y - x) \int_0^1 [\alpha'(x + t(y - x)) - \alpha'(x)] \, dt = (y - x) \int_0^1 \alpha'(x + t(y - x)) \, dt$, where $\alpha'$ denotes the derivative of $\alpha$. 39
Proof of Theorems 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 \( \hat{J}_n \equiv b^{d_X} \sum_{i=1}^{n} \left\{ \int \left[ \hat{G}_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) - G_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) \right] d\Delta(z) \right\}^2 \pi_i, \alpha_n (\tau | u) \equiv \hat{G}_{p,b}^{-1}(\tau | u) - G_{p,b}^{-1}(\tau | u), \alpha_{n1} (\tau; u) \equiv \alpha_n (\tau | u) - \alpha_n (\tau; u), \tau_{iz} \equiv G_{n}(Y_i | X_i, z) \), and \( \tilde{\tau}_{iz} \equiv \hat{G}_{p,b}(Y_i | X_i, z) \). Noting that \( a^2 - b^2 = (a - b)^2 + 2(a - b) b \), we have

\[
\begin{align*}
\hat{J}_n &= \hat{J}_n + (\hat{J}_n - \hat{J}_n) \\
&= \hat{J}_n + b^{d_X} \sum_{i=1}^{n} \left\{ \int \left[ \hat{G}_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) - G_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) \right] d\Delta(z) \right\}^2 \pi_i \\
& \quad + 2 b^{d_X} \sum_{i=1}^{n} \left\{ \int \left[ \hat{G}_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) - G_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) \right] d\Delta(z) \int G_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) d\Delta(z) \pi_i \\
&= \hat{J}_n + b^{d_X} \sum_{i=1}^{n} \left\{ \int \alpha_n(\tilde{\tau}_{iz} | x^*, z) d\Delta(z) \right\}^2 \pi_i + 2 b^{d_X} \sum_{i=1}^{n} \left\{ \int \alpha_{n2}(\tilde{\tau}_{iz} | x^*, z) d\Delta(z) \right\}^2 \pi_i \\
& \quad + 2 b^{d_X} \sum_{i=1}^{n} \left\{ \int \alpha_n(\tilde{\tau}_{iz} | x^*, z) d\Delta(z) \int G_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) d\Delta(z) \pi_i \\
& \quad + 2 b^{d_X} \sum_{i=1}^{n} \left\{ \int \alpha_{n2}(\tilde{\tau}_{iz} | x^*, z) d\Delta(z) \int G_{p,b}^{-1}(\tilde{\tau}_{iz} | x^*, z) d\Delta(z) \pi_i \\
& \equiv \hat{J}_n + \hat{J}_{n1} + \hat{J}_{n2} + \hat{J}_{n3} + \hat{J}_{n4} + \hat{J}_{n5}, \quad \text{say.} \ (B.8)
\end{align*}
\]

Since it is straightforward to show that \( \hat{\beta}_n - \beta_n = o_P(1) \) and \( \hat{\sigma}_n^2 - \sigma_n^2 = o_P(1) \), it suffices to prove the theorem by showing that

\[
\hat{J}_n = J_n + o_P(1), \quad \text{B.9}
\]

and

\[
J_n = \hat{J}_n \overset{d}{=} \mathcal{N}(c_{\theta_0}, \sigma_n^2), \quad \text{B.10}
\]

where \( J_n \equiv b^{d_X} \sum_{i=1}^{n} \left\{ \int \frac{e_i^{1:p} S_{p,b}(X_i | z) - V_{p,b}(Y_i; X_i, z)}{g_n(G_{n}^{-1}(\tau_{iz} | x^*, z) | x^*, z)} + e_i^{1:p} S_{p,b}(\tau_{iz} | x^*, z) - V_{p,b}(\tau_{iz} | x^*, z) d\Delta(z) \right\}^2 \pi_i, \)

We prove (B.9) by showing that

\[
\begin{align*}
\hat{J}_n &= b^{d_X} \sum_{i=1}^{n} \left\{ \int \frac{e_i^{1:p} S_{p,b}(X_i | z) - V_{p,b}(Y_i; X_i, z)}{g_n(G_{n}^{-1}(\tau_{iz} | x^*, z) | x^*, z)} \right\}^2 \pi_i + c_{\theta_0} + o_P(1), \quad \text{B.11} \\
\hat{J}_{n1} &= b^{d_X} \sum_{i=1}^{n} \left\{ \int e_i^{1:p} S_{p,b}(\tau_{iz} | x^*, z) - V_{p,b}(\tau_{iz} | x^*, z) d\Delta(z) \right\}^2 \pi_i + o_P(1), \quad \text{B.12} \\
\hat{J}_{n4} &= \hat{J}_{n4} + o_P(1), \quad \text{B.13} \\
\hat{J}_{ns} &= o_P(1) \text{ for } s = 2, 3, 5, \quad \text{B.14}
\end{align*}
\]
where \( \tilde{J}_{n4} \equiv b^d x \sum_{i=1}^{n} \int \frac{e_{i, p} S_{p,b}(X_i, z)^{-1} V_{p,b}(Y_i; X_i, z)}{g_n(G_n^{-1}(\tau_{ix}|x^*, z)|x^*, z)} e_{i, p} S_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz}; x^*, z) d\Delta(z) \pi_i. \) To show (B.11), write

\[
\tilde{J}_n = b^d x \sum_{i=1}^{n} \left[ \int G_n^{-1}(\tau_{iz}|x^*, z)d\Delta(z) \right]^2 \pi_i \\
+2b^d x \sum_{i=1}^{n} \left\{ \int \left[ G_n^{-1}(\hat{\tau}_{iz}|x^*, z) - G_n^{-1}(\tau_{iz}|x^*, z) \right] d\Delta(z) \right\}^2 \pi_i \\
+2b^d x \sum_{i=1}^{n} \int G_n^{-1}(\hat{\tau}_{iz}|x^*, z)d\Delta(z) \int \left[ G_n^{-1}(\hat{\tau}_{iz}|x^*, z) - G_n^{-1}(\tau_{iz}|x^*, z) \right] d\Delta(z) \pi_i
\]

\[
\equiv \tilde{J}_{n1} + \tilde{J}_{n2} + 2\tilde{J}_{n3}, \text{ say.}
\]

Under \( H_1(c_n) \) with \( c_n = n^{-1/2}b^{-d}x/2, \) \( \tilde{J}_{n1} = n^{-1}\sum_{i=1}^{n} \Theta_n(Y_i; X_i, A_i) \pi_i \xrightarrow{P} c\Theta_0 \) by Remark 5.2. Noting that

\[
G_n^{-1}(\hat{\tau}_{iz}|x^*, z) - G_n^{-1}(\tau_{iz}|x^*, z) = \frac{\hat{\tau}_{iz} - \tau_{iz}}{g_n(G_n^{-1}(\tau_{iz}|x^*, z)|x^*, z)} + \hat{R}_i(z), \quad (B.15)
\]

where \( \hat{R}_i(z) = -g_0(G_n^{-1}(\tau_{iz}|x^*, z)|x^*, z)^2 \left( g'(\tau_{iz} - \tau_{iz}) \right)^2, \) \( g'_n(\cdot|\cdot) \) denotes the derivative of \( g_n(\cdot|\cdot) \) with respect to \( \cdot \), and \( \hat{\tau}_{iz} \) lies between \( \tau_{iz} \) and \( \hat{\tau}_{iz} \), we have that under \( H_1(c_n) \),

\[
\tilde{J}_{n3} = n^{-1/2}b^d x/2 \sum_{i=1}^{n} \Theta_n(Y_i; X_i, A_i) \int \left[ G_n^{-1}(\hat{\tau}_{iz}|x^*, z) - G_n^{-1}(\tau_{iz}|x^*, z) \right] d\Delta(z) \pi_i
\]

\[
\equiv \tilde{J}_{n31} + \tilde{J}_{n32}, \text{ say.}
\]

Observe that \( \hat{R}_i(z) = O_P(\left( n^{-1} b^{-d} \log n \right) b^{2(p+1)}) \) uniformly in \( z \) by Lemma C.1(b), we have \( \tilde{J}_{n32} = n^{-1/2}b^d x/2O_P\left( n^{-1} b^{-d} \log n \right) b^{2(p+1)} = o_P(1). \) By Lemma C.1(a), the fact that \( \mathbb{B}_{p,b}(y; x, z) = O_P(b^{p+1}) \) uniformly in \( (y, x, z) \in \mathbb{R} \times \mathbb{X} \times \mathbb{Z}_0 \), and Assumption C.6, we have \( \tilde{J}_{n31} = n^{-1/2}b^d x/2 \sum_{i=1}^{n} \Theta_n(Y_i; X_i, A_i) \times \int \frac{e_{i, p} S_{p,b}(X_i, z)^{-1} V_{p,b}(Y_i; X_i, z)}{g_n(G_n^{-1}(\tau_{iz}|x^*, z)|x^*, z)} d\Delta(z) \pi_i + o_P(1). \) 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, \( \tilde{J}_{n3} = o_P(1) \) under \( H_1(c_n) \).

Using (B.15), we decompose \( \tilde{J}_{n2} \) as follows

\[
\tilde{J}_{n2} = b^d x \sum_{i=1}^{n} \left[ \int \frac{\hat{\tau}_{iz} - \tau_{iz}}{g_n(G_n^{-1}(\tau_{iz}|x^*, z)|x^*, z)} d\Delta(z) \right]^2 \pi_i \\
+2b^d x \sum_{i=1}^{n} \int \frac{\hat{\tau}_{iz} - \tau_{iz}}{g_n(G_n^{-1}(\tau_{iz}|x^*, z)|x^*, z)} d\Delta(z) \int \frac{\hat{\tau}_{iz} - \tau_{iz}}{g_n(G_n^{-1}(\tau_{iz}|x^*, z)|x^*, z)} d\Delta(z) \pi_i
\]

\[
\equiv \tilde{J}_{n21} + \tilde{J}_{n22} + 2\tilde{J}_{n23}, \text{ say.}
\]

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

\[
\tilde{J}_{n21} = b^d x \sum_{i=1}^{n} \left[ \int \frac{e_{i, p} S_{p,b}(X_i, z)^{-1} V_{p,b}(Y_i; X_i, z)}{g_n(G_n^{-1}(\tau_{iz}|x^*, z)|x^*, z)} d\Delta(z) \right]^2 \pi_i + o_P(1) = O_P(1),
\]
and \( J_{n22} = nb^{d_x} \mathcal{O}_P(n^{-2b^{-2d}}(\log n)^2 + b^{d(p+1)}) = \mathcal{O}_P(n^{-1}b^{-3d/2}(\log n)^2 + nb^{d(p+1)+d_x}) = \mathcal{O}_P(1) \). Then \( J_{n23} = \mathcal{O}_P(1) \) by Cauchy-Schwarz inequality. Consequently, (B.11) follows.

By Lemma C.5, (B.12) holds. With (B.12), it is standard to show that \( J_{n1} = \mathcal{O}_P(1) \). Using (B.15) we can decompose \( J_{n4} \) as

\[
J_{n4} = b^{d_x} \sum_{i=1}^{n} \int \alpha_{n1}(\hat{\tau}_{iz}; x^*, z) d\Delta(z) \int G_{n}^{-1}(\tau_{iz}|x^*, z) d\Delta(z) \pi_i
+ b^{d_x} \sum_{i=1}^{n} \int \alpha_{n1}(\hat{\tau}_{iz}; x^*, z) d\Delta(z) \int \frac{\hat{\tau}_{iz} - \tau_{iz}}{g_n(G_{n}^{-1}(\tau_{iz}|x^*, z)|x^*, z)} d\Delta(z) \pi_i
+ b^{d_x} \sum_{i=1}^{n} \int \alpha_{n1}(\hat{\tau}_{iz}; x^*, z) d\Delta(z) \int \tilde{R}_i(z) d\Delta(z) \pi_i \equiv \sum_{k=1}^{3} J_{n4s}, \text{ say.}
\]

Using arguments as used in the analysis of \( J_{n31} \) and \( J_{n32} \), we can readily show that \( J_{n4s} = \mathcal{O}_P(1) \) for \( s = 1, 3 \). For \( J_{n42} \), we can apply Lemmas C.1 and C.2 to obtain \( J_{n42} = J_{n4} + \mathcal{O}_P(1) \), where \( J_{n4} \) is defined after (B.14).\(^{13}\) Thus (B.13) follows.

We now show (B.14). By Lemma C.2(a), \( J_{n2} = nb^{d_x}[\mathcal{O}_P(\nu_k^2) + \mathcal{O}_P(b^{d(p+1) + n^{-1}b^{-d_x}})] = \mathcal{O}_P(1) \).

By the fact that \( J_{n1} = \mathcal{O}_P(1) \) and Cauchy-Schwarz inequality, \( J_{n3} = \mathcal{O}_P(1) \). For \( J_{n5} \), we have \( J_{n5} = nb^{d_x}[\mathcal{O}_P(\nu_k^2) + \mathcal{O}_P(b^{p+1} + n^{-1/2}b^{-d_x/2})] \mathcal{O}_P(n^{-1/2}b^{-d_x/2} \sqrt{\log n + b^{p+1}}) = \mathcal{O}_P(1) \). Consequently, (B.9) follows.

To show (B.10), let

\[
\eta_{1k}(\tau; x, z) = e_{1, p}^{\tau} \hat{S}_{p,b}(x, z) \mu_{p,b}(X_k - x, Z_k - z) / g_n (G_{n}^{-1}(\tau|x^*, z)|x^*, z),
\eta_{2k}(\tau; x, z) = e_{1, p}^{\tau} \hat{S}_{p,b}(\tau; x, z) \mu_{p,b}(X_k - x, Z_k - z) K_b(X_k - x, Z_k - z),
\]

and \( \zeta_0(W_i, W_k; z) \equiv \eta_{1k}(\tau_{iz}; X_i, z) \tilde{Y}_i(W_k) + \eta_{2k}(\tau_{iz}; x^*, z) \psi_{\tau_{iz}}(Y_k - G_{n}^{-1}(\tau_{iz}|U_k)) \). Then

\[
\frac{e_{1, p}^{\tau} \hat{S}_{p,b}(X_i; z)^{-1} V_{p,b}(Y_i; z)}{g_n(G_{n}^{-1}(\tau_{iz}|x^*, z)|x^*, z)} + e_{1, p}^{\tau} \hat{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz}; x^*, z) = \frac{1}{n} \sum_{k=1}^{n} \zeta_0(W_i, W_k; z).
\]

It follows that \( J_n = b^{d_x} \sum_{i=1}^{n} \left[ \int n^{-1} \sum_{k=1}^{n} \varphi(W_i, W_k; z) d\Delta(z) \right]^{2} \pi_i = n^{-2b^{d_x}} \sum_{i=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \zeta(W_i, W_{i_2}, W_{i_3}) \), where \( \zeta(W_i, W_{i_2}, W_{i_3}) \equiv \int \zeta_0(W_i, W_{i_2}, W_{i_3}) \zeta_0(W_i, W_{i_2}, W_{i_3}) d\Delta(z) \pi_i \). Let \( \varphi(w_i, w_{i_2}) \equiv E[\zeta(W_i, w_{i_2}, w_{i_3})] \), and \( \zeta(w_i, w_{i_2}, w_{i_3}) \equiv E[\zeta(W_i, w_{i_2}, w_{i_3}) - \varphi(w_{i_2}, w_{i_3})] \). Then we can decompose \( J_n \) as \( J_n = J_{n1} + J_{n2} + J_{n3} \), where

\[
J_{n1} = n^{-1} b^{d_x} \sum_{i=1}^{n} \sum_{i_2=1}^{n} \varphi(W_i, W_{i_2}) \text{ and } J_{n2} = n^{-2} b^{d_x} \sum_{i=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \zeta(W_i, W_{i_2}, W_{i_3}).
\]

Consider \( J_{n2} \) first. Write \( E[\zeta_{n2}^2] = n^{-4} b^{2d_x} \sum_{i=1}^{n} b^{d(p+1)} E[\zeta(W_i, W_{i_2}, W_{i_3})] \zeta(W_i, W_{i_2}, W_{i_3})]. \) Noting that \( E[\zeta(W_i, w_{i_2}, w_{i_3})] = E[\zeta(w_i, w_{i_2}, w_{i_3})] = E[\zeta(w_i, w_{i_2}, w_{i_3})] = 0 \), \( E[\zeta(W_i, W_{i_2}, W_{i_3})] \zeta(W_i, W_{i_2}, W_{i_3})] \)

\(^{13}\) Using the expressions for \( V_{p,b} \) and \( \hat{S}_{p,b} \), we can write \( J_{n4} \) as a third order U-statistic. By straightforward moment conditions, we can verify that \( E[J_{n4}^2] = o(1) \). Despite the asymptotic negligibility of \( J_{n42} \), we keep it in our asymptotic analysis, as it will simplify notation in other places.

42
= 0 if there are more than three distinct elements in \( \{i_1, \ldots, i_6\} \). With this, it is easy to show that 
\[ E \left[ J_n^2 \right] = O(n^{-1} b^{-2d} + n^{-2} b^{-3d} + n^{-3} b^{-4d}) = o(1) \]. Hence \( J_n \sim o_P(1) \) by Chebyshev inequality.

For \( J_n \), let \( \varphi(W_i, W_j) = \int \int \iota_0(\tilde{w}, W_i; z) \zeta_0(\tilde{w}, W_j; \tilde{z}) \pi(\tilde{x}, \tilde{y}) d\Delta(z) d\Delta(\tilde{z}) dG_n(\tilde{w}) \). Then \( J_n = n^{-1} b^{-d} \sum_{i=1}^n \varphi(W_i, W_i) + 2n^{-1} b^{-d} \sum_{1 \leq i \leq j \leq n} \varphi(W_i, W_j) \). The proof is based on Pollard (1991), Fan et al. (1994), and some standard results in the empirical process theory. Define the following three empirical processes:

\[
L_n(r) = \sum_{i=1}^n \left[ \rho_r \left( Y^*_i(r) - \mu_{p,b,i}(r) \frac{\theta}{\sqrt{n b^d}} \right) - \rho_r \left( Y^*_i(r) \right) \right] \tilde{K}_{b,i}(r), \tag{B.16}
\]

\[
\Upsilon_{n,1}(r) = \sum_{i=1}^n \psi_r(Y^*_i(r)) \mu'_p, b, i(r) \frac{\theta}{\sqrt{n b^d}} \tilde{K}_{b,i}(r), \tag{B.17}
\]

\[
\Upsilon_{n,2}(r) = \sum_{i=1}^n \Pi_i(r) \tilde{K}_{b,i}(r), \tag{B.18}
\]

where \( \psi_r(u) \equiv r - 1 \{u \leq 0\} \) which corresponds to \( l'(u) \) in Fan et al. (1994) by their notation and behaves like the derivative of \( \rho_r(u) \) by our notation, and

\[
\Pi_i(r) \equiv \rho_r \left( Y^*_i(r) - \mu'_{p,b,i}(r) \frac{\theta}{\sqrt{n b^d}} \right) - \rho_r \left( Y^*_i(r) \right) - \psi_r(Y^*_i(r)) \mu'_{p,b,i}(r) - \frac{\theta}{\sqrt{n b^d}} \tilde{K}_{b,i}(r). 
\]

Note that we have suppressed the dependence of \( L_n, \Upsilon_{n,1}, \) and \( \Upsilon_{n,2} \) on \( \theta \) for notational simplicity. Let \( R_n(r) = \Upsilon_{n,2}(r) - E_n[\Upsilon_{n,2}(r) | \varpi] \), where \( \varpi = (\varpi_1, \ldots, \varpi_n) \) and \( E_n \) denotes the expectation for empirical processes. By construction, we have the following relationship

\[
L_n(r) = \Upsilon_{n,1}(r) + \Upsilon_{n,2}(r) = \Upsilon_{n,1}(r) + R_n(r) + E_n[L_n(r) | \varpi] - E_n[\Upsilon_{n,1}(r) | \varpi]. \tag{B.19}
\]

For any finite \( \theta \), Lemma C.7 below shows that \( R_n(\tilde{r}) = o_P(1) \). Lemma C.8 below shows that \( E_n[L_n(\tilde{r}) | \varpi] = \frac{1}{2} \theta' S_p(r; u) \theta(1 + o_P(1)) \), where \( S_p(r; u) \equiv \sum i g \left( G^{-1}(r | u) | u \right) g(u) \). Lemma C.9 below shows that \( \Upsilon_{n,1}(\tilde{r}) = \sum_{i=1}^n \psi_r(Y^*_i(r_0)) \mu'_{p,b,i}(r_0) \frac{\theta}{\sqrt{n b^d}} \tilde{K}_{b,i}(r_0) + o_P(1) \). Combining these results
yields
\[ L_n(\hat{r}) = \frac{1}{\sqrt{nbd^2}} \sum_{i=1}^{n} \psi_r(Y_i^*(r_0)) \mu_{p,b,i}(r_0) \hat{K}_{b,i}(r_0) + \frac{1}{2} \theta S_p(\tau; u) \theta + o_P(1). \]

It is easy to show that \( \frac{1}{\sqrt{nbd^2}} \psi_r(Y_i^*(r_0)) \mu_{p,b,i}(r_0) \hat{K}_{b,i}(r_0) \) has bounded second moment. Then by the convexity lemma of Pollard (1991), the \( o_P(1) \) term in above equation is \( o_P(1) \) uniformly in \( \theta \) such that \( \|\theta\|_2 \leq M \) for any \( M > 0 \). Therefore, by the convexity lemma again, we have \( \hat{\theta} = -S_p(\tau; u)^{-1} \frac{1}{\sqrt{nbd^2}} \sum_{i=1}^{n} \psi_r(Y_i^*(r_0)) \times \mu_{p,b,i}(r_0) \hat{K}_{b,i}(r_0) + o_P(1) \). By the way we choose bandwidth \( b \) as in Assumption C.6, using the result from Theorem 3.1 in Su and White (2012) and the stronger than usual convergence rate of \( \hat{r} \), the above \( o_P(1) \) can be shown to be \( o_P(b^{d_x}/2) \). Noting that \( \theta \) is a vector stacking \( \sqrt{nbd^2}(\alpha_j - \frac{b}{j})D^jG^{-1}(\tau|u) \), we have
\[
\hat{G}^{-1}(\tau|u) - G^{-1}(\tau|u) = \frac{1}{\sqrt{nbd^2}} \psi_r(\hat{\theta})^T = e_1^T S_p(\tau; u)^{-1} \frac{1}{n} \sum_{i=1}^{n} \psi_r(Y_i^*(r_0)) \mu_{p,b,i}(r_0) K_{b,i}(r_0) + o_P(b^{d_x})^{-1/2}. \]

Everything else follows again from Theorem 3.1 in Su and White (2012). ■

**Proof of Proposition 5.6.** This is a direct result from Mammen et al. (2012, Theorem 1), where we only need to replace their local linear estimator by our local polynomial estimator and this extension is very straightforward. It is not hard to see that the assumptions in Mammen et al. (2012) are all satisfied here, and as in Lemma C.10, the residual for the linear expansion in their final results is \( o_P((nb^{d_x})^{-1/2}) \).

Apply the results of Theorem 1 in Mammen et al. (2012) to here, we have \( \hat{G}_{p,b}(y|u) - \hat{G}_{p,b}(y|u) = g(y|u) \hat{A}(u) + o_P((nb^{d_x})^{-1/2}) \), where
\[
\hat{A}(u) = \psi_r(\hat{\theta})^T E_n \left[ \frac{\hat{K}_{b,i}(r_0)}{pdf} \mu_{p,b,i}(r_0) (U_i(\hat{r}) - U_i(r_0)) \right]. \tag{B.20}
\]
\( \sqrt{nbd^2} \hat{A}(u) \) has the same structure as the influence term in \( \Xi_{n,0} \) in equation (D.9). By the same line of analysis, we can show that \( \sqrt{nbd^2} \hat{A}(u) = o_P(1) \), so \( \hat{G}_{p,b}(y|u) - \hat{G}_{p,b}(y|u) = o_P((nb^{d_x})^{-1/2}) \). ■

**Sketch of the Proof of Equation (5.7).** Let \( \hat{G}_{p,b}^*, \hat{G}_{p,b}^{-1} \) and \( \hat{J}_{n,i} \) be defined as \( \hat{G}_{p,b}, \hat{G}_{p,b}^{-1} \) and \( \hat{J}_{n,i} \), with \( \mathbb{W}_n \) replacing \( \mathbb{W} \). Define \( S(C) \equiv \{ \sup_{(y,u) \in \mathbb{R}^2} |\hat{G}_{p,b}(y|u) - G(y|u)| \leq Cn^{-1/2b^{d_x}d}(\log n)^{1/2} + b^{d_x+1} \}
\sup_{(\tau,u) \in T \times \mathbb{L}_d} |\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u)| \leq Cn^{-1/2b^{d_x}d}(\log n)^{1/2} + b^{d_x+1} \}, \) where \( T = [\epsilon_0, 1-\epsilon_0] \) 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 \( C_p \) such that \( P(S(C)) \leq \epsilon \) for sufficiently large \( n \) where \( S(C) \) is the complement of \( S(C) \). Noting that
\[
P(T_n^* \leq t| \mathbb{W}_n) = P(T_n^* \leq t| \mathbb{W}_n \cap S(C)) P(S(C)) + P(T_n^* \leq t| \mathbb{W}_n \cap S^c(C)) P(S^c(C))
\]
and that the second term in the above expression can be made arbitrarily small for sufficiently large \( n \), it suffices to prove \((i)\) by showing that \( P(T_n^* \leq t| \mathbb{W}_n \cap S(C)) \rightarrow \Phi(t) \) for all \( t \in \mathbb{R} \). Conditional on \( \mathbb{W}_n \cap S(C) \), \( \hat{A}_i \) 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

\[
\hat{J}_n = n^{-1}b^{d_x} \sum_{i=1}^{n} \varphi^*(W_i^*, W_i^*) + 2n^{-1}b^{d_x} \sum_{1 \leq i < j \leq n} \varphi^*(W_i^*, W_j^*) + o_P(1)
\]

\[
\equiv \mathbb{B}_{J_n} + \mathbb{V}_{J_n} + o_P(1),
\]

where \(P^*\) denotes probability conditional on \(\mathbb{W}_n \cap \mathcal{S}(C)\), and \(\varphi^*\) is defined analogously to \(\varphi\) 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 \(W_i^*\)'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 \(N(0, \sigma_j^2)\) conditional on \(\mathbb{W}_n\), where

\[
\sigma_j^2 \equiv 2\lim_{n \to \infty} E^* [\varphi^*(W_1^*, W_2^*)^2].
\]

The asymptotic bias and variance terms can be estimated analogously as \(\hat{\mathbb{B}}_{J_n}\) and \(\hat{\sigma}_j^2\) in the paper. The asymptotic normality of \(T_n^*\) conditional on \(\mathbb{W}_n \cap \mathcal{S}(C)\) then follows.

For (ii), let \(z_\alpha^*\) denote the \(1 - \alpha\) conditional quantile of \(T_n^*\) given \(\mathbb{W}_n\), i.e., \(P(T_n^* \geq z_\alpha^* | \mathbb{W}_n) = \alpha\). By choosing \(N_B\) sufficiently large, the approximation error of \(z_\alpha^*\) to \(z_\alpha\) can be made arbitrarily small and negligible. By (i), \(z_\alpha^* \to z_\alpha\) in probability where \(z_\alpha\) is the \(1 - \alpha\) quantile of the standard normal distribution.

Then, in view of Theorem 5.1, \(T_n\) diverges to \(\infty\) 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 \(\mathbb{H}_1\).

C  Some technical lemmas

Lemma C.1 Suppose that Assumptions C.1-C.4 and C.6 hold. Let \(\beta(y|u)\) be a vector that stacks \(\frac{1}{b}D^j G(y|u)\), 0 \(\leq |j| \leq p\), 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) \(\hat{\beta}(y|u) - \beta(y|u) = \overline{S}_{p,b}(u)^{-1} [V_{p,b}(y|u) + \mathbb{B}_{p,b}(y|u)] + o_P(\nu_b^2 + \nu_b b^p + 1),\)

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

Lemma C.2 Suppose that Assumptions C.1-C.4 and C.6 hold. Let \(T\) be any compact subset of \((0, 1)\). Then uniformly in \((\tau, u) \in \mathbb{T} \times \mathcal{U}_0\),

(a) \(\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u) = e_{p} \overline{S}_{p,b}(\tau; u)^{-1} V_{p,b}(\tau; u) + O_P(\nu_b^2 + \nu_b b^{p+1} + o_P(n^{-1/2}b^{-d/2})),\)

(b) \(\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u) = e_{p} \overline{S}_{p,b}(\tau; u)^{-1} V_{p,b}(\tau; u) + b^{p+1} e_{p} \overline{S}_{p,b}(\tau; u)^{-1} V_{p,b}(\tau; u) + O_P(\nu_b^2) + o_P(b^{p+1} + n^{-1/2}b^{-d/2}),\)

(c) \(\hat{G}_{p,b}^{-1}(\tau|u) - G^{-1}(\tau|u) = e_{p} \overline{S}_{p,b}(\tau; u)^{-1} V_{p,b}(\tau; u) + 1 + o_P(1)) + b^{p+1} e_{p} \overline{S}_{p,b}(\tau; u)^{-1} V_{p,b}(\tau; u) + o_P(b^{p+1} + n^{-1/2}b^{-d/2}),\)

where \(S_p(\tau; u) \equiv S_{p,g}(G^{-1}(\tau|u)|u) g(u)\) is the limit of \(\overline{S}_{p,b}(\tau; u) \equiv E[S_{p,b}(\tau; u)].\)

Lemma C.3 Suppose that Assumptions C.1-C.4 and C.6 hold. Then

\(\sup_{\tau \in \mathbb{T} \cap |\tau| \leq M_n} \sup_{u \in \mathcal{U}_0} \sqrt{n b^{d_x}} \|V_{p,b}((\tau|u) - V_{p,b}((\tau|u)) = o_P(1).\)

By Lemmas C.1 and C.3, with probability approaching 1 we have

\(\sup_{(a, x, z) \in A_H \times X_0 \times Z_0} \sqrt{n b^{d_x}} \left| V_{p,b} \left( \hat{G}_{p,b}(a|x^*, z); x, z \right) - V_{p,b} \left( G(a|x^*, z); x, z \right) \right| = o_P(1).\)

Lemma C.4 Suppose that Assumptions C.1-C.6 hold. Then for any \(\delta_n = O(\nu_b)\), we have

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

(b) \(\hat{m}_H(x, a + \delta_n) - \hat{m}_H(x, a) = \lambda_H^*(x, a) \delta_n + o_P(\delta_n + n^{-1/2}b^{-d/2})\),

where \(\lambda_H^*(x, a) = \int_{\mathbb{P}} g(a|z) dH(z).\)
Lemma C.5 Suppose the conditions in Theorem 5.3 hold. Then $\hat{J}_{n1} = J_{n1} + o_P(1)$, where $J_{n1} = \varphi_{i
u} \sum_{i=1}^{n} [\int e_{i
u}^{p} S_{p,b}(\tau_{i\nu}; x^{*}, z)^{-1} V_{p,b}(\tau_{i\nu}; x^{*}, z) d\Delta(z)]^{2} \pi_{i}$.

Lemma C.6 Suppose the conditions in Proposition 5.5 hold. Then $P(\hat{\sigma} \in C_{r_{M}}(\Omega_{M})) \rightarrow 1$.

Lemma C.7 Suppose the conditions in Proposition 5.5 hold. Then for any finite $\theta$, we have $R_{n}(\bar{\hat{\sigma}}) = o_P(1)$.

Lemma C.8 Suppose the conditions in Proposition 5.5 hold. Then for any finite $\theta$, we have $E_{n}[\psi_{n,1}(\bar{\hat{\sigma}}) | \varpi] = \frac{1}{1 + o_P(1)} \sum_{i=1}^{n} \psi_{i}(Y_{i,1}(\rho_{0})) \mu_{p,b,i}(\rho_{0}) \theta + o_P(1)$.

Lemma C.9 Suppose the conditions in Proposition 5.5 hold. Then for any finite $\theta$, $\tau_{n,1}(\bar{\hat{\sigma}}) = \sum_{i=1}^{n} \psi_{i}(Y_{i,1}(\rho_{0})) \mu_{p,b,i}(\rho_{0}) \theta + o_P(1)$.

Lemma C.10 Suppose the conditions in Proposition 5.5 hold. Then for any finite $\theta$, $\psi_{n,1}(\bar{\hat{\sigma}}) = \psi_{n,1}(\rho_{0}) + o_P(1)$.

REFERENCES


48
Supplementary Appendix to
“Testing for Monotonicity in Unobservables under Unconfoundedness”
(Not for publication)
Stefan Hoderlein\textsuperscript{a}, Liangjun Su\textsuperscript{b}, Halbert White\textsuperscript{c}, and Thomas Tao Yang\textsuperscript{a}
\textsuperscript{a} Department of Economics, Boston College
\textsuperscript{b} School of Economics, Singapore Management University
\textsuperscript{c} Department of Economics, University of California, San Diego

This appendix provides proofs for the technical lemmas and additional 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 $[S_{p,b}(u)]^{-1} S_{p,b}(u) = I_N$ where $I_N$ is an $N \times N$ identity matrix, by (4.2) we obtain the following standard bias and variance decomposition:

$$\hat{\beta}(y|u) - \beta(y|u) = [S_{p,b}(u)]^{-1}V_{p,b}(y|u) + [S_{p,b}(u)]^{-1}B_{p,b}(y|u).$$

By Theorems 2 and 4 in Masry (1996) with some modification to account for the non-compact support of the kernel function,\textsuperscript{14}

$$S_{p,b}(u) = S_{p,b}(u) + O_P(\nu_b),$$
$$V_{p,b}(y|u) = O_P(\nu_b),$$
$$B_{p,b}(y|u) - B_{p,b}(y|u) = O_P(\nu_b^b + 1),$$

where the probability orders hold uniformly in $u \in U_0$. 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,

$$S_{p,b}(u)^{-1} = \{S_{p,b}(u) + [S_{p,b}(u) - S_{p,b}(u)]\}^{-1} = [S_{p,b}(u)]^{-1} + O_P(\nu_b).$$

It follows that

$$\hat{\beta}(y|u) - \beta(y|u) = \{S_{p,b}(u)^{-1} + O_P(\nu_b)\}\{V_{p,b}(y|u) + O_P(\nu_b b + 1)\} = S_{p,b}(u)^{-1} \times [V_{p,b}(y|u) + B_{p,b}(y|u) + O_P(\nu_b) + b + 1].$$

Recall that $G_{p,b}(y|x, z) = e_{1,p}(y|u)$ where $e_{1,p}$ is defined after (4.3). Noting that uniformly in $(y, u) \in \mathbb{R} \times U_0$, $S_{p,b}(u) = g(u) S_{p} + O(b)$, and $B_{p,b}(y|u) = b^p + g(u) B_{p}(y|u) + O(b^p)$, we have

$$G_{p,b}(y|u) - G(y|u) = b^p e_{1,y} S_{p}^{-1} B_{p} + O(b^p).$$

Proof of Lemma C.2. Noting that $S_{p,b}(\tau; u) - \tilde{S}_{p,b}(\tau; u) = O_P(\nu_b)$ and $\tilde{V}_{p,b}(\tau; u) = O_P(\nu_b + b + 1)$ by the proof of (b) below, (a) follows from Theorem 2.1 of Su and White (2012). To prove (b), write

$$\tilde{V}_{p,b}(\tau; u) = V_{p,b}(\tau; u) + R_{p,b}(\tau; u),$$

where $R_{p,b}(\tau; u) \equiv \frac{1}{n} \sum_{i=1}^{n} 1(Y_i \leq G^{-1}(\tau|U_i)) - 1(Y_i \leq \beta_b(\tau; u)\mu_{iu}) K_{b}(U_i - u) \mu_{iu}$ and $\tilde{\mu}_{iu} \equiv \mu_{p,b}(U_i - u)$. 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

$$E[R_{p,b}(\tau; u)] = E \{G[G^{-1}(\tau|U_i) | U_i] - G(\beta_b(\tau, u)\mu_{iu} | U_i) K_{b}(U_i - u) \mu_{iu} \}$$

$$= E \{g[G^{-1}(\tau|U_i) | U_i] [G^{-1}(\tau|U_i) - \beta_b(\tau, u)\mu_{iu}] K_{b}(U_i - u) \mu_{iu} \} + o(1)$$

$$= b^p g[G^{-1}(\tau|u)|y(g(u) B_{p} + 1) \{1 + o(1)\}.$$\textsuperscript{14}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 $(\tau, u)$. Thus (b) follows. For (c), it suffices to show that $\sup_{(\tau,x)\in T \times \mathcal{X}} \| S_{p,b}(\tau; u) - S_p(\tau; u) \| = O_P(n^{-1/2}b^{-d/2}\sqrt{\log n} + b) = o_P(1)$. The proof is similar to but simpler than that of Corollary 2 in Masyr (1996) because we only need convergence in probability, whereas Masyr proved almost sure convergence. ■

If $G(a|x^*,z) \in T_0 = [\tau, \tilde{\tau}] \subset (0, 1)$ for $x^* \in X_0$ and all $(a, z) \in A_H \times \mathcal{Z}_0$, by Lemma C.1, $\hat{G}_{p,b}(a|x^*,z) \in T'_0$ with probability approaching 1 for sufficiently large $n$, where $T'_0 \equiv [\tau - \epsilon, \tilde{\tau} + \epsilon] \subset (0, 1)$ for some $\epsilon > 0$. Then the result in Lemma C.2 holds uniformly in $(\tau, u) \in T'_0 \times \mathcal{U}_0$.

Proof of Lemma C.3. Let $W(\tilde{\tau}, \tau; u) = \omega'(V_{p,b}(\tilde{\tau}; u) - V_{p,b}(\tau; u))$ where $\omega \in \mathbb{R}^{N_r}$ with $\| \omega \| = 1$. We need to show that

$$
\sup_{\tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} \sup_{u \in \mathcal{U}_0} |W(\tilde{\tau}, \tau; u)| = o_p(\vartheta_n) \text{ with } \vartheta_n = n^{-1/2}b^{-d/2}.
$$

(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 $W(\tilde{\tau}, \tau; u) = (nb)^{-d}\sum_{i=1}^n a_{i,u} \tilde{\tau} - 1 \{Y_i \leq G^{-1}(\tilde{\tau}U_i)\} - \tau + 1 \{Y_i \leq G^{-1}(\tau U_i)\}$, 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 $\sup_{\tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} \sup_{u \in \mathcal{U}_0} |W^+(\tilde{\tau}, \tau; u)| = o_p(\vartheta_n)$ and $\sup_{\tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} \sup_{u \in \mathcal{U}_0} |W^-(\tilde{\tau}, \tau; u)| = o_p(\vartheta_n)$. 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 $\tau_j - \tau_{j-1} \leq e_n$, we can cover the compact set $T'_0$ by $T_j = [\tau_{j-1}, \tau_j]$ for $j = 1, \ldots, n_1$, where $T_0 = \tilde{\tau} - \epsilon$ and $T_{n_1} = \tilde{\tau} + \epsilon$. Similarly, we can select $n_2 = O(\tilde{a}d/e_n^d)$ grid points $u_1, \ldots, u_{n_2}$ to cover the compact set $\mathcal{U}_0$ by $\mathcal{U} = \{u : \|u - u_i\| \leq e_nb\}$, $l = 1, \ldots, n_2$. Observe that $\sup_{\tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} \sup_{u \in \mathcal{U}_0} |W^+(\tilde{\tau}, \tau; u)| \leq W_{n1} + W_{n2}$, where

$$
W_{n1} \equiv \max_{1 \leq j \leq n_2} \sup_{\tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} |W^+(\tilde{\tau}, \tau; u)|,
$$

and

$$
W_{n2} \equiv \max_{1 \leq j \leq n_2} \sup_{u \in \mathcal{U}} \sup_{1 \leq \tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} |W^+(\tilde{\tau}, \tau; u) - W^+(\tilde{\tau}, \tau; u_t)|.
$$

Furthermore,

$$
W_{n1} \leq \max_{1 \leq j \leq n_2} \max_{1 \leq \tilde{k} \leq n_1} \sup_{1 \leq \tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} \sup_{|\tau_j - \tau_k| \leq M_{\mathcal{U}_0}} |W^+(\tau_j, \tau_k; u)|
$$

$$
+ \max_{1 \leq j \leq n_2} \max_{1 \leq k \leq n_1} \sup_{\tilde{\tau}, \tau \in T'_0, |\tilde{\tau} - \tau| \leq M_{\mathcal{U}_0}} \sup_{|\tau_j - \tau_k| \leq M_{\mathcal{U}_0}} |W^+(\tilde{\tau}, \tau; u) - W^+(\tau_j, \tau_k; u)|
$$

$\equiv W_{n11} + W_{n12}$, say.

Let $\zeta_{i,u}(\tau_j, \tau_k) = a_{i,u}^+ |\tau_j - 1 \{Y_i \leq G^{-1}(\tau_j U_i)\} - \tau_k + 1 \{Y_i \leq G^{-1}(\tau_k U_i)\} |$. Noting that $|\zeta_{i,u}(\tau_j, \tau_k)| \leq C \cdot E[\zeta_{i,u}(\tau_j, \tau_k)] = 0$ and $E[\zeta_{i,u}(\tau_j, \tau_k)^2] \leq Cb^4 \nu_0$ as $|\tau_j - \tau_k| \leq M_{\mathcal{U}_0}$, we apply the Bernstein inequality (e.g., Serfling, 1980, p.95) and Assumption C6. to obtain

$$
P(W_{n11} > \vartheta_n e_0) \leq C_1 n_2 \nu_0 n_1^{3/2} \max_{1 \leq j \leq n_2} \sup_{1 \leq k \leq n_1} \sup_{|\tau_j - \tau_k| \leq M_{\mathcal{U}_0}} P(W^+(\tau_j, \tau_k; u) > \vartheta_n e_0)
$$

$$
\leq 2C_1 n_2 \nu_0 n_1^{1/2} \exp \left( - \frac{n^2b^2d \vartheta_n^2 e_0^2}{2C_2 \nu_0 b^2 \vartheta_n^2 e_0} \right)
$$

$$
= O \left( n_2 \nu_0 n_1^{1/2} \exp \left( - \frac{b \vartheta_n e_0^2}{C_4 (n^{-1/2}b^{-d/2}\sqrt{\log n + n^{-1/2}b^{-d/2}e_0})} \right) \right) = o(1),
$$

2
where \( C_i, i = 1, 2, 3, 4, \) are positive constants. Thus \( W_{n1} = o_P(\vartheta_n). \) By the monotonicity of the indicator and quantile functions and the nonnegativity of \( a_{i, u}^+, \) we can readily show that

\[
W_{n2} = \max_{1 \leq j \leq n} \sup_{|\tau_j - \tau| \leq M \nu_2} \left( \frac{1}{nb^n} \sum_{i=1}^{n} a_{i, u}^+ [\bar{\tau} - 1] \{ Y_i \leq G^{-1}(\tau[U_i]) \} - \tau \right)
\]

\[
+ 1 \{ Y_j \leq G^{-1}(\tau(U_j)) \} - a_{i, u}^+ [\bar{\tau} + 1] \{ Y_j \leq G^{-1}(\tau(U_j)) \} - \tau + 1 \{ Y_j \leq G^{-1}(\tau(U_j)) \} \right|
\]

\[
\leq \max_{1 \leq j \leq n} \sup_{|\tau_j - \tau| \leq M \nu_2} \left( \frac{1}{nb^n} \sum_{i=1}^{n} a_{i, u}^+ [\bar{\tau} - 1] \{ Y_i \leq G^{-1}(\tau(U_i)) \} - \tau + 1 \{ Y_j \leq G^{-1}(\tau(U_j)) \} \right|
\]

\[
+ \max_{1 \leq j \leq n} \sup_{|\tau_j - \tau| \leq M \nu_2} \left( \frac{1}{nb^n} \sum_{i=1}^{n} a_{i, u}^+ [\bar{\tau} - 1] \{ Y_j \leq G^{-1}(\tau(U_j)) \} - \tau - 1 \{ Y_j \leq G^{-1}(\tau(U_j)) \} \right|
\]

\[= O_p(n^{-1/2}) = o_P(\vartheta_n). \]

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

\[
|K(u_2) - K(u_1)| \leq \delta K^*(u_1),
\]

(D.5)

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

and \( K_{iu}^* \equiv K^*((U_i - u)/b), \) and

\[
\left| \left( \frac{U_i - u}{b} \right) - \left( \frac{U_i - u_i}{b} \right) \right| \leq \left( \frac{U_i - u}{b} \right) |K_{iu} - K_{iu}| + \left( \frac{U_i - u_i}{b} \right) - \left( \frac{U_i - u_i}{b} \right) K_{iu} = e_n K_{iu}^* \]

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

\[
|a_{i, u}^+ - a_{i, u_i}^+| = |K_{iu}^* \mu_{p, b}(U_i - u) - K_{iu}^* \mu_{p, b}(U_i - u_i)| \leq C e_n (K_{iu}^* + K_{iu}^*)
\]

It follows that

\[
W_{n2} = \max_{1 \leq j \leq n} \sup_{u \in \mathcal{U}} \sup_{|\tau_j - \tau| \leq M \nu_2} \|W^+ (\bar{\tau}, \tau; u) - W^+ (\bar{\tau}, \tau; u)\|
\]

\[
\leq 2 \max_{1 \leq j \leq n} \sup_{u \in \mathcal{U}} (nb^n)^{-1} \sum_{i=1}^{n} |a_{i, u}^+ - a_{i, u_i}^+|
\]

\[
\leq Ce_n \max_{1 \leq j \leq n} \sup_{u \in \mathcal{U}} (nb^n)^{-1} \sum_{i=1}^{n} (K_{iu}^* + K_{iu}) = O_P(e_n) = o_P(\vartheta_n).
\]

Thus we have proved that \( \sup_{\tau, \tau \in \mathcal{T}_\varepsilon} |\bar{\tau}_n| \leq M \nu_2 \), \( \sup_{u \in \mathcal{U}} |W^+ (\bar{\tau}, \tau; u)| = o_P(\vartheta_n). \)

\textbf{Proof of Lemma C.4.} By Lemma C.1,

\[
\widehat{G}_{p, b}(a + \delta_n|u) - \widehat{G}_{p, b}(a|u) = [G(a + \delta_n|u) - G(a|u)]
\]

\[
+ e_{1, p} \mathbb{E}_{p, b}(a + \delta_n|u) - \mathbb{E}_{p, b}(a|u) + e_{1, p} \mathbb{E}_{p, b}(a + \delta_n|u) - \mathbb{E}_{p, b}(a|u) + o_P(n^{-1/2} + \nu b^{p+1}).
\]

\text{Proof of Lemma C.4.} By Lemma C.1,
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 \( o(b^{p+1}) = o_p(n^{-1/2}b^{-d\xi/2}) \) uniformly in \( u \in U_0 \) by the fact that \( B_{p,b} (a; u) = b^{p+1}B_p G_{p+1} (a|u) g (u) + o (b^{p+1}) \) uniformly in \( u \) and \( S_{p,b} (u) = S_p g (u) + o (1) \), and the continuity of \( G_{p+1} \). Analogously to the proof of Lemma C.3, we can show that \( V_{p,b} (a + \delta_n; u) - V_{p,b} (a; u) = o_p(n^{-1/2}b^{-d\xi/2}) \) uniformly in \( u \in U_0 \).

Thus (a) follows.

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

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

and

\[
D_{n2} = \int \left[ G^{-1}(\hat{G}_{p,b}(a + \delta_n|x^*, z)|x, z)) - G^{-1}(\hat{G}_{p,b}(a + \delta_n|x^*, z)|x, z)) \right] dH (z)
- \int \left[ G^{-1}(\hat{G}_{p,b}(a|x^*, z)|x, z)) - G^{-1}(\hat{G}_{p,b}(a|x^*, z)|x, z)) \right] dH (z).
\]

For \( D_{n1} \), we have

\[
D_{n1} = \int \frac{\hat{G}_{p,b}(a + \delta_n|x^*, z) - \hat{G}_{p,b}(a|x^*, z)}{g \left( G^{-1}(\hat{G}_{p,b}(a|x^*, z)|x, z)) \right) |x, z)} dH (z) + o_p \left( \delta_n + n^{-1/2}b^{-d\xi/2} \right)
= \int \frac{g(a|x^*, z)\delta_n}{g \left( G^{-1}(\hat{G}_{p,b}(a|x^*, z)|x, z)) \right) |x, z)} dH (z) + o_p \left( \delta_n + n^{-1/2}b^{-d\xi/2} \right)
= D_{n1}^* (x, a) \delta_n + o_p \left( \delta_n + n^{-1/2}b^{-d\xi/2} \right),
\]

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

\[
D_{n2} \approx b^{p+1} \int e_{1,p} S_{p+1} [ G_{p+1}^{-1} (G(a + \delta_n|x^*, z)|u) - G_{p+1}^{-1} (G(a|x^*, z)|u) ] dH (z)
+ \int e_{1,p} S_p (G(a + \delta_n|x^*, z); x, z) - \hat{V}_{p,b} (G(a + \delta_n|x^*, z); x, z)
- S_p (G(a|x^*, z); x, z) - \hat{V}_{p,b} (G(a|x^*, z); x, z) ] dH (z)
\equiv D_{n21} + D_{n22}, \text{ say.}
\]

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

\[
D_{n22,1} = \int e_{1,p} S_p (G(a + \delta_n|x^*, z); x, z) - \hat{V}_{p,b} (G(a + \delta_n|x^*, z); x, z)
- S_p (G(a|x^*, z); x, z) - \hat{V}_{p,b} (G(a|x^*, z); x, z) ] dH (z),
\]

\[
D_{n22,2} = \int e_{1,p} S_p (G(a + \delta_n|x^*, z); x, z) - S_p (G(a|x^*, z); x, z) - \hat{V}_{p,b} (G(a|x^*, z); x, z) ] dH (z).
\]

One can readily show that \( D_{n22,1} = o_p(n^{-1/2}b^{-d\xi/2}) \) and \( D_{n22,2} = o_p(n^{-1/2}b^{-d\xi/2}) \) by standard moment calculations and the dominated convergence theorem, and (b) follows. \( \blacksquare \)

**Proof of Lemma C.5.** To prove the result, we define \( \tilde{J}_{n1} \) analogously as \( \tilde{J}_{n1} \) with \( \tau_{iz} \) replaced by \( \tilde{\tau}_{iz} \): \( \tilde{J}_{n1} = b^{d\xi} \sum_{i=1}^n \left[ \int e_{1,p} S_p (\tilde{\tau}_{iz}; x^*, z)^{-1} V_{p,b} (\tilde{\tau}_{iz}; x^*, z) d\Delta (z) \right] \pi_i \). It suffices to show (i) \( \tilde{J}_{n1} =
$J_{n1} + o_P(1)$, and (ii) $J_{n1} = J_{n1} + o_P(1)$. To prove (i), let $d_{n1}(\tau; u) = e^\prime_{\tau, p} S_{p,b}(\tau; x, z)^{-1} [V_{p,b}(\tau; x, z) - V_{p,b}(\tau; x, z)]$. Then we have $J_{n1} - J_{n1} = D_{n1} + 2D_{n2}$, where

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

As $d_{n1}(\tau; u) = O_P(b^{d+1})$ uniformly in $(\tau, u) \in T_0 \times U_0$, we have $D_{n1} = nb^{d_x} O_P(b^{2d+1}) = o_P(1)$. For $D_{n2}$, we get $D_{n2} = \bar{D}_{n2} + o_P(1)$ using Lemmas C.1 and C.3, where $\bar{D}_{n2} = b^d x \sum_{i=1}^n \int d_{n1}(\tau_{iz}; x^*, z) d\Delta(z) \int e^\prime_{\tau, p} \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz}; x^*, z) d\Delta(z) \pi_i$. Standard moment calculations give $D_{n2} = nb^{d_x} O_P(b^{d+1}) \times O_P(n^{-1/2} b^{-d_x}/2) = o_P(1)$, so (i) holds.

Next, we show (ii). Let $d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) = e^\prime_{\tau, p} \bar{S}_{p,b}(\hat{\tau}_{iz}; x^*, z)^{-1} V_{p,b}(\hat{\tau}_{iz}; x^*, z) - e^\prime_{\tau, p} \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} \times V_{p,b}(\tau_{iz}; x^*, z)$, $d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) = e^\prime_{\tau, p} \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} [V_{p,b}(\hat{\tau}_{iz}; x^*, z) - V_{p,b}(\tau_{iz}; x^*, z)]$, and $\bar{R}_{n2} = d_{n2} - \bar{d}_{n2}$. Then uniformly in $z \in Z_0$ and conditional on $\hat{\tau}_{iz} \in T_0^*$,

\[
\bar{R}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) = e^\prime_{\tau, p} \left[ \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} - \bar{S}_{p,b}(\hat{\tau}_{iz}; x^*, z)^{-1} \right] V_{p,b}(\hat{\tau}_{iz}; x^*, z) \quad \text{(D.6)}
\]

Decompose $J_{n1} - J_{n1}$ as

\[
J_{n11} + 2J_{n12} \equiv b^d x \sum_{i=1}^n \left[ \int d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) d\Delta(z) \right]^2 \pi_i
\]

\[
+ 2b^d x \sum_{i=1}^n \int [d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z)] d\Delta(z) \int e^\prime_{\tau, p} \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} V_{p,b}(\tau_{iz}; x^*, z) d\Delta(z) \pi_i.
\]

Further decompose $J_{n11}$ as $J_{n11} = J_{n11,a} + J_{n11,b} + 2J_{n11,c}$, say, with

\[
J_{n11,a} + J_{n11,b} + 2J_{n11,c} \equiv b^d x \sum_{i=1}^n \left[ \int d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) d\Delta(z) \right]^2 \pi_i
\]

\[
+ b^d x \sum_{i=1}^n \left[ \int \bar{R}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) d\Delta(z) \right]^2 \pi_i
\]

\[
+ 2b^d x \sum_{i=1}^n \sum_{j=1}^n \int d_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) d\Delta(z) \int \bar{R}_{n2}(\hat{\tau}_{iz}, \tau_{iz}; x^*, z) d\Delta(z) \pi_i.
\]

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

\[
P \left( \left| J_{n11,a} \right| \geq \theta_n \epsilon \right) \leq P \left( \left| J_{n11,a} \right| \geq \theta_n \epsilon, \sup_{z} \max_{1 \leq i \leq n} |\hat{\tau}_{iz} - \tau_{iz}| \leq M \nu_b \right) + \epsilon/2,
\]

and showing $J_{n11,a} = o_P(1)$ is equivalent to showing that the first term in the last expression is $o(1)$. Conditional on $\sup_{z} \max_{1 \leq i \leq n} |\hat{\tau}_{iz} - \tau_{iz}| \leq M \nu_b$ and $\tau_{iz} \in T_0 \subset (0, 1)$, by Lemma C.3

\[
J_{n11,a} = b^d x \sum_{i=1}^n \left[ \int e^\prime_{\tau, p} \bar{S}_{p,b}(\tau_{iz}; x^*, z)^{-1} [V_{p,b}(\hat{\tau}_{iz}; x^*, z) - V_{p,b}(\tau_{iz}; x^*, z)] d\Delta(z) \right]^2 \pi_i
\]

\[
\leq C n b^d x \sup_{z} \sup_{t \leq \tau, |\hat{\tau} - \tau| \leq M \nu_b} \| V_{p,b}(\hat{\tau}; u) - V_{p,b}(\tau; u) \|^2 = o_P(1),
\]

5
By (D.6), \( \tilde{J}_{n1,b} = nb^{d/2}O_P((n^{-1}b^{-d} \log n + b^{2(p+1)})n^{-1}b^{-d} \log n) = O_P(n^{-1}b^{-3d/2} \log n)^2 + b^{2(p+1)-d/2} \log n) = o_P(1) \). By Cauchy-Schwarz (CS hereafter) inequality, \( \tilde{J}_{n1,c} = o_P(1) \), so \( \tilde{J}_{n1} = o_P(1) \).

Analogously to the determination of the probability order of \( J_{n1} \), we can show that \( \tilde{J}_{n1, a} + o_P(1) \), where \( \tilde{J}_{n1, a} = b^{d/2} \sum_{i=1}^n f_{\sigma_\beta} (\tilde{\tau}_{i}; \tau_i; x^*, z) d\Delta (z) \int e_1 \nu \tilde{\lambda}_{b, \beta} (\tilde{\tau}_{i}; x^*, z) d\Delta (z) \phi_{\beta} \). By the CS inequality and Lemmas C.1 and C.3, \( \tilde{J}_{n1, a} = o_P(1) \). Thus \( \tilde{J}_{n1} = o_P(1) \). 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.** Following the proof of Theorem 2 in Fan et al. (1994), one can easily show that \( R_n(r_0) = o_P(1) \). It suffices to show that \( R_n(\tilde{r}) - R_n(r_0) = o_P(1) \). We prove this by applying Theorem 2.11.9 in van der Vaart and Wellner (1996, VW hereafter) to the series \( z^{(1)}_{ni}(r) \), where \( z^{(1)}_{ni}(r) = \Pi_i (r, \theta) \hat{K}_{b,i}(r) - E_n \Pi_i (r, \theta) \hat{K}_{b,i}(r) \). Note that \( R_n(r) \equiv \sum_{i=1}^n z^{(1)}_{ni}(r) \). By standard results in nonparametric estimation literature (e.g., Andrews, 1995), \( ||\tilde{r} - r||_\infty = o_P(1) \). Lemma C.6 shows that \( \tilde{r} \in \mathcal{C}_{\beta}^{(1)} (\Omega_\beta) \) with probability approaching 1. Given these two results, we are ready to verify the conditions of Theorem 2.11.9 in VW.

First, note that

\[
\sup_r \left( \left| \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r) \mathbb{1} \{ \sup_r |z^{(1)}_{ni}(r)| > c \}) \right| \right) \leq n \sup_r \left( \left| \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r)^2) \right| / c \right) \text{ for any } c > 0.
\]

Since \( \psi_r(\cdot) \) plays similar role as the first derivative of \( \rho_c(\cdot) \), by essentially the same argument as used for the analysis of \( R_n(\theta) \) in Fan et al. (1994, p. 441) and \( R_{t,n}(\theta) \) in Pollard (1991), we have\(^{15}\)

\[
n \sup_r \left( \left| \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r)^2) \right| \right) \leq n \sup_r \left( \left| \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r)^2) \right| \right) \leq n \sup_r \left( \left| \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r)^2) \right| \right) = o(1),
\]

where the last equality holds by the change of variable and the fact that \( \mathbb{1} \{ |Y^*_r(r)| \leq o_p(1) \} \overset{P}{\to} 0 \).

The sup operator can go through by the variable transformation in \( K \) and finiteness of \( r \). So we have

\[
\sup_{||r_1 - r_2||_2 = o(1)} \left( \frac{1}{n} \sum_{i=1}^n \left[ \left( \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r_1) - z^{(1)}_{ni}(r_2))^2 \right) \right] \right)
\]

\[
\leq 2 \sup_{||r_1 - r_2||_2 = o(1)} \left( n \sup_r \left( \left| \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r_1)^2) \right| \right) + n \sup_r \left( \left| \frac{1}{n} \sum_{i=1}^n (z^{(1)}_{ni}(r_2)^2) \right| \right) = o(1),
\]

where the last equality follows by the same reasoning as used in the first step.

Third, since \( ||r_1 - r_2||_2 \leq ||r_1 - r_2||_\infty \), the covering number using \( ||\cdot||_2 \) is smaller than the covering number with norm being \( ||\cdot||_\infty \). Then we have \( \log N_{\frac{n}{2}}(\epsilon, \mathcal{C}^{(1)}_{\beta} (\Omega_\beta), ||\cdot||_2) < C\epsilon^{-\frac{\mathcal{C}^{(1)}_{\beta}(\Omega_\beta)}{2}} \). By Assumption that \( d_\epsilon/\phi < 2 \), we have \( \int_0^1 \log N_{\frac{n}{2}}(\epsilon, \mathcal{C}^{(1)}_{\beta} (\Omega_\beta), ||\cdot||_2) d\epsilon < \infty \).

Hence, the conditions of Theorem 2.11.9 in VW are satisfied and \( \sum_{i=1}^n (z^{(1)}_{ni}(r) - E_n (z^{(1)}_{ni}(r)) \) is asymptotic stochastic equicontinuous with respect to the \( ||\cdot||_2 \) norm. Therefore, we have stochastic equicontinuity. Since

\(^{15}\)Note that \( \Pi_i (r, \theta) \equiv 0 \), for \( |Y^*_r(r)| > \mu_{p,b,i} \sqrt{n} \).

\(^{16}\)This formula is for scalar function. The multi-dimension function \( r \) compared with a scalar function will only affect \( C \) in the covering number, which is not essential here.
\[ \| \hat{r} - r \|_2 \leq \| \hat{r} - r \|_\infty = o(1), \text{ by stochastic equicontinuity we have } \sum_{i=1}^{n} [z_{ni}^{(1)}(\hat{r}) - z_{ni}^{(1)}(r)] - E_n[z_{ni}^{(1)}(\hat{r}) + E_n[z_{ni}^{(1)}(r)] = o_P(1). \text{ In view of the fact that } E_n[z_{ni}^{(1)}(r)] \equiv 0 \text{ and } R_n(r) \equiv \sum_{i=1}^{n} z_{ni}^{(1)}(r), \text{ we obtain } R_n(\hat{r}) - R_n(r_0) = o_P(1). \]

Proof of Lemma C.8. The result could be obtained from the same arguments as used in Su and White (2012) and Fan et al. (1994, see equation (4.3) in particular). Note that \( g(G^{-1}(\tau|u)|u) \) corresponds to \( \varphi''(0) \) in Fan et al.’s notation.

Proof of Lemma C.9. We investigate \( \Upsilon_{n,1}(\hat{r}) \) by using the empirical process theory. However, the second step in the proof of Lemma C.7 for verifying Theorem 2.11.9 of van der Vaart and Wellner (1996) will only hold with some additional high level assumptions given in Assumption C.9; see also the conditions in Mammen et al. (2012). To avoid tedious proof, we use the latter authors’ asymptotic results directly.

By Lemma C.10 and results in Mammen et al. (2012), we have \( \Upsilon_{n,1}(\hat{r}) = \Upsilon_{n,1}(r_0) + E_n[\Upsilon_{n,1}(\hat{r}) - E_n[\Upsilon_{n,1}(r_0)] + o_P(1) \).

With this, we only need to show that \( E_n[\Upsilon_{n,1}(\hat{r})] - E_n[\Upsilon_{n,1}(r_0)] \) as follows:

\[
\begin{align*}
E_n[\Upsilon_{n,1}(\hat{r})] - E_n[\Upsilon_{n,1}(r_0)] &= E_n \left\{ \frac{\theta^t}{\sqrt{n}b^d} \sum_{i=1}^{n} \left[ \hat{\psi}_{r,i} \hat{\mu}_i \hat{K}_i - \psi_{r,i} \mu_i K_i \right] \right\} \\
&= E_n \left\{ \frac{\theta^t}{\sqrt{n}b^d} \sum_{i=1}^{n} \left[ \left( \hat{\psi}_{r,i} - \psi_{r,i} \right) \mu_i K_i + \psi_{r,i} \left( \hat{\mu}_i K_i - \mu_i K_i \right) + \left( \hat{\psi}_{r,i} - \psi_{r,i} \right) \left( \hat{\mu}_i K_i - \mu_i K_i \right) \right] \right\} \\
&= \Xi_{n,0} + \Xi_{n,1} + \Xi_{n,2}, \text{ say,}
\end{align*}
\]

where short notations are defined in obvious ways: \( \psi_{r,i} = \psi_r(Y_i^*(r_0)) \), \( \mu_i = \mu_{p,b,i}(r_0) \), \( K_i = \hat{K}_{b,i}(r_0) \), and similarly for \( \psi_{r,i}, \hat{\mu}_i, \) and \( \hat{K}_i \) with \( r_0 \) replaced by \( \hat{r} \). We prove the claim by showing that \( \Xi_{n,0}, \Xi_{n,1}, \) and \( \Xi_{n,2} \) are all \( o_P(1) \).

First, we study \( \Xi_{n,2} \). By the Taylor expansion of \( \hat{\mu}_i \hat{K}_i - \mu_i K_i \) at \( r_0 \) and taking the expectation to cancel out \( 1/b^d \), we can readily show that \( \Xi_{n,2} = O_P \left( \sqrt{n}b^d \| \hat{r} - r \|_\infty /b \right) = o_P(1) \).

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 \( E[\psi_r(Y_i^*(r_0)) \mu_i K_i \mid X_i, \varpi_i] = O_P \left( b^{p+1} \right) \). With this and following the analysis of \( \Xi_{n,2} \), we can show that \( \Xi_{n,1} = O_P \left( \sqrt{n}b^d (b^{p+1}) \| \hat{r} - r \|_\infty /b \right) = o_P(1) \).

For \( \Xi_{n,0} \), by the IID assumption we have

\[
\Xi_{n,0} = E_n \left[ \theta^t \sqrt{\frac{n}{b^d}} \mu_{p,b,i}(r_0) \hat{K}_{b,i}(r_0) E_n[\psi_r(Y_i^*(\hat{r})) - \psi_r(Y_i^*(r_0))] \right] (X_i, \varpi_i). \tag{D.7}
\]
Note that
\[
E_n \left( \hat{\psi}_{\tau,i} - \psi_{\tau,i} \mid X_i, \omega_i \right) = E_n \left[ -I \left( Y_i^* (\hat{\tau}) \leq 0 \right) + I \left( Y_i^* (r_0) \leq 0 \right) \right] U_i (r_0 (\omega_i)) \]
(D.8)
\[
= -G \left( \sum_{0 \leq \|j\| \leq p} \frac{1}{j!} D_j G^{-1} (\tau \mid u) (U_i (\hat{\tau}) - u)^j \bigg| U_i (r_0 (\omega_i)) \right) + G (\beta_b (\tau \mid u) \mid U_i (r_0 (\omega_i)))
\]
\[
= -g (\beta_b (\tau \mid u) \mid U_i (r_0)) \left\{ \sum_{0 \leq \|j\| \leq p} \frac{1}{j!} D_j G^{-1} (\tau \mid u) \frac{\partial U^j}{\partial U} \bigg|_{U = U_i (r_0 (\omega_i)) - u} \right\} (U_i (\hat{\tau}) - U_i (r_0)) + \Xi_{n,3}
\]
\[
= -g (\beta_b (\tau \mid u) \mid U_i (r_0)) \left\{ \sum_{0 \leq \|j\| \leq p} \frac{1}{j!} D_j G^{-1} (\tau \mid u) \frac{\partial U^j}{\partial Z} \bigg|_{U = U_i (r_0 (\omega_i)) - u} \right\} (\hat{\tau} - r_0) + \Xi_{n,3}
\]
\[
= -\chi (r_0 (\omega_i)) (\hat{\tau} (\omega_i) - r_0 (\omega_i)) + \Xi_{n,3},
\]
where \( \Xi_{n,3} \) is the residual term from the Taylor expansion, \( \frac{\partial U^j}{\partial Z} \) is the first order derivative of \( U \) with respect to \( Z \) (the second component of \( U \)) and the fourth equality holds by the fact that \( U_i (\hat{\tau}) - U_i (r_0) \equiv \left( X_i \hat{\tau} (\omega_i) \right) - \left( X_i r_0 (\omega_i) \right) = \left( \hat{\tau} (\omega_i) - r_0 (\omega_i) \right), \chi (r_0 (\omega_i)) \) is defined as \( \chi (r_0 (\omega_i)) \equiv g (\beta_b (\tau \mid u) \mid U_i (r_0)) \left\{ \sum_{0 \leq \|j\| \leq p} \frac{1}{j!} D_j G^{-1} (\tau \mid u) \frac{\partial U^j}{\partial Z} \bigg|_{U = U_i (r_0 (\omega_i)) - u} \right\} \).

It is easy to show that \( \Xi_{n,3} \) is asymptotically negligible. Substituting (D.8) back into (D.7) yields
\[
\Xi_{n,0} = -E_n \left[ \theta' \left( \frac{n}{\beta} \mu_{p,b,i} (r_0) \tilde{K}_{b,i} (r_0) \chi (r_0 (\omega_i)) (\hat{\tau} (\omega_i) - r_0 (\omega_i)) \right) + o (1) \right],
\]
(D.9)
\[
= -\int_{\Omega_i} \theta' \left( \frac{n}{\beta} \mu_{p,b,i} (r_0 (\omega)) \chi (r_0 (\omega)) \tilde{K}_{b,i} (\omega_i - \omega) \mu_{p,b,i} (r_0) F_\omega (d\omega) \right) + o (1),
\]
where \( \hat{\tau} (\omega_i) \equiv E \left[ \mu_{p,b} (r_0 (\omega)) \chi (r_0 (\omega)) \frac{\tilde{K}_{b,i} (r_0 (\omega))}{\beta} \bigg| \omega \right] \). The influence term in \( \Xi_{n,0} \) could be analyzed analogously as in the proof of Proposition 1 in Mammen et al. (2012). We decompose \( \hat{\tau} (\omega) - r_0 (\omega) \) in the same way as in equation (A.24) in Mammen et al. (2012). Following the same lines of proof as theirs and by the assumption that \( \hat{\beta} \equiv \alpha_P \left( n \beta d_X \right) \), the bias term is asymptotically negligible and we can obtain a similar expression as equation (A.26) in Mammen et al. (2012):
\[
\Xi_{n,0} = -\frac{1}{n \sqrt{d_X}} \sum_{i=1}^{n} \left( \frac{1}{n} \int_{\Omega_i} \theta' \left( r_0 (\omega_i) \right) c_1 \phi \tilde{K}_{b,i} (\omega_i) \mu_{p,b,i} (r_0) F_\omega (d\omega) \right) + o (1).
\]
Let \( t = (\omega_i - \omega) \). Using \( \hat{\beta} = o (b) \), we have \( \hat{\tau} \left( r_0 (\omega_i) + \hat{b} \right) = t \left( r_0 (\omega_i) \right) \{1 + o(\hat{b}/b)\} \) and
\[
\Xi_{n,0} = -\frac{1}{n \beta d_X} \sum_{i=1}^{n} \epsilon_i \theta' \left( r_0 (\omega_i) \right) c_1 \phi \tilde{K}_{b,i} (\omega_i) \mu_{p,b,i} (r_0) F_\omega \left( d\omega \right) + o (1)
\]
\[
\equiv \Xi_{n,0} + o (1).
\]
We now inspect the term \( \hat{\tau} (r_0 (\omega_i)) \). It is not hard to see that
\[
\hat{\tau} \left( r_0 (\omega_i) \right) \equiv E \left[ \mu_{p,b} (r_0 (\omega)) \chi (r_0 (\omega)) \frac{\tilde{K}_{b} (r_0 (\omega))}{\beta} \bigg| \omega = \omega_i \right]
\]
\[
\equiv \frac{\mu_{p,b} (r_0 (\omega_i)) \chi (r_0 (\omega_i)) \tilde{K}_{b} (r_0 (\omega_i)) f_{X|\omega=\omega_i} (x) \{1 + o (1)\},
\]
where $\overline{r}_{p,b} (r_0 (\omega))$ are the term left after doing integration of $\mu_{p,b} (r_0 (\omega))$ over the $X$ variable, $\overline{K}_b (r_0 (\omega))$ is the kernel function of the $Z$ variable only. Thus $E [\varphi (r_0 (\omega))]^2 = O(b^d)$. Since all terms other than $\varphi (r_0 (\omega))$ in $\Xi_{n,4}$ have finite second moments, by the IID assumption and the conditions that $E (\varepsilon_i | \omega_i) = 0$ and $E (\varepsilon_i^2) < \infty$, we have that $\text{var}(\Xi_{n,4}) = O(b^{2d}) = o(1)$, implying that $\Xi_{n,4} = o_P (1)$ by Markov inequality. So we have $\Xi_{n,0} = o_P (1)$.

**Proof of Lemma C.10.** To comply with the notations in Mammen et al. (2012). Let $\bar{z}^{(2)} (r) \equiv \psi_r (Y_i^* (r)) \mu_{p,b,i} (r) \times \bar{K}_i (r) / b^d$ and $\overline{\Upsilon}_{n,1} (r) \equiv \frac{1}{n} \sum_{i=1}^n \bar{z}^{(2)} (r)$. Then we have $\Upsilon_{n,1} (r) = \sqrt{nb^d} \overline{\Upsilon}_{n,1} (r) \theta$. $\overline{\Upsilon}_{n,1} (r)$ is of similar structure as the term in Mammen et al.’s Lemma 4. Using their result, we have

$$
\overline{\Upsilon}_{n,1} (\hat{r}) - \overline{\Upsilon}_{n,1} (r_0) - E_n [\Upsilon_{n,1} (\hat{r})] + E_n [\Upsilon_{n,1} (r_0)] \equiv O_P (\Xi_{n,5})
$$

where\(^{17}\) $\Xi_{n,5} = O_p \left( \frac{(nb^d)^{-1/2} \| \hat{r} - r_0 \|_\infty^{\varepsilon / 2}}{b^{\varepsilon}} \right) = O_p \left( (nb^d)^{1/2} \| \hat{r} - r_0 \|_\infty^{\varepsilon / 2} \right)$. By Markov inequality, we have

$$
\sqrt{nb^{2d}} \left\{ \overline{\Upsilon}_{n,1} (\hat{r}) - \overline{\Upsilon}_{n,1} (r_0) - E_n [\Upsilon_{n,1} (\hat{r})] + E_n [\Upsilon_{n,1} (r_0)] \right\} \equiv O_P (1)
$$

as desired. ■

### E 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 $Y_i = m (X_i, A_i) = (1 + 0.1 X_i^2) A_i$ and $\gamma (X_i, A_i) = X_i (1 - A_i)$.

First, it is easy to verify that under Assumption A.2,

$$
G (y|x, z) = P \left( (1 + 0.1 X_i^2) A_i \leq y | X_i = x, Z_i = z \right) = F_{A|Z} \left( \frac{y}{v (x)} \right).
$$

$$
G^{-1} (\tau|x, z) = \nu (x) F_{A|Z}^{-1} \left( \nu (x) \tau \right),
$$

$$
G^{-1} (G (y|x, z)) = F_{A|Z}^{-1} \left( F_{A|Z} \left( \frac{y}{v (x)} \right) \right) = \frac{y}{v (x)},
$$

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

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

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

Noting that $P \left( (1 + 0.1 X_i^2) A_i + c_n x e^{-A_i} \leq 0 \right) \to 0$ for any $x \in \mathcal{X}$ as $c_n \to 0$, without loss of generality we focus on the case where $y > 0$. We consider two cases: (1) $x > 0$, and (2) $x \leq 0$. Note that $m_n (x, a) = v (x) a + c_n x e^{-a}$.

First, we consider case (1). Noting that

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

\(^{17}\)The rate of the residual term is $O \left( n^{-k} \right)$ in Mammen et al. (2012)’s notation. The corresponding terms in our paper to $n^{-1} \left( 1 - e^{-r/\min_n} \right) n^{-1} \max \left( \delta, \gamma \right)$ in $O \left( n^{-k} \right)$ are $\sqrt{\max_n \left( \frac{1 - e^{-r}}{2 \beta} \right)}, \| \hat{r} - r_0 \|_\infty^{\varepsilon / 2}$, respectively.
\(m_n(x,a)\) is strictly increasing in \(a\) when \(a > \ln \left( \frac{c_n x}{v(x)} \right)\) and strictly decreasing in \(a\) when \(a < \ln \left( \frac{c_n x}{v(x)} \right)\). In addition, when \(a = \ln \left( \frac{c_n x}{v(x)} \right)\), we have
\[
m_n(x,a) = v(x) \ln \left( \frac{c_n x}{v(x)} \right) + c_n x e^{-\ln \left( \frac{c_n x}{v(x)} \right)} = v(x) \left[ \ln \left( \frac{c_n x}{v(x)} \right) + 1 \right]< 0 \text{ for sufficiently small } c_n.
\]
With these, we can easily argue that for any \(y > 0\) and \(x > 0\), the equation \(m_n(x,a) = y\) has exactly two solutions, \(a_{1n}(x,y)\) and \(a_{2n}(x,y)\), of different signs, so that we can assume that \(a_{1n}(x,y) > 0\) and \(a_{2n}(x,y) < 0\) without loss of generality, and obtain that
\[
G_n(y|x,z) = P(m_n(x,A) \leq y | Z = z) = P(a_{2n}(x,y) \leq A \leq a_{1n}(x,y) | Z = z)
= F_{A|Z}(a_{1n}(x,y) | z) - F_{A|Z}(a_{2n}(x,y) | z).
\]
\tag{E.1}
In addition, it is easy to argue that \(a_{1n}(x,y)\) is bounded above from infinity for any fixed \(x > 0\) and \(y > 0\) 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 \(v(x) a + c_n x e^{-a} = y\) as
\[
a + \frac{c_n x}{v(x)} e^{-a} = \frac{y}{v(x)}.
\]
By the implicit function theorem, we can write the positive solution in terms of \(\frac{c_n x}{v(x)}\) and \(\frac{y}{v(x)}\):
\[
a_{1n}(x,y) = \varphi \left( \frac{y}{v(x)}, \frac{c_n x}{v(x)} \right),
\]
where \(\varphi\) 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
\[
a_{1n}(x,y) = \varphi \left( \frac{y}{v(x)}, 0 \right) + \varphi_2 \left( \frac{y}{v(x)}, 0 \right) \frac{c_n x}{v(x)} + o(c_n)
\]
\tag{E.3}
where \(\varphi_2 (\cdot, \cdot)\) denotes the partial derivative of \(\varphi\) with respect to its second argument. In addition, (E.2) implies
\[
a_{1n}(x,y) = \frac{y}{v(x)} - \frac{c_n x}{v(x)} e^{-a_{1n}(x,y)}
= \frac{y}{v(x)} - \frac{c_n x}{v(x)} e^{-[\varphi \left( \frac{y}{v(x)}, 0 \right) + \varphi_2 \left( \frac{y}{v(x)}, 0 \right) \frac{c_n x}{v(x)} + o(c_n)]}
= \frac{y}{v(x)} - \frac{c_n x}{v(x)} e^{-\varphi \left( \frac{y}{v(x)}, 0 \right)} + o(c_n)
\]
\tag{E.4}
Comparing (E.3) with (E.4), we can conclude that \(\varphi \left( \frac{y}{v(x)}, 0 \right) = \frac{y}{v(x)}\), implying that
\[
a_{1n}(x,y) = \frac{y}{v(x)} - \frac{c_n x}{v(x)} e^{-\frac{y}{v(x)}} + o(c_n).
\]
\tag{E.5}
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 (E.2) by setting \(\alpha = e^a\) and rewriting (E.2) in terms of \(\alpha\):
\[
\ln(\alpha) a + \frac{c_n x}{v(x)} = \frac{y}{v(x)} \alpha.
\]
\tag{E.6}
The negative solution \( a_{2n}(x, y) \) to (E.2) corresponds to a positive solution \( \alpha_{2n}(x, y) \) to (E.6). 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

\[
\alpha_{2n}(x, y) = \psi \left( \frac{y}{v(x)} \right) c_n x + o(c_n) \tag{E.7}
\]

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

\[
\alpha_{2n}(x, y) = \psi \left( \frac{y}{v(x)} \right) 0 + \psi_2 \left( \frac{y}{v(x)} \right) 0 c_n x + o(c_n) \tag{E.8}
\]

where \( \psi_2(\cdot, \cdot) \) denotes the partial derivative of \( \psi \) 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 \( \psi \left( \frac{y}{v(x)} \right) 0 = 0 \) and \( \psi_2 \left( \frac{y}{v(x)} \right) > 0 \) in (E.8). Consequently, we have

\[
a_{2n}(x, y) = \ln \left( \psi_2 \left( \frac{y}{v(x)} \right) 0 c_n x + o(c_n) \right) \tag{E.9}
\]

where \( \psi_2 \left( \frac{y}{v(x)} \right) > 0 \).

Combining (E.1), (E.5) and (E.9), we have

\[
G_n(y|x, z) = \left. F_{A|Z} \left( \frac{y}{v(x)} \right) - \frac{c_n x}{v(x)} e^{-\frac{y}{\varepsilon(x)}} + o(c_n) \right| z = \left. F_{A|Z} \left( \ln \left( \psi_2 \left( \frac{y}{v(x)} \right) 0 c_n x + o(c_n) \right) \right) \right| z
\]

\[
= \left. F_{A|Z} \left( \frac{y}{v(x)} \right) \right| z - \left. \phi \left( \frac{y}{v(x)} \right) 0 c_n x + o(c_n), z \right| z + o(c_n)
\]

\[
= \left. F_{A|Z} \left( \frac{y}{v(x)} \right) \right| z - \left. \frac{c_n x}{v(x)} e^{-\frac{y}{\varepsilon(x)}} - \phi_1(0, z) \psi_2 \left( \frac{y}{v(x)} \right) 0 c_n x + o(c_n) \right| z
\]

where the second equality follows from (5.5) and the last equality follows from Taylor expansions, and \( \phi_1(\cdot, \cdot) \) is the partial derivative of \( \phi \) with respect to its first element. Therefore, we have for \( y > 0 \) and \( x > 0 \).

\[
G_n^{-1}(G_n(y|x, z)|x^*, z) - G_n^{-1}(G(y|x, z)|x^*, z) = \frac{-c_n}{v(x)}
\]

\[
= \left. \frac{c_n}{v(x)} e^{-\frac{y}{\varepsilon(x)}} + \phi_1(0, z) \psi_2 \left( \frac{y}{v(x)} \right) 0 c_n x + o(1) \right| z
\]

(E.10)

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 \( y > 0 \), and \( x \leq 0, \frac{\partial m_n(x, a)}{\partial a} = v(x) - c_n x e^{-a} > 0 \) \( \forall a \in \mathcal{A} \), and the equation

\[
v(x) a + c_n x e^{-a} = y
\]

has a unique positive solution, which we continue to denote as \( a_{1n}(x, y) \). As in case (1), the result in (E.5) continues to hold, and we have

\[
G_n(y|x, z) = P(m_n(x, A) \leq y|Z = z) = P(-\infty \leq A \leq a_{1n}(x, y)|Z = z)
\]

\[
= \left. F_{A|Z} \left( a_{1n}(x, y) \right) \right| z.
\]

(E.11)
Then for any $y > 0$ and $x \leq 0$, we have

$$G_n^{-1}(G_n(y|x,z)|x^*, z) - G_n^{-1}(G(y|x,z)|x^*, z) = c_n \left[ \frac{x}{v(x)} e^{-\frac{x}{v(x)}} + o(1) \right]. \quad (E.12)$$

It follows that for any $y > 0$, we have

$$\Theta_n^\dagger(y; x, z) = \left[ -\frac{x}{v(x)} e^{-\frac{x}{v(x)}} + \omega(y; x, z) + o(1) \right] 1\{x > 0\} + \left[ \frac{x}{v(x)} e^{-\frac{x}{v(x)}} + o(1) \right] 1\{x \leq 0\}. \quad (E.13)$$

where

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

(E.14)

REFERENCES


