Testing Generalized Regression Monotonicity

Yu-Chin Hsu†

Institute of Economics

Academia Sinica

Chu-An Liu‡

Institute of Economics

Academia Sinica

Xiaoxia Shi§

Department of Economics

University of Wisconsin at Madison

First version: December 15, 2011

This version: June 12, 2016

† E-mail: ychsu@econ.sinica.edu.tw. 128 Academia Road, Section 2, Nankang, Taipei, 115 Taiwan.

‡ E-mail: caliu@econ.sinica.edu.tw. 128 Academia Road, Section 2, Nankang, Taipei, 115 Taiwan.

§ E-mail: xshi@ssc.wisc.edu. 1180 Observatory Dr, Social Sciences Building ♯6428, Madison, WI 53706-1320 USA.

Acknowledgements: Yu-Chin Hsu gratefully acknowledges the research support from Ministry of Science and Technology of Taiwan (MOST103-2628-H-001-001-MY4) and Career Development Award of Academia Sinica, Taiwan. All errors and omissions are our own responsibility.

Abstract

We propose a test for a generalized regression monotonicity (GRM) hypothesis. The GRM hypothesis is the sharp testable implication of the monotonicity of certain latent structures, as we show in this paper. Examples include the monotone instrumental variable assumption of Manski and Pepper (2000) and the monotonicity of the conditional mean function when only interval data are available for the dependent variable. These instances of latent monotonicity can be tested using our test. Moreover, the GRM hypothesis includes regression monotonicity and stochastic monotonicity as special cases. Thus, our test also serves as an alternative to existing tests for those hypotheses. We show that our test controls the size uniformly over a broad set of data generating processes asymptotically, is consistent against fixed alternatives, and has nontrivial power against some $n^{-1/2}$ local alternatives.

JEL classification: C01, C12, C21

Keywords: Generalized regression monotonicity, hypothesis testing, monotone instrumental variable, interval outcome, uniform size control.

1 Introduction

In this paper, we construct a test for the generalized regression monotonicity (GRM) hypothesis defined as:

$$H_0: \ E_P[f^{(1)}(W,\tau)|X=x_1, Z=z] \geq E_P[f^{(2)}(W,\tau)|X=x_2, Z=z], \quad \forall x_1, x_2 \in \mathcal{X} \text{ and } x_1 \geq x_2, \quad \forall z \in \mathcal{Z} \text{ and } \tau \in \mathcal{T}, \tag{1.1}$$

where $W=(Y',X',Z')'$ are observed random variables generated from a distribution P, E_P denotes the expectation under P, and $f^{(1)}(W,\tau)$ and $f^{(2)}(W,\tau)$ are known real valued functions indexed by $\tau \in \mathcal{T}$ where $\mathcal{T}$ can be either finite or infinite. The random variables Y, X, and Z are of dimensions $d_y \geq 1$, $d_x \geq 1$, and $d_z \geq 0$, respectively.[1] The sets $\mathcal{X}$ and $\mathcal{Z}$ are the support sets of X and Z, respectively. Without loss of generality, we assume that $\mathcal{X} \subseteq [0,1]^{d_x}$ and $\mathcal{Z} \subseteq [0,1]^{d_z}$.[2]

The null hypothesis in (1.1) is the sharp testable implications of the monotonicity of certain latent structures. One example is the monotonicity of potential outcomes in an instrumental variable, better known as the monotone instrumental variable (MIV) assumption after Manski and Pepper (2000). The MIV assumption has been recogonized as a useful identification tool in Manski and Pepper (2000, 2009), Kreider and Pepper (2007), Kreider and Hill (2009), and Gunderson, Kreider, and Pepper (2012).[3] However, a test for MIV validity has not been developed.[4]

Another example is the monotonicity of the conditional mean of an interval-observed dependent variable. The interval data problem is wide-spread in empirical research either due to survey design, where people are asked to choose from several brackets rather than to report their actual value of a variable, or due to some inherent missing data problems, for example, potential wage for females. As a result, regressions using interval data as

[1] If $d_z = 0$, then there is no Z in the model.

[2] A strictly monotone transformation can always be applied to bring the support of each component to $[0,1]$ without changing the information content of the inequalities.

[3] A Stata command for bounding treatment effects under the MIV and other related assumptions is developed by McCarthy, Millimet, and Roy (2015).

[4] Chetverikov (2013) develops a test for the related monotone treatment response and the monotone treatment selection assumptions. Kitagawa (2015) develops a test for IV validity in the context of local average treatment effect.

the dependent variable are unavoidable sometimes. Manski and Tamer (2002) provide econometrics tools for estimation in such situations, but a nonparametric test for the monotonicity of the regression function has not been specifically considered. We show that the sharp testable implications of both the MIV assumption and the latent regression monotonicity are in the form of GRM, and our test can be used for these hypotheses.

The GRM hyothesis also includes regression monotonicity and stochastic monotonicity as special cases. Thus, our test also offers an alternative to existing tests of those. Regression monotonicity arises in a lot of problems in economics. For example, many comparative static hypotheses directly take the form of regression monotonicity. In addition, Chetverikov (2013) shows that regression monotonicity is the testable implication of the monotone treatment response assumption and monotone treatment selection assumption introduced in Manski and Pepper (2000). Existing tests for regression monotonicity have been proposed by Ghosal, Sen, and van der Vaart (2000), Hall and Heckman (2000), and Chetverikov (2013). Testing stochastic monotonicity is useful for bounding parameters in a selection model and for assessing the stationarity of a Markov process. See Lee, Linton, and Whang (2009) and Seo (2015) for details and further applications. Existing tests for stochastic monotonicity include Lee, Linton, and Whang (2009), Delgado and Escanciano (2012), and Seo (2015). We compare these existing tests to our test in Section 2.4.

To test the GRM, we adapt Andrews and Shi's (2013a, AS hereafter) instrumental function approach to transform the conditional inequality hypothesis into an inequality hypothesis that involves only unconditional moments without loss of information content of the original inequality hypothesis. The adaption is needed because each of our inequalities involves conditional moments evaluated at two *different* values of the conditioning variable, for which AS' approach does not apply.

After the transformation, we approximate each unconditional moment by its sample counterpart, and construct a Cramér-von Mises type test. Since our problem involves moment inequalities, we employ the generalized moment selection method (GMS) to improve the power of the test as in AS, and propose both a bootstrap GMS critical value and a multiplier GMS critical value. We show that our test has uniform asymptotic size over a broad set of data generating processes, is consistent against fixed alternatives, and has nontrivial local power against some $n^{-1/2}$-local alternatives. We conduct Monte-Carlo simulations for two examples to examine the finite-sample properties of our test.

A different test from ours for the GRM may be constructed by verifying the conditions in Lee, Song, and Whang's (2016) recent paper. Comparing to such a test, our test has the advantage of not requiring a non-parametric estimator of the conditional moments.

The rest of this paper is organized as follows. In Section 2, we give five motivating examples for testing GRM. We introduce the modified instrumental function approach, and propose our test in Section 3. Uniform size property and power properties of our tests are given in Section 4 and Section 5, respectively. Section 6 reports Monte-Carlo simulation results, and Section 7 concludes. All mathematical proofs are deferred to the Appendix.

We adopt the following convention in the paper: for $x_1, x_2 \in R^{d_x}$ with $d_x \geq 2$, we say that $x_1 \geq x_2$ iff $x_{1s} \geq x_{2s}$ for all $s = 1, \ldots, d_x$, where x_{js} is the sth element of vector x_j. Also, we say that $x_1 > x_2$ iff $x_{1s} \geq x_{2s}$ for all $s = 1, \ldots, d_x$, and $x_{1k} > x_{2k}$ for some $k \in \{1, \ldots, d_x\}$. Finally, $x_1 \gg x_2$ iff $x_{1s} > x_{2s}$ for all $s = 1, \ldots, d_x$.

2 Examples of GRM

Hypotheses given in (1.1) are of interest in a wide array of econometrics problems. We give several examples below.

2.1 Testing MIV

Example 2.1. *The MIV condition proposed by Manski and Pepper (2000) has been used to obtain tighter identification in a selection model. One can test the MIV condition by testing a hypothesis of the form of* H_0 *in (1.1). To fix ideas, let* D *be a binary treatment and* $(Y(0), Y(1))$ *be the potential outcomes. The variable* $Y(0)$ *is only observed when* $D = 0$*, and* $Y(1)$ *is only observed when* $D = 1$*. Let* X *be a monotone IV in the sense of Manski and Pepper (2000):*

$$H_0^{MIV} : E[Y(d)|X = x_1] \geq E[Y(d)|X = x_2], \quad \text{for all } x_1 \geq x_2, \text{ for } d = 0, 1. \tag{2.1}$$

Suppose that $Y(0)$ and $Y(1)$ are known to lie in the deterministic interval $[y_l, y_u]$. Then the MIV condition in (2.1) implies the following hypothesis:

$$\begin{aligned} H_0^{GRM} : E[f^{(1)}(Y,\tau)|X = x_1] &\geq E[f^{(2)}(Y,\tau)|X = x_2], \\ &\textit{for all } x_1 \geq x_2, \quad \textit{for } \tau = 1 \textit{ and } 2, \end{aligned} \tag{2.2}$$

and

$$\begin{aligned} f^{(1)}(Y,1) &= YD + y_u \cdot (1-D), \qquad f^{(1)}(Y,2) = y_u D + Y \cdot (1-D), \\ f^{(2)}(Y,1) &= YD + y_l \cdot (1-D), \qquad f^{(2)}(Y,2) = y_l D + Y \cdot (1-D). \end{aligned} \tag{2.3}$$

In this example, X can be a vector. Additional control variables Z may be present.

As shown in the following theorem, H_0^{MIV} implies H_0^{GRM}, and thus should be rejected if the latter is rejected. The theorem also shows that H_0^{GRM} is the sharp, that is, the strongest, testable implication of H_0^{MIV}. The proof of the theorem is given in the appendix.

Theorem 2.1. (i) *Suppose that the distribution of $(Y(1), Y(0), D, X)$ satisfies H_0^{MIV}, and $Y(1), Y(0) \in [y_\ell, y_u]$. Then the distribution of (Y, D, X) satisfies H_0^{GRM}.*
(ii) *Suppose that $Y \in [y_\ell, y_u]$, and the distribution of (Y, D, X) satisfies H_0^{GRM}. Then there exists $(Y(1), Y(0))$ such that $Y = DY(1) + (1-D)Y(0)$, $y_\ell \leq Y(1), Y(0) \leq y_u$, and the distribution of $(Y(1), Y(0), D, X)$ satisfies H_0^{MIV}.*

Even though H_0^{GRM} is the sharp testable implication of H_0^{MIV}, the two are not equivalent. The sharpness only guarantees that H_0^{MIV} can not be ruled out based on observed data whenever H_0^{GRM} holds. It can certainly happen that H_0^{MIV} is violated when H_0^{GRM} holds, in which case our test asymptotically detects the violation with probability less than or equal to size. In general, the power of our test depends on the functions $E[y_u \cdot (1-D)|X = x]$ and $E[y_\ell \cdot (1-D)|X = x]$. The closer these two functions are, the more likely for us to detect the violation of the MIV assumption.

2.2 Testing Regression Monotonicity with Interval-Observed Dependent Variable

Example 2.2. *Consider a dependent variable Y and covariate vectors X and Z. The researcher is interested to know whether $E[Y|X = x, Z = z]$ is monotonically increasing* *in x. However, Y is not observed. Instead, Y is known to lie in the observed random interval $[Y_\ell, Y_u]$, as considered in Manski and Tamer (2002). Thus, one cannot directly test the null hypothesis:*

$$H_0^{LRM} : E[Y|X = x_1, Z = z] \geq E[Y|X = x_2, Z = z] \; \forall x_1 \geq x_2, \; \forall z, \tag{2.4}$$

where LRM stands for "latent regression monotonicity." We show that H_0^{LRM} can be tested through a GRM type hypothesis:

$$H_0^{GRM} : E[Y_u|X = x_1, Z = z] \geq E[Y_\ell|X = x_2, Z = z] \; \forall x_1 \geq x_2, \; \forall z. \tag{2.5}$$

We show in the next theorem that H_0^{GRM} in (2.5) is the sharp testable implication of H_0^{LRM}. The proof of this theorem is given in the appendix.

Theorem 2.2. (i) *Suppose that the distribution of (Y, X, Z) satiafies H_0^{LRM}, and that $Y \in [Y_u, Y_\ell]$. Then H_0^{GRM} in* (2.5) *holds.*
(ii) *Suppose that the distribution of (Y_u, Y_ℓ, X, Z) satisfies H_0^{GRM} in* (2.5). *Then, there exists a random variable Y such that $Y \in [Y_\ell, Y_u]$ everywhere, and that the distribution of (Y, X, Z) satisfies H_0^{LRM}.*

The remarks below Theorem 2.1 apply here as well, with H_0^{MIV} replaced by H_0^{LRM}, and with $E[y_u \cdot (1-D)|X = x]$ and $E[y_\ell \cdot (1-D)|X = x]$ replaced by $E[Y_u|X = x, Z = z]$ and $E[Y_\ell|X = x, Z = z]$, respectively.

2.3 Other examples

The GRM hypothesis includes the hypotheses of regression monotonicity, stochastic monotonicity, and higher-order stochastic monotonicity as special cases, as we describe in details now.

Example 2.3. *Suppose that $f^{(1)}(W, \tau) = f^{(2)}(W, \tau) = Y$ and Y is a scalar. Then H_0 in (1.1) reduces to:*

$$H_0 : \; E[Y|X = x, Z = z] \text{ is weakly increasing in } x \in \mathcal{X}, \forall z \in \mathcal{Z}. \tag{2.6}$$

This is the usual regression monotonicity hypothesis. Testing H_0 is a nonparametric version of testing the sign of a regression coefficient in a linear regression model. For *example, if Y is the survival of a patient and X is the daily dose of a certain drug given to the patient. Then H_0 implies that there is a monotone relationship between the daily dose and the survival rate as the dose varies in a chosen range $\mathcal{X}$. Note that if $d_z = 0$, then H_0 is the regression monotonicity hypothesis studied in Ghosal, Sen, and van der Vaart (2000) and Chetverikov (2013). See Chetverikov (2013) for more testing problems that can be formulated as (2.6) with $d_z = 0$.*

Example 2.4. *Suppose that $f^{(1)}(Y,\tau) = f^{(2)}(Y,\tau) = -1(Y \leq \tau)$ for $\tau \in R$ and $d_z = 0$. Then H_0 reduces to:*

$$H_0: \; F_{Y|X}(y|x) \text{ is non-increasing in } x \in \mathcal{X} \text{ for all } y \in R, \tag{2.7}$$

where $F_{Y|X}(y|x)$ denotes the conditional distribution of Y conditioning on $X = x$. Then H_0 is the stochastic monotonicity hypothesis studied in Lee, Linton, and Whang (2009), Delgado and Escanciano (2012), and Seo (2015).

Example 2.5. *Suppose that $f^{(1)}(Y,\tau) = f^{(2)}(Y,\tau) = -\frac{1}{(j-1)!}1(Y \leq \tau)(\tau - Y)^{j-1}$ for $\tau \in R$ and $d_z = 0$. Then H_0 reduces to:*

$$H_0: \mathcal{I}_j(y; F_{Y|X}(s|x)) \text{ is non-increasing in } x \in \mathcal{X} \text{ for all } y \in R, \tag{2.8}$$

where $\mathcal{I}_j(.;F)$ is the function that integrates the function F to order $j-1$ so that,

$$\begin{aligned}
\mathcal{I}_1(y;F) =& F(y), \\
\mathcal{I}_2(y;F) =& \int_0^y F(t)dt = \int_0^y \mathcal{I}_1(t;F)dt, \\
&\vdots \\
\mathcal{I}_j(y;F) =& \int_0^y \mathcal{I}_{j-1}(t;F)dt.
\end{aligned}$$

Therefore, H_0 is the higher-order stochastic monotonicity hypothesis. Shen (2015) studies the conditional higher-order stochastic monotonicity at a fixed point of $X = x$. Our test covers the uniform version of Shen's hypothesis.

2.4 Discussions

1. When Z contains only discrete random variables, the tests proposed in Ghosal, Sen, and van der Vaart (2000) and Chetverikov (2013) are applicable to Example 2.3,

and the tests proposed in Lee, Linton, and Whang (2009), Delgado and Escanciano (2012), and Seo (2015) are applicable to Example 2.4. These tests do not apply when Z contains continuous random variables. In addition, the tests of Ghosal, Sen, and van der Vaart (2000), Lee, Linton, and Whang (2009), and Delgado and Escanciano (2012) rely on least-favorable case critical value, and can have poor power when the data generating process is not close to the least-favorable case. None of the tests in the five papers mentioned apply to Examples 2.1 and 2.2, where $f^{(1)}(Y,\tau) \neq f^{(2)}(Y,\tau)$.

2. Chetverikov (2013) considers a testable implication of the monotone treatment selection and monotone treatment response assumptions of Manski and Pepper (2000), which, in the notation of Example 2.1, is

$$E[Y|X=x_1] \geq E[Y|X=x_2], \quad \text{for all } x_1 \geq x_2. \tag{2.9}$$

This is a special case of Example 2.3.

3. As we mentioned in Introduction, the only testing framework that covers Examples 2.1 and 2.2 is Lee, Song, and Whang's (2016). To be specific, let $\tilde{x} = (x_1, x_2, z)$, $\widetilde{\mathcal{X}} = \{\mathcal{X} \times \mathcal{X} \times \mathcal{Z} |\ x_1 \geq x_2\}$, $q_{\tau,1}(\tilde{x}) = E[f^{(1)}(Y,\tau)|X = x_1, Z = z]$ and $q_{\tau,2}(\tilde{x}) = E[f^{(2)}(Y,\tau)|X = x_2, Z = z]$, and let $\nu_\tau(\tilde{x}) = q_{\tau,2}(\tilde{x}) - q_{\tau,1}(\tilde{x})$. Then, (1.1) can be rewritten into Lee, Song and Whang's (2016) framework:

$$H_0: \ \nu_\tau(\tilde{x}) \leq 0 \ \text{ for all } (\tilde{x},\tau) \in \widetilde{\mathcal{X}} \times \mathcal{T}. \tag{2.10}$$

Lee, Song, and Whang's (2016) conditions for the validity of their test may cover hypothesis (2.10) under suitable primitive conditions. We do not aim to provide those primitive conditions in this paper because we take a different approach toward testing the GRM hypothesis. Unlike their approach, ours does not require preliminary nonparametric estimation.

3 Proposed Test

3.1 Model Transformation

In order to form a test statistic, we transform the conditional inequality hypothesis into an inequality hypothesis that involves only unconditional moments. The transformation should preserve all the information content of the original inequality hypothesis because otherwise the resulting test has no power against some fixed alternatives. The most closely related approach in the literature is AS, where they transform conditional moment inequalities into unconditional ones using an infinite set of instrumental functions. Our problem is more complicated because our inequalities involve conditional moments evaluated at different values of the conditioning variable.

We propose a modification to AS's instrumental function approach. The basic idea of our modified approach is to use two different instrumental functions on the two sides of the inequalities. To be specific, we find a set, $\mathcal{G}$, of $g = (g_x^{(1)}, g_x^{(2)}, g_z)$ such that (1.1) is equivalent to

$$H_0 : \nu_P(\tau, g) \equiv m_P^{(2)}(\tau, g) w_P^{(1)}(g) - m_P^{(1)}(\tau, g) w_P^{(2)}(g) \leq 0, \quad \text{for all } \tau \in \mathcal{T} \text{and for all } g \in \mathcal{G}, \tag{3.1}$$

where, for $j = 1$ and 2,

$$m_P^{(j)}(\tau, g) = E_P[f^{(j)}(Y, \tau) g_x^{(j)}(X) g_z(Z)], \quad w_P^{(j)}(g) = E_P[g_x^{(j)}(X) g_z(Z)]. \tag{3.2}$$

Like in AS, we also would like the set $\mathcal{G}$ to be simple enough in order for certain uniform central limit theorem to apply.

We consider two possible $\mathcal{G}$ choices, for both of which, we define the following notation:

$$\begin{aligned} C_{x,r} &\equiv \Big(\prod_{j=1}^{d_x} [x_j, x_j + r] \Big) \cap \mathcal{X} \quad \text{for } x \in \mathcal{X} \text{ and } r \in (0, 1], \\ C_{z,r} &\equiv \Big(\prod_{j=1}^{d_z} [z_j, z_j + r] \Big) \cap \mathcal{Z} \quad \text{for } z \in \mathcal{Z} \text{ and } r \in (0, 1]. \end{aligned} \tag{3.3}$$

For $\ell = (x_1, x_2, z, r) \in \mathcal{X}^2 \times \mathcal{Z} \times (0, 1]$, define

$$g_{x,\ell}^{(1)} = 1(x \in C_{x_1,r}), \quad g_{x,\ell}^{(2)} = 1(x \in C_{x_2,r}), \quad g_{z,\ell} = 1(z \in C_{z,r}). \tag{3.4}$$

The first $\mathcal{G}$ we consider is the set of the indicator functions of countable hypercubes:

$$\mathcal{G}_{\text{c-cube}} = \left\{ g_\ell \equiv \left(g_{x,\ell}^{(1)}, g_{x,\ell}^{(2)},\ g_{z,\ell}\right) : \ell \in \mathcal{L}_{\text{c-cube}} \right\}, \text{ where} \tag{3.5}$$
$$\mathcal{L}_{\text{c-cube}} = \Big\{ (x_1, x_2, z, r) :\ r = q^{-1},\ q \cdot (x_1, x_2, z) \in \{0, 1, 2, \cdots, q-1\}^{2d_x + d_z}, \\ x_1 \geq x_2, , \text{ and } q = q_0, q_0 + 1, \cdots \Big\},$$

and q_0 is a natural number.

The second $\mathcal{G}$ that we consider is the set of the indicator functions of a continuum of hypercubes:

$$\mathcal{G}_{\text{cube}} = \{ g_\ell : \ell \in \mathcal{L}_{\text{cube}} \}, \text{ where} \tag{3.6}$$
$$\mathcal{L}_{\text{cube}} = \{ (x_1, x_2, z, r) : x_1, x_2 \in [0, 1-r]^{2d_x + d_z},\ x_1 \geq x_2,\ r \in (0, \bar{r}] \},$$

for some $0 < \bar{r} < 1$.

Because there is a one-to-one mapping between $\mathcal{G}_{\text{cube}}$ (or $\mathcal{G}_{\text{c-cube}}$) and the set of indices $\mathcal{L}_{\text{cube}}$ (or $\mathcal{L}_{\text{c-cube}}$), for the remainder of the paper, we will use ℓ to stand for g_ℓ when used inside a function to simplify notation. For example, $\nu_P(\tau, g_\ell)$ will be written as $\nu_P(\tau, \ell)$, $m_P^{(j)}(\tau, g_\ell)$ as $m_P^{(j)}(\tau, \ell)$, and $w_P^{(j)}(g_\ell)$ as $w_P^{(j)}(\ell)$.

Both $\mathcal{G}_{\text{c-cube}}$ and $\mathcal{G}_{\text{cube}}$ are Vapnik-Cĕrvonenkis (VC) sets, and thus will guarantee the application of a uniform central limit theorem. Both are also rich enough to capture all the information provided by (1.1), which is shown in the following lemma.

Assumption 3.1. *Suppose that for $j = 1$ and 2, $E_P[f^{(j)}(Y, \tau) | X = x, Z = z]$ is continuous on $\mathcal{X} \times \mathcal{Z}$ for all $\tau \in \mathcal{T}$ under distribution P.*

Lemma 3.1. *Suppose Assumption 3.1 holds. Then for $\mathcal{G} = \mathcal{G}_{\text{c-cube}}$ or $\mathcal{G} = \mathcal{G}_{\text{cube}}$, H_0 in (1.1) is equivalent to that in (3.1).*

The proof for Lemma 3.1 is relatively straightforward because we assume continuity of $E_P[f^{(j)}(Y, \tau) | X = x, Z = z]$, unlike in AS. The continuity assumption allows X and Z to be discretely distributed and is reasonably weak. Also note that if $f^{(1)}(Y, \tau) = f^{(2)}(Y, \tau)$, we can drop the instrumental functions with $x_1 = x_2$ because for those g's, for all P, $m_P^{(1)}(\tau, \ell) = m_P^{(2)}(\tau, \ell)$ and $w_P^{(1)}(\ell) = w_P^{(2)}(\ell)$, which implies that $\nu_P(\tau, \ell) = 0$.

3.2 Estimation of $\nu_P(\tau, \ell)$

In the following, all results hold for both $\mathcal{G}_{\text{c-cube}}$ and $\mathcal{G}_{\text{cube}}$, so for notational simplicity, we suppress the subscripts "c-cube" and "cube" and just write $\mathcal{G}$ and $\mathcal{L}$ unless necessary. Suppose, we have an i.i.d. sample of size n

Now that we have transformed the conditional inequalities into unconditional inequalities, we are ready to introduce the test statistic. Define, for $j = 1, 2$,

$$\begin{aligned} m_i^{(j)}(\tau, \ell) &= m^{(j)}(W_i, \tau, \ell) = f^{(j)}(Y_i, \tau) g_{x,\ell}^{(j)}(X_i) g_{z,\ell}(Z_i) \\ w_i^{(j)}(\ell) &= w^{(j)}(W_i, \ell) = g_{x,\ell}^{(j)}(X_i) g_{z,\ell}(Z_i). \end{aligned} \tag{3.7}$$

Let the sample means of them be

$$\hat{m}_n^{(j)}(\tau, \ell) = \frac{1}{n} \sum_{i=1}^n m_i^{(j)}(\tau, \ell), \quad \hat{w}_n^{(j)}(\ell) = \frac{1}{n} \sum_{i=1}^n w_i^{(j)}(\ell). \tag{3.8}$$

We estimate $\nu_P(\tau, \ell)$ by its sample analogue:

$$\hat{\nu}_n(\tau, \ell) = \hat{m}_n^{(2)}(\tau, \ell) \hat{w}_n^{(1)}(\ell) - \hat{m}_n^{(1)}(\tau, \ell) \hat{w}_n^{(2)}(\ell). \tag{3.9}$$

As we mentioned above, the simplicity of $\mathcal{G}_{\text{c-cube}}$ and $\mathcal{G}_{\text{cube}}$, along with a manageability condition on $\mathcal{T}$ (given later) makes sure that $\sqrt{n}(\hat{\nu}_n(\tau, \ell) - \nu_P(\tau, \ell))$ satisfies a functional central limit theorem.

3.3 Test Statistic

Here we define the test statistic $\widehat{T}_n$ for our test. Let $\hat{\sigma}_n^2(\tau, \ell)$ be

$$\begin{aligned} &\hat{\sigma}_n^2(\tau, \ell) \\ =&\frac{1}{n} \sum_{i=1}^n \Big\{ \hat{w}_n^{(1)}(\ell) \big(m_i^{(2)}(\tau, \ell) - \hat{m}_n^{(2)}(\tau, \ell) \big) + \hat{m}_n^{(2)}(\tau, \ell) \big(w_i^{(1)}(\ell) - \hat{w}_n^{(1)}(\ell) \big) \\ &\qquad - \hat{w}_n^{(2)}(\ell) \big(m_i^{(1)}(\tau, \ell) - \hat{m}_n^{(1)}(\tau, \ell) \big) - \hat{m}_n^{(1)}(\tau, \ell) \big(w_i^{(2)}(\ell) - \hat{w}_n^{(2)}(\ell) \big) \Big\}^2, \end{aligned} \tag{3.10}$$

which is an estimator for the asymptotic variance of $\sqrt{n}(\hat{\nu}_n(\tau, \ell) - \nu_P(\tau, \ell))$. Note that $\hat{\sigma}_n^2(\tau, \ell)$ may be close to 0 with non-negligible probability for some $(\tau, \ell) \in \mathcal{T} \times \mathcal{L}$. This is not desirable because the inverse of it needs to be consistent for its population counterpart uniformly over $\mathcal{T} \times \mathcal{L}$ for the test statistics considered below. In consequence, as in AS, we consider a modification, denoted as $\hat{\sigma}^2_{\epsilon,n}(\tau,\ell)$, that is bounded away from 0. For some fixed $\epsilon > 0$, define $\hat{\sigma}^2_{\epsilon,n}(\tau,\ell)$ as

$$\hat{\sigma}^2_{\epsilon,n}(\tau,\ell) = \max\{\hat{\sigma}^2_n(\tau,\ell), \epsilon\}, \quad \text{for all } (\tau,\ell) \in \mathcal{T} \times \mathcal{L}. \tag{3.11}$$

Note that unlike AS, the $\hat{\sigma}^2_{\epsilon,n}(\tau,\ell)$ in (3.11) is not scale-equivariant to the moment conditions, meaning that our test statistic defined below is not scale-invariant. It is hard to get scale-equivariance in our case due to the presence of τ. See Andrews and Shi (2015) for the use of non-scale-equivariant weights as well.

Let Q be a probability measure on $\mathcal{T} \times \mathcal{L}$, and our test statistic is defined as

$$\widehat{T}_n = \int \max\left\{\sqrt{n}\frac{\hat{\nu}_n(\tau,\ell)}{\hat{\sigma}_{\epsilon,n}(\tau,\ell)}, 0\right\}^2 dQ(\tau,\ell). \tag{3.12}$$

We only consider the measures such that $Q(\tau,\ell) = Q_{\mathcal{T}}(\tau)Q_{\mathcal{L}}(\ell)$ for measures $Q_{\mathcal{T}}$ on $\mathcal{T}$ and $Q_{\mathcal{L}}$ on $\mathcal{L}$ because such measures are sufficient for our purpose in all cases that we can think of. We require that the support of Q equal $\mathcal{T} \times \mathcal{L}$. The support condition is needed to ensure that there is no information loss in the aggregation, and is formally stated in the next assumption. Let d_τ be a metric on $\mathcal{T}$ and d_ℓ be a metric on $\mathcal{L}$. Let $B_c(\tau_*) = \{\tau \in \mathcal{T} : d_\tau(\tau,\tau_*) \leq c\}$, and $B_c(\ell_*) = \{\ell \in \mathcal{L} : d_\ell(\ell,\ell_*) \leq c\}$.

Assumption 3.2. *For any $c > 0$, any $\tau \in \mathcal{T}$, and any $\ell \in \mathcal{L}$,* (a) $Q_{\mathcal{T}}(B_c(\tau)) > 0$, *and* (b) $Q_{\mathcal{L}}(B_c(\ell)) > 0$.

We give some examples of Q that satisfies Assumption 3.2. Because we only consider product measures, we can choose $Q_{\mathcal{T}}$ and $Q_{\mathcal{L}}$ separately. For $Q_{\mathcal{T}}$, if $\mathcal{T}$ is a singleton or a finite set as in Examples 2.1-2.3, we let $Q_{\mathcal{T}}$ assign equal weight on each element in $\mathcal{T}$. If $\mathcal{T}$ contains a continuum of elements as in Examples 2.4 and 2.5, and $\mathcal{T}$ has a finite support, e.g., $[a,b]$, which would be true if we know in advance that Y has support on $[a,b]$, we can let $Q_{\mathcal{T}}$ be a uniform distribution on $[a,b]$. If $\mathcal{T}$ has support on the whole real line, we can let $Q_{\mathcal{T}}$ be from a standard normal distribution. For $Q_{\mathcal{L}}$, if $\mathcal{L} = \mathcal{L}_{\text{c-cube}}$, we can let $Q_{\mathcal{L}}$ assign weight $\propto q^{-2}$ on each q where $\propto$ stands for "is proportional to," and, for each q, let $Q_{\mathcal{L}}$ assign equal weight on each instrumental function with $r = q^{-1}$.[5] If $\mathcal{L} = \mathcal{L}_{\text{cube}}$, we can let the marginal of $Q_{\mathcal{L}}$ on $(0,\bar{r}]$ be a uniform distribution and conditional on each r, let $Q_{\mathcal{L}}$ induce a uniform distribution on $\{(x_1,x_2,z) \in [1-r]^{2d_x+d_z} : x_1 \geq x_2\}$.[6]

[5] Note that for each q, there are $(q(q+1)/2)^{d_x} \cdot q^{d_z}$ of instrumental functions with $r = q^{-1}$.

[6] There are many choices of Q satisfying Assumption 3.2. Different choices of Q will not affect the

3.4 Generalized Moment Selection

We define the critical value for our test. Note that our null hypothesis involves inequality constraints. It is well known that if one obtains critical values based on the least favorable configuration (LFC) where all inequalities are assumed to be binding, the power of the test may be poor when the data generating process is not local to the LFC. We employ the generalized moment selection (GMS) approach in AS to achieve better power property. Let $\{\kappa_n : n \geq 1\}$ be a sequence of positive numbers that diverges to infinity as $n \to \infty$ and $\{B_n : n \geq 1\}$ be a non-decreasing sequence of positive numbers that diverges to infinity as $n \to \infty$ as well. Let the GMS function $\phi_n(\tau, \ell)$ be

$$\phi_n(\tau, \ell) = -B_n \cdot 1\Big(\sqrt{n}\frac{\hat{\nu}_n(\tau, \ell)}{\hat{\sigma}_{\epsilon,n}(\tau, \ell)} < -\kappa_n\Big) \quad \text{for all } (\tau, \ell) \in \mathcal{T} \times \mathcal{L}. \tag{3.13}$$

Assumption 3.3. (GMS) *Assume that $\kappa_n \to \infty$, $B_n \to \infty$, $n^{-1/2}\kappa_n \to 0$, and $\kappa_n^{-1} B_n \to 0$ as $n \to \infty$.*

Assumption 3.3 imposes conditions on κ_n and B_n sequences, and is a combined version of Assumptions GMS1 and GMS2 of AS.

3.5 Null Distribution Approximation

Before defining the critical values, we provide two approaches to approximating the process $\widehat{\Phi}_n(\cdot) \equiv \sqrt{n}(\hat{\nu}_n(\cdot) - \nu_P(\cdot))$. We first introduce the multiplier method based on the conditional multiplier central limit theorem in Chapter 2.9 of van der Vaart and Wellner (1996). Let $\{U_i : i \geq 1\}$ be a sequence of i.i.d. random variables that is independent of the whole sample path $\{W_i : n \geq 1\}$ such that $E[U] = 0$, $E[U^2] = 1$, and $E[|U|^{\delta_1}] < C$ for some $2 < \delta_1 < \delta$ and $C < \infty$ where δ is the constant in Assumption 4.1 below. Define $\widehat{\Phi}_n^u(\tau, \ell)$ as

$$\begin{aligned}&\widehat{\Phi}_n^u(\tau, \ell)\\ =&\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i\Big\{\hat{w}_n^{(1)}(\ell)\big(m_i^{(2)}(\tau, \ell) - \hat{m}_n^{(2)}(\tau, \ell)\big) + \hat{m}_n^{(2)}(\tau, \ell)\big(w_i^{(1)}(\ell) - \hat{w}_n^{(1)}(\ell)\big)\end{aligned}$$

uniform asymptotic size property and the consistency against fixed alternatives of our test. However, our tests based on different choices of Q will have different power in finite samples and asymptotically against local alternatives. To discuss the properties of our tests equipped with different choices of Q is an interesting topic that we do not pursue this in this paper.

$$- \hat{w}_n^{(2)}(\ell)\big(m_i^{(1)}(\tau,\ell) - \hat{m}_n^{(1)}(\tau,\ell)\big) - \hat{m}_n^{(1)}(\tau,\ell)\big(w_i^{(2)}(\ell) - \hat{w}_n^{(2)}(\ell)\big)\Big\}. \tag{3.14}$$

Next we describe the bootstrap method to approximate $\widehat{\Phi}_n(\cdot)$. Let $\{W_i^b : i \leq n\}$ be an i.i.d. bootstrap sample drawn from the empirical distribution of $\{W_i : i \leq n\}$. Let $m_i^{(j)b}(\tau,\ell) = m^{(j)}(W_i^b,\tau,\ell)$ and $w_i^{(j)b}(\ell) = w^{(j)}(W_i^b,\ell)$ for $j = 1$ and 2. Define

$$\begin{aligned}
&\hat{\nu}_n^b(\tau,\ell) = \hat{m}_n^{(2)b}(\tau,\ell)\hat{w}_n^{(1)b}(\ell) - \hat{m}_n^{(1)b}(\tau,\ell)\hat{w}_n^{(2)b}(\ell),\\
&\hat{m}_n^{(j)b}(\tau,\ell) = \frac{1}{n}\sum_{i=1}^{n} m_i^{(j)b}(\tau,\ell), \quad \hat{w}_n^{(j)b}(\ell) = \frac{1}{n}\sum_{i=1}^{n} w_i^{(j)b}(\ell).
\end{aligned} \tag{3.15}$$

Finally, define the bootstrap process $\widehat{\Phi}_n^b(\cdot)$ as

$$\widehat{\Phi}_n^b(\cdot) = \sqrt{n}(\hat{\nu}_n^b(\tau,\ell) - \hat{\nu}_n(\tau,\ell)). \tag{3.16}$$

Let the critical value statistics be

$$\widehat{T}_n^u = \int \max\left\{\frac{\widehat{\Phi}^u(\tau,\ell)}{\hat{\sigma}_\epsilon(\tau,\ell)} + \phi_n(\tau,\ell), 0\right\}^2 dQ(\tau,\ell), \tag{3.17}$$

$$\widehat{T}_n^b = \int \max\left\{\frac{\widehat{\Phi}^b(\tau,\ell)}{\hat{\sigma}_\epsilon(\tau,\ell)} + \phi_n(\tau,\ell), 0\right\}^2 dQ(\tau,\ell). \tag{3.18}$$

We call $\widehat{T}_n^u$ the multiplier statistic and $\widehat{T}_n^b$ the bootstrap statistic. The conditional distributions (given the original sample) of them asymptotically provide upper bounds for the null distribution of our test statistic.

3.6 GMS Critical Value

We are ready to define the multiplier GMS critical value $\hat{c}_\eta^u$ and the bootstrap GMS critical value, $\hat{c}_\eta^b$:

$$\hat{c}_\eta^u = \sup\left\{q \middle| P^u\big(\widehat{T}_n^u \leq q\big) \leq 1 - \alpha + \eta\right\} + \eta, \tag{3.19}$$

$$\hat{c}_\eta^b = \sup\left\{q \middle| P^b\big(\widehat{T}_n^b \leq q\big) \leq 1 - \alpha + \eta\right\} + \eta, \tag{3.20}$$

where $\eta > 0$ is an arbitrarily small positive number, e.g., 10^{-6}, and P^u and P^b denote the multiplier probability measure and bootstrap probability measure, respectively. Note that $\hat{c}_\eta^u$ and $\hat{c}_\eta^b$ are defined as the $(1 - \alpha + \eta)$-th quantiles of the multiplier null distribution and bootstrap null distribution plus η, respectively. AS call the constant η an infinitesimal uniformity factor that is used to avoid the problems that arise due to the presence of the infinite-dimensional nuisance parameter $\nu_P(\tau, \ell)$ and to eliminate the need for complicated and difficult-to-verify uniform continuity and strictly-monotonicity conditions on the large sample distribution functions of the test statistic.

3.7 Decision Rule

The decision rule is the following:

$$\text{Reject } H_0 \text{ in (3.1) if } \widehat{T}_n > \hat{c}_\eta, \tag{3.21}$$

where $\hat{c}_\eta$ can be $\hat{c}^u_\eta$ or $\hat{c}^b_\eta$.

4 Uniform Asymptotic Size

In this section, we show that our test has correct asymptotic size uniformly over a broad set of distributions. We impose conditions on $\{f^{(j)}(\tau, W) : \tau \in \mathcal{T}\}$ for $j = 1$ and 2 to regulate the complexity of them. It ensures that the empirical process $\widehat{\Phi}_n(\cdot)$ and its multiplier and bootstrap counterparts satisfy the functional central limit theorem under drifting sequence of distributions.

Let the collection of distributions of our interest be denoted as $\mathcal{P}$.

Assumption 4.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space equipped with probability distribution $\mathbb{P}$. Let $\mathcal{P}$ denote the collection of distributions P such that:*
(a) $\max\{|f^{(1)}(\tau, w)|, |f^{(2)}(\tau, w)|\} \le F(w)$ *for all $w \in \mathcal{W}$, for all $\tau \in \mathcal{T}$ for some envelope function $F(w)$.*
(b) $E_P F^\delta(W_{n,i}) \le C < \infty$ *for all $P \in \mathcal{P}$ for some $\delta > 2$.*
(c) *the processes $\{f^{(j)}(\tau, W_{n,i}) : \tau \in \mathcal{T}, i \le n, 1 \le n\}$ for $j = 1$ and 2 are manageable with respect to the envelope function $F(W_{n,i})$ where $\{W_{n,i} : i \le n, 1 \le n\}$ is a row-wise i.i.d. triangular array with $W_{n,i} \sim P_n$ for any sequence $\{P_n \in \mathcal{P}\}$.*

The manageability condition in Assumption 4.1(c) is from Definition 7.9 of Pollard (1990); see Pollard (1990) for more details. Assumption 4.1(c) is not restrictive. For example, if $\mathcal{T}$ is finite as in Examples 2.1-2.3, or if $\{f^{(j)}(\tau, \cdot) : \tau \in \mathcal{T}\}$ is a VC class as in Examples 2.4 and 2.5, Assumption 4.1(c) holds. Assumption 4.1(b) implies that

$|E_P[m^{(j)}(\tau,\ell)]| \leq M$ for some $M > 0$ for all (τ,ℓ) uniformly over $P \in \mathcal{P}$. This ensures that the asymptotic covariance kernel of $\sqrt{n}\big(\hat{\nu}_n(\tau,\ell) - \nu_P(\tau,\ell)\big)$ is uniformly bounded for all $P \in \mathcal{P}$.

To establish the uniform asymptotic size, we introduce some notation. Define

$$
\begin{aligned}
h_{1,P}(\tau,\ell) &= E_P(\ddot{w}(W,\tau,\ell)), \text{ and} \\
h_{2,P}\big((\tau_1,\ell_1),(\tau_2,\ell_2)\big) &= Cov_P(\ddot{m}(\tau_1,\ell_1),\ddot{m}(\tau_2,\ell_2)), \text{ where} \\
\ddot{w}(W,\tau,\ell) &= (-w^{(2)}(W,\ell), w^{(1)}(W,\ell), -m^{(2)}(W,\tau,\ell), m^{(1)}(W,\tau,\ell))', \\
\ddot{m}(W,\tau,\ell) &= (m^{(1)}(W,\tau,\ell), m^{(2)}(W,\tau,\ell), w^{(1)}(W,\ell), w^{(2)}(W,\ell))'. \qquad (4.1)
\end{aligned}
$$

We define $h_{1,P}(\tau,\ell)$ and $h_{2,P}((\tau_1,\ell_1),(\tau_2,\ell_2))$ this way so that under suitable assumptions, we have

$$
\begin{aligned}
& Cov_P\Big(\sqrt{n}\big(\hat{\nu}_n(\tau_1,\ell_1) - \nu_P(\tau_1,\ell_1)\big), \sqrt{n}\big(\hat{\nu}_n(\tau_2,\ell_2) - \nu_P(\tau_2,\ell_2)\big)\Big) \\
\approx\ & h_{1,P}(\tau_1,\ell_1)' \cdot h_{2,P}\big((\tau_1,\ell_1),(\tau_2,\ell_2)\big) \cdot h_{1,P}(\tau_2,\ell_2). \qquad (4.2)
\end{aligned}
$$

Also, $h_{1,P}(\tau,\ell)$ determines $\nu_P(\tau,\ell)$ because

$$
\nu_P(\tau,\ell) = E_P[m^{(2)}(W,\tau,\ell)]E_P[w^{(1)}(W,\ell)] - E_P[m^{(1)}(W,\tau,\ell)]E_P[w^{(2)}(W,\ell)]. \quad (4.3)
$$

Let

$$
\begin{aligned}
& \mathcal{H}_1 = \{h_{1,P}(\cdot) : P \in \mathcal{P}\}, \quad \mathcal{H}_2 = \{h_{2,P}(\cdot,\cdot) : P \in \mathcal{P}\}, \\
& \mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2. \qquad (4.4)
\end{aligned}
$$

On the space of $\mathcal{H}$, we use the metric d defined by

$$
\begin{aligned}
& d(h^{(1)},h^{(2)}) = \max\{d_1(h_1^{(1)},h_1^{(2)}), d_2(h_2^{(1)},h_2^{(2)})\}, \\
& d_1(h_1^{(1)},h_1^{(2)}) = \sup_{(\tau,\ell)\in\mathcal{T}\times\mathcal{L}} \|h_1^{(1)}(\tau,\ell) - h_1^{(2)}(\tau,\ell)\|, \qquad (4.5) \\
& d_2(h_2^{(1)},h_2^{(2)}) = \sup_{(\tau_1,\ell_1),(\tau_2,\ell_2)\in\mathcal{T}\times\mathcal{L}} \|h_2^{(1)}\big((\tau_1,\ell_1),(\tau_2,\ell_2)\big) - h_2^{(2)}\big((\tau_1,\ell_1),(\tau_2,\ell_2)\big)\|,
\end{aligned}
$$

where $\|\cdot\|$ denotes the Euclidean norms. For notational simplicity, we use d to denote d_1 and d_2 as well, and we suppress (τ,ℓ) whenever there is no confusion. For example, let $h_{1,P}$ denote $h_{1,P}(\cdot)$, and $h_{2,P}$ denote $h_{2,P}(\cdot,\cdot)$. For any $h \in \mathcal{H}$, define $h_{2,\nu} = h_1' \cdot h_2 \cdot h_1$ and for any P, define $h_{2,\nu,P}$ as $h_{1,P}' \cdot h_{2,P} \cdot h_{1,P}$. Let $\mathcal{H}_{2,\nu} \equiv \{h_{2,\nu}: \ h_{2,\nu} = h_1' \cdot h_2 \cdot h_1, \ h \in \mathcal{H}\}$. The metric d on the space $\mathcal{H}_{2,\nu}$ is defined similarly.

Let $\mathcal{P}^0$ denote the collection of null distributions in P. We impose the following conditions on $\mathcal{P}^0$.

Assumption 4.2. *The set $\mathcal{P}_0$ satisfies:*

(a) $\mathcal{P}^0 \subseteq \mathcal{P}$.

(b) *The null hypothesis H_0 defined in* (3.1) *holds under any $P \in \mathcal{P}^0$.*

(c) $\mathcal{H}^0 \equiv \{(h_{1,P}, h_{2,P}) : P \in \mathcal{P}^0\}$ *is a compact subset of $\mathcal{H}$ under the metric d defined in equation* (4.5).

Let $\mathcal{H}_{2,\nu}^0 \equiv \{h_{2,\nu}: \ h_{2,\nu} = h_1' \cdot h_2 \cdot h_1, \ h \in \mathcal{H}^0\}$. The compactness of $\mathcal{H}^0$ in Assumption 4.2(c) implies the compactness of $\mathcal{H}_{2,\nu}^0$. The compactness of $\mathcal{H}_{2,\nu}^0$ is necessary for us to obtain the uniform asymptotic size over $\mathcal{P}^0$. This is assumed in AS, Donald and Hsu (2016), and Hsu (2016) as well. The following theorem summarizes the uniform asymptotic size of our test. Additional notation are needed. Let

$$
\begin{aligned}
\mathcal{T}^o(P) \equiv \{\tau \in \mathcal{T}: \ &\exists\, x_{1\ell,\tau} \ll x_{1u,\tau}, \ x_{2\ell,\tau} \ll x_{2u,\tau}, \ z_{\ell,\tau} \ll z_{u,\tau}, \\
&x_{1\ell,\tau} \leq x_{2\ell,\tau}, \ x_{1u,\tau} \leq x_{2u,\tau}, \text{and for some constant } C_\tau \in R \\
&E_P[f^{(1)}(Y,\tau)|X = x_1, Z = z] = E_P[f^{(2)}(Y,\tau)|X = x_2, Z = z] = C_\tau, \\
&\text{for all } x_1 \in [x_{1\ell,\tau}, x_{1u,\tau}], x_2 \in [x_{2\ell,\tau}, x_{2u,\tau}], \text{ and } z \in [z_{\ell,\tau}, z_{u,\tau}].\} \qquad (4.6)
\end{aligned}
$$

$$
\mathcal{L}^o(\tau, P) \equiv \{\ell \in \mathcal{L}: \ \nu_P(\ell, \tau) = 0\} \tag{4.7}
$$

$$
(\mathcal{TL})^o(P) \equiv \{(\tau, \ell) \in \mathcal{T} \times \mathcal{L}: \ \nu_P(\tau, \ell) = 0\} = \{(\tau, \ell): \ \ell \in \mathcal{L}^o(\tau, P)\}. \tag{4.8}
$$

The set $\mathcal{T}^o(P)$ denotes the collection of τ's such that the inequalities are binding over a hypercube of (x_1, x_2, z) under P. The set $\mathcal{L}^o(\tau, P)$ denotes the collection of ℓ's such that the unconditional moment defined by ℓ is binding at τ, and $(\mathcal{TL})^o(P)$ the set of (τ, ℓ) such that the unconditional moment with (τ, ℓ) is binding. Under Assumption 3.2, it is straightforward to see that if $\tau \in \mathcal{T}^o(P)$, then $\int_{\mathcal{L}^o(\tau,P)} dQ_{\mathcal{L}}(\ell) > 0$, and that if $\int_{\mathcal{T}^o(P)} dQ_{\mathcal{T}}(\tau) > 0$, then $\int_{(\mathcal{TL})^o(P)} dQ(\ell, \tau) > 0$.

Let $\Psi_{h_{2,\nu}}$ denote the mean-zero Gaussian process with covariance kernel function $h_{2,\nu}$. Let $\sigma^2_{\epsilon, h_{2,\nu}}(\tau, \ell) = \max\{h_{2,\nu}((\tau, \ell), (\tau, \ell)), \epsilon\}$.

Assumption 4.3. *There exists a* $P_c \in \mathcal{P}^0$ *such that* $\int_{(\mathcal{T}\mathcal{L})^o(P_c)} \max\left\{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)/\sigma_{\epsilon,h_{2,\nu}}(\tau,\ell), 0\right\}^2 dQ(\tau,\ell)$ *is non-degenerate.*

We restate the conditions on the multipliers $\{U_i : i \geq 1\}$ in the following assumption.

Assumption 4.4. *Let* $\{U_i : i \geq 1\}$ *be a sequence of i.i.d. random variables independent with the original sample such that* $E[U] = 0$, $E[U^2] = 1$, *and* $E[|U|^{\delta_1}] < C$ *for some* $2 < \delta_1 < \delta$ *and some* $C > 0$ *where* δ *is the same as in Assumption* 4.1.

Assumption 4.4 is needed for the multiplier method only. For the rest of the paper, we implicitly assume that Assumption 4.4 holds for all the results associated with multiplier method.

Theorem 4.1. *Suppose that Assumptions* 3.1*,* 3.3*, and* 4.1*-*4.2 *hold, and that* $\alpha < 1/2$*. Let* $\hat{c}_\eta$ *be either* $\hat{c}^u_\eta$ *or* $\hat{c}^b_\eta$*. Then*
(i) $\limsup_{n\to\infty} \sup_{P\in\mathcal{P}^0} P_P(\widehat{T}_n > \hat{c}_\eta) \leq \alpha$;
(ii) *if Assumption* 4.3 *also holds, then* $\lim_{\eta\to 0} \limsup_{n\to\infty} \sup_{P\in\mathcal{P}^0} P_P(\widehat{T}_n > \hat{c}_\eta) = \alpha$.

Theorem 4.1(i) shows that our test has correct uniform asymptotic size over $\mathcal{P}^0$ defined by Assumptions 4.1-4.2. This result is similar to Theorem 2(a) of AS. Theorem 4.1(ii) shows that our test is at most infinitesimally conservative asymptotically when there exists at least one P_c that is at the boundary of the null hypothesis in the sense that the limiting distribution of $\widehat{T}_n$ is non-degenerate under P_c, which our Assumption 4.3 guarantees.

5 Power Properties

In this section, we show the consistency of our test against fixed alternatives and we show that our test has non-trivial local power against some $n^{-1/2}$-local alternatives.

5.1 Power against Fixed Alternatives

Define the collections of τ's at which the the null hypothesis is violated as

$$\mathcal{T}^a(P) \equiv \Big\{\tau : E_P[f^{(1)}(W,\tau)|X = x_1, Z = z] < E_P[f^{(2)}(W,\tau)|X = x_2, Z = z],$$

$$\text{for some } z \in \mathcal{Z} \text{ and } x_1, x_2 \in \mathcal{X} \text{ with } x_1 \geq x_2.\Big\} \tag{5.1}$$

The following assumption specifies the fixed alternatives we consider.

Assumption 5.1. *The distribution $P_* \in \mathcal{P}$ satisfies:*

(a) $\mathcal{T}^a(P_*)$ *contains* $B_c(\tau_*)$ *for some* $c > 0$ *and some* $\tau_* \in \mathcal{T}$*,*

(b) *Assumption* 3.1 *holds under* P_**, and*

(c) *Assumption* 4.1 *holds with* P_* *in place of* P_n *and* $P \in \mathcal{P}$*.*

Assumption 5.1(a) together with Assumption 3.2 ensures that $\mathcal{T}^a(P_*)$ has strictly positive measure under Q. This automatically holds when $\mathcal{T}$ is finite and $\mathcal{T}^a(P_*)$ is non-empty. The following theorem shows the consistency of our test against the fixed alternatives satisfying Assumption 5.1.

Theorem 5.1. *Suppose that Assumptions* 3.1-3.3 *and* 5.1*, and* $\alpha < 1/2$*. Then we have* $\lim_{n\to\infty} P_{P_*}(\widehat{T}_n > \hat{c}_\eta) = 1$*.*

The proof is done by showing that $\widehat{T}_n$ diverges to positive infinity, and that $\hat{c}_\eta$ is bounded in probability.

5.2 Asymptotic Local Power

We consider the local power of our tests in this section. Consider a sequence of $P_n \in \mathcal{P}\backslash\mathcal{P}^0$ that converges to some $P_c \in \mathcal{P}^0$ under the Kolmogove-Smirnov metric where $A\backslash B \equiv \{x : x \in A \text{ but } x \notin B\}$ for any two sets A and B. The $n^{-1/2}$-local alternatives are defined in the following assumptions.

Let P_{xz} denote the marginal distribution of (X, Z) under P.

Assumption 5.2. *The sequence* $\{P_n \in \mathcal{P}\backslash\mathcal{P}^0 : \ n \geq 1\}$ *satisfies:*

(a) *for some* $P_c \in \mathcal{P}^0$ *that satisfies the non-degeneracy in Assumption* 4.3*,*

$$E_{P_n}[f^{(1)}(W,\tau)|X,Z] = E_{P_c}[f^{(1)}(W,\tau)|X,Z] + \gamma\delta_1(X,Z,\tau)/\sqrt{n},$$
$$E_{P_n}[f^{(2)}(W,\tau)|X,Z] = E_{P_c}[f^{(2)}(W,\tau)|X,Z] + \gamma\delta_2(X,Z,\tau)/\sqrt{n}.$$

where $\gamma > 0$ *is a constant, and* δ_1 *and* δ_2 *are two functions.*

(b) $P_{n,xz} = P_{c,xz}$ *for all* $n \geq 1$*.*

(c) *for* $j = 1$ *and* 2, $\delta_j(x, z, \tau)$ *is continuous on* $\mathcal{X} \times \mathcal{Z}$ *for all* $\tau \in \mathcal{T}$.

(d) $\delta_1(x_1, z, \tau) \leq \delta_2(x_2, z, \tau)$ *for all* $x_1, x_2 \in \mathcal{X}$ *such that* $x_1 \geq x_2$, $z \in \mathcal{Z}$ *and for all* $\tau \in \mathcal{T}$.

(e) *for some* $\tau \in \mathcal{T}^o(P_c)$, $\delta_1(x_1, z, \tau) < \delta_2(x_2, z, \tau)$ *for some* $x_1 \in (x_{1\ell,\tau}, x_{1u,\tau})$, $x_2 \in (x_{2\ell,\tau}, x_{2u,\tau})$ *such that* $x_1 > x_2$, *and some* $z \in (z_{\ell,\tau}, z_{u,\tau})$, *where* $x_{1\ell,\tau}$, $x_{1u,\tau}$, $x_{2\ell,\tau}$, $x_{2u,\tau}$, $z_{\ell,\tau}$, *and* $z_{u,\tau}$ *are some values satisfying the conditions defining* $\mathcal{T}^o(P_c)$ *in* (4.6).

(f) $d(h_{P_n}, h_{P_c}) \to 0$.

Assumption 5.2(a) requires that for $j = 1, 2$, the difference between the conditional mean of $f^{(j)}(W, \tau)$ on X and Z under P_n and that under P_c is of order $n^{-1/2}$. Assumption 5.2(b) requires that the marginal distribution of X and Z remains the same along the sequence. With some minor modifications of our proof, this condition can be relaxed. Assumption 5.2(c) along with Assumption 3.1 ensures that the conditional means of $f^{(1)}(W, \tau)$ and $f^{(2)}(W, \tau)$ under P_n are continuous on $\mathcal{X}$ and $\mathcal{Z}$. Assumption 5.2(e) ensures that the null hypothesis does not hold under P_n for $n \geq 1$, i.e., $P_n \notin \mathcal{P}^0$. Assumption 5.2(f) implies that $d(h_{2,\nu,P_n}, h_{2,\nu,P_c}) \to 0$, which specifies the asymptotic covariance kernel of $\sqrt{n}(\hat{\nu}_n(\cdot) - \nu_{P_c}(\cdot))$.

Let $x_{1\ell,\tau}$, $x_{1u,\tau}$, $x_{2\ell,\tau}$, $x_{2u,\tau}$, $z_{\ell,\tau}$, and $z_{u,\tau}$ be the values specified in Assumption 5.2(e). Define $\mathcal{T}^+(P_c)$ as

$$\begin{aligned}\mathcal{T}^+(P_c) \equiv \{\tau \in \mathcal{T}^o(P_c) :\ & \delta_1(x_1, z, \tau) < \delta_2(x_2, z, \tau) \text{ for some } x_1 \in (x_{1\ell,\tau}, x_{1u,\tau}), \\ & x_2 \in (x_{2\ell,\tau}, x_{2u,\tau}) \text{ such that } x_1 > x_2, \text{ and some } z \in (z_{\ell,\tau}, z_{u,\tau})\}. \quad (5.2)\end{aligned}$$

Assumption 5.3. *Assume that* $\int_{\mathcal{T}^+(P_c)} dQ_{\mathcal{T}} > 0$ *where* $\mathcal{T}^+(P_c)$ *is defined in* (5.2).

Assumption 5.3 holds if $\mathcal{T}^+(P_c)$ contains an open ball around τ_* for some τ_* in $\mathcal{T}$ by Assumption 3.2.

The following theorem shows the local power of our test.

Theorem 5.2. *Suppose Assumptions* 3.1-3.3, 5.2, *and* 5.3 *hold, and* $\alpha < 1/2$. *Then*

(i) $\lim_{\eta \to 0} \liminf_{n \to \infty} P_{P_n}(\widehat{T}_n > \hat{c}_\eta) \geq \alpha$.

(ii) $\lim_{\gamma \to \infty} \liminf_{n \to \infty} P_{P_n}(\widehat{T}_n > \hat{c}_\eta) = 1$.

Part (i) of the theorem shows the near asymptotic unbiasedness of our test against the $n^{-1/2}$-local alternatives defined by Assumptions 5.2 and 5.3. Part (ii) of the theorem implies that as long as the $n^{-1/2}$-local alternative defined in Assumption 5.2 is far enough from the null (that is, γ is large enough), the asymptotic power of our test is strictly greater than size.

6 Monte Carlo Simulation

To implement our test, one needs to pick several user-chosen parameters in advance. In this section, we first make suggestions on how to pick these parameters. We then report Monte Carlo results for two examples. The first example is a test of the monotone instrumental variable assumption. The second example is a test of regression monotonicity, as also considered in Chetverikov (2013).

6.1 Implementation

We make the following suggestions.

1. **Support of Covariates:** Transform the support of each covariate, X_j, to unit interval by applying the following mapping. If X_j has support $[a, b]$, then define $X_j^* = (X - a)/(b - a)$. If X_j has support on the whole real line, define $X_j^* = \Phi(\hat{\sigma}_j^{-1}(X_j - \hat{\mu}_j))$ where $\hat{\sigma}_j$ is the sample standard deviation of X_{ji}'s, $\hat{\mu}_j$ is the sample mean of X_{ji}'s, and $\Phi(\cdot)$ is the standard normal cdf function. Apply the same mapping to each Z_j.

2. **Instrumental functions:** Use the countable hypercube instrumental functions on the new conditioning variables:

$$\mathcal{G}_{\text{c-cube}} = \Big\{ g_\ell \equiv \big(g^{(1)}_{x^*,\ell}, g^{(2)}_{x^*,\ell},\ g_{z^*,\ell}\big) : \ell \in \mathcal{L}_{\text{c-cube}} \Big\}, \text{ where} \tag{6.1}$$
$$\mathcal{L}_{\text{c-cube}} = \Big\{ (x_1^*, x_2^*, z^*, r) : r = q^{-1},\ q \cdot (x_1^*, x_2^*, z^*) \in \{0, 1, 2, \cdots, q-1\}^{2d_x + d_z}, \\ x_1 \geq x_2, , \text{ and } q = 2, 3, \cdots, q_1 \Big\},$$

where q_1 is a natural number and is picked such that the expected sample size of the smallest cube is around 15 as suggested by AS.[7]

[7] The expected sample size of the smallest cube is roughly equal to $n \cdot (q_1)^{-(d_x + d_z)}$.

3. **Selection of τ's:** If $\mathcal{T}$ is of finite elements as in Examples 2.1-2.3, use all elements in $\mathcal{T}$. If $\mathcal{T}$ contains a continuum of elements as in Examples 2.4 and 2.5, pick a finite number of τ's and allow the number of τ's grows with sample size. For Examples 2.4 and 2.5, we specifically suggest to consider the finite subset $\{y_1, \ldots, y_n\}$ of $\mathcal{T}$ that is also used in Lee, Linton and Whang (2009).

4. $Q(\tau, \ell)$**:** the distribution $Q_{\mathcal{T}}$ assigns equal weights on $\mathcal{T}$, and the distribution $Q_{\mathcal{L}}$ assigns weight $\propto q^{-2}$ on each q and for each q, $Q_{\mathcal{L}}$ assigns equal weight on each instrumental function with $r = q^{-1}$. Recall that for each q, there are $(q(q+1)/2)^{d_x} \cdot q^{d_z}$ of instrumental functions with $r = q^{-1}$.

5. ϵ**,** κ_n**,** B_n**,** η**:** Based on the experiments in the simulations, we suggest to set $\epsilon = 10^{-6}$, $\kappa_n = 0.15 \cdot \ln(n)$, $B_n = 0.85 \cdot \ln(n)/\ln\ln(n)$, and $\eta = 10^{-6}$.

For both of the Monte Carlo examples below, we consider samples of sizes $n = 100$, 200, and 500. For q_1, we set $q_1 = 6$ when $n = 100$, $q_1 = 13$ when $n = 200$, and $q_1 = 33$ when $n = 500$. The expected sample sizes of the smallest cube are 16.6, 15.3 and 15.1, respectively. All our simulation results are based on 500 simulation repetitions, and for each repetition, the critical value is approximated by 500 bootstrap replications. Nominal size of the test is set to be 10%.

6.2 Testing the Monotone Instrumental Variable Assumption

We then consider the finite-sample performance of our test for Example 2.1. Without loss of generality, we assume that $Y(0) = 0$. Then, we only need to consider the null that $E[f^{(1)}(Y, 1)|X = x_1] \geq E[f^{(2)}(Y, 1)|X = x_2]$ when $x_1 \geq x_2$.

By the Remarks below Theorem 2.1, violations of H_0^{MIV} are not always statistically detectable (regardless of sample size). It is also mentioned there that the closer $E[y_u \cdot (1-D)|X = x]$ and $E[y_\ell \cdot (1-D)|X = x]$ are, the more likely for us to detect the violation of MIV. Note that

$$E[y_u(1 - D)|X = x] = E[y_u|D = 0, X = x]P(D = 0|X = x).$$

We can see that the smaller $P(D = 0|X = x)$ is and the smaller the gap between y_u and y_ℓ, the more likely for us to detect the violation. Therefore, we consider the cases where the MIV assumption is violated and control these two factors to make the violation statistically detectable or not detectable.

Case (1): Let

$$Y(1) = -2X + \epsilon,\ X \sim Uni[0,1],$$
$$\epsilon \sim Uni[-0.1, 0.1],\ D = 1(U \leq 0.8),\ \text{and } U \sim Uni[0,1],$$

where X, ϵ, U are mutually independent, and $Uni[a,b]$ stands for the uniform distribution on the interval $[a,b]$. Here $y_u = 0.1$ and $y_\ell = -2.1$. In this case, the MIV is violated, and it is detectable because H_0^{GRM} is also violated.

Case (2): Let

$$Y(1) = -2X + \epsilon,\ X \sim Uni[0,1],$$
$$\epsilon \sim Uni[-1, 1],\ D = 1(U \leq 0.5),\ \text{and } U \sim Uni[0,1],$$

where X, ϵ, U are mutually independent. Here $y_u = 1$ and $y_\ell = -3$. In this case, we can verify that the MIV is violated, but the violation is not statistically detectable because H_0^{GRM} is not violated.

Case (3): Let

$$Y(1) = -2X + \epsilon,\ X \sim Uni[0,1],$$
$$\epsilon \sim Uni[-0.1, 0.1],\ D = 1(U \leq 0.2 + 0.8X),\ \text{and } U \sim Uni[0,1],$$

where X, ϵ, U are mutually independent. Here $y_u = 0.1$ and $y_\ell = -2.1$. In this case, we can verify that the MIV is violated, and the violation is detectable because H_0^{GRM} is also violated.

Case (4): Let

$$Y(1) = -2X + \epsilon,\ X \sim Uni[0,1],$$
$$\epsilon \sim Uni[-1, 1],\ D = 1(U \leq 0.9 - 0.8X),\ \text{and } U \sim Uni[0,1],$$

where X, ϵ, U are mutually independent. Here $y_u = 1$ and $y_\ell = -3$. In this case, we can verify that the MIV is violated, but the violation is not statistically detectable because H_0^{GRM} is not violated.

In the simulations we consider two possibilities: (a) y_u and y_ℓ are known, and (b) y_u and y_ℓ are unknown but we replace y_u and y_ℓ with $\max_i Y_i$ and $\min_i Y_i$, respectively. Note that $\max_i Y_i \xrightarrow{p} y_u$ and $\min_i Y_i \xrightarrow{p} y_\ell$ at a faster rate than $n^{-1/2}$ which implies that the estimation effects of $\max_i Y_i$ and $\min_i Y_i$ can be ignored asymptotically. On the other hand, $\max_i Y_i \leq y_u$ and $\min_i Y_i \geq y_\ell$, so we expect that the power of case (b) be better than (a) when the violation is statistically detectable.

Table 1 shows the rejection probabilities for our test, and it confirms our theoretical findings. The rejection probabilities are greater than the nominal size 0.1 in cases (1) and (3) where the GRM and the MIV are both violated. In these cases, the rejection probabilities of the multipler version (GMS-u) and the bootstrap version (GMS-b) are similar, both increases with the sample size, and both are higher when y_u and y_ℓ are estimated. Neither version of our test has any power in cases (2) and (4). This is consistent with Theorem 2.1, which says that no test can have power greater than size in those cases because the sharp testable implication of MIV is not violated.

6.3 Testing Regression Monotonicity

We next consider a Monte Carlo demonstration of our test for a regression monotonicity example. We use the same designs as in Chetverikov (2013), where there is no Z in the model and X is a scalar. Let X be a uniform distribution on $[-1, 1]$ and ξ be a normal distribution or uniform distribution with mean zero and standard deviation equal to σ_ξ. The variable Y is generated as

$$Y = c_1 X - c_2 \phi(c_3 X) + \xi, \tag{6.2}$$

where $c_1, c_2, c_3 \geq 0$ and $\phi(\cdot)$ is the pdf of the standard normal distribution. As in Chetverikov (2013), we consider four set of parameters:

Case (1): $c_1 = c_2 = c_3 = 0$ and $\sigma_\xi = 0.05$.

Case (2): $c_1 = c_3 = 1$, $c_2 = 4$ and $\sigma_\xi = 0.05$.

Case (3): $c_1 = 1$, $c_2 = 1.2$, $c_3 = 5$ and $\sigma_\xi = 0.05$.

Case (4): $c_1 = 1$, $c_2 = 1.5$, $c_3 = 4$ and $\sigma_\xi = 0.1$.

Table 1: Rejection Probabilities of our Test for MIV ($\alpha = 0.1$, number of simulation repetitions = 500, critical value simulation draws =500)

		y_u and y_ℓ are known		y_u and y_ℓ are unknown	
Cases	n	GMS-u	GMS-b	GMS-u	GMS-b
(1): H_0^{MIV} violated	100	0.912	0.918	0.954	0.954
H_0^{GRM} violated	200	0.994	0.992	1.000	1.000
	500	1.000	1.000	1.000	1.000
(2): H_0^{MIV} violated	100	0.000	0.000	0.000	0.000
H_0^{GRM} holds	200	0.000	0.000	0.000	0.000
	500	0.000	0.000	0.000	0.000
(3): H_0^{MIV} violated	100	0.230	0.218	0.306	0.292
H_0^{GRM} violated	200	0.472	0.472	0.584	0.596
	500	0.910	0.910	0.944	0.950
(4): H_0^{MIV} violated	100	0.000	0.000	0.000	0.000
H_0^{GRM} holds	200	0.000	0.000	0.000	0.000
	500	0.000	0.000	0.000	0.000

It can be verified that H_0 holds in Cases (1) and (2), and H_1 holds in Cases (3) and (4). Table 2 shows the rejection probabilities for our test with both the multiplier critical value (GMS-u) and the bootstrap critical value (GMS-b). The columns of CS-SD and IS-SD are taken from Chetverikov (2013). CS-SD refers to the step-down procedure with consistent sigma estimator and IS-SD refers to the step-down procedure with inconsistent sigma estimator. For details of the procedures CS-SD and IS-SD, see Chetverikov (2013).

As we can see from Table 2, our test controls the size well in Cases (1) and (2), and the rejection rates increase with the sample size in Cases (3) and (4). The performance of our tests is comparable to the tests proposed by Chetverikov's (2013).

Table 2: Rejection Probabilities of Our Test (GMS-u, GMS-b) and Chetverikov's (2013) test (CS-SD, IS-SD) for Regression Monotonicity ($\alpha = 0.1$, number of simulation repetitions = 500, critical value simulation draws = 500)

		normal				uniform			
Case	n	GMS-u	GMS-b	CS-SD	IS-SD	GMS-u	GMS-b	CS-SD	IS-SD
(1)	100	0.106	0.100	0.128	0.164	0.100	0.088	0.122	0.201
	200	0.118	0.116	0.114	0.149	0.126	0.136	0.121	0.160
	500	0.090	0.090	0.114	0.133	0.118	0.110	0.092	0.117
(2)	100	0.000	0.000	0.008	0.024	0.000	0.000	0.007	0.033
	200	0.002	0.002	0.010	0.017	0.002	0.004	0.010	0.024
	500	0.004	0.004	0.007	0.016	0.000	0.000	0.011	0.021
(3)	100	0.008	0.008	0.433	0.000	0.010	0.008	0.449	0.000
	200	0.706	0.674	0.861	0.650	0.712	0.678	0.839	0.617
	500	0.996	0.996	0.997	0.995	1.000	1.000	0.994	0.990
(4)	100	0.156	0.164	0.223	0.043	0.152	0.156	0.217	0.046
	200	0.408	0.378	0.506	0.500	0.386	0.342	0.478	0.456
	500	0.884	0.880	0.826	0.822	0.904	0.890	0.846	0.848

7 Conclusion

In this paper, we construct a test for the hypothesis of generalized regression monotonicity. The GRM is the sharp testable implication of monotonicity in certain latent structures. Examples include the monotone instrumental variable assumption, and the monotonicity of a nonparametric mean regression function when the dependent variable is only observed with interval values. The GRM also includes regression monotonicity and stochastic monotonicity as special cases. Our tests are shown to have uniform size control asymptotically, to be consistent against fixed alternatives, and to have nontrivial local power against some $n^{-1/2}$-local alternatives.

For future studies, it would be interesting to extend our tests to allow for the cases in which X or/and Z include generated regressors or single indexes. Another direction is to test the nonparametric generalized regression monotonicity in the form of

$$H_0: \ E_P[f^{(1)}(W,\tau)|X = x_1, Z = z_o] \geq E_P[f^{(2)}(W,\tau)|X = x_2, Z = z_o],$$
$$\forall x_1, x_2 \in \mathcal{X} \text{ and } x_1 \geq x_2, \quad \text{and } \tau \in \mathcal{T},$$

where z_o defines a specific subpopulation of our interest. A test of this hypothesis may be developed in the spirit of Andrews and Shi (2014).

APPENDIX

A Auxiliary Lemmas

For any covariance kernel function h, let Ψ_h denote the mean-zero Gaussian process with covariance kernel function h. Define

$$\hat{\chi}_P(\tau,\ell) \equiv \begin{pmatrix} \sqrt{n}\big(\hat{m}^{(1)}(\tau,\ell) - m_P^{(1)}(\tau,\ell)\big) \\ \sqrt{n}\big(\hat{m}^{(2)}(\tau,\ell) - m_P^{(2)}(\tau,\ell)\big) \\ \sqrt{n}\big(\hat{w}^{(1)}(\ell) - w_P^{(1)}(\ell)\big) \\ \sqrt{n}\big(\hat{w}^{(2)}(\ell) - w_P^{(2)}(\ell)\big) \end{pmatrix},$$

$$\widehat{\Phi}_P(\tau,\ell) \equiv \sqrt{n}(\hat{\nu}_n(\tau,\ell) - \nu_P(\tau,\ell)). \tag{A.1}$$

When P_{a_n} is in place of P, we have a_n in place of n in previous notations. Also, define

$$\hat{h}_{1,P}(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \ddot{w}(W_i,\cdot), \qquad \widehat{\ddot{m}}_P(\cdot) = \frac{1}{n}\sum_{i=1}^{n} \ddot{m}(W_i,\cdot),$$

$$\hat{h}_{2,P}(\cdot,\cdot) = \frac{1}{n}\sum_{i=1}^{n} \big(\ddot{m}(W_i,\cdot) - \widehat{\ddot{m}}_P(\cdot)\big)\big(\ddot{m}(W_i,\cdot) - \widehat{\ddot{m}}_P(\cdot)\big)',$$

$$\hat{h}_P = (\hat{h}_{1,P}, \hat{h}_{2,P}), \qquad \hat{h}_{\nu,P} = \hat{h}_{1,P}' \cdot \hat{h}_{2,P} \cdot \hat{h}_{1,P}.$$

Lemma A.1. *Suppose Assumption* 4.1 *holds. For a sequence* $\{P_{a_n} \in \mathcal{P} : n \geq 1\}$ *for a subsequence* $\{a_n\}$ *of* $\{n\}$, *suppose that* $d(h_{P_{a_n}}, h) \to 0$ *for some* $h \in \mathcal{H}$. *Then we have:*
(i) $d(\hat{h}_{P_{a_n}}, h) \xrightarrow{p} 0$, *and*
(ii) $\hat{\chi}_{P_{a_n}}(\tau,\ell) \Rightarrow \Psi_{h_2}$.

The following lemma summarizes revelent results regarding $\hat{\nu}_n(\tau,\ell)$.

Lemma A.2. *Suppose Assumption* 4.1 *holds. For a sequence* $\{P_{a_n} \in \mathcal{P} : n \geq 1\}$ *for a subsequence* $\{a_n\}$ *of* $\{n\}$, *suppose that* $d(h_{P_{a_n}}, h) \to 0$ *for some* $h \in \mathcal{H}$. *Then we have:*
(i) $d(h_{2,\nu,P_{a_n}}, h_{2,\nu}) \to 0$,
(ii) $d(\hat{h}_{2,\nu,P_{a_n}}, h_{2,\nu}) \xrightarrow{p} 0$,
(iii) $\widehat{\Phi}_{P_{a_n}}(\tau,\ell) \Rightarrow \Psi_{h_{2,\nu}}$,
(iv) $\widehat{\Phi}^u_{P_{a_n}} \Rightarrow \Psi_{h_{2,\nu}}$ *conditional on sample path with probability one,*
(v) $\widehat{\Phi}^b_{P_{a_n}} \Rightarrow \Psi_{h_{2,\nu}}$ *conditional on sample path with probability one,*
(vi) $\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} |\hat{\sigma}^{-1}_{\epsilon,a_n}(\tau,\ell) - \sigma^{-1}_{\epsilon,h_{2,\nu}}(\tau,\ell)| \xrightarrow{p} 0$ *where* $\sigma^2_{\epsilon,h_{2,\nu}}(\tau,\ell) = \max\{h_{2,\nu}((\tau,\ell),(\tau,\ell)),\epsilon\}$,
(vii) $\widehat{\Phi}^u_{P_{a_n}}(\cdot)/\hat{\sigma}_{\epsilon,a_n}(\cdot) \Rightarrow \cdot\Psi_{h_{2,\nu}}(\cdot)/\sigma_{\epsilon,h_{2,\nu}}(\cdot)$ *conditional on sample path with probability one, and*
(viii) $\widehat{\Phi}^b_{P_{a_n}}(\cdot)/\hat{\sigma}_{\epsilon,a_n}(\cdot) \Rightarrow \Psi_{h_{2,\nu}}(\cdot)/\sigma_{\epsilon,h_{2,\nu}}(\cdot)$ *conditional on sample path with probability one.*

Lemma A.3. *Suppose Assumptions* 3.3 *and* 4.1 *hold. For a sequence* $\{P_{a_n} \in \mathcal{P} : n \geq 1\}$ *for a subsequence* $\{a_n\}$ *of* $\{n\}$*, suppose that* (a) $d(h_{P_{a_n}}, h) \to 0$ *for some* $h \in \mathcal{H}$*, and that* (b) $\nu_{P_{a_n}}(\tau, \ell) = \nu_{P_c}(\tau, \ell) + \delta(\tau, \ell)/\sqrt{n}$ *for some* $P_c \in \mathcal{P}^0$ *and some function* $\delta : \mathcal{T} \times \mathcal{L} \to R$*. Then we have:*

(i) $\widehat{T}_n \xrightarrow{d} \int_{(\mathcal{T}\mathcal{L})^o(P_c)} \max\left\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)+\delta(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\right\}^2 dQ(\tau, \ell)$.

(ii) $\widehat{T}_n^u \xrightarrow{d} \int_{(\mathcal{T}\mathcal{L})^o(P_c)} \max\left\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\right\}^2 dQ(\tau, \ell)$ *conditional on almost all paths of the original sample.*

(iii) $\widehat{T}_n^b \xrightarrow{d} \int_{(\mathcal{T}\mathcal{L})^o(P_c)} \max\left\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\right\}^2 dQ(\tau, \ell)$ *conditional on almost all paths of the original sample.*

B Proof of Theorems

Proof of Theorem 2.1. First, we show part (i). Observe that, for $x_1 \geq x_2$,

$$
\begin{aligned}
&E[f^{(1)}(Y,1)|X = x_1] = E[YD + y_u \cdot (1 - D)|X = x_1] \\
=&E[Y(1)D + Y(1)(1 - D) + (y_u - Y(1)) \cdot (1 - D)|X = x_1] \\
=&E[Y(1) + (y_u - Y(1)) \cdot (1 - D)|X = x_1] \\
\geq&E[Y(1)|X = x_1] \\
\geq&E[Y(1)|X = x_2] \\
\geq&E[Y(1)D + Y(1)(1 - D) + (y_l - Y(1)) \cdot (1 - D)|X = x_2] \\
=&E[YD + (y_l) \cdot (1 - D)|X = x_2] = E[f^{(2)}(Y,1)|X = x_2], \qquad \text{(B.1)}
\end{aligned}
$$

where the second line holds because $YD = Y(1)D$, the fourth line holds because $y_u - Y(1) \geq 0$ by assumption, and by similar arguments the last two lines hold. Similarly, $E[f^{(1)}(Y,2)|X = x_1] \geq E[f^{(2)}(Y,2)|X = x_2]$ when $x_1 \geq x_2$. Part (i) follows.

We show part (ii) by construction. Let

$$
\begin{aligned}
I_1(x) &= \sup_{a \leq x} E[DY|X = a] + y_\ell E[1 - D|X = a] \\
I_2(x) &= \sup_{a \leq x} E[(1 - D)Y|X = a] + y_\ell E[D|X = a]. \qquad \text{(B.2)}
\end{aligned}
$$

Let

$$
\begin{aligned}
Y(1) &= DY + (1 - D)\frac{I_1(X) - E(DY|X)}{1 - E(D|X)} \\
Y(0) &= (1 - D)Y + D\frac{I_2(X) - E[(1 - D)Y|X]}{E(D|X)}. \qquad \text{(B.3)}
\end{aligned}
$$

It is easy to see that

$$E[Y(1)|X] = I_1(X), \text{ and } E[Y(0)|X] = I_2(X). \tag{B.4}$$

By the construction of $I_1(\cdot)$ and $I_2(\cdot)$, they are increasing. Thus, H_0^{MIV} is satisfied with $(Y(1), Y(0), D, X)$.

Now we only need to verify that $Y(1), Y(0)$ are bounded between y_ℓ and y_u. Consider the derivation:

$$\begin{aligned}
& y_\ell \leq Y(0) \leq y_u \\
\Leftrightarrow\quad & y_\ell \leq \frac{I_1(X) - E(DY|X)}{1 - E(D|X)} \leq y_u \\
\Leftrightarrow\quad & E[DY|X] + E[(1-D)y_\ell|X] \leq I_1(X) \leq E[DY|X] + E[(1-D)y_u|X] \\
\Leftrightarrow\quad & I_1(X) \leq E[DY|X] + E[(1-D)y_u|X] \\
\Leftrightarrow\quad & E[DY|X] + E[(1-D)y_\ell|X = a] \leq I_1(X) \leq E[DY|X] + E[(1-D)y_u|X = x] \\
& \hspace{22em} \forall a \leq x \\
\Leftarrow\quad & H_0^{GRM},
\end{aligned} \tag{B.5}$$

where the first $\Leftrightarrow$ holds because $Y \in [y_\ell, y_u]$, the second follows from rearranging terms, the third and the fourth $\Leftrightarrow$'s follow from the definition of $I_1(X)$, and the $\Leftarrow$ holds by the definition of H_0^{GRM}. Similarly, we can show that, H_0^{GRM} implies $y_\ell < Y(1) \leq y_u$ as well. Therefore, the constructed $Y(0)$ and $Y(1)$ are bounded between y_ℓ and y_u. $\square$

Proof of Theorem 2.2. Part (i) is obvious, thus its proof is omitted.

We show part (ii) by construction. Let $f_u(x, z)$ denote $E[Y_u|X = x, Z = z]$, and let $f_\ell(x, z)$ denote $E[Y_\ell|X = x, Z = z]$. Let

$$f(x, z) = \sup_{a \leq x} f_\ell(a, z). \tag{B.6}$$

Then, by definition, $f(x, z)$ is increasing in x for any z. And by H_0^{GRM} in (2.5), $f(x, z) \in [f_\ell(x, z), f_u(x, z)]$ for all x, z. Let

$$\lambda(x, z) = \frac{f_u(x, z) - f(x, z)}{f_u(x, z) - f_\ell(x, z)}, \tag{B.7}$$

where $0/0$ is taken to be 0. Then, $\lambda(x, z) \in [0, 1]$ for all x, z. Let

$$Y = Y_\ell \lambda(X, Z) + Y_u(1 - \lambda(X, Z)). \tag{B.8}$$

By construction, $Y \in [Y_\ell, Y_u]$. Also it is elementary that

$$E[Y|X,Z] = f_\ell(X,Z)\lambda(X,Z) + f_u(X,Z)(1-\lambda(X,Z)) = f(X,Z). \tag{B.9}$$

That means that the distribution of (Y,X,Z) satisfies H_0^{LRM}. This concludes the proof of part (ii). $\square$

Proof of Theorem 4.1. We prove the case for $\hat{c}_\eta = \hat{c}_\eta^u$. The proof for $\hat{c}_\eta = \hat{c}_\eta^b$ is similar and we omit it. Our proof is similar to that of Theorem 6.3 of Donald and Hsu (2016) and that of Hsu (2016). Let $\mathcal{H}_{1,\nu}$ denote the set of all functions from $\mathcal{T} \times \mathcal{L}$ to $[-\infty, 0]$. Let $h_\nu = (h_{1,\nu}, h_{2,\nu})$, where $h_{1,\nu} \in \mathcal{H}_{1,\nu}$ and $h_{2,\nu} \in \mathcal{H}_{2,\nu}$, and define

$$T(h_\nu) = \int \max\Big\{\frac{\Phi_{h_{2,\nu}}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu}}(\tau,\ell)} + h_{1,\nu}(\tau,\ell), 0\Big\}^2 dQ(\tau,\ell).$$

Define $c_0(h_{1,\nu}, h_{2,\nu}, 1-\alpha)$ as the $(1-\alpha)$-th quantile of $T(h_\nu)$.

Similar to Lemma A2 of AS, we can show that for any $\xi > 0$,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}^0} P\Big(\widehat{T}_n > c_0(h_{1,\nu,n}^P, h_{2,\nu,P}, 1-\alpha) + \xi\Big) \le \alpha, \tag{B.10}$$

where $h_{1,\nu,n}^P = \sqrt{n}\nu_P(\cdot,\cdot)$ and $h_{1,\nu,n}^P$ belongs to $\mathcal{H}_{1,\nu}$ under $P \in \mathcal{P}^0$. Also, similar to Lemma A3 of AS, we can show that for all $\alpha < 1/2$

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} P\Big(c_0(\phi_n, h_{2,\nu,P}, 1-\alpha) < c_0(h_{1,\nu,n}^P, h_{2,\nu,P}, 1-\alpha)\Big) = 0. \tag{B.11}$$

As a result, to complete the proof of Theorem 4.1, it suffices to show that for all $0 < \delta < \eta$

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}^0} P\Big(\hat{c}_{n,\eta}^u < c_0(\phi_N, h_{2,P}, 1-\alpha) + \xi\Big) = 0. \tag{B.12}$$

Let $\{P_n \in \mathcal{P}^0 : n \ge 1\}$ be a sequence for which the probability in the statement of (B.12) evaluated at P_n differs from its supremum over $P \in \mathcal{P}^0$ by δ_n or less, where $\delta_n > 0$ and $\lim_{n\to\infty}\delta_n = 0$. By the definition of lim sup, such sequence always exists. Therefore, it is equivalent to show that for $0 < \xi < \eta$,

$$\lim_{n\to\infty} P\Big(\hat{c}_{n,\eta}^u < c_0(\phi_n, h_{2,\nu,P}, 1-\alpha) + \xi\Big) = 0, \tag{B.13}$$

where $\hat{c}_{n,\eta}^u$ denotes the critical value under P_n. To be more specific, we know that quantity on the left hand side exists, but we want to show that it is 0. Given that $\mathcal{H}^0$ is a compact set, there exists a subsequence k_n of n such that $h_{P_{k,n}} \to h^*$ for some $h^* \in \mathcal{H}^0$ and this implies that $h_{2,\nu,P_{k_n}}$ converges to $h_{2,\nu}^*$. By Lemma A.2, $\cdot\Psi_{P_{k_n}}^u(\cdot)/\hat{\sigma}_{\epsilon,k_n}(\cdot) \Rightarrow \Psi_{h_{2,\nu}^*}(\cdot)/\sigma_{\epsilon,h_{2,\nu}^*}(\cdot)$ conditional on sample path in probability. Then there exists a further subsequence m_n of k_n such that $\Psi_{P_{m_n}}^u(\cdot)/\hat{\sigma}_{\epsilon,m_n}(\cdot) \Rightarrow \Psi_{h_{2,\nu}^*}(\cdot)/\sigma_{\epsilon,h_{2,\nu}^*}(\cdot)$ conditional on sample path almost surely.

For any $\omega \in \left\{\omega \in \Omega: \ \Psi^u_{P_{k_n}}(\cdot)/\hat{\sigma}_{\epsilon,k_n}(\cdot) \Rightarrow \cdot \, \Psi_{h^*_{2,\nu}}(\cdot)/\sigma_{\epsilon,h^*_{2,\nu}}(\cdot)\right\} \equiv \Omega_1$, by the same argument for Theorem 1 of AS we can show that for any constant $a_{m_n} \in R$ which may depend on h_1 and P and for any $0 < \xi_1$,

$$\limsup_{n\to\infty} \sup_{h_{1,\nu}\in\mathcal{H}_{1,\nu}} P_u\Big(\int \max\Big\{ \frac{\Psi^u_{P_{m_n}}(\tau,\ell)}{\hat{\sigma}_{\epsilon,m_n}(\tau,\ell)}(\omega) + h_{1,\nu}(\tau,\ell), 0\Big\}^2 dQ(\tau,\ell) \le a_{m_n}\Big) \\ - P\big(T(h_\nu) \le a_{m_n} + \xi_1\big) \le 0. \tag{B.14}$$

(B.14) is similar to (12.28) in AS. By (B.14) and by the similar argument for Lemma A5 of AS, we have that for all $0 < \xi < \xi_1 < \eta$,

$$\liminf\nolimits_{n\to\infty} \hat{c}^u_{m_n,\eta}(\omega) \ge c_0(\phi_{m_n}, h_{2,\nu,P_{m_n}}, 1-\alpha) + \xi_1. \tag{B.15}$$

Therefore, for any $\omega \in \Omega_1$, (B.15) holds. Given that $P(\Omega_1) = 1$, we have that for all $0 < \xi < \xi_1 < \eta$

$$P\Big(\big\{\omega : \liminf\nolimits_{n\to\infty} \hat{c}^u_{m_n,\eta}(\omega) \ge c_0(\phi_{m_n}, h_{2,\nu,P_{m_n}}, 1-\alpha) + \xi_1\big\}\Big) = 1,$$

which implies that

$$\lim\nolimits_{n\to\infty} P(\hat{c}^u_{m_n,\eta} < c_0(\phi_{m_n}, h_{2,\nu,P_{m_n}}, 1-\alpha) + \delta) = 0. \tag{B.16}$$

Note that for any convergent sequence A_n, if there exists a subsequence A_{m_n} converging to A, then A_n converges to A as well. Therefore, (B.16) is sufficient for (B.13). Theorem 4.1(a) is shown by combining (B.10), (B.11) and (B.12).

To show Theorem 4.1(ii), note that, under the P_c specified in Assumption 4.3, Lemma A.3 (i) implies that

$$\widehat{T}_n \xrightarrow{d} \int_{(\mathcal{TL})^o(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell). \tag{B.17}$$

this limiting distribution is non-degenerate by Assumption 4.3. Let $H(a)$ denote the CDF of the limiting distribution defined in (B.17). By Davydov, Lifshits and Smorodina (1995), $H(a)$ is continuous and strictly increasing on $a \in [0,\infty)$ with $H(0) > 1/2$ under Assumption 4.3. Therefore, the $(1-\alpha)$ quantile of the limiting distribution defined in (B.17) is strictly greater than 0 when $\alpha \le 1/2$, and we denote it as $c_0(1-\alpha)$. Also, $c_0(1-\alpha)$ is continuous on $\alpha \in (0, 1/2]$.

By the same proof for part (i), it is true that $\hat{c}^u_\eta \xrightarrow{p} c_0(1-\alpha+\eta)+\eta$, and by the continuity of the limiting distribution, we have $\lim_{\eta\to 0} c_0(1-\alpha+\eta)+\eta \to c_0(1-\alpha)$. Therefore, $\lim_{\eta\to 0}\limsup_{n\to\infty} P(\widehat{T}_n > \hat{c}^u_\eta) = \alpha$. $\square$

Proof of Theorem 5.1. Assumptions 3.2(a) and 5.1 together implies that

$$\int_{\mathcal{T}^a(P^*)} dQ(\tau) > 0. \tag{B.18}$$

For any $\tau \in \mathcal{T}^a(P_*)$, there exist $x_1 \geq x_2$ and z such that $E_P[f^{(1)}(W,\tau)|X = x_1, Z = z] > E_P[f^{(2)}(W,\tau)|X = x_2, Z = z]$. By continuity, there exist $r^* > 0$ such that $x_1^* - r^* \geq x_2^* + r^*$, that for all $x_1 \in [x_1^* - 2r^*, x_1^* + 2r^*]$, $x_2 \in [x_2^* - 2r^*, x_2^* + 2r^*]$ and $z \in [z^* - 2r^*, z + 2r^*]$, and that $E_P[f^{(1)}(W,\tau)|X = x_1, Z = z] > E_P[f^{(2)}(W,\tau)|X = x_2, Z = z]$. Then for all $\ell = (x_1, x_2, z, r)$ such that $x_1 \in [x_1^* - r^*, x_1^* + r^*]$, $x_2 \in [x_2^* - r^*, x_2^* + r^*]$, $z \in [z^* - r^*, z + r^*]$ and $r \leq r^*$, we have $\nu_{P_*}(\tau, \ell) > 0$. Let

$$\mathcal{L}^*(\tau) = [x_1^* - r^*, x_1^* + r^*] \times [x_2^* - r^*, x_2^* + r^*] \times [z^* - r^*, z + r^*] \times (0, r^*]. \tag{B.19}$$

By assumption, we have $\int_{\mathcal{L}^*(\tau)} Q(\ell) > 0$ and this implies

$$\int_{\mathcal{L}^*(\tau)} \max\Big\{\frac{\nu_{P_*}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_*}}(\tau,\ell)}, 0\Big\}^2 Q(\ell) = \int_{\mathcal{L}^*(\tau)} \Big(\frac{\nu_{P_*}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_*}}(\tau,\ell)}\Big)^2 Q(\ell) > 0 \tag{B.20}$$

because $\nu_{P_*}(\tau, \ell) > 0$ when $\ell \in \mathcal{L}^*(\tau)$. Next we have

$$\begin{aligned} A^* &\equiv \int \max\Big\{\frac{\nu_{P_*}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_*}}(\tau,\ell)}, 0\Big\}^2 Q(\tau,\ell) \\ &\geq \int_{\mathcal{T}^a(P_*)} \int_{\mathcal{L}^*(\tau)} \Big(\frac{\nu_{P_*}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_*}}(\tau,\ell)}\Big)^2 Q(\ell)Q(\tau) > 0. \end{aligned} \tag{B.21}$$

Note that $n^{-1}\widehat{T}_n \xrightarrow{p} A^* > 0$ under P_*. Therefore, $\widehat{T}_n$ diverges to positive infinity in probability, but $\hat{c}_\eta^u$ is bounded in probability. Therefore, $\lim_{n\to\infty} P(\widehat{T}_n > \hat{c}_\eta^u) = 1$. The proof for the bootstrap critical value is the same and thus omitted. □

Proof of Theorem 5.2. Note that

$$\begin{aligned} m_{P_n}^{(j)}(\tau,\ell) &= E_{P_n}[f^{(j)}(Y,\tau) g_{x,\ell}^{(j)}(X) g_{z,\ell}(Z)] \\ &= E_{P_{n,xz}}\big[E_{P_n}[f^{(j)}(Y,\tau)|X,Z] \cdot g_{x,\ell}^{(j)}(X) g_{z,\ell}(Z)\big] \\ &= E_{P_{c,xz}}\big[E_{P_c}[f^{(j)}(Y,\tau)|X,Z] \cdot g_{x,\ell}^{(j)}(X) g_{z,\ell}(Z)\big] \\ &\quad + E_{P_{c,xz}}\big[\gamma\delta_j(X,Z,\tau) \cdot g_{x,\ell}^{(j)}(X) g_{z,\ell}(Z)\big]/\sqrt{n} \\ &= m_{P_c}^{(j)}(\tau,\ell) + \gamma\delta_j(\tau,\ell)/\sqrt{n}, \end{aligned} \tag{B.22}$$

where $\delta_{(j)}(\tau,\ell) = E_{P_{c,xz}}\big[\delta_j(X,Z,\tau) \cdot g_{x,\ell}^{(j)}(X) g_{z,\ell}(Z)\big]$. The third equality holds because of Assumptions 5.2(a) and (b). Also,

$$w_{P_n}^{(j)}(\tau,\ell) = E_{P_n}[g_{x,\ell}^{(j)}(X) g_{z,\ell}(Z)] = E_{P_{n,xz}}[g_{x,\ell}^{(j)}(X) g_{z,\ell}(Z)]$$

$$= E_{P_{c,xz}}[g_{x,\ell}^{(j)}(X)g_{z,\ell}(Z)] = w_{P_c}^{(j)}(\tau,\ell) \tag{B.23}$$

where the third equality holds because $P_{n,xz} = P_{c,xz}$. Therefore,

$$\begin{aligned}\nu_{P_n}(\tau,\ell) &= m_{P_n}^{(2)}(\tau,\ell)w_{P_n}^{(1)}(\tau,\ell) - m_{P_n}^{(1)}(\tau,\ell)w_{P_n}^{(2)}(\tau,\ell)\\ &= m_{P_c}^{(2)}(\tau,\ell)w_{P_c}^{(1)}(\tau,\ell) - m_{P_c}^{(1)}(\tau,\ell)w_{P_c}^{(2)}(\tau,\ell)\\ &\qquad + \gamma\big(\delta_{(2)}(\tau,\ell)w_{P_c}^{(1)}(\tau,\ell) - \delta_{(1)}(\tau,\ell)w_{P_c}^{(2)}(\tau,\ell)\big)/\sqrt{n}\\ &= \nu_{P_c}(\tau,\ell) + \gamma\delta_\nu(\tau,\ell)/\sqrt{n},\end{aligned} \tag{B.24}$$

where $\delta_\nu(\tau,\ell) \equiv \delta_{(2)}(\tau,\ell)w_{P_c}^{(1)}(\tau,\ell) - \delta_{(1)}(\tau,\ell)w_{P_c}^{(2)}(\tau,\ell)$. Under Assumption 5.2 (d), we have

$$\delta_\nu(\tau,\ell) \geq 0 \ \forall \tau,\ell. \tag{B.25}$$

In addition, under Assumptions 5.2(e) and 5.3, we have,

$$\int_{(\mathcal{TL})^+(P_c)} dQ(\tau,\ell) > 0, \text{ where } (\mathcal{TL})^+(P_c) = \{(\tau,\ell) \in (\mathcal{TL})^o(P_c) : \delta_\nu(\tau,\ell) > 0\}. \tag{B.26}$$

Under the local alternative sequence $\{P_n\}_{n\geq 1}$, using B.24, Lemma A.3(i) shows that

$$\widehat{T}_n \xrightarrow{d} \int_{(\mathcal{TL})^o(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell) + \gamma\delta_\nu(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell). \tag{B.27}$$

Also, Lemma A.3(ii) shows that the critical value statistic

$$\widehat{T}_n^u \xrightarrow{d} \int_{(\mathcal{TL})^o(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) \tag{B.28}$$

conditional on almost all sample paths. Note that the limiting distribution defined in (B.28) is identical to that in (B.17). We denote its cumulative distribution function as $H(a)$.

We consider two cases, depending on whether the limiting distribution defined in (B.28) is degenerate or not. First, suppose that it is non-degenerate. By the proof for part (ii) of Theorem 4.1, we have that $H(a)$ is continuous and strictly increasing on $a \in [0,\infty)$. We also have that, the $(1-\alpha)$ quantile of the right-hand-side of (B.28), $c_0(1-\alpha)$, satisfies: $c_0(1-\alpha) > 0$ if $\alpha < 1/2$, and it is continuous on $\alpha \in (0, 1/2)$. Because $\delta_\nu(\tau,\ell) \geq 0$ for all τ and ℓ, we have that the limiting distribution of the test statistic defined in (B.27) is non-degenerate, strictly increasing on $[0,\infty)$, and first order stochastically dominant to that in (B.28). It follows that

$$\begin{aligned}&\lim_{\eta\to 0}\lim_{n\to\infty} P(\widehat{T}_n \geq \widehat{c}_\eta^u)\\ &= \lim_{\eta\to 0} P\Big(\int_{(\mathcal{TL})^o(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell) + \gamma\delta_\nu(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) \geq c_0(1-\alpha+\eta)+\eta\Big)\\ &\geq \lim_{\eta\to 0} P\Big(\int_{(\mathcal{TL})^o(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) \geq c_0(1-\alpha+\eta)+\eta\Big)\end{aligned}$$

$$= \alpha, \tag{B.29}$$

where the first equality holds because the test statistic defined in (B.27) is non-degenerate and strictly increasing on $[0, \infty)$, and the first inequality holds because limiting distribution of the test statistic defined in (B.27) first order stochastically dominates that in (B.28). The last equality holds because the distribution defined in (B.28) is continuous and strictly increasing on $[0, \infty)$. This shows part (i) of the theorem for the non-degenerate case.

We now show part (ii) for the non-degenerate case. Consider the derivation

$$\begin{aligned}
&\lim_{\gamma\to\infty}\lim_{n\to\infty} P(\widehat{T}_n \geq \widehat{c}_\eta^u) \\
&= \lim_{\gamma\to\infty} P\Big(\int_{(\mathcal{TL})^o(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)+\gamma\delta_\nu(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) \geq c_0(1-\alpha+\eta)+\eta)\Big) \\
&\geq \lim_{\gamma\to\infty} P\Big(\int_{(\mathcal{TL})^+(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)+\gamma\delta_\nu(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) \geq c_0(1-\alpha+\eta)+\eta)\Big) \\
&= 1,
\end{aligned} \tag{B.30}$$

where the last equality holds by (B.26). This shows part (ii) of the theorem for the non-degenerate case.

Now we consider the the second case, where the limiting distribution in (B.28) is degenerate. The limiting distribution in (B.28) is degenerate iff the measure of $\{(\tau,\ell) \in (\mathcal{TL})^o(P_c) : h_{2,\nu,P_c}((\tau,\ell),(\tau,\ell)) > 0\}$. Let

$$S = \{(\tau,\ell) \in (\mathcal{TL})^o(P_c) : h_{2,\nu,P_c}((\tau,\ell),(\tau,\ell)) = 0\}. \tag{B.31}$$

Equation (B.26) implies that $\int_{(\mathcal{TL})^o(P_c)} dQ(\tau,\ell) > 0$ because $(\mathcal{TL})^+(P_c) \subseteq (\mathcal{TL})^o(P_c)$. That and the degeneracy of the limiting distribution in (B.28) imply that the limiting distribution in (B.27) reduces to

$$\begin{aligned}
\int_S \max\Big\{\frac{\delta_\nu(\tau,\ell)}{\sqrt{\epsilon}}, 0\Big\}^2 dQ(\tau,\ell) &= \int_S \frac{\delta_\nu^2(\tau,\ell)}{\epsilon} dQ(\tau,\ell) \\
&\geq \int_{(\mathcal{TL})^+(P_c)} \frac{\delta_\nu^2(\tau,\ell)}{\epsilon} dQ(\tau,\ell) \\
&> 0.
\end{aligned} \tag{B.32}$$

where the strict inequality holds by (B.26).

Because the limiting distribution in (B.28) is degenerate, $\widehat{c}_\eta^u \xrightarrow{p} c_0(1-\alpha+\eta)+\eta = \eta$. Therefore, for η is small enough that $\int_{(\mathcal{TL})^+} \delta_\nu^2(\tau,\ell)\epsilon^{-1} dQ(\tau,\ell) > \eta$,

$$\lim_{n\to\infty} P(\widehat{T}_n \geq \widehat{c}_\eta^u) = 1 \tag{B.33}$$

This shows both part (i) and part (ii) of the theorem for the degenerate case. □

C Proof of Lemmas

Proof of Lemma 3.1. We first show that (3.1) implies (1.1) by contradiction. For this direction, we show the case for $\mathcal{L}_{\text{c-cube}}$ and given that $\mathcal{L}_{\text{c-cube}}$ is a subset of $\mathcal{L}_{\text{cube}}$, so the case for $\mathcal{L}_{\text{cube}}$ follows.

Suppose that (1.1) is not true, then there exist $x_1 > x_2$, $\tau \in \mathcal{T}$ and z such that $E_P[f^{(1)}(Y,\tau)|X = x_1, Z = z] < E_P[f^{(2)}(Y,\tau)|X = x_2, Z = z]$. By continuity, there exist $[x_{1\ell}, x_{1u}]$, $[x_{2\ell}, x_{2u}]$ and $[z_l, z_u]$ with $x_{1\ell} \ll x_{1u}$, $x_{2\ell} \ll x_{2u}$, $z_l \ll z_u$, $x_{1\ell} \geq x_{2\ell}$, $x_{1u} \geq x_{2u}$ such that

$$E_P[f^{(1)}(Y,\tau)|X = x_1, Z = z] < E_P[f^{(2)}(Y,\tau)|X = x_2, Z = z]$$
$$\text{for all } x_1 \in [x_{1\ell}, x_{1u}], x_2 \in [x_{2\ell}, x_{2u}], z \in [z_l, z_u]. \tag{C.1}$$

Given that rational numbers are dense and $x_{1\ell} \geq x_{2\ell}$, $x_{1u} \geq x_{2u}$, we can find x_1^*, x_2^*, z^* and a natural number q^* that is large enough such that

$$q^* \cdot (x_1, x_2, z) \in \{0, 1, \ldots, (q^*)^{-1}\}^{2d_x + d_z},$$
$$x_1^* \leq x_2^*,$$
$$[x_1^*, x_1^* + (q^*)^{-1}] \subseteq [x_{1\ell}, x_{1u}],\ [x_2^*, x_2^* + (q^*)^{-1}] \subseteq [x_{2\ell}, x_{2u}],\ [z^*, z^* + (q^*)^{-1}] \subseteq [z_\ell, z_u].$$

Let $\ell^* = (x_1, x_2, z, (q^*)^{-1})$ and it is obvious that $\ell^* \in \mathcal{L}_{\text{c-cube}}$. Equation (C.1) implies that

$$E_P\Big[f^{(1)}(Y,\tau)\big|X \in C_{x_1^*, r_x^*}, Z \in C_{z^*, r_z^*}\Big] < E_P\Big[f^{(2)}(Y,\tau)\big|X \in C_{x_2^*, r_x^*}, Z \in C_{z^*, r_z^*}\Big], \tag{C.2}$$

which is equivalent to

$$\frac{m_P^{(1)}(\tau, \ell^*)}{w_P^{(1)}(\ell^*)} = \frac{E_P[f^{(1)}(Y,\tau) g_{x,\ell^*}^{(1)}(X) g_{z,\ell^*}(Z)]}{E_P[g_{x,\ell^*}^{(1)}(X) g_{z,\ell^*}(Z)]}$$
$$< \frac{E_P[f^{(2)}(Y,\tau) g_{x,\ell^*}^{(2)}(X) g_{z,\ell^*}(Z)]}{E_P[g_{x,\ell^*}^{(2)}(X) g_{z,\ell^*}(Z)]} = \frac{m_P^{(2)}(\tau, \ell^*)}{w_P^{(2)}(\tau, \ell^*)}, \tag{C.3}$$

Therefore, there exist $\tau \in \mathcal{T}$ and $\ell^* \in \mathcal{L}_{\text{c-cube}}$ such that

$$\nu_P(\tau, \ell^*) = m_P^{(2)}(\tau, \ell^*) w_P^{(1)}(\ell^*) - m_P^{(1)}(\tau, \ell^*) w_P^{(2)}(\ell^*) > 0, \tag{C.4}$$

i.e., (3.1) is violated.

Next, we show that (1.1) implies (3.1). It is sufficient to show the $\mathcal{L}_{\text{cube}}$ case since $\mathcal{L}_{\text{c-cube}}$ is a subset of $\mathcal{L}_{\text{cube}}$. For notational simplicity, we consider the case where $d_x = 1$ and $d_z = 0$, and the proof for cases where $d_x \geq 2$ and/or $d_z \geq 1$ is similar. When $d_x = 1$ and $d_z = 0$, we have $\ell = (x_1, x_2, r)$. Note that we only need to consider those ℓ's such that $E[g_{x,\ell}^{(1)}] = P(X \in C_{x_1, r_x}) > 0$ and $E[g_{x,\ell}^{(2)}] = P(X \in C_{x_2, r_x}) > 0$ because $E[g_{x,\ell}^{(1)}] = P(X \in C_{x_1, r_x}) = 0$ implies that $m_P^{(1)}(\tau, \ell) = 0$ and $w_P^{(1)}(\tau, \ell) = 0$ for all $\tau \in \mathcal{T}$. This further implies that $\nu_P(\tau, \ell) = 0$ for all $\tau \in \mathcal{T}$. For any $\ell \in \mathcal{L}$ such that $E[g_{x,\ell}^{(1)}] > 0$ and $E[g_{x,\ell}^{(2)}] > 0$, there are three different cases to consider:

First. $x_1 = x_2$.

Second. $x_1 > x_2$, and $x_1 \geq x_2 + r$.

Third. $x_1 > x_2$, and $x_1 < x_2 + r$.

For the first case, clearly, $g^{(1)}_{x,\ell} = g^{(2)}_{x,\ell}$ and

$$\nu_P(\tau,\ell) = E_P\big[\big(f^{(2)}(Y,\tau) - f^{(1)}(Y,\tau)\big)g^{(1)}_{x,\ell}(X)\big] \cdot E_P[g^{(1)}_{x,\ell}(X)]. \tag{C.5}$$

By (1.1), we have

$$E[f^{(1)}(Y,\tau)|X = x] \geq E[f^{(2)}(Y,\tau)|X = x] \quad \text{for all } x \in [x_1 - r, x_1 + r], \tag{C.6}$$

and by law of iterated expectations,

$$\begin{aligned} E_P\big[\big(f^{(2)}(Y,\tau) - f^{(1)}(Y,\tau)\big)g^{(1)}_{x,\ell}(X)\big] &= E_{P_x}\Big[E_P\big[(f^{(2)}(Y,\tau) - f^{(1)}(Y,\tau))|X\big]g^{(1)}_{x,\ell}(X)\Big] \\ &\leq 0. \end{aligned} \tag{C.7}$$

This implies that $\nu_P(\tau,\ell) \leq 0$ for the first case.

For the second case, we have $x_1' \geq x_2'$ for all $x_1' \in [x_1, x_1 + r]$ and $x_2' \in [x_2, x_2 + r]$. By (1.1),

$$E[f^{(1)}(Y,\tau)|X = x_1'] \geq E[f^{(2)}(Y,\tau)|X = x_2'] \quad \text{for all } x_1' \in [x_1, x_1 + r],\ x_2' \in [x_2, x_2 + r]. \tag{C.8}$$

It follows that

$$\begin{aligned} \frac{E_P\big[f^{(1)}(Y,\tau)g^{(1)}_{x,\ell}(X)\big]}{E_P\big[g^{(1)}_{x,\ell}(X)\big]} &= E_P\big[f^{(1)}(Y,\tau)|X \in [x_1, x_1 + r]\big] \\ &\geq E_P\big[f^{(2)}(Y,\tau)|X \in [x_2, x_2 + r]\big] = \frac{E_P\big[f^{(2)}(Y,\tau)g^{(2)}_{x,\ell}(X)\big]}{E_P\big[g^{(2)}_{x,\ell}(X)\big]}. \end{aligned} \tag{C.9}$$

This implies that $\nu_P(\tau,\ell) \leq 0$ for the second case.

For the third case, it is true that $x_1 + r > x_2 + r > x_1 > x_2$. Therefore, $[x_2, x_2 + r] = [x_2 + r, x_1 + r] \cup [x_1 + r, x_2]$ and $[x_1 + r, x_1] = [x_1 + r, x_2] \cup [x_2, x_1]$. By the similar argument in the first case and the second case,

$$\begin{aligned} E_P\big[f^{(1)}(Y,\tau)|X \in [x_2 + r, x_1 + r]\big] &\geq E_P\big[f^{(2)}(Y,\tau)|X \in [x_1, x_2 + r]\big] \\ E_P\big[f^{(1)}(Y,\tau)|X \in [x_2 + r, x_1 + r]\big] &\geq E_P\big[f^{(2)}(Y,\tau)|X \in [x_2, x_1]\big] \\ E_P\big[f^{(1)}(Y,\tau)|X \in [x_1, x_2 + r]\big] &\geq E_P\big[f^{(2)}(Y,\tau)|X \in [x_1, x_2 + r]\big] \\ E_P\big[f^{(1)}(Y,\tau)|X \in [x_1, x_2 + r]\big] &\geq E_P\big[f^{(2)}(Y,\tau)|X \in [x_2, x_1]\big]. \end{aligned} \tag{C.10}$$

These imply that

$$
\begin{aligned}
&E_P\big[f^{(1)}(Y,\tau)|X\in[x_2+r,x_1+r]\big]\\
&\geq \frac{P(X\in[x_1,x_2+r])}{P(X\in[x_2,x_2+r])}E_P\big[f^{(2)}(Y,\tau)|X\in[x_1,x_2+r]\big]+\\
&\quad \frac{P(X\in[x_2,x_1])}{P(X\in[x_2,x_2+r])}E_P\big[f^{(2)}(Y,\tau)|X\in[x_2,x_1]\\
&=E_P\big[f^{(2)}(Y,\tau)|X\in[x_2,x_1],\ \text{and}
\end{aligned}
\tag{C.11}
$$

$$
E_P\big[f^{(1)}(Y,\tau)|X\in[x_1,x_2+r]\big]\geq E_P\big[f^{(2)}(Y,\tau)|X\in[x_2,x_1].
\tag{C.12}
$$

It follows that

$$
\begin{aligned}
&E_P\big[f^{(1)}(Y,\tau)|X\in[x_1,x_1+r]\big]\\
&=\frac{P(X\in[x_2+r,x_1+r])}{P(X\in[x_1,x_1+r])}E_P\big[f^{(1)}(Y,\tau)|X\in[x_2+r,x_1+r]\big]+\\
&\quad \frac{P(X\in[x_2+r,x_1+r])}{P(X\in[x_1,x_1+r])}E_P\big[f^{(1)}(Y,\tau)|X\in[x_1,x_2+r]\big]\\
&\geq E_P\big[f^{(2)}(Y,\tau)|X\in[x_2,x_1].
\end{aligned}
\tag{C.13}
$$

This implies that $\nu_P(\tau,\ell)\leq 0$ for the third case.

This completes the proof for Lemma 3.1. □

D Proofs of Auxiliary Lemmas

Proof of Lemma A.1. For notational simplicity, we prove it for the sequence $\{n\}$ and all of the arguments go through with $\{a_n\}$ in place of $\{n\}$.

We apply Lemma E2 of Andrews and Shi (2013b; AS2 hereafter) to show part (i). It is sufficient to show that every element of $\hat{h}_{P_n}$ converges to h uniformly. Note that $\{m^{(j)}(\omega,W_{n,i},\tau,\ell):\tau\in\mathcal{T},\ \ell\in\mathcal{L},\ i\leq n,\ n\geq 1\}$ is manageable with respect to envelopes $\{(F_{n,1}(\omega),\dots,F_{n,n}(\omega)):\ n\geq 1\}$ because $m^{(j)}(W,\tau,\ell)=f^{(j)}(W,\tau)\cdot g^{(j)}_{x,\ell}(X)\cdot g_{z,\ell}(Z)$, and $\{f^{(j)}(\omega,W_{n,i},\tau):\ \tau\in\mathcal{T},\ i\leq n,\ n\geq 1\}$, $\{g^{(j)}_{x,\ell}(\omega,X_{n,i}):\ \ell\in\mathcal{L},\ i\leq n,\ n\geq 1\}$ and $\{g_{z,\ell}(\omega,Z_{n,i}):\ \ell\in\mathcal{L},\ i\leq n,\ n\geq 1\}$ are manageable with respect to envelopes $\{(F_{n,1}(\omega),\dots,F_{n,n}(\omega)):\ n\geq 1\}$, $\{(1,\dots,1):\ n\geq 1\}$ and $\{(1,\dots,1):\ n\geq 1\}$ respectively. By Assumption 4.1, there exists $0<\eta\, n^{-1}\sum_{i=1}^n E_{P_n}F_{n,i}^{1+\eta}\leq M$ for some $0<M<\infty$ for all $n\geq 1$. Then by Lemma E2 AS2, we have

$$
\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}}\Big|\frac{1}{n}\sum_{i=1}^n m^{(j)}(W_{n,i},\tau,\ell)-E_{P_n}m^{(j)}(W,\tau,\ell)\Big|\xrightarrow{p}0.
\tag{D.1}
$$

Similar arguments apply to $w^{(j)}(W,\ell)$. This shows that $d(\hat{h}_{P_n,1},h_1)\xrightarrow{p}0$.

The proof for $d(\hat{h}_{P_n,2}, h_2) \xrightarrow{p} 0$ is identical to the proof of Lemma A1(b) of AS2 after we replace their D_F with identity matrix and their $\widehat{\Sigma}_n(\theta, g, g*)$ with $\hat{h}_{2,P}(\cdot,\cdot)$, so we omit it for brevity. This completes part (i).

The proof for Part (ii) is a non-standardized version of Lemma A1(a) of AS2 and the proof is identical to that for Lemma A1(a) of AS2. We omit it for brevity. $\square$

Proof of Lemma A.2. For notational simplicity, we prove it for the sequence $\{n\}$ and all of the arguments go through with $\{a_n\}$ in place of $\{n\}$. Part (i) follows from the fact that $d(h_{P_n}, h) \to 0$ and the definitions of h_{2,ν,P_n} and $h_{2,\nu}$. Part (ii) follows from Lemma A.1(i).

For part (iii), note that uniformly over $(\tau, \ell) \in \mathcal{T} \times \mathcal{L}$,

$$\begin{aligned}
&\sqrt{n}\big(\hat{m}_n^{(2)}(\tau,\ell)\hat{w}_n^{(1)}(\ell) - E_{P_n}[m^{(2)}(\tau,\ell)]E_{P_n}[w^{(1)}(\ell)]\big) \\
&= E_{P_n}[w^{(1)}(\ell)] \cdot \sqrt{n}(\hat{m}_n^{(2)}(\tau,\ell) - E_{P_n}[m^{(2)}(\tau,\ell)]) \\
&\qquad + E_{P_n}[m^{(2)}(\tau,\ell)] \cdot \sqrt{n}(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)]) \\
&\qquad + \sqrt{n}(\hat{m}_n^{(2)}(\tau,\ell) - E_{P_n}[m^{(2)}(\tau,\ell)])(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)]) \\
&= E_{P_n}[w^{(1)}(\ell)] \cdot \sqrt{n}(\hat{m}_n^{(2)}(\tau,\ell) - E_{P_n}[m^{(2)}(\tau,\ell)]) \\
&\quad + E_{P_n}[m^{(2)}(\tau,\ell)] \cdot \sqrt{n}(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)]) + o_p(1),
\end{aligned} \tag{D.2}$$

where the $o_p(1)$ in the last line follows from Lemma A.1(ii). Similar expansion applies to $\sqrt{n}\big(\hat{m}_n^{(1)}(\tau,\ell)\hat{w}_n^{(2)}(\ell) - E_{P_n}[m^{(1)}(\tau,\ell)]E_{P_n}[w^{(2)}(\ell)]\big)$. Therefore, uniformly over $(\tau, \ell) \in \mathcal{T} \times \mathcal{L}$,

$$\begin{aligned}
\widehat{\Phi}_{P_{a_n}}(\tau,\ell) &= \sqrt{n}(\hat{\nu}_n(\tau,\ell) - \nu_{P_n}(\tau,\ell)) \\
&= E_{P_n}[w^{(1)}(\ell)] \cdot \sqrt{n}(\hat{m}_n^{(2)}(\tau,\ell) - E_{P_n}[m^{(2)}(\tau,\ell)]) \\
&\quad + E_{P_n}[m^{(2)}(\tau,\ell)] \cdot \sqrt{n}(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)]) \\
&\quad - E_{P_n}[w^{(2)}(\ell)] \cdot \sqrt{n}(\hat{m}_n^{(1)}(\tau,\ell) - E_{P_n}[m^{(1)}(\tau,\ell)]) \\
&\quad - E_{P_n}[m^{(1)}(\tau,\ell)] \cdot \sqrt{n}(\hat{w}_n^{(2)}(\ell) - E_{P_n}[w^{(2)}(\ell)]) + o_p(1) \\
&= h_{P_n,1} \cdot \hat{\chi}_{P_n}(\tau,\ell) + o_p(1).
\end{aligned} \tag{D.3}$$

By Lemma A.1(ii) and the fact that $d(h_{P_n}, h) \to 0$, we have $h_{P_n,1} \cdot \hat{\chi}_{P_n}(\tau,\ell) \Rightarrow \Psi_{h_{2,\nu}}$. Equation (D.3) is equivalent to that $\sup_{(\tau,\ell)\in\mathcal{T}\times\mathcal{L}} |\widehat{\Phi}_{P_{a_n}}(\tau,\ell) - h_{P_n,1} \cdot \hat{\chi}_{P_n}(\tau,\ell)| \xrightarrow{p} 0$ and by Lemma 1.10.2 of van der Vaart and Wellner (1996), this suffices to show that $\widehat{\Phi}_{P_{a_n}}(\tau,\ell) = \sqrt{n}(\hat{\nu}_n(\tau,\ell) - \nu_{P_n}(\tau,\ell)) \Rightarrow \Psi_{h_{2,\nu}}$.

For part (iv), we define $\beta_n(W_i, \tau, \ell)$ as

$$\beta_n(W_i,\tau,\ell) = E_{P_n}[w^{(1)}(\ell)] \cdot \big(m_i^{(2)}(\tau,\ell) - E_{P_n}[m^{(2)}(\tau,\ell)]\big)$$

$$
\begin{aligned}
&+ E_{P_n}[m^{(2)}(\tau,\ell)] \cdot \big(w_i^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)]\big) \\
&- E_{P_n}[w^{(2)}(\ell)] \cdot \big(m_i^{(1)}(\tau,\ell) - E_{P_n}[m^{(1)}(\tau,\ell)]\big) \\
&- E_{P_n}[m^{(1)}(\tau,\ell)] \cdot \big(w_i^{(2)}(\ell) - E_{P_n}[w^{(2)}(\ell)]\big),
\end{aligned} \tag{D.4}
$$

and we denote it as $\beta_{n,i}(\tau,\ell)$. It is straightforward to see that $\widehat{\Phi}_{P_{a_n}}(\tau,\ell) = n^{-1/2}\sum_{i=1}^n \beta_{n,i}(\tau,\ell) + o_p(1)$ from (D.3). Also, define

$$
\begin{aligned}
\hat{\beta}_{n,i}(\tau,\ell) = {}& \hat{w}_n^{(1)}(\ell) \cdot (m_i^{(2)}(\tau,\ell) - \hat{m}_n^{(2)}(\tau,\ell)) + \hat{m}_n^{(2)}(\tau,\ell) \cdot (w_i^{(1)}(\ell) - \hat{w}_n^{(1)}(\ell)) \\
& - \hat{w}_n^{(2)}(\ell) \cdot (m_i^{(1)}(\tau,\ell) - \hat{m}_n^{(1)}(\tau,\ell)) - \hat{m}_n^{(1)}(\tau,\ell) \cdot (w_i^{(2)}(\ell) - \hat{w}_n^{(2)}(\ell)),
\end{aligned} \tag{D.5}
$$

which is the sample counterpart of $\beta_{n,i}(\tau,\ell)$. It is true that $\widehat{\Phi}^u_{P_n} = n^{-1/2}\sum_{i=1}^n U_i \cdot \hat{\beta}_{n,i}(\tau,\ell)$.

Because Ψ_{h_ν} is Borel measurable and separable, then by Section 1.12 (page 73) of van der Vaart and Wellner (1996), $\widehat{\Phi}^u_{P_n} \Rightarrow \Psi_{h_{2,\nu}}$ conditional on sample path with probability one iff $\sup_{g\in BL_1} |E_u g(\widehat{\Phi}^u_{P_n}) - E[g(\Psi_{h_{2,\nu}})]| \xrightarrow{p} 0$ where BL_1 denotes the set of all real functions on $\ell^\infty(\mathcal{T}\times\mathcal{L})$ with a Lipschitz norm bounded by 1 and E_u denotes the expectation w.r.t. U_i's. Then by Lemma 1.9.2 of van der Vaart and Wellner (1996), $\sup_{g\in BL_1} |E_u g(\widehat{\Phi}^u_{P_n}) - E[g(\Psi_{h_{2,\nu}})]| \xrightarrow{p} 0$ iff for any subsequence $\{b_n\}$ of $\{n\}$, there exists a further subsequence of $\{k_n\}$ such that $\sup_{g\in BL_1} |E_u g(\widehat{\Phi}^u_{P_n}) - E[g(\Psi_{h_{2,\nu}})]| \xrightarrow{a.s.} 0$ which is equivalent to that for any subsequence $\{b_n\}$ of $\{n\}$, there exists a further subsequence of $\{k_n\}$ such that $\widehat{\Phi}^u_{P_{k_n}} \Rightarrow \Psi_{h_{2,\nu}}$ conditional on sample path almost surely. Hence, to show part (iv), it is sufficient to show that for any subsequence $\{b_n\}$ of $\{n\}$, there exists a further subsequence of $\{k_n\}$ such that $\widehat{\Phi}^u_{P_{k_n}} \Rightarrow \Psi_{h_{2,\nu}}$ conditional on sample path almost surely.

First, let $M_g > 1$ be some constant such that $E_{P_n}[m_n^{(j)}(W,\tau,\ell)] \le M_g$ and $E_{P_n}[w_n^{(j)}(W,\ell)] \le M_g$ for all $n \ge 1$. Such M_g exists because of Assumption 4.1(b). Under Assumption 4.1 and by law of large number (LLN), we have $n^{-1}\sum_{i=1}^n (F_{n,i}+M_g)^2 - E_{P_n}[(F_{n,i}+M_g)^2] \xrightarrow{p} 0$. Also, by LLN, we have $n^{-1}\sum_{i=1}^n (F_{n,i}+M_g)^{\delta_1} - E_{P_n}[(F_{n,i}+M_g)^{\delta_1}] \xrightarrow{p} 0$ where δ is as defined in Assumption 4.1 and it is true that $\limsup_{n\to\infty} E_{P_{k_n}}[(F_{k_n,i}+M_g)^{\delta_1}] < \infty$.

As a result, for any subsequence $\{b_n\}$ of $\{n\}$, there exists a further subsequence of $\{k_n\}$ such that

$$
\begin{aligned}
& d(\hat{h}_{P_{k_n}}, h) \xrightarrow{a.s.} 0, \\
& \frac{1}{k_n}\sum_{i=1}^{k_n} (F_{k_n,i}+M_g)^2 - E_{P_{k_n}}[(F_{k_n,i}+M_g)^2] \xrightarrow{a.s.} 0, \text{ and} \\
& \frac{1}{k_n}\sum_{i=1}^{k_n} (F_{k_n,i}+M_g)^{\delta} - E_{P_{k_n}}[(F_{k_n,i}+M_g)^{\delta}] \xrightarrow{a.s.} 0.
\end{aligned} \tag{D.6}
$$

Define

$$
\Omega_1 \equiv \Big\{\omega\in\Omega:\ d(\hat{h}_{P_{k_n}}, h)(\omega) \to 0,
$$

$$\frac{1}{k_n}\sum_{i=1}^{k_n}(F_{k_n,i}+M_g)^2(\omega)-E_{P_{k_n}}[(F_{k_n,i}+M_g)^2]\to 0, \text{ and}$$
$$\frac{1}{k_n}\sum_{i=1}^{k_n}(F_{k_n,i}+M_g)^{\delta_1}(\omega)-E_{P_{k_n}}[(F_{k_n,i}+M_g)^{\delta_1}]\to 0\Big\}. \tag{D.7}$$

By construction, $P(\Omega_1)=1$. We show that $k_n^{-1/2}\sum_{i=1}^{k_n}U_i\cdot\hat{\beta}_{k_n,i}(\tau,\ell)(\omega)\Rightarrow\Psi_{h_{2,\nu}}(\tau,\ell)$ for all $\omega\in\Omega_1$. First, we re-write

$$\begin{aligned}
&\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\cdot\hat{\beta}_{k_n,i}(\tau,\ell)(\omega)\\
&=\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\cdot\beta_{k_n,i}(\tau,\ell)(\omega)\\
&\quad+(\hat{w}_n^{(1)}(\ell)-E_{P_n}[w^{(1)}(\ell)])(\omega)\cdot\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\cdot(m_i^{(2)}(\tau,\ell)-E_{P_n}[m^{(2)}(\tau,\ell)])(\omega)\\
&\quad+(\hat{m}_n^{(2)}(\tau,\ell)-E_{P_n}[m^{(2)}(\tau,\ell)])(\omega)\cdot\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\cdot(w_i^{(1)}(\ell)-E_{P_n}[w^{(1)}(\ell)])(\omega)\\
&\quad-(\hat{w}_n^{(2)}(\ell)-E_{P_n}[w^{(2)}(\ell)])(\omega)\cdot\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\cdot(m_i^{(1)}(\tau,\ell)-E_{P_n}[m^{(1)}(\tau,\ell)])(\omega)\\
&\quad-(\hat{m}_n^{(1)}(\tau,\ell)-E_{P_n}[m^{(1)}(\tau,\ell)])(\omega)\cdot\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\cdot(w_i^{(2)}(\ell)-E_{P_n}[w^{(2)}(\ell)])(\omega)\\
&\quad+(\hat{w}_n^{(1)}(\ell)-E_{P_n}[w^{(1)}(\ell)])^2(\omega)\cdot\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i+(\hat{m}_n^{(2)}(\tau,\ell)-E_{P_n}[m^{(2)}(\tau,\ell)])^2(\omega)\times\\
&\quad\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\\
&\quad-(\hat{w}_n^{(2)}(\ell)-E_{P_n}[w^{(2)}(\ell)])^2(\omega)\cdot\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i-(\hat{m}_n^{(1)}(\tau,\ell)-E_{P_n}[m^{(1)}(\tau,\ell)])^2(\omega)\times\\
&\quad\frac{1}{\sqrt{k_n}}\sum_{i=1}^{k_n}U_i\\
&=\ A+B_1+B_2-B_3-B_4+C_1+C_2-C_3-C_4,
\end{aligned}\tag{D.8}$$

where A, B_j's and C_j's are defined term by term. It is sufficient for us to show that $A\Rightarrow\Psi_{h_{2,\nu}}$, and B_j's and C_j's are all $o_p(1)$ uniformly over $\tau\in\mathcal{T}$ and $\ell\in\mathcal{L}$.

We use Theorem 10.6 (functional central limit theorem) of Pollard (1990) to show $A\Rightarrow\Psi_{h_{2,\nu}}$. Define

$$g_{k_n,i}(\tau,\ell)=\frac{U_i}{\sqrt{k_n}}\beta_{k_n,i}(\tau,\ell)(\omega), \tag{D.9}$$
$$G_{k_n,i}=\frac{U_i}{\sqrt{k_n}}4M_g(F_{k_n,i}(\omega)+M_g).$$

By Lemma E1 of AS2 and the manageability of the every element of $\beta_{k_n,i}(\tau,\ell)$, we have that $\{g_{k_n,i}(\tau,\ell,\omega_u) : \tau \in \mathcal{T},\ \ell \in \mathcal{L},\ i \leq k_n,\ n \geq 1\}$ is manageable with respect to envelopes $\{(G_{k_n,1}(\omega_u), \ldots, G_{k_n,k_n}(\omega_u) : n \geq 1\}$. Hence, (i) of Theorem 10.6 of Pollard (1990) holds. Let $\zeta_{k_n}(\tau,\ell) = \sum_{i=1}^{k_n} g_{k_n,i}(\tau,\ell)$. By definition, $E_u[\zeta_{k_n}(\tau_1,\ell_1)\zeta_{k_n}(\tau_2,\ell_2)] = h'_{1,P_{k_n}}\tilde{h}_{2,P_{k_n}}h_{1,P_{k_n}}((\tau_1,\ell_1), (\tau_2,\ell_2))$ where

$$\tilde{h}_{2,P_{k_n}} = \frac{1}{k_n}\sum_{i=1}^{k_n}\big(\ddot{m}(W_i,\cdot) - E_{P_{k_n}}[\ddot{m}(\cdot)]\big)\big(\ddot{m}(W_i,\cdot) - E_{P_{k_n}}[\ddot{m}(\cdot)]\big)'. \tag{D.10}$$

Also,

$$\tilde{h}_{2,P_{k_n}} = \hat{h}_{2,P_{k_n}}(\omega) - \big(\widehat{\ddot{m}}_{P_{k_n}}(\cdot) - E_{P_{k_n}}[\ddot{m}(\cdot)]\big)\big(\widehat{\ddot{m}}_{P_{k_n}}(\cdot) - E_{P_{k_n}}[\ddot{m}(\cdot)]\big)'(\omega). \tag{D.11}$$

Equation (D.11) and $d(\hat{h}_{P_{k_n}}(\omega), h) \to 0$ imply that $d(\tilde{h}_{2,P_{k_n}}(\omega), h_2) \to 0$. By assumption, we have that $d(h_{1,P_{k_n}}, h_1) \to 0$, so $\tilde{h}_{2,\nu,P_{k_n}} \equiv h'_{1,P_{k_n}}\tilde{h}_{2,P_{k_n}}h_{1,P_{k_n}} \to h_{2,\nu}$. That is, (ii) of Theorem 10.6 of Pollard (1990) hold. Note that $\sum_{i=1}^{k_n}(F_{k_n,i}(\omega) + M_g)^2 - E_{P_{k_n}}[(F_{k_n,i}(\omega) + M_g)^2] \to 0$ and $E_{P_{k_n}}[(F_{k_n,i}(\omega) + M_g)^2] < C$ for some constant C. These imply that, for some constant C, $\limsup_{n\to\infty} k_n^{-1}\sum_{i=1}^{k_n}(F_{k_n,i}(\omega) + M_g)^2 < C$. Also, consider the derivation

$$\limsup_{n\to\infty} E_u\Big[\sum_{i=1}^{k_n} G^2_{k_n,i}\Big] = \limsup_{n\to\infty}\frac{1}{k_n}\sum_{i=1}^{k_n}\big[4M_g(F_{k_n,i}(\omega) + M_g)\big]^2 < 16M_g^2 C < \infty. \tag{D.12}$$

That is, part (iii) of Theorem 10.6 of Pollard (1990) holds. By a similar argument of (16.39) of AS, we have, for any $\epsilon \in (0,\infty)$,

$$\begin{aligned}\sum_{i=1}^{k_n} E_u\big[G^2_{k_n,i}(\omega)\cdot 1(G_{k_n,i} > \epsilon)\big] &\leq \epsilon^{-\delta_1}\sum_{i=1}^{k_n} E_u\Big[\Big|\frac{U_i}{\sqrt{k_n}}\ddot{F}_{k_n,i}(\omega)\Big|^{\delta_1}\Big] \\ &\leq \frac{C}{k_n^{\delta/2-1}\epsilon^{\delta_1}}\cdot\frac{1}{k_n}\sum_{i=1}^{k_n}\big|(F_{k_n,i}(\omega) + M_g)\big|^{\delta_1} \\ &\to 0, \end{aligned} \tag{D.13}$$

where the C in the second inequality comes from $E[|U|^{2+\delta_1}] < C$ and the convergence result in the last line holds because $k_n^{-\delta_1/2+1} \to 0$ and $\limsup_{n\to\infty} k_n^{-1}\sum_{i=1}^{k_n}|F_{k_n,i}(\omega) + M_g|^{\delta_1}(\omega) < \infty$. That is, (iv) of Theorem 10.6 of Pollard (1990) holds. Note that

$$\begin{aligned}&\rho_{k_n}((\tau_1,\ell_1),(\tau_2,\ell_2)) \\ &= \sum_{i=1}^{k_n} E_u\big[g^2_{k_n,i}(\tau_1,\ell_1) + g^2_{k_n,i}(\tau_2,\ell_2) - 2g_{k_n,i}(\tau_1,\ell_1)g_{k_n,i}(\tau_2,\ell_2)\big] \\ &= \frac{1}{k_n}\sum_{i=1}^{k_n}\beta^2_{k_n,i}(\tau_1,\ell_1)(\omega) + \beta^2_{k_n,i}(\tau_2,\ell_2)(\omega) - 2\beta_{k_n,i}(\tau_1,\ell_1)(\omega)\beta_{k_n,i}(\tau_2,\ell_2)(\omega) \\ &= \tilde{h}_{2,\nu,P_{k_n}}((\tau_1,\ell_1),(\tau_1,\ell_1)) + \tilde{h}_{2,\nu,P_{k_n}}((\tau_2,\ell_2),(\tau_2,\ell_2)) - 2\tilde{h}_{2,\nu,P_{k_n}}((\tau_1,\ell_1),(\tau_2,\ell_2))\end{aligned}$$

$$
\begin{aligned}
&\to h_{2,\nu}((\tau_1,\ell_1),(\tau_1,\ell_1)) + h_{2,\nu}((\tau_2,\ell_2),(\tau_2,\ell_2)) - 2h_{2,\nu}((\tau_1,\ell_1),(\tau_2,\ell_2)) \\
&\equiv \rho((\tau_1,\ell_1),(\tau_2,\ell_2)), \qquad \text{(D.14)}
\end{aligned}
$$

uniformly over $(\tau_1,\ell_1),(\tau_2,\ell_2) \in \mathcal{T} \times \mathcal{L}$. This is sufficient for (v) of Theorem 10.6 of Pollard (1990). Therefore, we have $\zeta_{k_n} \Rightarrow \Psi_{h_{2,\nu}}$ by Theorem 10.6 of Pollard (1990).

For B_1 term, notice that by the same argument for A, we have

$$
\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} U_i \cdot \big(m_i^{(2)}(\tau,\ell) - E_{P_n}[m^{(2)}(\tau,\ell)]\big)(\omega) \Rightarrow \Psi_{h_2(1,1)}, \qquad \text{(D.15)}
$$

where $h_2(1,1)$ denote the $(1,1)$-th element of h_2. Note that $\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} |\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)])|(\omega) \to 0$, so it is true that

$$
\begin{aligned}
\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} |B_1| \leq \sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} &\big|(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)])(\omega)\big| \times \\
&\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} \Big|\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} U_i \cdot (m_i^{(2)}(\tau,\ell) - E_{P_n}[m^{(2)}(\tau,\ell)])(\omega)\Big| \\
&= o(1) \cdot O_p(1) = o_p(1). \qquad \text{(D.16)}
\end{aligned}
$$

Therefore, $B_j = o_p(1)$ for all $j = 1, \ldots, 4$. For C_1 term, we have

$$
\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} U_i = O_p(1) \qquad \text{(D.17)}
$$

and $\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} \big|(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)])(\omega)\big|^2 = o(1)$, so

$$
\begin{aligned}
&\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} \Big|(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)])^2(\omega) \frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} U_i\Big| \\
=&\Big|\frac{1}{\sqrt{k_n}} \sum_{i=1}^{k_n} U_i\Big| \cdot \sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} \Big|(\hat{w}_n^{(1)}(\ell) - E_{P_n}[w^{(1)}(\ell)])\Big|^2(\omega) \\
=&O_p(1) \cdot o(1) = o_p(1). \qquad \text{(D.18)}
\end{aligned}
$$

Similarly, $C_j = o_p(1)$ for $j = 2, 3$ and 4.

These results are sufficient to show that ${k_n}^{-1/2} \sum_{i=1}^{k_n} U_i \cdot \hat{\beta}_{k_n,i}(\tau,\ell)(\omega) \Rightarrow \Psi_{h_{2,\nu}}$ for all $\omega \in \Omega_1$ with $P(\Omega_1) = 1$. This shows $\widehat{\Phi}_{P_n}^u \Rightarrow \Psi_{h_{2,\nu}}$ conditional on sample path with probability one.

For part (v), let

$$
\hat{\chi}_{P_n}^b(\tau,\ell) \equiv \begin{pmatrix} \sqrt{n}\big(\hat{m}_n^{(1)b}(\tau,\ell) - \hat{m}^{(1)}(\tau,\ell)\big) \\ \sqrt{n}\big(\hat{m}_n^{(2)b}(\tau,\ell) - \hat{m}^{(2)}(\tau,\ell)\big) \\ \sqrt{n}\big(\hat{w}_n^{(1)b}(\tau,\ell) - \hat{w}^{(1)}(\ell)\big) \\ \sqrt{n}\big(\hat{w}_n^{(2)b}(\tau,\ell) - \hat{w}^{(2)}(\ell)\big) \end{pmatrix}. \qquad \text{(D.19)}
$$

By part 8 of Lemma D.2 of Bugni, Canay and Shi (2015), we have $\hat{\chi}^b_{P_n} \Rightarrow \Phi_{h_2}$ conditional on almost all sample paths. Next, by the same arguments for part (iii), we can show part (v).

To show part (vi), we have $\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} |\hat{h}_{2,\nu,P_{a_n}}((\tau,\ell),(\tau,\ell)) - h_{2,\nu}((\tau,\ell),(\tau,\ell))| \xrightarrow{p} 0$ from part (ii). By the fact that $\max\{a,\epsilon\}$ is a continuous function, it follows that

$$\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} |\max\{\hat{h}_{2,\nu,P_{a_n}}((\tau,\ell),(\tau,\ell)),\epsilon\} - \max\{h_{2,\nu}((\tau,\ell),(\tau,\ell)),\epsilon\}| \xrightarrow{p} 0, \tag{D.20}$$

so $\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} |\sigma^2_{\epsilon,a_n}(\tau,\ell) - \sigma^2_{\epsilon}(\tau,\ell)| \xrightarrow{p} 0$. Given that $\epsilon > 0$, we have $\sigma^2_\epsilon(\tau,\ell) \geq \epsilon > 0$ for all τ,ℓ. Hence, it follows that $\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} |\sigma^{-1}_{\epsilon,a_n}(\tau,\ell) - \sigma^{-1}_{\epsilon}(\tau,\ell)| \xrightarrow{p} 0$ and this shows part (vi).

Part (vii) follows from parts (iv) and (vi), and part (viii) follows from parts (v) and (vi). □

Proof of Lemma A.3. To show part (i), for $\iota > 0$, define $(\mathcal{TL})^\iota(P_c) = \{(\tau,\ell) : \nu_{P_c}(\tau,\ell) \geq -\iota \cdot \sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)\}$ and $(\mathcal{TL})^\iota(P_c)^c$ denote the complement of $(\mathcal{TL})^\iota(P_c)$. Note that by Lemma A.2(v)-(vi) and condition (b) of the present lemma, we have

$$\sup_{\tau\in\mathcal{T},\ell\in\mathcal{L}} \left| \frac{\hat{\nu}_n(\tau,\ell)}{\hat{\sigma}_{\epsilon,n}(\tau,\ell)} - \frac{\nu_{P_c}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)} \right| \xrightarrow{p} 0. \tag{D.21}$$

Then it follows that, with probability approaching one,

$$\sup_{(\tau,\ell)\in(\mathcal{TL})^\iota(P_c)^c} \frac{\hat{\nu}_n(\tau,\ell)}{\hat{\sigma}_{\epsilon,n}(\tau,\ell)} \leq -\iota/2. \tag{D.22}$$

This implies that

$$\int_{(\mathcal{TL})^\iota(P_c)^c} \max\Big\{\sqrt{n}\frac{\hat{\nu}_n(\tau,\ell)}{\hat{\sigma}_{\epsilon,n}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) = o_p(1). \tag{D.23}$$

Therefore,

$$\begin{aligned} \widehat{T}_n &= \int_{(\mathcal{TL})^\iota(P_c)} \max\Big\{\sqrt{n}\frac{\hat{\nu}_n(\tau,\ell)}{\hat{\sigma}_{\epsilon,n}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) + o_p(1) \\ &\leq \int_{(\mathcal{TL})^\iota(P_c)} \max\Big\{\sqrt{n}\frac{\hat{\nu}_n(\tau,\ell) - \nu_{P_n}(\tau,\ell) + \delta(\tau,\ell)/\sqrt{n}}{\hat{\sigma}_{\epsilon,n}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) + o_p(1) \end{aligned} \tag{D.24}$$

where the equality holds by the previous equation. The inequality holds because condition (a) holds and $\nu_{P_c}(\tau,\ell) \leq 0$ and $\max\{a^2, 0\}$ is non-decreasing in a. Therefore,

$$\limsup_{n\to\infty} P(\widehat{T}_n \leq t) \leq P\Big(\int_{(\mathcal{TL})^\iota(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell) + \delta(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) \leq a\Big). \tag{D.25}$$

Note that ι is any arbitrary positive number, so letting $\iota \to 0$ and using the facts that $\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}$ is a tight Gaussian process and that $\int_{(\mathcal{TL})^\iota(P_c)\setminus(\mathcal{TL})^o(P_c)} dQ(\tau,\ell) \to 0$, we have, for any $t \in R$,

$$\limsup_{n\to\infty} P(\widehat{T}_n \leq t) \leq P\Big(\int_{(\mathcal{TL})^o(P_c)} \max\Big\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell) + \delta(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\Big\}^2 dQ(\tau,\ell) \leq t\Big). \tag{D.26}$$

On the other hand, we have

$$\widehat{T}_n \geq \int_{(\mathcal{T}\mathcal{L})^o(P_c)} \max\left\{\sqrt{n}\frac{\hat{\nu}_n(\tau,\ell)}{\hat{\sigma}_{\epsilon,n}(\tau,\ell)}, 0\right\}^2 dQ(\tau,\ell). \tag{D.27}$$

It follows that, for all $t \in R$,

$$\liminf_{n\to\infty} P(\widehat{T}_n \leq t) \geq P\Big(\int_{(\mathcal{T}\mathcal{L})^o(P_c)} \max\left\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell) + \delta(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\right\}^2 dQ(\tau,\ell) \leq t\Big). \tag{D.28}$$

Equations (D.26) and (D.28) together imply that, for all $t \in R$,

$$\lim_{n\to\infty} P(\widehat{T}_n \leq t) = P\Big(\int_{(\mathcal{T}\mathcal{L})^o(P_c)} \max\left\{\frac{\Phi_{h_{2,\nu,P_c}}(\tau,\ell) + \delta(\tau,\ell)}{\sigma_{\epsilon,h_{2,\nu,P_c}}(\tau,\ell)}, 0\right\}^2 dQ(\tau,\ell) \leq t\Big). \tag{D.29}$$

This concludes the proof of part (i).

Part (ii) and part (iii) can be proved following the same steps, except one uses parts (vii) and (viii) of Lemma A.2 instead of (v) and (vi) of that lemma, and one eliminates $\delta(\tau,\ell)$ using Assumption 3.3. Details are omitted for brevity. □

References

Andrews, D. W. K., and X. Shi (2013a). "Inference Based on Conditional Moment Inequalities," *Econometrica*, **81**, 609–666.

——— (2013b). "Supplement to 'Inference Based on Conditional Moment Inequalities'," *Econometrica Supplemental Material*, **81**.

——— (2014). "Nonparametric Inference Based on Conditional Moment Inequalities," *Journal of Econometrics*, **179**, 31–45.

——— (2015). "Inference Based on Many Conditional Moment Inequalities," Working Paper.

Bugni, F. A., I. A. Canay, and X. Shi (2015). "Specification Tests for Partially Identified Models Defined by Moment Inequalities," *Journal of Econometrics*, **185**, 259–282.

Chetverikov, D. (2013). "Testing Regression Monotonicity in Econometric Models," Working Paper.

Davydov, Y. A., M. A. Lifshits, and N. V. Smorodina (1995). "Local Properties of Distributions of Stochastic Functionals," Translations of Mathematical Monographs Vol. 173. American Mathematical Society: Providence, Rhode Island.

Delgado, M. A., and J. C. Escanciano (2012). "Distribution-free Tests of Stochastic Monotonicity," *Journal of Econometrics*, **170**, 68–75.

Donald, S. G., and Y.-C. Hsu (2016). "Improving the Power of Tests of Stochastic Dominance," *Econometric Reviews*, **35**, 553–585.

Ghosal, S., A. Sen, and A. van der Vaart (2000). "Testing Monotonicity of Regression," *Annals of Statistics*, **28**, 1054–1082.

Gundersen, C., B. Kreider, and J. Pepper (2012). "The Impact of the National School Lunch Program on Child Health: A Nonparametric Bounsd Analysis," *Journal of Econometrics*, **166**, 79–91.

Hall, P., and N. E. Heckman (2000). "Testing for Monotonicity of a Regression Mean by Calibrating for Linear Functions," *Annals of Statistics*, **28**, 20–39.

Hsu, Y.-C. (2016). "Consistent Tests for Conditional Treatment Effects," Working paper.

Kitagawa, T. (2015). "A Test for Instrument Validity," *Econometrica*, **83**, 2043–2063.

Kreider, B. and S. C. Hill (2009). "Partially Identifying Treatment Effects with an Application to Covering the Uninsured," *Journal of Human Resources*, **44**, 409–449.

——— and J. V. Pepper (2007). "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of American Statistical Association*, **102**, 432–441.

Lee, S., O. Linton, and Y.-J. Whang (2009). "Testing for Stochastic Monotonicity," *Econometrica*, **77**, 585–602.

———, K. Song, and Y.-J. Whang (2016). "Testing for a General Class of Functional Inequalities," Working paper.

McCarthy, I., D. L. Millimet, and M. Roy (2015). "Bounding Treatment Effects: A Command for the Partial Identification of the Average Treatment Effect with Endogenous and Misreported Treatment Assignment," *Stata Journal*, **15**, 411–436.

Manski, C. F., and J. V. Pepper (2000). "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, **68**, 997–1010.

——— (2009). "More on Monotone Instrumental Variables," *Econometrics Journal*, **12**, S200–S216.

——— and E. Tamer (2002). "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, **70**, 519–546.

Pollard, D. (1990). *Empirical Processes: Theory and Application, CBMS Conference Series in Probability and Statistics*, Vol. 2. Hayward, CA: Institute of Mathematical Statistics.

Shen, S. (2015). "Estimation and Inference of Distributional Partial Effects: Theory and Applications," Working paper.

Seo, J. (2015). "Tests of Stochastic Monotonicity with Improved Size and Power Properties," Working paper.

van der Vaart, A. W., and J. A. Wellner (1996). *Weak Convergence and Empirical Processes: With Application to Statistics*, New York: Springer-Verlag.