Model Selection Tests for Conditional Moment Inequality Models

Yu-Chin Hsu[†]

Institute of Economics
Academia Sinica

Xiaoxia Shi[‡]

Department of Economics
University of Wisconsin at Madison

First version: December 15, 2011
This version: May 24, 2013

[†] ychsu@econ.sinica.edu.tw

[‡] xshi@ssc.wisc.edu.

Abstract

In this paper, we propose a Vuong (1989)-type model selection test for conditional moment inequality models. The test uses a new average generalized empirical likelihood (AGEL) criterion function designed to incorporate full restriction of the conditional model. We also introduce a new adjustment to the test statistic making it asymptotically pivotal whether the candidate models are nested or nonnested. The test uses simple standard normal critical value and is shown to be asymptotically similar, to be consistent against all fixed alternatives and to have nontrivial power against $n^{-1/2}$-local alternatives. Monte Carlo simulations demonstrate that the finite sample performance of the test is in accordance with the theoretical prediction.

JEL classification: C12, C52

Keywords: Asymptotic size, Model selection test, Conditional moment inequalities, Partial identification, Generalized empirical likelihood

1 Introduction

Conditional moment inequality (CMI) models have been increasingly recognized as a convenient and useful statistical formulation of many nonstandard economic programs. Such models are shown to arise naturally in models with missing data (e.g. Manski and Pepper (2000)), in games with multiple equilibria (e.g. Ciliberto and Tamer (2009)), in large scale dynamic games (e.g. Pakes, Porter, Ho and Ishii (2007)) and in differentiated product demand models with measurement error (Gandhi, Lu and Shi (2012)). Methods for parameter inference for such models have become abundant by now, however, there has been no method available to assist practitioners to choose between different CMI specifications. This paper fills in this gap by proposing a Vuong (1989)-type model selection test that allows one to choose between two CMI models according to their distance to the true data distribution. We show that the test has correct asymptotic size no matter the candidate models are nested, overlapping or strictly nonnested, is consistent against all fixed alternatives and has nontrivial $n^{-1/2}$-local power.

Our test is set up in a CMI context allowing for partial identification, but it is worth noting that we do not require partial identification. In fact, our test has general applicability to the problem of model selection among partially identified CMI models, point identified CMI models, partially identified conditional moment equality models and point identified conditional moment equality models, as well as selection across these four types of models. To the best of our knowledge, our test is the first model selection test available for any of these model selection testing scenarios.

From a technical point of view, this paper builds upon the empirical process arguments in Shi (2009a). However, the conditional models are substantially different from the unconditional models studied in Shi (2009a). In particular, we face two new challenges. First of all, the exponential tilting criterion function used in Shi (2009a) allows only finite number of moment inequalities and thus does not apply to the CMI models, where the CMI's imply infinite number of unconditional moment inequalities. Instead, we propose a new criterion function, namely the average generalized empirical likelihood (AGEL) criterion function. To form the AGEL function, we first transform the CMI's into equivalent infinite number of unconditional moment inequalities following Andrews and Shi (2013; AS hereafter), and then take a weighted average of the generalized empirical likelihood of the models defined by each (set of) unconditional moment inequality (-ies). The weighted average defines the AGEL function and preserves the full model restriction of the CMI model.

Secondly, an essential feature of the unconditional moment inequalities transformed from the CMI's is that many of the moment functions have variance arbitrarily close to zero. This creates new challenges in proving the convergence rate of the GEL nuisance parameters (which is a necessary step in deriving the asymptotic distribution of our test statistic) that is not faced in Shi (2009a). We deal with this by introducing a careful truncation of the weighted average.

One challenge that we share with Shi (2009a) is the possible degeneracy of the asymptotic distribution of the pseudo-likelihood ratio statistic. We propose a different solution than Shi (2009a). Specifically, we introduce a new adjustment to the pseudo-likelihood ratio statistic, making it asymptotically pivotal (in fact, standard normal) under the null regardless of the true data distribution or the relationship between the candidate models. As a result, our test simply uses the standard normal critical value and achieves asymptotic similarity. The adjustment simplifies the split sample idea of Yachew (1992) and achieves the same purpose as the latter.

Besides Shi (2009a), this paper is related to a large literature on model selection test following Vuong (1989). A literature review can be found in the introduction of Shi (2009a). We only note here that none of those papers deal with conditional moment (either equality or inequality) models. This paper is also related to the empirical likelihood approaches for parameter inference in conditional moment *equality* models proposed in Donald, Imbens and Newey (2003) and Kitamura, Tripathi and Ahn (2004). Our AGEL criterion function differs from both. We choose the AGEL criterion function primarily for tractability in the context of (potentially) partially identified conditional moment inequality models.

The rest of the paper is organized as follows. Section 2 presents the model selection problem of our interest and gives several motivating examples. Section 3 establishes the AGEL criterion function to measure the distance between the CMI models and the true distribution. The model selection test is proposed in Section 4. Section 5 summarizes the uniform asymptotic size property of our test and and Section 6 summarizes the power properties against fixed alternative and local alternatives. Section 7 discusses possible extensions in three different directions. Section 8 presents Monte-Carlo simulation results for a missing data example and Section 9 concludes. All mathematical proofs are contained in the Appendix.

2 Model Selection problems

We consider two conditional moment inequality/equality models $\mathcal{P}_1 = \bigcup_{\theta_1 \in \Theta_1} \mathcal{P}_{1,\theta_1}$ and $\mathcal{P}_2 = \bigcup_{\theta_2 \in \Theta_2} \mathcal{P}_{2,\theta_2}$, where $\mathcal{P}_{1,\theta_1}$ and $\mathcal{P}_{2,\theta_2}$ are the set of distributions that are consistent with the moment conditions for parameters θ_1 and θ_2, respectively:[1]

$$\mathcal{P}_{1,\theta_1} = \left\{ \begin{array}{l} P : E_P[m_{1,j}(W,\theta_1)|X] = 0 \text{ a.s. } [P_x] \text{ for } j = 1, ..., p_1, \\ \qquad E_P[m_{1,j}(W,\theta_1)|X] \geq 0 \text{ a.s. } [P_x] \text{ for } j = p_1 + 1, ..., k_1. \end{array} \right\}$$

$$\mathcal{P}_{2,\theta_2} = \left\{ \begin{array}{l} P : E_P[m_{2,j}(W,\theta_2)|X] = 0 \text{ a.s. } [P_x] \text{ for } j = 1, ..., p_2, \\ \qquad E_P[m_{2,j}(W,\theta_2)|X] \geq 0 \text{ a.s. } [P_x] \text{ for } j = p_2 + 1, ..., k_2. \end{array} \right\} \tag{2.1}$$

[1]We start with models with the same conditioning variables and later extend our theory to allow the conditioning variables to differ.

In the above equation, $\{W_i = (Y_i', X_i') \in \mathcal{W}\}_{i=1}^n$ is a random sample generated from P_0, a generic true distribution on $\mathcal{W}$. Also, $Y_i \in \mathcal{Y} \subseteq R^{d_y}$, $X_i \in \mathcal{X} \subseteq R^{d_x}$, $\mathcal{W} = \mathcal{Y} \times \mathcal{X}$. The notation E_P denotes the expectation under the distribution P and P_x the marginal distribution of X implied by P. The true distribution P_0 may or may not belong to either model. For $s = 1$ and 2, $m_s \equiv (m_{s,1}, ..., m_{s,p_s}, m_{s,p_s+1}, ..., m_{s,k_s})'$ are R^{k_s} -valued moment functions known up to the finite-dimensional parameters $\theta_s \in \Theta_s \subset R^{d_{\theta_s}}$. Model $\mathcal{P}_s$ is called **correctly specified** if $P_0 \in \mathcal{P}_s$ and is called **misspecified** otherwise. The parameters θ_s may or may not be point-identified.

The goal of this paper is to compare models $\mathcal{P}_1$ and $\mathcal{P}_2$ and select the one that is closer to the true distribution P_0 in terms of a pseudo-distance measure. Let $d_{\mathcal{L}}(\mathcal{P}_s, P_0)$ be the pseudo-distance measure that will be defined later. We want to construct model selection tests for the null hypothesis

$$H_0 : d_{\mathcal{L}}(\mathcal{P}_1, P_0) = d_{\mathcal{L}}(\mathcal{P}_2, P_0). \tag{2.2}$$

Now, we give a few illustrative examples of model selection problems in the context of conditional moment inequalities. The examples extend those in Shi (2009a).

Example 1 (Interval Outcome in Regression Models). Consider the regression models with interval outcomes from Manski (2005). It is of interest to select different regressors or functional forms for the regression functions. To be more specific, let Y be a latent random variable (e.g. wealth or income) that is not perfectly observed. Instead, we observe an upper bound and a lower bound on Y, say $\overline{Y}$ and $\underline{Y}$, respectively. For a vector of covariates X and a function $r_1(X, \theta_1)$ that is known up to a finite-dimensional parameter θ, let $Y = r_1(X, \theta_1) + \varepsilon$. Suppose Z is a vector of instrument variables such that $E(\varepsilon|Z) = 0$ a.s. $[P_z]$. Then, the model $\mathcal{P}_1$ is

$$\begin{aligned} \mathcal{P}_1 = \{P : & E_P[\overline{Y} - r_1(X, \theta_1)|Z] \geq 0 \ \text{ a.s. } [P_z] \ \text{ and} \\ & E_P[r_1(X, \theta_1) - \underline{Y}|Z] \geq 0 \ \text{ a.s. } [P_z], \quad \theta_1 \in \Theta_1.\} \end{aligned} \tag{2.3}$$

The distribution P are defined on the space of the *observed* random variables ($\overline{Y}$, $\underline{Y}$, X, Z). For model $\mathcal{P}_2$, the $r_1(X, \theta_1)$, θ_1 and Θ_1 in (2.3) are replaced with $r_2(X, \theta_2)$, θ_2 and Θ_2, respectively. A model selection test is to determine whether $\mathcal{P}_1$ or $\mathcal{P}_2$ is closer to the true distribution.

Example 2 (Interval Regressor in Regression Models). Consider the regression models with interval regressors from Manski (2005). Let Y be a dependent variable and X be a set of covariates. Also, let v be a regressor that is not observed perfectly but we observe an upper bound and a lower bound on v, say $\overline{v}$ and $\underline{v}$. Assume that $E(Y|X, v) = f(x, v, \theta)$, where f is a function known up to the finite-dimensional parameter θ. Manski (2005) assumes that f is weakly *increasing* in v, and obtains the following moment inequality model:

$$\mathcal{P}_1 = \{P : E_P[Y - f(x, \underline{v}, \theta_1)|X] \geq 0 \ \text{ a.s. } [P_x], \ \text{ and}$$

$$E_P[f(x,\overline{v},\theta_1) - Y|X] \geq 0 \;\; \text{a.s. } [P_x], \quad \theta_1 \in \Theta_1.\} \tag{2.4}$$

The distribution P are defined on the space of the *observed* random variables $(Y, \overline{v}, \underline{v}, X)$.

On the other hand, if we assume that f is weakly *decreasing* in v, we have a different moment inequality model:

$$\begin{aligned}\mathcal{P}_2 = \{P : &E_P[Y - f(x,\underline{v},\theta_2)|X] \leq 0 \;\; \text{and} \\ &E_P[f(x,\overline{v},\theta_2) - Y|X] \leq 0 \;\; \text{a.s. } [P_x], \quad \theta_2 \in \Theta_2.\}\end{aligned} \tag{2.5}$$

By comparing models $\mathcal{P}_1$ and $\mathcal{P}_2$, one can determine which model is closer to the true data distribution. If one has the prior information that one of the model is true, the test then helps one determine the sign of $\partial f / \partial v$.

Example 3 (Entry Game – Cross-firm Effect). Consider the entry game model from Andrews, Berry and Jia (2004) and Ciliberto and Tamer (2009). Consider a 2×2 entry game with the following payoff matrix:

		Firm 2: 0	Firm 2: 1
Firm 1	0	$0,\ 0$	$0,\ X_2'\beta_2 - \varepsilon_2$
	1	$X_1'\beta_1 - \varepsilon_1,\ 0$	$X_1\theta_1 + a_1 - \varepsilon_1,\ X_2\theta_2 + a_2 - \varepsilon_2$

The observable random variables are the market characteristics $X \equiv (X_1, X_2)'$ and the game outcome Y. The variable Y may take four values: $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$, where the first number in the parenthesis is the equilibrium action of firm 1 and the second number, the equilibrium action of firm 2. The coefficients β_1 and β_2 are the marginal effects of the characteristics X on profits, and ε_1 and ε_2 are the unobserved components of the firms' profits. The parameters a_1 and a_2 are the cross-firm effects, which are the effects of the firms on their opponents' profit when they are on the market at the same time.

Let $F_{\varepsilon_1,\varepsilon_2}(\cdot,\cdot;\theta_\varepsilon)$ denote the joint c.d.f. of ε_1 and ε_2, $F_{\varepsilon_1}(\cdot;\beta_\varepsilon)$ the marginal c.d.f. of ε_1, and $F_{\varepsilon_2}(\cdot;\theta_\varepsilon)$ the marginal c.d.f. of ε_2. The c.d.f.s are known to the econometrician up to the finite-dimensional parameter β_ε. Assume that the firms have full information about their own and their opponents' payoffs and play a simultaneous-move Nash game.

Andrews, Berry and Jia (2004) assume $a_1 \leq 0$ and $a_2 \leq 0$ and obtain the following moment inequality model:

$$\begin{aligned}\mathcal{P}_1 = \{P : &E_P[p_j(X,\theta_1) - 1(Y = j)|X] = 0, \text{ for } j = (0,0) \text{ or } (1,1), \\ &E_P[p_j(X,\theta_1) - 1(Y = j)|X] \geq 0, \text{ for } j = (0,1), \text{ or } (1,0) \;\; \text{a.s. } [P_x],\end{aligned}$$

$$\theta_1 \equiv (\beta_1', \beta_2', a_1, a_2, \beta_\varepsilon') \in \Theta_1.\} \tag{2.6}$$

where

$$\begin{aligned}
p_{(0,0)}(X, \theta_1) &= 1 - F_{\varepsilon_1}(X_1'\beta_1; \theta_\varepsilon) - F_{\varepsilon_2}(X_2'\beta_2; \beta_\varepsilon) + F_{\varepsilon_1,\varepsilon_2}(X_1'\theta_1, X_2'\beta_2; \theta_\varepsilon), \\
p_{(0,1)}(X, \theta_1) &= F_{\varepsilon_2}(X_2'\beta_2; \beta_\varepsilon) - F_{\varepsilon_1,\varepsilon_2}(X_1'\beta_1 + a_1, X_2'\beta_2; \beta_\varepsilon), \\
p_{(1,0)}(X, \theta_1) &= F_{\varepsilon_1}(X_1'\beta_1; \beta_\varepsilon) - F_{\varepsilon_1,\varepsilon_2}(X_1'\beta_1, X_2'\beta_2 + a_2; \beta_\varepsilon), \\
p_{(1,1)}(X, \theta_1) &= F_{\varepsilon_1,\varepsilon_2}(X_1'\beta_1 + a_1, X_2'\beta_2 + a_2; \beta_\varepsilon).
\end{aligned} \tag{2.7}$$

If we assume the cross-firm effects have different signs as in Andrews, Berry and Jia (2004), we obtain a different moment inequality model:

$$\begin{aligned}
\mathcal{P}_2 = \{P : &E_P[p_j(X, \theta_2) - 1(Y = j)|X] \geq 0, \text{ for } j = (0,0) \text{ or } (1,1), \\
&E_P[p_j(X, \theta_2) - 1(Y = j)|X] = 0, \text{ for } j = (0,1), \text{ or } (1,0) \text{ a.s. } [P_x], \\
&\theta_2 \equiv (\beta_1', \beta_2', a_1, a_2, \beta_\varepsilon') \in \Theta_2.\}
\end{aligned} \tag{2.8}$$

where p_j, $j = (0,0)$, $(1,1)$, $(0,1)$ and $(1,0)$ are defined in (2.7).

A model selection test comparing the two models can determine which sign of the cross-firm effects is more consistent with the data. Such test is useful especially when there are reasons to be unsure about the signs of a_1 and a_2 in some markets. For example, in a shopping center, one retail stores may worry about the other store stealing its business, but on the other hand may benefit from the casual shoppers that the other store attracts to the shopping center. The overall sign of the cross-firm effect then becomes an empirical question.

3 Preliminaries

3.1 Pseudo-distance Measure

To define our pseudo-distance measure, we first use AS's method to transform the conditional moment equality/inequality to infinitely many number of unconditional moment equalities/inequalities without loss of information. That is, we choose a collection of instrument functions $\mathcal{G} = \{g_\ell : \mathcal{X} \to [0,1] : \ell \in \mathcal{L}\}$ for an index set $\mathcal{L}$ to make sure, for $s = 1, 2$, and every $\theta_s \in \Theta_s$,

$$\mathcal{P}_{s,\theta_s} = \left\{ \begin{array}{ll} P: & E_P[m_{s,j}(W, \theta_s) g_\ell(X)] = 0 \text{ a.s. for } j = 1, \cdots, p_s \\ & E_P[m_{s,j}(W, \theta_s) g_\ell(X)] \geq 0 \text{ a.s. for } j = p_s + 1, \cdots, k_s, \text{ for all } \ell \in \mathcal{L}. \end{array} \right\} \tag{3.1}$$

AS give a list of $\mathcal{G}$ sets that can ensure that the above equation hold. Here, we impose an assumption on $\mathcal{X}$ and focus on two types of $\mathcal{G}$ for simplicity.

Assumption 3.1 *$\mathcal{X}$ is a Cartesian product of compact intervals, $\mathcal{X} = \prod_{j=1}^{d_x}[x_{\ell j}, x_{uj}]$ and without loss of generality, we assume $\mathcal{X} = \prod_{j=1}^{d_x}[0,1]$.*

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

$$\begin{aligned}
\mathcal{G}_{\text{c-cube}} &= \{g_\ell(\cdot) = 1(\cdot \in C_\ell) : \ell \equiv (x,r) \in \mathcal{L}_{\text{c-cube}}\}, \text{ where} \\
C_\ell &= \left(\times_{j=1}^{d_x}[x_j, x_j + r]\right) \cap \mathcal{X} \text{ and} \\
\mathcal{L}_{\text{c-cube}} &= \left\{(x, (2q)^{-1}) : \ 2q \cdot x \in \{0, 1, 2, \cdots, 2q-1\}^{d_x}, \text{ and } q = q_0, q_0+1, \cdots\right\},
\end{aligned} \tag{3.2}$$

where q_0 is a natural number. Note that for each q, $\{C_\ell : \ell \in \mathcal{L}_{\textit{c-cube}}$ and $r = (2q)^{-1}\}$ forms a partition of $\mathcal{X}$. The set $\mathcal{G}_{\text{c-cube}}$ is an example given in AS and Lemma 1 of AS guarantees that it satisfies equation (3.1)

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

$$\mathcal{G}_{\text{cube}} = \{g_\ell(\cdot) = 1(\cdot \in C_\ell) : \ell \in \mathcal{L}_{\text{cube}}\}, \text{ where } \mathcal{L}_{\text{cube}} = \{(x,r) : x \in [0, 1-r]^{d_x}, r \in (0, \bar{r}]\}, \tag{3.3}$$

for some $\bar{r} > 0$. The set $\mathcal{G}_{\text{cube}}$ is similar to AS's $\mathcal{G}_{\text{box}}$ except that all edges of a hypercube in $\mathcal{G}_{\text{cube}}$ are of the same length. Note that $\mathcal{G}_{\text{c-cube}}$ is a subset of $\mathcal{G}_{\text{cube}}$. Therefore, $\mathcal{G}_{\text{cube}}$ also guarantees (3.1)

We will define the pseudo-distance from $\mathcal{P}_s$ to P_0 to be the infimum of the pseudo-distance from $\mathcal{P}_{s,\theta_s}$ to P_0 over $\theta_s \in \Theta_s$. Thus, we need to define the latter distance first. To do so, we first define the ℓ-th supermodel of $\mathcal{P}_{s,\theta_s}$ as

$$\mathcal{P}_{s,\theta_s,\ell} = \left\{ \begin{array}{ll} P: & E_P[m_{s,j}(W,\theta_s)g_\ell(X)] = 0 \text{ a.s. for } j = 1, \cdots, p_s \\ & E_P[m_{s,j}(W,\theta_s)g_\ell(X)] \geq 0 \text{ a.s. for } j = p_s+1, \cdots, k_s, \end{array} \right\} \tag{3.4}$$

for each $\ell \in \mathcal{L}$. The term "supermodel" is used to indicate the fact that $\mathcal{P}_{s,\theta_s} \subseteq \mathcal{P}_{s,\theta_s,\ell}$. Then, for each $\theta_s \in \Theta_s$ and each $\ell \in \mathcal{L}$, $\mathcal{P}_{s,\theta_s,\ell}$ is an unconditional moment inequality model covered in Shi (2009a).

Like in Shi (2009a), we can define the GEL distance from $\mathcal{P}_{s,\theta_s,\ell}$ to P_0

$$d(\mathcal{P}_{s,\theta_s,\ell}, P_0) = \Psi\left(\sup_{\gamma_s \in \Gamma_s(\theta_s)} E_{P_0}\left[\kappa\left(\gamma_s' m_s(X,\theta_s) g_\ell(X)\right)\right]\right), \tag{3.5}$$

where $\kappa(\cdot) : \mathcal{K} \to R$ is a strictly concave function defined on a subset $\mathcal{K}$ of R, $\Psi(\cdot) : R \to R$ is a strictly increasing function and $\Gamma_s(\theta_s) = \{\gamma_s \in R^{p_s} \times R_+^{k_s - p_s} : \gamma_s' m_s(x,\theta_s) \in \mathcal{K} \text{ for all } x \in \mathcal{X}\}$. The functions $\kappa(\cdot)$ and $\Psi(\cdot)$ are user chosen and determine which of the GEL distances one is using. The common choices of the GEL distances include the empirical likelihood (EL), the exponential tilting (ET) and the continuous updating GMM (CUE), which corresponds to ($\kappa(y) = \log(1-y)$, $\mathcal{K} = (-\infty, 1)$, $\Psi(\kappa) = \kappa$), ($\kappa(y) = 1 - e^y$, $\mathcal{K} = R$, $\Psi(\kappa) = -\log(1-\kappa)$), and ($\kappa(y) = (1-(y+1)^2)/2$, $\mathcal{K} =$

R, $\Psi(\kappa) = \kappa$), respectively.[2] Other choices can be used as well, as long as the following assumption is satisfied:

Assumption 3.2 (i) $\kappa(\cdot)$ *is strictly concave, three times continuously differentiable with* $\kappa(0) = 0$, *and* $\kappa'(0) = \kappa''(0) = -1$.
(ii) $\Psi(\cdot)$ *is strictly increasing and is twice continuously differentiable with* $\Psi(0) = 0$ *and* $\Psi'(0) = 1$.

We then define the pseudo-distance from $\mathcal{P}_{s,\theta_s}$ to P_0 to be a weighted average of $d(\mathcal{P}_{s,\theta_s,\ell}, P_0)$ across $\ell \in \mathcal{L}$:

$$d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0) = \int_{\mathcal{L}} d(\mathcal{P}_{s,\theta_s,\ell}, P_0) dF(\ell), \tag{3.6}$$

where $F(\ell)$ is a probability measure whose support contains $\mathcal{L}$. Because the new pseudo-distance is an average of the GEL distances, we refer to it by the average generalized empirical likelihood distance (AGEL). Finally, we define the distance from $\mathcal{P}_s$ to P_0 as

$$d_{\mathcal{L}}(\mathcal{P}_s, P_0) = \inf_{\theta_s \in \Theta_s} d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0) = \inf_{\theta_s \in \Theta_s} \int_{\mathcal{L}} d(\mathcal{P}_{s,\theta_s,\ell}, P_0) dF(\ell). \tag{3.7}$$

Note that different choices of $F(\ell)$, $\kappa(\cdot)$ or $\Psi(\cdot)$ will produce different pseudo-distance measures. If both $\mathcal{P}_1$ and $\mathcal{P}_2$ are misspecified, different choices of the pseudo-distance measures do not necessarily agree on which model is closer to P_0. On the other hand, in Lemma A.6 in the appendix, we show that $d_{\mathcal{L}}(\mathcal{P}_s, P_0) \geq 0$ and $d_{\mathcal{L}}(\mathcal{P}_s, P_0) = 0$ iff $P_0 \in \mathcal{P}_s$. This holds for general choices of $F(\ell)$, $\kappa(\cdot)$ or $\Psi(\cdot)$. This implies that if $P_0 \in \mathcal{P}_1$ and $P_0 \notin \mathcal{P}_2$, then $0 = d_{\mathcal{L}}(\mathcal{P}_1, P_0) < d_{\mathcal{L}}(\mathcal{P}_2, P_0)$, i.e. $\mathcal{P}_1$ is closer to P_0 than $\mathcal{P}_2$ no matter what $F(\ell)$, $\kappa(\cdot)$ or $\Psi(\cdot)$ to use. This is an important property because it means that the model selection test can determine (up to statistical error) which model is correctly specified when one has the prior information that one of them is.

3.2 A Uniqueness Assumption

Next we introduce a uniqueness assumption that extends the unique pseudo-true distribution assumption in Shi (2009a). This assumption allows the model selection test to be of a simple form.

Let the optimal value of Lagrange multiplier γ for each ℓ and each θ_s be

$$\gamma^*_{s,\ell,P_0}(\theta_s) = \arg \max_{\gamma_s \in \Gamma_s(\theta_s)} E_{P_0}\left[\kappa\left(\gamma_s' m_s(W, \theta_s) g_\ell(X)\right)\right]. \tag{3.8}$$

Assumptions that we impose later for our main results will guarantee that $\gamma^*_{s,\ell,P_0}(\theta_s)$ exists (that is, is finite) and is unique, for every $\ell \in \mathcal{L}$ and $\theta_s \in \Theta_s$. With $\gamma^*_{s,\ell}(\theta_s, P_0)$ defined, we can write

$$d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0) = \int_{\mathcal{L}} \kappa\left(\gamma^{*\prime}_{s,\ell,P_0}(\theta_s) m_s(W, \theta_s) g_\ell(X)\right) dF(\ell). \tag{3.9}$$

[2] In the CUE case, the distance measure is equivalent to the the GMM criterion function with the continuously updating weighting matrix being the inverse of the non-recentered covariance matrix.

Our assumptions will also guarantee that $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0)$ is continuous in $\theta_s \in \Theta_s$ and that Θ_s is compact. Given those, the following set is well defined:

$$\Theta_s^*(P_0) = \arg\min_{\theta_s \in \Theta_s} d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0). \tag{3.10}$$

We call this set the *Pseudo-true Set* as an analogue to the "pseudo-true value" in potentially misspecified point identified models. If the model is correctly specified, that is, if $P_0 \in \mathcal{P}_s$, then $\Theta_s^*(P_0)$ is the identified set. One of the important features of conditional moment inequality model is that their identified set can contain more than one point. We respect this feature and allow (but do not require) $\Theta_s^*(P_0)$ to contain more than one value.

While we allow the pseudo-true set to be multi-valued, one uniqueness condition is needed to give the model selection test that we propose in the next section a simple form. This uniqueness condition is stated below:

Assumption 3.3 $\gamma^*_{s,\ell,P_0}(\theta_s)' m_s(W,\theta_s) g_\ell(X) = \gamma^*_{s,\ell,P_0}(\theta_s^*)' m_s(W,\theta_s^*) g_\ell(X)$ *a.s.* $[P_0]$ *for all* $\ell \in \mathcal{L}$ *and* $\theta_s, \theta_s^* \in \Theta_s^*(P_0)$.

Assumption 3.3 is similar to and serves the same purpose as the unique pseudo-true distribution assumption in Shi (2009a), although it is difficult to give it a pseudo-true distribution interpretation in the conditional models here. Like the unique pseudo-true distribution assumption, Assumption 3.3 is automatically satisfied if the model is correctly specified (still partially identified) because $\gamma^*_{s,\ell,P_0}(\theta_s) = 0$ for all $\theta_s \in \Theta_s^*(P_0)$ and for all $\ell \in \mathcal{L}$ in that case. It is also automatically satisfied when (a) the model is misspecified and $\Theta_s^*(P_0)$ is a singleton, and (b) $\Theta_s^*(P_0)$ is not a singleton, but we can reparametrize the model so that the new parameter has a unique pseudo-true value.

Because Assumption 3.3 is automatically satisfied in the above cases, it is innocuous in the following important model selection testing scenarios:

- conventional nested testing, in which the correct specification of the bigger (less restrictive) model is maintained,

- nonnested testing in which one has the prior knowledge that one of the models is correctly specified,

- nested or nonnested testing in which the conditional moment inequality models are point identified. Point identified CMI models are important special cases of CMI models, and their use has been studied in Moon and Schorfheide (2009), among others, and

- nested or nonnested testing for point identified conditional moment equality models.

In the cases other than those, Assumption 3.3 can be restrictive, and we discuss a way to relax this assumption in Section 7.

4 Model Selection Test

In this section, we introduce our model selection test. The test is based on the pseudo-likelihood ratio statistic and uses a standard normal critical value. We introduce both a 2-sided version and a 1-sided version of the test.

The pseudo-likelihood ratio statistic is the sample analogue estimator of $LR_{P_0} := d_{\mathcal{L}}(\mathcal{P}_1, P_0) - d_{\mathcal{L}}(\mathcal{P}_2, P_0)$: [3]

$$\widehat{LR}_n = \hat{d}_{\mathcal{L}}(\mathcal{P}_1, P_0) - \hat{d}_{\mathcal{L}}(\mathcal{P}_2, P_0) \tag{4.1}$$

where

$$\hat{d}_{\mathcal{L}}(\mathcal{P}_s, P_0) = \min_{\theta_s \in \Theta_s} \int_{\mathcal{L}_{r_n}} \Psi \left[n^{-1} \sum_{i=1}^{n} \kappa(\hat{\gamma}_{s,\ell,n}(\theta_s)' m_s(W_i, \theta_s) g_\ell(X_i)) \right] dF(\ell) \tag{4.2}$$

with $\mathcal{L}_{r_n} = \{\ell \in \mathcal{L} : r \geq r_n\}$ for r_n being a positive sequence that converges to zero as $n \to 0$, and

$$\hat{\gamma}_{s,\ell,n}(\theta_s) = \arg \min_{\gamma_s \in R^{p_s} \times R_+^{k_s - p_s}} n^{-1} \sum_{i=1}^{n} \kappa(\gamma_s' m_s(W_i, \theta_s) g_\ell(X_i)). \tag{4.3}$$

Notice that in $\hat{d}_{\mathcal{L}}(\mathcal{P}_s, P_0)$, the integrations is over $\mathcal{L}_{r_n}$ instead of $\mathcal{L}$. We use the trimming argument to control the estimation accuracy of $\hat{\gamma}_{s,\ell,n}(\cdot)$ when the last element of ℓ is small. We pick r_n to balance the estimation accuracy of $\hat{\gamma}_n$ and the approximation quality of $d_{\mathcal{L}_{r_n}}(\mathcal{P}_s, P_0)$ for $d_{\mathcal{L}}(\mathcal{P}_s, P_0)$.

The pseudo-likelihood ratio statistic $\widehat{LR}_n$ can be shown to satisfy $n^{1/2}(\widehat{LR}_n - LR_{P_0}) \to_d N(0, \omega^2_{P_0})$ under regularity conditions for

$$\begin{aligned}
&\omega^2_{P_0} = E_{P_0}(\Lambda^*_{P_0,i})^2, \text{ where} \\
&\Lambda^*_{P_0,i} = \int_{\mathcal{L}} \Big\{ \Psi'(M^*_{1,\ell,P_0}) \big[\kappa\big(\gamma^*_{1,\ell,P_0}(\theta^*_1)' m_1(W, \theta^*_1) g_\ell(X)\big) - M^*_{1,\ell,P_0} \big] \\
&\qquad\qquad - \Psi'(M^*_{2,\ell,P_0}) \big[\kappa\big(\gamma^*_{2,\ell,P_0}(\theta^*_2)' m_2(W, \theta^*_2) g_\ell(X)\big) - M^*_{2,\ell,P_0} \big] \Big\} dF(\ell), \text{ for} \\
&M^*_{s,\ell,P_0} = E_{P_0}\big[\kappa\big(\gamma^*_{s,\ell,P_0}(\theta^*_s)' m_s(W, \theta^*_s) g_\ell(X)\big) \big], \text{ and} \\
&\theta^*_s \in \Theta^*_s(P_0), \quad \text{for } s = 1 \text{ and } 2.
\end{aligned} \tag{4.4}$$

This weak convergence result can be used to build a hypothesis test when two additional problems are solved.

The first problem is standard – the asymptotic variance $\omega^2_{P_0}$ needs to be estimated and used to studentize $\widehat{LR}_n$. To estimate $\omega^2_{P_0}$, we use $\hat{\omega}^2_n = \sup_{\theta_s \in \widehat{\Theta}_{s,n},\ s=1,2} \hat{\omega}^2_n(\theta_1, \theta_2)$, where

$$\widehat{\Theta}_{s,n} = \arg \min_{\theta_s \in \Theta_s} \int_{\mathcal{L}_{r_n}} \Psi[\widehat{M}_{s,\ell,n}(\theta_s)] dF(\ell), \text{ and}$$

[3]Notice that here the population "pseudo"-likelihood ratio has the opposite interpretation as the likelihood ratio in a parametric model in that here $LR_{P_0} > 0$ means Model 2 is better (closer to P_0) than Model 1.

$$\hat{\omega}_n^2(\theta_1,\theta_2) = \frac{1}{n}\sum_{i=1}^{n}\Big[\int_{\mathcal{L}_{r_n}} \Psi'(\widehat{M}_{1,\ell,n}(\theta_1))\big[\kappa\big(\hat{\gamma}_{1,\ell,n}(\theta_1)'m_1(W_i,\theta_1)g_\ell(X_i)\big) - \widehat{M}_{1,\ell,n}(\theta_1)\big]$$
$$- \Psi'(\widehat{M}_{2,\ell,n}(\theta_2))\big[\kappa\big(\hat{\gamma}_{2,\ell,n}(\theta_2)'m_2(W_i,\theta_2)g_\ell(X_i)\big) - \widehat{M}_{2,\ell,n}(\theta_2)\big]\Big\}dF(\ell)\Big]^2, \text{ for}$$
$$\widehat{M}_{s,\ell,n}(\theta_s) = n^{-1}\sum_{i=1}^{n}\kappa(\hat{\gamma}_{s,\ell,n}(\theta_s)'m_s(W_i,\theta_s)g_\ell(X_i)), \text{ for } s=1,2. \tag{4.5}$$

The set $\widehat{\Theta}_{s,n}$ is not necessarily singleton, which is why we define $\hat{\omega}_n^2$ to be a supremum over points in $\widehat{\Theta}_{1,n}\times\widehat{\Theta}_{2,n}$. However, we note that in theory, we can use $\hat{\omega}_n^2(\hat{\theta}_{1,n},\hat{\theta}_{1,n})$ as $\hat{\omega}_n^2$ for any $\hat{\theta}_{s,n}\in\widehat{\Theta}_{s,n}$. In practice, different points $(\theta_1',\theta_2')'\in\widehat{\Theta}_{1,n}\times\widehat{\Theta}_{2,n}$ typically produce the same $\hat{\omega}_n^2(\theta_1,\theta_2)$.

The second and trickier problem is the possibility that $\omega_{P_0}^2 = 0$. This possibility happens when $\Lambda_0^* = 0$ $a.s.$, in particular, when both models are correctly specified. When it happens, the noise in the estimation of $\Theta_s^*(P_0)$ will dominate and cause the studentized $\widehat{LR}_n$ not to have a simple asymptotic distribution. This problem is the same as that studied in Shi (2009b) in regular point-identified models, but unfortunately does not have a neat solution as that in Shi (2009b) due to the partial identification and the inequality constraints.

Our solution to this problem is to introduce some extra randomness to $\widehat{LR}_n$ so that the noise in the estimation of $\Theta_s^*(P_0)$ will be dominated by this extra randomness when $\omega_{P_0}^2 = 0$.[4] Specifically, we introduce an auxiliary random variable, $U\sim N(0,1)$, which is independent from the original sample, and let our test statistic be

$$\widehat{T}_n = (\hat{\omega}_n^2+\hat{\sigma}_n^2)^{-1/2}(\sqrt{n}\widehat{LR}_n + \hat{\sigma}_n U), \tag{4.6}$$

where $\hat{\sigma}_n$ is a data-dependent scalar that is asymptotically independent of $\sqrt{n}(\widehat{LR}_n - LR_{P_0})$, $\hat{\omega}_n$ and U. The scalar $\hat{\sigma}_n$ should be bounded away from zero in probability when $\omega_{P_0}^2 = 0$ and converges to zero when $\omega_{P_0}^2 > 0$. A suitable choice is given in at the end of Section 5 below.

In the next section, we show that under H_0 and regularity conditions and with suitable choices of $\hat{\sigma}_n$, $\widehat{T}_n \to_d N(0,1)$. Thus, our model selection test uses the $N(0,1)$ quantile as critical value. Specifically, our *two-sided model selection test* of level α is defined as

$$\varphi_n^{\text{2-sided}}(\alpha) = 1\{|\widehat{T}_n| > z_{\alpha/2}\}, \tag{4.7}$$

where $z_{\alpha/2}$ is the $1-\alpha/2$ quantile of $N(0,1)$. We select Model 1 if $\varphi_n^{\text{2-sided}}(\alpha)=1$ and $\widehat{T}_n > 0$ and select Model 2 if $\varphi_n^{\text{2-sided}}(\alpha)=1$ and $\widehat{T}_n<0$. Our *one-sided model selection test* of level α (for H_0

[4]The use of extra randomness works in a very similar fashion as the sample-splitting technique used in Yachew (1992), but instead of implicitly add noise to the test statistic by sample-splitting, we add the noise explicitly. The advantage of our approach is that the amount of noise added can be easily controlled and in particular can be made to vanish with sample size when it is not needed, that is when $\omega_{P_0}^2 > 0$.

vs. $H_1 : d_{\mathcal{L}}(\mathcal{P}_1, P_0) - d_{\mathcal{L}}(\mathcal{P}_2, P_0) > 0$) is defined as

$$\varphi_n^{\text{1-sided}}(\alpha) = 1\{\widehat{T}_n > z_\alpha\}. \tag{4.8}$$

The one-sided test should be used when one has prior knowledge that $LR_{P_0} \geq 0$. One notable example where such prior knowledge is available is when it is known that Model $\mathcal{P}_2$ nests Model $\mathcal{P}_1$.

5 Asymptotic Size

In this section, we show that the model selection test we proposed above has correct asymptotic size, that is

$$AsySZ(\alpha) := \limsup_{n\to\infty} SZ_n(\alpha) := \limsup_{n\to\infty} \sup_{P\in\mathcal{F}_0} E_P\varphi_n(\alpha) = \alpha, \tag{5.1}$$

where $\varphi_n(\alpha) = \varphi_n^{\text{2-sided}}(\alpha)$ or $\varphi_n^{\text{1-sided}}(\alpha)$ and $\mathcal{F}_0$ is the set of true data distributions under which H_0 and regularity conditions hold and is defined in Assumption 5.3 below. In fact, we will show that our test is asymptotically similar, that is, not only $AsySZ(\alpha) = \alpha$ but also

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{F}_0} E_P\varphi_n(\alpha) = \alpha. \tag{5.2}$$

To begin, we need additional assumptions. The first assumption is on the parameter space and the moment functions:

Assumption 5.1 *For $s = 1, 2$, assume that:*
(i) *Θ_s is compact, and*
(ii) *for all $w \in \mathcal{W}$, $m_s(w, \theta_s)$ is three times continuously differentiable in θ_s.*

Next, we impose regularity conditions on the data generating process. These conditions will specify the set $\mathcal{F}_0$ on which the asymptotic size is defined. To introduce the regularity conditions, some additional notations are needed. Let

$$\mathcal{M}_{s,\ell,P_0}(\gamma_s, \theta_s) = E_{P_0}\kappa(\gamma_s' m(W_i, \theta_s) g_\ell(X_i)). \tag{5.3}$$

Let $eig_{\max}(A)$ denote the biggest eigenvalue of a matrix A. For a positive number M, let Γ_M^s denote $N_M(0_{k_s}) \cap (R^{p_s} \times R_+^{k_s - p_s})$, where $N_M(0_{k_s})$ is a closed ball in R^{k_s} centered at the origin with radius M. Let $\phi_s = (\gamma_s', \theta_s')'$. Let "$\wedge$" and "$\vee$" denote the minimum and the maximum operator, respectively. Let $N_\varepsilon(\Theta_s^*(P_0)) = \bigcup_{\theta_s \in \Theta_s^*(P_0)} N_\varepsilon(\theta_s)$. Let $p_{P_0,\ell} = E_{P_0}(g_\ell(X))$ for all $\ell \in \mathcal{L}$, which is the probability of X in C_ℓ under P_0. Finally, let the *left* Hausdorff distance from a subset A_1 of a Euclidean space to another, A_2, be defined as

$$\rho_{lh}(A_1, A_2) = \sup_{a\in A_1} \inf_{a'\in A_2} ||a - a'||. \tag{5.4}$$

We call it the *left* Hausdorff distance because the symmetrized version of ρ_{lh} is the Hausdorff distance: $\rho_h(A_1, A_2) = \max\{\rho_{lh}(A_1, A_2), \rho_{lh}(A_2, A_1)\}$.

Assumption 5.2 summarizes the regularity conditions that we hold as the maintained hypothesis, and Assumption 5.3 defines the null hypothesis.

Assumption 5.2 *For positive constants M and δ, the set $\mathcal{F}$ is the set of P_0 such that for $s = 1, 2$, Assumption 3.3 holds, and*

(i) $\{W_i\}_{i=1}^n$ *is an i.i.d. sample drawn from P_0,*

(ii) $d_{\mathcal{L}}(\mathcal{P}_s, P_0) - d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0) < -\delta \cdot (\rho_{lh}^2(\theta_s, \Theta_s^*(P_0)) \wedge \delta)$,

(iii) $\gamma_s' m_s(w, \theta_s) \in \mathcal{K}$ *for all $w \in \mathcal{W}$ and for all $\phi_s \in \Gamma_M^s \times \Theta_s$,*

(iv) $\sup_{\theta_s \in \Theta_s, \ell \in \mathcal{L}} \|\gamma_{s,\ell,P_0}^*(\theta_s)\| \leq M - \delta$,

(v) $\inf_{\phi_s \in \Gamma_M^s \times \Theta_s} eig_{\max}\Big(E_{P_0}\Big[\frac{\partial^2 \kappa(\gamma_s' m_s(W,\theta_s))}{\partial \gamma_s \partial \gamma_s'} \Big| X\Big]\Big) < -\delta$, *a.s.* $[P_{0,x}]$

(vi) *almost surely in* $[P_{0,x}]$,

$$E_{P_0}\Bigg[\sup_{\phi_s \in \Gamma_M^s \times \Theta_s} \Big\{ |\kappa(\gamma_s' m_s(W, \theta_s))|^{2+\delta} + \Big\|\frac{\partial \kappa(\gamma_s' m_s(W, \theta_s))}{\partial \phi_s}\Big\|^{2+\delta} + \Big\|\frac{\partial^2 \kappa(\gamma_s' m_s(W, \theta_s))}{\partial \phi_s \partial \phi_s}\Big\|^{2+\delta} + \sum_{j=1}^{k_s} \Big\|\frac{\partial^3 \kappa(\gamma_s' m_s(W, \theta_s))}{\partial \gamma_{s,j} \partial \gamma_s \partial \phi_s'}\Big\| \Big\} \Big| X\Bigg] < M$$

(vii) $E_{P_0}(\omega_{P_0}^{-1} \Lambda_{P_0,i}^*)^{2+\delta} < M$, *if* $\omega_{P_0} > 0$, *and*

(viii) $P_{0,x}$ *is absolutely continuous on $\mathcal{X}$ w.r.t the Lebesgue measure with density $f(x)$ such that $\delta \leq f(x) \leq M$ for all $x \in \mathcal{X}$.*

Assumption 5.3 *The set* $\mathcal{F}_0 = \{P_0 \in \mathcal{F} : d_{\mathcal{L}}(\mathcal{P}_1, P_0) = d_{\mathcal{L}}(\mathcal{P}_2, P_0)\}$.

Assumption 5.2(ii) is a global identifictaion condition, which can be weakened at the expense of more stringent condition on the adjustment factor $\hat{\sigma}_n$. Assumption 5.2(iv) basically requires the models to be not too misspecified.[5] Assumption 5.2(v) is a full-rank condition, which is needed for $\gamma_{s,\ell,P_0}^*(\theta_s)$ to be uniquely defined. This assumption is standard in the GEL literature, although it is not standard in the moment inequality literature. Assumption 5.2(viii) will make sure that p_{ℓ,P_0} is proportional to the volume of C_ℓ, which is useful for us to characterize the effect of the truncation. Assumption 5.2(viii) requires that all covariates be continuous. However, at the expense of additional

[5]This does not mean that we do not allow for global misspecification, but rather means that the global misspecification cannot be unbounded.

notation, we can deal with the case where X has both continuous and discrete components. We present two ways to do so in Section 7.

The next assumption is on the measure on $\mathcal{L}$ used in the integral, and on the truncation of $\mathcal{L}$. In the assumption, L^c for a subset L of $\mathcal{L}$ denote the complement of L relative to $\mathcal{L}$.

Assumption 5.4 (i) *$\sqrt{n}r_n^{d_x+1} \to 0$, $\sqrt{n}r_n^{d_x} \to \infty$ as $n \to \infty$, and*

(ii) *The support of $F(\ell)$ is $\mathcal{L}$ and $\int_{\mathcal{L}_{r_n}^c} dF(\ell) = O(r_n)$.*

For $\mathcal{L}_{\text{cube}}$, any $F(\ell)$ with bounded density satisfies Assumption 5.4(ii). For $\mathcal{L}_{\text{c-cube}}$, one $F(\ell)$ that satisfies Assumption 5.4(ii) has a probability mass function: $f(x,r) \propto r^3$, where $\propto$ stands for "is proportional to".[6]

Under the assumptions above, we can characterize the asymptotic behavior of $\widehat{LR}_n$ and $\hat{\omega}_n$. The result is summarized in the following lemma. In the lemma, $t_n = (n^{1/2}r_n^{d_x+1}) \vee (n^{-1/2}r_n^{-d_x})$. By the assumption above, $t_n \to 0$ as $n \to \infty$.

Lemma 5.1 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2 *and* 5.4 *hold. Then for any subsequence $\{a_n\}$ of $\{n\}$ and any sequences $\{P_{a_n} \in \mathcal{F}\}$ such that $t_{a_n}^{-2}\omega_{P_{a_n}}^2 \to v_\infty \in [0,\infty]$,*
(a) *if $v_\infty \in [0,\infty)$, then $t_{a_n}^{-1}\sqrt{a_n}(\widehat{LR}_{a_n} - LR_{P_{a_n}}) = O_p(1)$, $t_{a_n}^{-2}\hat{\omega}_{a_n}^2 \to_p v_\infty$, and*
(b) *if $v_\infty = \infty$, $\sqrt{a_n}(\widehat{LR}_{a_n} - LR_{P_{a_n}})/\omega_{P_{a_n}} \to_d N(0,1)$ and $\hat{\omega}_{a_n}^2/\omega_{P_{a_n}}^2 \xrightarrow{p} 1$.*

From Lemma 5.1, we see that the asymptotic behavior of $\widehat{LR}_n$ and $\hat{\omega}_n^2$ depend on the speed at which $\omega_{P_{a_n}}^2$ converges to zero. These different behaviors suggest the conditions that extra noise introduced in $\widehat{T}_n$ should satisfy. In particular, when $\omega_{P_{a_n}}^2$ converges to zero faster than $t_{a_n}^2$, we can only derive rate at which $\sqrt{a_n}(\widehat{LR}_{a_n} - LR_{P_{a_n}})$ and $\hat{\omega}_{a_n}$ converge to zero. In this case, we would like $\hat{\sigma}_{a_n}$ to dominate these terms so that we will still have $\hat{T}_{a_n}$ converging to the $N(0,1)$ distribution. The assumption below summarizes the requirement on $\hat{\sigma}_n$:

Assumption 5.5 (i) *$\hat{\sigma}_n \in [0,1]$ and is independent of U, and for any subsequence $\{a_n\}$ of n and any sequence $\{P_{a_n} \in \mathcal{F}\}_{n\geq 1}$ such that $t_{a_n}^{-2}\omega_{P_{a_n}}^2 \to v_\infty \in [0,\infty]$,*

(ii) *if $v_\infty \in [0,\infty)$, we have $\hat{\sigma}_{a_n} \to_p 1$, and*

(iii) *if $v_\infty = \infty$, we have $\Pr_{P_{a_n}}(\sqrt{a_n}(\widehat{LR} - LR_{P_{a_n}})/\omega_{P_{a_n}} \leq x|\hat{\sigma}_{a_n}) \to_p \Phi(x)$ for all $x \in R$ and $\Pr_{P_{a_n}}(|\omega_{P_{a_n}}/\hat{\omega}_{a_n} - 1| > \varepsilon|\hat{\sigma}_{a_n}) \to_p 0$ for all $\varepsilon > 0$, where $\Phi(x)$ is the cdf of $N(0,1)$.*

Now, we are ready to show that the model selection test we propose in the previous section is asymptotically similar: that is, for any sequences $\{P_n \in \mathcal{F}_0\}$, $\lim_{n\to\infty} E_{P_n}\varphi_n(\alpha) = \alpha$ for $\varphi_n(\alpha) = \varphi_n^{\text{2-sided}}(\alpha)$ or $\varphi_n^{\text{1-sided}}(\alpha)$.

[6] With such an $F(\ell)$, one can show that $\int_{\mathcal{L}_{r_n}} dF(\ell) \propto 1 - \frac{2r_n^{-1}(2r_n^{-1}-1)}{(2r_n^{-1}+1)^2} \propto r_n^{-1}$, where the first $\propto$ follows from Cauchy's proof for the Basel problem.

Theorem 5.1 *Suppose Assumptions* 3.1*,* 3.2*,* 5.1*,* 5.2*,* 5.3*,* 5.4 *and* 5.5 *hold. Then for* $\varphi_n(\alpha) = \varphi_n^{\text{2-sided}}(\alpha)$ *or* $\varphi_n^{\text{1-sided}}(\alpha)$ (a) *equation* (5.1) *holds, and* (b) *equation* (5.2) *holds.*

We conclude this section by giving a data-dependent choice of $\hat{\sigma}_n$ that satisfies Assumption 5.5. For a positive integer b_n, let $\hat{\omega}^*_{b_n}$ be estimated according to (4.5) based on a subsample of size b_n. To be more specific, we first re-estimate $\gamma^*_{s,\ell,P_0}(\theta_s)$ according to (4.3) using the subsample. Then re-estimate $\Theta_s(P_0)$ according to the first line of (4.5) with the r_n replaced with r_{b_n} accordingly. Finally, compute $\hat{\omega}^*_{b_n}$ according to the second line of (4.5) with those re-estimated parameters and using the subsample. We define

$$\hat{\sigma}_n = \exp(-t_{b_n}^{-2}(\hat{\omega}^*_{b_n})^2). \tag{5.5}$$

The following lemma shows that $\hat{\sigma}_n$ defined this way satisfies Assumption 5.5 if the subsample size grows with n but at a slower rate than n.

Lemma 5.2 *Suppose Assumptions* 3.1*,* 3.2*,* 5.1*,* 5.2 *and* 5.4 *hold.. Also suppose* $b_n/n \to 0$*. Then* $\hat{\sigma}_n$ *defined in equation* (5.5) *satisfies Assumption* 5.5*.*

6 Power Properties

In this section, we show that our model selection test is consistent against all fixed alternatives and it has nontrivial power against all $n^{-1/2}$-local alternatives.

First, we show the fixed alternative result. Theorem 6.1 below shows that our model selection test rejects H_0 with probability approaching one under a fixed data distribution such that H_0 is violated.

Theorem 6.1 *Suppose Assumptions* 3.1*,* 3.2*,* 5.1*,* 5.2*,* 5.3*,* 5.4 *and* 5.5 *hold. Then for any* $P_* \in \mathcal{F}\backslash\mathcal{F}_0$*, we have* $\lim_{n\to\infty} E_{P_*}\varphi_n(\alpha) = 1$ *for* $\varphi_n(\alpha) = \varphi_n^{\text{2-sided}}(\alpha)$ *or* $\varphi_n^{\text{1-sided}}(\alpha)$*.*

Next, we show the local power result. The following assumption specifies the $n^{-1/2}$-local alternatives we consider. Without loss of generality, we assume that the model $\mathcal{P}_1$ is closer to the sequence of true distribution than the model $\mathcal{P}_2$.

Assumption 6.1 *The sequence of true data distribution* $\{P_{n,*}\}_{n\geq 1}$ *satisfies* $P_{n,*} \in \mathcal{F}\backslash\mathcal{F}_0$ *for all* $n \geq 1$*,* (i) $\omega_{P_{n,*}} > 0$ *for all* $n \geq 1$*,* $\omega_{P_{n,*}} \to \omega_\infty \geq 0$*, and* (ii) $\sqrt{n}LR_{P_{n,*}} \to h_1 > 0$.

Theorem 6.2 below shows that our test has non-trivial power against $n^{-1/2}$-local alternatives satisfying Assumption 6.1. It is also straightforward to see from the theorem that when the local alternatives approach the global alternatives in that $h_1 \to \infty$, the asymptotic power increases to one.

Theorem 6.2 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2, 5.3, 5.4, 5.5 *and* 6.1 *hold. Then,*
(a) $\liminf_{n\to\infty} E_{P_{n,*}}\varphi_n^{\text{2-sided}}(\alpha) \geq 1 - \Phi(z_{\alpha/2} - h_1/\sqrt{\omega_\infty^2 + 1}) + \Phi(-z_{\alpha/2} - h_1/\sqrt{\omega_\infty^2 + 1})$ *and*
(b) $\liminf_{n\to\infty} E_{P_{n,*}}\varphi_n^{\text{1-sided}}(\alpha) \geq 1 - \Phi(z_\alpha - h_1/\sqrt{\omega_\infty^2 + 1})$.

7 Extensions

In this section, we extend our method in three different directions: (i) to allow for discrete variables in the conditioning sets, (ii) to allow the conditioning sets in competing models to be different, and (iii) to relax Assumption 3.3.

7.1 Discrete Variable in Conditioning Set

Here we discuss two methods to deal with discrete conditioning variables. For both methods, we discuss the case where there is only one discrete conditioning variable and it is a binary variable taking values in $\{0,1\}$. More general discrete variables can be incorporated similarly.

Let $W = (Y, X, Z)$ where Z is a binary variable and X are continuous variables. Let the conditional moment inequality/equality models be $\mathcal{P}_s = \bigcup_{\theta_s\in\Theta_s}\mathcal{P}_{s,\theta_s}$ for $s = 1$ and 2

$$\mathcal{P}_{s,\theta_s} = \left\{ \begin{array}{l} P : E_P[m_{s,j}(W,\theta_s)|X,Z] = 0 \text{ a.s. } [P_{xz}] \text{ for } j = 1,...,p_s, \\ \quad E_P[m_{s,j}(W,\theta_s)|X,Z] \geq 0 \text{ a.s. } [P_{xz}] \text{ for } j = p_s+1,...,k_s \end{array} \right\} \tag{7.1}$$

where P_{xz} is the marginal distribution of (X, Z) implied by P.

The first method we consider is as follows. Define the instrument functions as $g_{\ell,z}(X,Z) = g_\ell(X)\cdot 1(Z = z)$ where $z \in \{0,1\}$, and $g_\ell(X)$ and ℓ are the same as before. Then $\mathcal{P}_{s,\theta_s}$ can be written as $\bigcap_{\ell\in\mathcal{L},z\in\{0,1\}}\mathcal{P}_{s,\theta_s,\ell,z}$, where

$$\mathcal{P}_{s,\theta_s,\ell,z} = \left\{ \begin{array}{l} P : E_P[m_{s,j}(W,\theta_s)g_{\ell,z}(X,Z)] = 0 \text{ for } j = 1,...,p_s, \\ \quad E_P[m_{s,j}(W,\theta_s)g_{\ell,z}(X,Z)] \geq 0 \text{ for } j = p_s+1,...,k_s. \end{array} \right\} \tag{7.2}$$

and $\mathcal{L}$ can be $\mathcal{L}_{\text{c-cube}}$ or $\mathcal{L}_{\text{cube}}$. Then all the results discussed above can be extended to this case with suitable modification of the regularity conditions.

In the second method, we treat the two values of Z as argumenting the number of conditional moment restrictions, that is, we write

$$\mathcal{P}_{s,\theta_s} = \left\{ \begin{array}{l} P : E_P[m_{s,j,z}(W,\theta_s)|X] = 0 \text{ a.s. } [P_x] \text{ for } j = 1,...,p_s, \text{ and } z = 1,2, \\ \quad E_P[m_{s,j,z}(W,\theta_s)|X] \geq 0 \text{ a.s. } [P_x] \text{ for } j = p_s+1,...,k_s, \text{ and } z = 1,2, \end{array} \right\} \tag{7.3}$$

where $m_{s,j,z}(W,\theta_s) = m_{s,j}(W,\theta_s)\cdot 1(Z = z)$. Transform the conditional moment restriction into unconditional ones, and we can again write $\mathcal{P}_{s,\theta_s} \equiv \bigcap_{\ell\in\mathcal{L}}\mathcal{P}_{s,\theta_s,\ell}$, where

$$\mathcal{P}_{s,\theta_s,\ell} = \left\{ \begin{array}{ll} P : & E_P[m_{s,j,z}(W,\theta_s)g_\ell(X)] = 0 \text{ for } j = 1,...,p_s, \text{ and } z = 1,2, \\ & E_P[m_{s,j,z}(W,\theta_s)g_\ell(X)] \geq 0 \text{ for } j = p_s+1,...,k_s, \text{ and } z = 1,2, \end{array} \right\} \tag{7.4}$$

and $\mathcal{L}$ can be $\mathcal{L}_{\text{c-cube}}$ or $\mathcal{L}_{\text{cube}}$. Then all the results discussed above can be extended to this case easily with suitable modification of the regularity conditions.

7.2 Competing Models with Difference Conditioning Sets

In the sections above, we require that the competing models have the same conditioning set. Here we show that with suitable modification, our method can easily allow for the cases where the conditioning sets for competing models are different.

For $s = 1$ and 2, consider $\mathcal{P}_s = \bigcup_{\theta_s \in \Theta_s} \mathcal{P}_{s,\theta_s}$ where

$$\mathcal{P}_{s,\theta_s} = \left\{ \begin{array}{l} P : E_P[m_{s,j}(W,\theta_s)|X_s] = 0 \text{ a.s. } [P_{x_s}] \text{ for } j = 1, ..., p_s, \\ \quad E_P[m_{s,j}(W,\theta_s)|X_s] \geq 0 \text{ a.s. } [P_{x_s}] \text{ for } j = p_s + 1, ..., k_s. \end{array} \right\} \tag{7.5}$$

In the above equation, $W = (Y, X_1, X_2)$ is generated from P_0 where $X_s \in \mathcal{X}_s \subseteq R^{d_s}$. The conditioning variables X_1 and X_2 can be the same, nest each other, overlap but be nonnested, or be disjoint.

Let $\ell_s = (x_s, r) \in [0,1]^{d_s} \times R$ and define $\mathcal{G}_s = \{g_{\ell_s}(X_s) : \ \ell_s \in \mathcal{L}_s\}$. Define distance from $\mathcal{P}_s$ to P_0 as

$$d_{\mathcal{L}_s}(\mathcal{P}_s, P_0) = \inf_{\theta_s \in \Theta_s} \int_{\mathcal{L}_s} d(\mathcal{P}_{s,\theta_s,\ell_s}, P_0) dF_s(\ell_s). \tag{7.6}$$

where $F_s(\ell_s)$ is a probability measure whose support contains $\mathcal{L}_s$ and $d(\mathcal{P}_{s,\theta_s,\ell_s}, P_0)$ is defined as in (3.5). Define $\mathcal{M}_{s,\ell_s,P_0}(\gamma_s,\theta_s)$, $\widehat{\mathcal{M}}_{s,\ell_s,P_0}(\gamma_s,\theta_s)$, γ^*_{s,ℓ_s,P_0}, $\hat{\gamma}_{s,\ell_s,n}$, $\mathcal{L}_{s,r_{sn}}$, and M^*_{s,ℓ_s,P_0} accordingly.

Define $t_{s,n} = n^{1/2} r_{sn}^{d_s+1} \vee n^{-1/2} r_{sn}^{d_s}$ and $t_n = t_{1,n} \vee t_{2,n}$. Assume that r_{sn} satisfies Assumption 5.4 for $s = 1, 2$. With suitable modification on the regularity conditions, under all sequences $\{P_n \in \mathcal{F}\}_{n=1}^{\infty}$, we have

$$\sqrt{n}(\widehat{LR}_n - LR_{P_n}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Lambda^*_{P_n,i} + O_p(t_n), \tag{7.7}$$

$$\begin{aligned} \Lambda^*_{P_n,i} = & \int_{\mathcal{L}_1} \Psi'(M^*_{1,\ell_1,P_n}) \big[\kappa\big(\gamma^*_{1,\ell_1,P_n}(\theta^*_1, P_n)' m_1(W_i,\theta^*_1) g_{\ell_1}(X_i)\big) - M^*_{1,\ell_1,P_n}\big] dF_1(\ell_1) \\ & - \int_{\mathcal{L}_2} \Psi'(M^*_{2,\ell_2,P_n}) \big[\kappa\big(\gamma^*_{2,\ell_2,P_n}(\theta^*_2)' m_2(W_i,\theta^*_2) g_{\ell_2}(X_i)\big) - M^*_{2,\ell_2,P_n}\big] dF_2(\ell_2), \text{ for} \\ & \theta^*_s \in \Theta^*_s(P_n), \quad \text{for } s = 1 \text{ and } 2. \end{aligned}$$

Define $\hat{\omega}^2_n = \sup_{(\theta_1',\theta_2')' \in \widehat{\Theta}_{1,n} \times \widehat{\Theta}_{2,n}} \hat{\omega}^2_n(\theta_1,\theta_2)$, where

$$\begin{aligned} & \hat{\omega}^2_n(\theta_1,\theta_2) \\ = & \frac{1}{n} \sum_{i=1}^{n} \Big[\int_{\mathcal{L}_{1,r_{1n}}} \Psi'(\widehat{M}_{1,\ell_1,n}(\theta_1)) \big[\kappa\big(\hat{\gamma}_{1,\ell_1,n}(\theta_1)' m_1(W_i,\theta_1) g_{\ell_1}(X_i)\big) - \widehat{M}_{1,\ell_1,n}(\theta_1)\big] dF_1(\ell_1) \\ & - \int_{\mathcal{L}_{2,r_{2n}}} \Psi'(\widehat{M}_{2,\ell_2,n}(\theta_2)) \big[\kappa\big(\hat{\gamma}_{2,\ell_2,n}(\theta_2)' m_2(W_i,\theta_2) g_{\ell_2}(X_i)\big) - \widehat{M}_{2,\ell_2,n}(\theta_2)\big] dF_2(\ell_2) \Big]^2. \end{aligned} \tag{7.8}$$

Then all the results discussed in the previous sections can be extended to this case easily.

7.3 The Uniqueness Assumption

First, we explain why the uniqueness assumption, Assumption 3.3 is useful. From Lemma A.8, we see that when Assumption 3.3 is satisfied, we can establish that

$$n^{1/2}(\widehat{LR}_n - LR_{P_0}) = n^{-1/2}\sum_{i=1}^{n}\Lambda^*_{P_0,i} + o_p(1). \tag{7.9}$$

If Assumption 3.3 is not satisfied, $\Lambda^*_{P_0,i}$ will not be well-defined. Instead, we can only define a function $\Lambda^*_{P_0,i}(\theta^*_1,\theta^*_2)$, which takes different values for different $\theta^*_s \in \Theta^*_s(P_0)$. Consequently, Lemma A.8 will establish

$$n^{1/2}(\widehat{LR}_n - LR_{P_0}) = n^{-1/2}\sum_{i=1}^{n}\Lambda^*_{P_0,i}(\theta^*_{1,n},\theta^*_{2,n}) + o_p(1), \tag{7.10}$$

where $\theta^*_{s,n}$ is the closest point in $\Theta^*_s(P_0)$ to $\hat{\theta}_{s,n}$, for the $\hat{\theta}_{s,n}$ used in constructing $\widehat{LR}_n$. The random sequence $\theta^*_{2,n}$ may not converge, and is correlated with the data that forms $\Lambda^*_{P_0,i}(\cdot,\cdot)$, causing $n^{-1/2}\sum_{i=1}^{n}\Lambda^*_{P_0,i}(\theta^*_{1,n},\theta^*_{2,n})$ not to have a normal asymptotic distribution. Losing the asymptotic normality destroys the simple structure of our test.

One way to preserve the simplicity of the test without the unique assumption is to estimate $\widehat{\Theta}_{s,n}$ from a separate sample, and then use a random point $\hat{\theta}_{s,n}$ from that set estimator to construct $\widehat{LR}_n$, $\hat{\omega}_n$ and $\hat{\sigma}_n$. Because the $\hat{\theta}_{s,n}$ is from a separate sample, its closest point in $\Theta^*_s(P_0)$, $\theta^*_{s,n}$, will be independent with the data that forms $\Lambda^*_{P_0,i}(\cdot,\cdot)$. Then all our derivations in the previous section can be done conditional on $\hat{\theta}_{s,n}$.

The natural way to come up with a separate sample is to split the original sample into two equal halves. One half is used to estimate the parameters and the other to construct the statistics.

8 Monte Carlo Simulation

In this section we report Monte Carlo results for a missing data example. Let Y_i be a binary variable that is observable only if a selection variable $D_i = 1$ and is missing if $D_i = 0$. Let $\overline{Y}_i = Y_iD_i + (1-D_i)$ and $\underline{Y}_i = Y_iD_i$. Then by definition $Y_i \in [\overline{Y}_i, \underline{Y}_i]$. Let X_{1i} and X_{2i} be two covariates. Suppose two candidates models are both Probit models but disagree on which of the two covariates is relevant. That is, for $j = 1, 2$:

$$\begin{aligned}\mathcal{P}_j = \{P : \; & E_P[\Phi(\theta_1 + \theta_2 X_{ji}) - \underline{Y}_i | X_{1i}, X_{2i}] \geq 0, \text{ and} \\ & E_P[\overline{Y}_i - \Phi(\theta_1 + \theta_2 X_{ji}) | X_{1i}, X_{2i}] \geq 0\},\end{aligned} \tag{8.1}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function (cdf).

Consider the following data generating process:

$$
\begin{aligned}
Y_i &= 1\{1 + 1.5^{1/2}(\theta_{21}X_{1i} + \theta_{22}X_{2i}) + u_i \geq 0\}, \\
D_i &= 1\{1.5 + 0.5(X_{1i} + X_{2i}) + v_i\},\ (u_i, v_i) \sim N(0, [1, 0.5; 0.5, 1]),
\end{aligned} \tag{8.2}
$$

where X_{1i} and X_{2i} are generated as follows: $X_{1i} = Z_{1i}\Phi(Z_{2i}), X_{2i} = Z_{2i}\Phi(Z_{1i})$ with $(Z_{1i}, Z_{2i})' \sim N(0, I_2)$. The parameter $(\theta_{21}, \theta_{22})$ determines which covariate(s) is (are) relevant for Y in the data generating process. When $\theta_{21} = \theta_{22}$, the two candidate models are equally good. In particular when $\theta_{21} = \theta_{22} = 0$, the two candidate models are both correctly specified. When $\theta_{21} > \theta_{22} \geq 0$, model $\mathcal{P}_1$ is better than model $\mathcal{P}_2$ and when $0 \leq \theta_{21} < \theta_{22}$, model $\mathcal{P}_2$ is better than model $\mathcal{P}_1$. We consider four configurations to investigate the size and the power properties of our test. The four configurations are: $(\theta_{21}, \theta_{22}) = (0, 0), (1, 1), (0, 1), (1, 1.5)$.

To implement our method, we first transform the variables X_{1i} and X_{2i} to the unit interval. Define $(X_{1i}^*, X_{2i}^*) = \Phi(\hat{\Sigma}_n^{-1/2} \cdot (X_{1i}, X_{2i})')$, where $\hat{\Sigma}_n$ is the sample covariance matrix of $(X_{1i}, X_{2i})'$ and Φ is the standard normal cdf function applied element by element. Then we treat the transformed variables X_{1i}^* and X_{2i}^* as the conditioning variables. Finally we use the countable hypercube instrumental functions on the new conditioning variables:

$$
\mathcal{G} = \left\{ 1\left((x_1^*, x_2^*) \in \left[\frac{b_1}{2q}, \frac{b_1+1}{2q}\right] \times \left[\frac{b_2}{2q}, \frac{b_2+1}{2q}\right] \right) : b_1, b_2 = 0, 1, \ldots, 2q-1,\ q = q_0, \ldots, q_1 \right\}. \tag{8.3}
$$

We use $q_0 = 1$ and $q_1 = 2$ for all the settings we consider. The probability measure on $\mathcal{G}_{\text{c-cube}}$ gives equal probability to the $b's$ given each r and gives each q a probability proportional to $1/q^2$. For $\hat{\sigma}_n$, we use $b_n = n/\log(n)$ and $t_n = n^{(1/2-(d_x+1)/(2d_x+1))} = n^{-1/10}$.

The results are reported in Table 1. The two numbers in the parentheses are respectively the probability of rejecting the null and selecting model $\mathcal{P}_1$ and that of rejecting the null and selecting model $\mathcal{P}_2$. As we can see, the selection probabilities (first two rows) for either model is close to 5% when the two models are equally good (the first two rows). It is worth noting that this is the case even when both models are correctly specified (the first row), in which case $\omega^2 = 0$. The selection probability of the better model is bigger than 5% and grows with the sample size when there is a better model among the two (the last two rows), suggesting that the test is consistent.

9 Conclusion

To sum up, we propose a model selection test for conditional moment inequality models. The test can be applied when the competing models are nested or nonnested. In all cases, the test is asymptotically similar, consistent against fixed alternatives and has non-trivial power against $n^{-1/2}$ local alternatives. This is the first such test for conditional moment inequality models.

Table 1: Null and Alternative Selection Probabilities ($\alpha = 10\%$)

$(\theta_{21}, \theta_{22})$	$n = 250$	$n = 500$	$n = 1000$
$(0, 0)$	$(.063, .049)$	$(.053, .049)$	$(.039, .054)$
$(1, 1)$	$(.063, .056)$	$(.057, .055)$	$(.056, .067)$
$(0, 1)$	$(.011, .249)$	$(.003, .323)$	$(.001, .454)$
$(1, 1.5)$	$(.007, .340)$	$(.000, .575)$	$(.000, .823)$

APPENDIX

A Auxiliary Lemmas

In this section, we collect all the auxiliary lemmas used to prove the main results in the text. The proofs of these auxiliary lemmas are deferred to Appendix C.

To begin, we first introduce some new notation. Let ϕ_s denote the combined parameter vector $(\gamma_s', \theta_s')'$. Let $\Phi_s = \Gamma_M^s \times \Theta_s$ for $s = 1$ and 2. For any two sets A and B, let $A \triangle B \equiv \{x : x \in A, x \in B, \text{ but } x \notin A \cap B\}$. Define a pseudo metric on $\mathcal{L}$ as

$$\rho_\ell(\ell_1, \ell_2) = \lambda(C_{\ell_1} \triangle C_{\ell_2})^{1/2}, \tag{A.1}$$

where C_ℓ is defined in the second line of (3.2) and $\lambda(\cdot)$ is the Lebesgue measure. Define a pseudo-metric on $\Phi_s \times \mathcal{L}$ as

$$\rho_s((\phi_{s1}, \ell_1), (\phi_{s2}, \ell_2)) = \|\phi_{s1} - \phi_{s2}\| + \rho_\ell(\ell_1, \ell_2), \tag{A.2}$$

for $(\phi_{s1}, \ell_1), (\phi_{s2}, \ell_2) \in \Phi_s \times \mathcal{L}$.

For $s = 1, 2$, let

$$\begin{aligned} \mathcal{M}_{s,\ell,P_0}(\gamma_s, \theta_s) &= E_{P_0} \kappa(\gamma_s' m(W_i, \theta_s) g_\ell(X_i)), \text{ and} \\ \widehat{\mathcal{M}}_{s,\ell,n}(\gamma_s, \theta_s) &= n^{-1} \sum_{i=1}^n \kappa(\gamma_s' m(W_i, \theta_s) g_\ell(X_i)). \end{aligned} \tag{A.3}$$

Lemma A.1 below shows some basic properties of stochastically equicontinuous empirical processes. This result is not new in the literature, but we state and prove it here for easy reference.

Lemma A.1 *Consider the triangular array of empirical processes $\{\nu_n(t) : t \in \mathcal{T}\}_{n=1}^\infty$. If* (i) *$(\mathcal{T}, \rho)$ is a totally bounded pseudo-metric space,* (ii) *$\nu_n(t)$ is stochastically equicontinuous w.r.t. ρ and* (iii) *for every $t \in \mathcal{T}$, $\|\nu_n(t)\| = O_p(1)$, then $\sup_{t \in \mathcal{T}} \|\nu_n(\phi)\| = O_p(1)$.*

Lemma A.2 below shows the stochastic equicontinuity of several empirical processes that form different parts of $\hat{T}_n$.

Lemma A.2 *Suppose Assumptions* 3.1, 3.2, 5.1 *and* 5.2 *hold. Then, under any sequence* $\{P_n \in \mathcal{F}\}_{n=1}^{\infty}$*, for* $s = 1$ *and 2,*
(a) *the triangular arrays of empirical process:*

$$\{\nu_{s,n}^0(\phi_s, \ell) := n^{1/2}(\widehat{\mathcal{M}}_{s,\ell,n}(\phi_s) - \mathcal{M}_{s,\ell,P_n}(\phi_s)) : \phi_s \in \Phi_s,\ \ell \in \mathcal{L}\}$$

is stochastically equicontinuous w.r.t. the pseudo-metric ρ_s *defined in* (A.2),
(b) $\sup_{\phi_s \in \Phi_s, \ell \in \mathcal{L}} |n^{1/2}(\widehat{\mathcal{M}}_{s,\ell,n}(\phi_s) - \mathcal{M}_{s,\ell,P_n}(\phi_s))| = O_p(1)$,
(c) *the triangular arrays of empirical processes*

$$\{\nu_{s,n}^1(\phi_s, \ell) := n^{1/2}(\partial\widehat{\mathcal{M}}_{s,\ell,n}(\phi_s)/\partial\phi_s - \partial\mathcal{M}_{s,\ell,P_n}(\phi_s)/\partial\phi_s) : \phi_s \in \Phi_s,\ \ell \in \mathcal{L}\} \text{ and}$$
$$\{\nu_{s,n}^2(\phi_s, \ell) := n^{1/2}(\partial^2\widehat{\mathcal{M}}_{s,\ell,n}(\phi_s)/\partial\phi_s\partial\phi_s' - \partial^2\mathcal{M}_{s,\ell,P_n}(\phi_s)/\partial\phi_s\partial\phi_s') : \phi_s \in \Phi_s,\ \ell \in \mathcal{L}\}$$

are stochastically equicontinuous w.r.t. the pseudo-metric ρ_s,
(d) $\sup_{\phi_s \in \Phi_s, \ell \in \mathcal{L}} \|n^{1/2}(\partial\widehat{\mathcal{M}}_{s,\ell,P_n}(\phi_s)/\partial\phi_s - \partial\mathcal{M}_{s,\ell,P_n}(\phi_s)/\partial\phi_s\| = O_p(1)$*, and*
$\sup_{\phi_s \in \Phi_s, \ell \in \mathcal{L}} \|n^{1/2}(\partial^2\widehat{\mathcal{M}}_{s,\ell,n}(\phi_s)/\partial\phi_s\partial\phi_s' - \partial^2\mathcal{M}_{s,\ell,P_n}(\phi_s)/\partial\phi_s\partial\phi_s')\| = O_p(1)$,
(e) *for any random mappings* $\{\phi_{s,n}^{(1)},\ \phi_{s,n}^{(2)} : \mathcal{L} \to \Phi_s\}_{n=1}^{\infty}$ *such that* $\sup_{\ell \in \mathcal{L}} \|\phi_{s,n}^{(1)}(\ell) - \phi_{s,n}^{(2)}(\ell)\| = o_p(1)$*, we have*

$$\sup_{\ell \in \mathcal{L}} |\widehat{\mathcal{M}}_{s,\ell,n}(\phi_{s,n}^{(1)}(\ell)) - \mathcal{M}_{s,\ell,P_n}(\phi_{s,P_n}^{(2)}(\ell))| \to_p 0$$
$$\sup_{\ell \in \mathcal{L}} \|\partial^2\widehat{\mathcal{M}}_{s,\ell,n}(\phi_{s,n}^{(1)}(\ell))/\partial\phi_s\partial\phi_s' - \partial^2\mathcal{M}_{s,\ell,P_n}(\phi_{s,n}^{(2)}(\ell))/\partial\phi_s\partial\phi_s'\| \to_p 0.$$

Lemma A.3 below shows the consistency of $\hat{\gamma}_{s,\ell,n}(\theta_s)$ for $\gamma_{s,\ell,P_n}^*(\theta_n)$ under drifting sequences of data distributions P_n.

Lemma A.3 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2 *and* 5.4 *hold. Under any sequence* $\{P_n \in \mathcal{F}\}_{n=1}^{\infty}$*, we have for* $s = 1$ *and* 2,
(a) $\sup_{\theta_s \in \Theta_s, \ell \in \mathcal{L}_{r_n}} \|\hat{\gamma}_{s,\ell,n}(\theta_s) - \gamma_{s,\ell,P_n}^*(\theta_s)\| \to_p 0$,
(b) $\sup_{\theta_s \in \Theta_s, \ell \in \mathcal{L}_{r_n}} p_{\ell,n}\big\|\hat{\gamma}_{s,\ell,n}(\theta_s) - \gamma_{s,\ell,P_n}^*(\theta_s)\big\| = O_p(n^{-1/2})$*, and* $\sup_{\theta_s \in \Theta_s, \ell \in \mathcal{L}_{r_n}} \big\|\hat{\gamma}_{s,\ell,n}(\theta_s) - \gamma_{s,\ell,P_n}^*(\theta_s)\big\| = O_p(n^{-1/2}r_n^{-d_x})$,
(c) *for any two random sequences* $\{\theta_{s,n}^{(1)}\}_{n=1}^{\infty}$ *and* $\{\theta_{s,n}^{(2)}\}_{n=1}^{\infty}$ *such that* $\big\|\theta_{s,n}^{(1)} - \theta_{s,n}^{(2)}\big\| = o_p(1)$*, we have that* $\sup_{\ell \in \mathcal{L}} \big\|\gamma_{s,\ell,P_n}^*(\theta_{s,n}^{(1)}) - \gamma_{s,\ell,P_n}^*(\theta_{s,n}^{(2)})\big\| = O_p(\big\|\theta_{s,n}^{(1)} - \theta_{s,n}^{(2)}\big\|)$*, and*
(d) *for the two random sequences in part (c), we have* $\sup_{\ell \in \mathcal{L}_{r_n}} \big\|\hat{\gamma}_{s,\ell,n}(\theta_{s,n}^{(1)}) - \gamma_{s,\ell,P_n}^*(\theta_{s,n}^{(2)})\big\| \to_p 0$.

Lemma A.4 below shows that under our assumptions the effect of the truncation of $\mathcal{L}$ is small.

Lemma A.4 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2 *and* 5.4 *hold. Uniformly over* $P_0 \in \mathcal{F}$ *and* $\theta_s \in \Theta_s$ *for* $s = 1, 2$,

$$r_n^{-d_x-1} \int_{\mathcal{L}_{r_n}^c} \Psi[\mathcal{M}_{s,\ell,P_0}(\gamma_{s,\ell,P_0}^*(\theta_s), \theta_s)]dF(\ell) = O_p(1) \tag{A.4}$$

Lemma A.5 below shows a full-rank condition for each super model $\mathcal{P}_{s,\ell,\theta_s}$ of $\mathcal{P}_{s,\theta_s}$. This full-rank condition guarantees that $\gamma_{s,\ell,P_0}^*(\theta_s)$ is uniquely defined for every θ_s and every ℓ.

Lemma A.5 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2 *and* 5.4 *hold. Under any* $P_0 \in \mathcal{F}$, *for all* $\ell \in \mathcal{L}$ *and for all* $\phi_s \in \Phi_s$,

$$eig_{\max}\left(E[\kappa''(\gamma_s' m_s(W,\theta_s))m_s(W,\theta_s)m_s(W,\theta_s)'g_\ell(X)]\right) \leq -p_{P_0,\ell}\cdot\delta \quad \textit{for } s = 1 \textit{ and } 2. \tag{A.5}$$

Lemma A.6 below shows an important property of our pseudo-distance measure: the pseudo-distance is zero when and only when intuitively it should be zero, that is when P_0 belongs to the model, or in other words, the model is correctly specified.

Lemma A.6 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2 *and* 5.4 *hold. For any* $P_0 \in \mathcal{F}$, *then* $P_0 \in \mathcal{P}_s$ *iff* $d_{\mathcal{L}}(\mathcal{P}_s, P_0) = 0$ *for* $s = 1, 2$.

Lemma A.7 below establishes the convergence rate of $\widehat{\Theta}_{1,n}$ and $\widehat{\Theta}_{2,n}$ w.r.t. the *left* Hausdorff distance.

Lemma A.7 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2 *and* 5.4 *hold. Then, under all sequences* $\{P_n \in \mathcal{F}\}_{n=1}^{\infty}$, *we have* $\max_{s=1,2}\rho_{lh}(\widehat{\Theta}_{s,n}, \Theta_s^*(P_n)) = O_p(n^{-1/2}r_n^{-d_x/2})$.

Lemma A.8 below shows a linear representation of $\sqrt{n}(\widehat{LR}_n - LR_{P_n})$ under a sequence of data distributions $\{P_n\}$. This lemma is a crucial step for establishing Lemma 5.1.

Lemma A.8 *Suppose Assumptions* 3.1, 3.2, 5.1, 5.2 *and* 5.4 *hold. Then, under all sequences* $\{P_n \in \mathcal{F}\}_{n=1}^{\infty}$,

$$\sqrt{n}(\widehat{LR}_n - LR_{P_n}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Lambda_{P_n,i}^* + O_p(t_n), \tag{A.6}$$

where $\Lambda_{P_n,i}^*$ *is defined in* (4.4) *and* t_n *is defined above Lemma* 5.1.

The following three lemmas prove a Kuhn-Tucker condition that is used repeatedly in the proof of our main results. These lemmas are taken from Chong and Żak (2001) with minor modification. Consider the following problem:

$$\text{maximize} f(x) \qquad \text{subject to} \qquad g(x) \geq 0, \tag{A.7}$$

where $f: R^d \to R$ and $g: R^d \to R^m$.[7] Let x^* satisfy $g(x^*) \geq 0$ and define $J(x^*) \equiv \{j : g_j(x^*)\}$ which is the set of the index of active inequality. We say that x^* is a regular point if the vectors $\partial g_j(x^*)/\partial x$ for $j \in J(x^*)$ are linear independent. We say that x^* is a feasible point if $g(x^*) \geq 0$. Define

$$L(x^*, \mu^*) = \frac{\partial^2 f(x^*)}{\partial x \partial x'} + \sum_{j=1}^{m}\mu_j^* \cdot \frac{\partial^2 g_j(x^*)}{\partial x \partial x'}, \tag{A.8}$$

$$T(x^*, \mu^*) = \{y : \partial g_j(x^*)/\partial x \cdot y = 0, j \in \widetilde{J}(x^*, \mu^*)\}, \tag{A.9}$$

$$\widetilde{J}(x^*, \mu^*) = \{j : g_j(x^*) = 0, \mu_j^* > 0\}. \tag{A.10}$$

It is obvious that $\widetilde{J}(x^*, \mu^*) \subseteq J(x^*)$.

[7] Note that here, $f(x)$, $g(x)$ and x here refer to some generic functions and their argument and do not refer to the same things as similar or the same symbols defined in the main sections of this paper. Since these new definitions only apply locally from here to the end of this section, there should be no confusion caused by this abuse of notations.

Lemma A.9 Karush-Kuhn-Tucker Theorem. *Let $f(x)$ and $g(x)$ are once continuously differentiable on x. Let x^* be a regular point and a local maximum for the problem defined in (A.7). Then there exists $\mu^* \in R^m$ such that*

1. $\mu^* \geq 0$,
2. $\partial f(x^*)/\partial x + \sum_{j=1}^m \mu_j^* \cdot \partial g_j(x^*)/\partial x = 0$, *and*
3. $g(x^*)' \cdot \mu^* = 0$.

Lemma A.10 Second-Order Sufficient Conditions. *Suppose f and g are twice continuously differentiable in x and there exists a feasible point $x^* \in R^d$ and $\mu^* \in \mathcal{R}^m$ such that*

1. $\mu^* \geq 0$, $\partial f(x^*)/\partial x + \sum_{j=1}^m \mu_j^* \cdot \partial g_j(x^*)/\partial x = 0$, $g(x^*)' \cdot \mu^* = 0$, *and*
2. *For all $y \in T(x^*, \mu^*)$ with $y \neq 0$, we have $y' L(x^*, \mu^*) y < 0$.*

Then, x^* is a strict local maximizer of problem defined in (A.7).

Lemma A.11 *Suppose in problem* (A.7)*, f and g are once continuously differentiable in x. Suppose g_j for $j = 1, \ldots, m$ are concave in x. If there exists a feasible point $x^* \in R^d$ and $\mu^* \in \mathcal{R}^m$ such that $\mu^* \geq 0$, $\partial f(x^*)/\partial x + \sum_{j=1}^m \mu_j^* \cdot \partial g_j(x^*)/\partial x = 0$, and $g(x^*)' \cdot \mu^* = 0$, then $\partial f(x^*)/\partial x \cdot (x - x^*) \leq 0$ for any x such that $g(x) \geq 0$.*

B Proof of Main Results

Proof of Lemma 5.1: Lemma 5.1 is stated in terms of subsequences $\{a_n\}_{n=1}^\infty$. For notational simplicity, we prove it for the sequence $\{n\}$. All the arguments go through with $\{a_n\}$ in place of $\{n\}$.

For (a), $t_n^{-1}\sqrt{n}(\widehat{LR}_n - LR_{P_n}) = O_p(1)$ follows simply from Lemma A.8 and the CLT.

For $t_n^{-2}\hat{\omega}_n^2$, observe that

$$\begin{aligned}
\hat{\omega}_n^2 = o_p\left(\frac{1}{n}\right) + \frac{1}{n}\sum_{i=1}^n \Bigg[\int_{\mathcal{L}_{r_n}} &\Psi'(\widehat{\mathcal{M}}_{1,\ell,n}(\hat{\theta}_{1,n}))\big[\kappa\big(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n})'m_1(W_i,\hat{\theta}_{1,n})g_\ell(X_i)\big) - \widehat{\mathcal{M}}_{1,\ell,n}\big] \\
&- \Psi'(\widehat{\mathcal{M}}_{2,\ell,n}(\hat{\theta}_{2,n}))\big[\kappa\big(\hat{\gamma}_{2,\ell,n}(\hat{\theta}_{2,n})'m_2(W_i,\hat{\theta}_{2,n})g_\ell(X_i)\big) - \widehat{\mathcal{M}}_{2,\ell,n}\big]\Big\}dF(\ell)\Bigg]^2 \\
= o_p\left(\frac{1}{n}\right) + \frac{1}{n}\sum_{i=1}^n \Bigg[\Lambda_{P_n,i}^* - \int_{\mathcal{L}_{r_n}^c} &\Big\{\Psi'(\mathcal{M}_{1,\ell,P_n}^*)\big[\kappa\big(\gamma_{1,\ell,P_n}^*(\theta_{1,n}^*)'m_1(W_i,\theta_1^*)g_\ell(X_i)\big) - \mathcal{M}_{1,\ell,P_n}^*\big] \\
&- \Psi'(\mathcal{M}_{2,\ell,P_n}^*)\big[\kappa\big(\gamma_{2,\ell,P_n}^*(\theta_{2,n}^*)'m_2(W_i,\theta_{2,n}^*)g_\ell(X_i)\big) - \mathcal{M}_{2,\ell,P_n}^*\big]\Big\}dF(\ell) \\
+ \int_{\mathcal{L}_{r_n}} &\Big\{\Psi'(\widehat{\mathcal{M}}_{1,\ell,n}(\hat{\theta}_{1,n}))\big[\kappa\big(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n})'m_1(W_i,\hat{\theta}_{1,n})g_\ell(X_i)\big) - \widehat{\mathcal{M}}_{1,\ell,n}(\hat{\theta}_{1,n})\big] \\
&- \Psi'(\mathcal{M}_{1,\ell,P_n}^*)\big[\kappa\big(\gamma_{1,\ell,P_n}^*(\theta_{1,n}^*)'m_1(W_i,\theta_{1,n}^*)g_\ell(X_i)\big) - \mathcal{M}_{1,\ell,P_n}^*\big]\Big\}dF(\ell) \\
- \int_{\mathcal{L}_{r_n}} &\Big\{\Psi'(\widehat{\mathcal{M}}_{2,\ell,n}(\hat{\theta}_{2,n}))\big[\kappa\big(\hat{\gamma}_{2,\ell,n}(\hat{\theta}_{2,n})'m_2(W_i,\hat{\theta}_{2,n})g_\ell(X_i)\big) - \widehat{\mathcal{M}}_{2,\ell,P_n}(\hat{\theta}_{2,n})\big] \\
&- \Psi'(\mathcal{M}_{2,\ell,P_n}^*)\big[\kappa\big(\gamma_{2,\ell,P_n}^*(\theta_{2,n}^*)'m_2(W_i,\theta_{2,n}^*)g_\ell(X_i)\big) - \mathcal{M}_{2,\ell,P_n}^*\big]\Big\}dF(\ell)\Bigg]^2
\end{aligned}$$

$$\equiv o_p\left(\frac{1}{n}\right) + \frac{1}{n}\sum_{i=1}^{n}[\Lambda^*_{P_n,i} - C_{1,i} + C_{2,i} - C_{3,i}]^2, \tag{B.1}$$

for some $\hat{\theta}_{s,n} \in \widehat{\Theta}_{s,n}$ and some $\theta^*_{s,n} \in \Theta^*_s(P_n)$, for $s = 1, 2$. By (B.1), we have

$$\begin{aligned} t_n^{-2}\hat{\omega}_n^2 &= t_n^{-2}\frac{1}{n}\sum_{i=1}^{n}[\Lambda^*_{P_n,i} - C_{1,i} + C_{2,i} - C_{3,i}]^2 + o_p(t_n^{-2}n^{-1}) \\ &\leq 4\Big(t_n^{-2}\frac{1}{n}\sum_{i=1}^{n}(\Lambda^*_{P_n,i})^2 + t_n^{-2}\frac{1}{n}\sum_{i=1}^{n}C_{1,i}^2 + t_n^{-2}\frac{1}{n}\sum_{i=1}^{n}C_{2,i}^2 + t_n^{-2}\frac{1}{n}\sum_{i=1}^{n}C_{3,i}^2\Big) + o_p(t_n^{-2}n^{-1}), \end{aligned} \tag{B.2}$$

because $\left[\sum_{i=1}^k a_i\right]^2 \leq k \cdot \sum_{i=1}^k a_i^2$ for any finite k. Therefore, to show (a), it is sufficient to show that all the four terms in the parenthesis in the last line of (B.2) are $O_p(1)$. Note that $t_n^{-2}n^{-1}\sum_{i=1}^n(\Lambda^*_{P_n,i})^2 \to_p v_\infty$ by Assumption 5.2(vii) and $E_{P_n}\left[t_n^{-2}n^{-1}\sum_{i=1}^n(\Lambda^*_{P_n,i})^2\right] = t_n^{-2}\omega^2_{P_n} \to v_\infty < \infty$. Next, note that

$$\begin{aligned} |C_{1,i}| &\leq C\int_{\mathcal{L}^c_{r_n}}\Big\{\sup_{\phi_1\in\Gamma^1_M\times\Theta_1}|\kappa(\gamma_1' m_1(W_i,\theta_1))| + \sup_{\phi_2\in\Gamma^2_M\times\Theta_2}|\kappa(\gamma_2' m_2(W_i,\theta_2))|\Big\}dF(\ell) \\ &\leq C\cdot r_n\Big(\sup_{\phi_1\in\Gamma^1_M\times\Theta_1}|\kappa(\gamma_1' m_1(W_i,\theta_1))| + \sup_{\phi_2\in\Gamma^2_M\times\Theta_2}|\kappa(\gamma_2' m_2(W_i,\theta_2))|\Big), \end{aligned} \tag{B.3}$$

where C is a generic positive number not dependent on P_n. The first inequality holds because for all ℓ, $|\kappa(\gamma_s' m_1(W_i,\theta_s)g_\ell(X_i))| \leq |\kappa(\gamma_s' m_1(W_i,\theta_s))|$ and $\Psi'(\widehat{\mathcal{M}}_{s,\ell,n}(\hat{\theta}_{1,n})) \leq C$ by assumptions. The second inequality holds by Assumption 5.4(ii). Therefore,

$$\begin{aligned} &E_{P_n}\Big[t_n^{-2}\frac{1}{n}\sum_{i=1}^{n}C_{1,i}^2\Big] \\ \leq &C\cdot r_n^2 t_n^{-2}\cdot E_{P_n}\Big[\sup_{\phi_1\in\Gamma^1_M\times\Theta_1}|\kappa(\gamma_1' m_1(W_i,\theta_1))|^2 + \sup_{\phi_2\in\Gamma^2_M\times\Theta_2}|\kappa(\gamma_2' m_2(W_i,\theta_2))|^2\Big] \\ \to &0, \end{aligned} \tag{B.4}$$

where the first line follows (B.3) and $(a+b)^2 \leq 2a^2 + 2b^2$. The last line follows from the definition of t_n and Assumption 5.4(i). Equation (B.4) implies that $t_n^{-2}n^{-1}\sum_{i=1}^n C_{1,i}^2 = o_p(1)$. Also, for all $\ell \in \mathcal{L}_{r_n}$,

$$\begin{aligned} &\Psi'(\widehat{\mathcal{M}}_{1,\ell,n}(\hat{\theta}_{1,n})\big[\kappa\big(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n})' m_1(W_i,\hat{\theta}_{1,n})g_\ell(X_i)\big) - \widehat{\mathcal{M}}_{1,\ell,n}\big] \\ &\qquad - \Psi'(\mathcal{M}^*_{1,\ell,P_n})\big[\kappa\big(\gamma^*_{1,\ell,P_n}(\theta^*_{1,n})' m_1(W_i,\theta^*_1)g_\ell(X_i)\big) - \mathcal{M}^*_{1,\ell,P_n}\big] \\ &= \Psi'(\widehat{\mathcal{M}}_{1,\ell,n}) - \Psi'(\mathcal{M}^*_{1,\ell,P_n})\big[\kappa\big(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n})' m_1(W_i,\hat{\theta}_{1,n})g_\ell(X_i)\big) - \widehat{\mathcal{M}}_{1,\ell,n}\big] \\ &+ \Psi'(\mathcal{M}^*_{1,\ell,P_n})\big[\kappa\big(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n})' m_1(W_i,\hat{\theta}_{1,n})g_\ell(X_i)\big) - \kappa\big(\gamma^*_{1,\ell,P_n}(\theta^*_{1,n})' m_1(W_i,\theta^*_1)g_\ell(X_i)\big) \\ &- \Psi'(\mathcal{M}^*_{1,\ell,P_n})(\widehat{\mathcal{M}}_{1,\ell,n} - \mathcal{M}^*_{1,\ell,P_n}) \\ &\equiv C_{21,i,\ell} + C_{22,i,\ell} - C_{23,i,\ell}. \end{aligned} \tag{B.5}$$

For any $\ell \in \mathcal{L}_{r_n}$,

$$\begin{aligned} &\big|\widehat{\mathcal{M}}_{1,\ell,n} - \mathcal{M}^*_{1,\ell,P_n}\big| \\ &\leq \big|\widehat{\mathcal{M}}_{1,\ell,n} - \mathcal{M}_{1,\ell,P_n}(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n}),\hat{\theta}_{1,n})\big| + \big|\mathcal{M}_{1,\ell,P_n}(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n}),\hat{\theta}_{1,n}) - \mathcal{M}^*_{1,\ell,P_n}\big| \\ &\leq \big|\widehat{\mathcal{M}}_{1,\ell,n} - \mathcal{M}_{1,\ell,P_n}(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n}),\hat{\theta}_{1,n})\big| + \Big\|\frac{\partial\mathcal{M}_{1,\ell,P_n}(\tilde{\gamma}_1,\tilde{\theta}_{1,n})}{\partial\phi_1'}\Big\| \end{aligned}$$

$$
\begin{aligned}
&\times \Big(\|\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n}) - \gamma^*_{1,\ell,P_n}(\hat{\theta}_{1,n})\| + \|\gamma^*_{1,\ell,P_n}(\hat{\theta}_{1,n}) - \gamma^*_{1,\ell,P_n}(\theta^*_{1,n})\| + \|\hat{\theta}_{1,n} - \theta^*_{1,n}\| \Big) \\
&= O_p(n^{-1/2}) + O_p(n^{-1/2}) + O_p(n^{-1/2} r_n^{-d_x/2}) + O_p(n^{-1/2} r_n^{-d_x/2}) \\
&= O_p(n^{-1/2} r_n^{-d_x/2}),
\end{aligned}
\tag{B.6}
$$

because

$$
\begin{aligned}
&|\widehat{\mathcal{M}}_{1,\ell,n} - \mathcal{M}_{1,\ell,P_n}(\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n}), \hat{\theta}_{1,n})| = O_p(n^{-1/2}), \\
&\Big\| \frac{\partial \mathcal{M}_{1,\ell,P_n}(\tilde{\gamma}_1, \tilde{\theta}_{1,n})}{\partial \phi_1'} \Big\| \cdot \|\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n}) - \gamma^*_{1,\ell,P_n}(\hat{\theta}_{1,n})\| \\
&= p_{\ell,n}^{-1} \Big\| \frac{\partial \mathcal{M}_{1,\ell,P_n}(\tilde{\gamma}_1, \tilde{\theta}_{1,n})}{\partial \phi_1'} \Big\| \cdot p_{\ell,n} \|\hat{\gamma}_{1,\ell,n}(\hat{\theta}_{1,n}) - \gamma^*_{1,\ell,P_n}(\hat{\theta}_{1,n})\| = O(1) \cdot O_p(n^{-1/2}) = O_p(n^{-1/2}), \\
&\Big\| \frac{\partial \mathcal{M}_{1,\ell,P_n}(\tilde{\gamma}_1, \tilde{\theta}_{1,n})}{\partial \phi_1'} \Big\| \cdot \|\gamma^*_{1,\ell,P_n}(\hat{\theta}_{1,n}) - \gamma^*_{1,\ell,P_n}(\theta^*_{1,n})\| \\
&= O_p(1) \cdot O_p(\|\hat{\theta}_{1,n} - \theta^*_{1,n}\|) = O_p(n^{-1/2} r_n^{-d_x/2}), \\
&\Big\| \frac{\partial \mathcal{M}_{1,\ell,P_n}(\tilde{\gamma}_1, \tilde{\theta}_{1,n})}{\partial \phi_1'} \Big\| \cdot \|\hat{\theta}_{1,n} - \theta^*_{1,n}\| = O_p(n^{-1/2} r_n^{-d_x/2}).
\end{aligned}
\tag{B.7}
$$

Similarly,

$$
|\Psi'(\widehat{\mathcal{M}}_{1,\ell,n}) - \Psi'(\mathcal{M}^*_{1,\ell,P_n})| = |\Psi''(\widetilde{\mathcal{M}}_{1,\ell,n}) \cdot (\widehat{\mathcal{M}}_{1,\ell,n} - \mathcal{M}^*_{1,\ell,P_n})| = O_p(n^{-1/2} r_n^{-d_x/2}).
\tag{B.8}
$$

Next,

$$
\begin{aligned}
&\Big| t_n^{-2} \frac{1}{n} \sum_{i=1}^n C_{2,i}^2 \Big| \\
=& \Big| \int_{\mathcal{L}_{r_n}} \int_{\mathcal{L}_{r_n}} \Big\{ t_n^{-2} \frac{1}{n} \sum_{i=1}^n (C_{21,i,\ell_1} + C_{22,i,\ell_1} - C_{23,i,\ell_1}) \cdot (C_{21,i,\ell_2} + C_{22,i,\ell_2} - C_{23,i,\ell_2}) \Big\} dF(\ell_1) dF(\ell_2) \Big| \\
\leq& \int_{\mathcal{L}_{r_n}} \int_{\mathcal{L}_{r_n}} \Big\{ t_n^{-2} \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^3 \sum_{k=1}^3 \big| C_{2j,i,\ell_1} \cdot C_{2k,i,\ell_2} \big| \Big\} dF(\ell_1) dF(\ell_2).
\end{aligned}
$$

To show $t_n^{-2} n^{-1} \sum_{i=1}^n C_{2,i}^2 = o_p(1)$, it is sufficient to show that $t_n^{-2} n^{-1} \sum_{i=1}^n \big| C_{2j,i,\ell_1} \cdot C_{2k,i,\ell_2} \big| = o_p(1)$ for all $j, k = 1, 2$, and 3 uniformly over $\ell_1, \ell_2 \in \mathcal{L}_{r_n}$. We have $t_n^{-2} n^{-1} \sum_{i=1}^n \big| C_{21,i,\ell_1} C_{21,i,\ell_2} \big| = o_p(1)$ by

$$
\begin{aligned}
&t_n^{-2} \frac{1}{n} \sum_{i=1}^n \big| C_{21,i,\ell_1} C_{21,i,\ell_2} \big| \\
&= t_n^{-2} \Big| \big(\Psi'(\widehat{\mathcal{M}}_{1,\ell_1,n}) - \Psi'(\mathcal{M}^*_{1,\ell_1,P_n}) \big) \Big\{ \frac{1}{n} \sum_{i=1}^n \big[\kappa\big(\hat{\gamma}_{1,\ell_1,n}(\hat{\theta}_{1,n})' m_1(W_i, \hat{\theta}_{1,n}) g_{\ell_1}(X_i) \big) - \widehat{\mathcal{M}}_{1,\ell_1,n} \big] \\
&\quad \times \big[\kappa\big(\hat{\gamma}_{1,\ell_2,n}(\hat{\theta}_{1,n})' m_1(W_i, \hat{\theta}_{1,n}) g_{\ell_2}(X_i) \big) - \widehat{\mathcal{M}}_{1,\ell_2,n} \big] \Big\} \big(\Psi'(\widehat{\mathcal{M}}_{1,\ell_2,n}) - \Psi'(\mathcal{M}^*_{1,\ell_2,P_n}) \big) \Big| \\
&= t_n^{-2} \cdot O_p(n^{-1/2} r_n^{-d_x/2}) \cdot O_p(1) \cdot O_p(n^{-1/2} r_n^{-d_x/2}) = O_p(n^{-1} r_n^{-d_x} t_n^{-2}) = o_p(1),
\end{aligned}
\tag{B.9}
$$

where $\Psi'(\widehat{\mathcal{M}}_{1,\ell_1,n}) - \Psi'(\mathcal{M}^*_{1,\ell_1,P_n}) = O_p(n^{-1/2} r_n^{-d_x/2})$ by (B.8) and the sample average is $O_p(1)$, and the last equality follows from Assumption 5.4 and the definition of t_n. Similar arguments as above show that

$t_n^{-2}n^{-1}\sum_{i=1}^n \left|C_{21,i,\ell_1}C_{23,i,\ell_2}\right| = o_p(1)$, $t_n^{-2}n^{-1}\sum_{i=1}^n \left|C_{23,i,\ell_1}C_{21,i,\ell_2}\right| = o_p(1)$, and $t_n^{-2}n^{-1}\sum_{i=1}^n \left|C_{23,i,\ell_1}C_{23,i,\ell_2}\right| = o_p(1)$. Also, note that

$$
\begin{aligned}
& t_n^{-2}\frac{1}{n}\sum_{i=1}^n \left|C_{21,i,\ell_1}C_{22,i,\ell_2}\right| \\
&= t_n^{-2}\Big|\big(\Psi'(\widehat{\mathcal{M}}_{1,\ell_1,n}) - \Psi'(M^*_{1,\ell_1,P_n})\big)\Big\{\frac{1}{n}\sum_{i=1}^n \big[\kappa\big(\hat\gamma_{1,\ell_1,n}(\hat\theta_{1,n})'m_1(W_i,\hat\theta_{1,n})g_{\ell_1}(X_i)\big) - \widehat{\mathcal{M}}_{1,\ell_1,n}\big] \\
&\quad \times \big[\Psi'(\mathcal{M}^*_{1,\ell_1,P_n})\big[\kappa\big(\hat\gamma_{1,\ell_2,n}(\hat\theta_{1,n})'m_1(W_i,\hat\theta_{1,n})g_{\ell_2}(X_i)\big) - \kappa\big(\gamma^*_{1,\ell_2,P_n}(\theta^*_{1,n})'m_1(W_i,\theta^*_{1,n})g_{\ell_2}(X_i)\big)\big]\Big\}\Big| \\
&= t_n^{-2}\Big|\big(\Psi'(\widehat{\mathcal{M}}_{1,\ell_1,n}) - \Psi'(M^*_{1,\ell_1,P_n})\big)\Big\{\frac{1}{n}\sum_{i=1}^n \big[\kappa\big(\hat\gamma_{1,\ell_1,n}(\hat\theta_{1,n})'m_1(W_i,\hat\theta_{1,n})g_{\ell_1}(X_i)\big) - \widehat{\mathcal{M}}_{1,\ell_1,n}\big] \\
&\quad \times \big[\Psi'(\mathcal{M}^*_{1,\ell_1,P_n})\big[\frac{\partial\kappa\big(\tilde\gamma_1'm_1(W_i,\tilde\theta_1)g_{\ell_2}(X_i)\big)}{\partial\phi_1}\big]\Big\}(\hat\phi_{1,\ell_2,n}(\hat\theta_{1,n}) - \phi^*_{1,\ell_2,P_n}(\theta^*_{1,n}))\Big| \\
&= t_n^{-2}\cdot O_p(n^{-1/2}r_n^{-d_x/2})\big(O_p(n^{-1/2}) + O_p(n^{-1/2}r_n^{-d_x/2})\big) = o_p(1),
\end{aligned}
\tag{B.10}
$$

where $\phi^*_{s,\ell,P_n}(\theta_s) = (\gamma^*_{s,\ell,P_n}(\theta_s)', \theta_s')'$, the second equality holds by a mean-value expansion and the last line holds by similar arguments in the proof of Lemma A.7 and in (B.9). By similar arguments, we have $t_n^{-2}n^{-1}\sum_{i=1}^n \left|C_{22,i,\ell_1}C_{21,i,\ell_2}\right| = o_p(1)$, $t_n^{-2}n^{-1}\sum_{i=1}^n \left|C_{22,i,\ell_1}C_{23,i,\ell_2}\right| = o_p(1)$, and $t_n^{-2}n^{-1}\sum_{i=1}^n \left|C_{23,i,\ell_1}C_{22,i,\ell_2}\right| = o_p(1)$. Last, we have

$$
\begin{aligned}
& t_n^{-2}\frac{1}{n}\sum_{i=1}^n \left|C_{22,i,\ell_1}C_{22,i,\ell_2}\right| \\
&= t_n^{-2}\Big|\Big\{\frac{1}{n}\sum_{i=1}^n \big[\Psi'(\mathcal{M}^*_{1,\ell_1,P_n})\kappa\big(\hat\gamma_{1,\ell_1,n}(\hat\theta_{1,n})'m_1(W_i,\hat\theta_{1,n})g_{\ell_1}(X_i)\big) - \kappa\big(\gamma^*_{1,\ell_1,P_n}(\theta^*_{1,n})'m_1(W_i,\theta^*_1)g_{\ell_1}(X_i)\big)\big] \\
&\quad \times \big[\Psi'(\mathcal{M}^*_{1,\ell_2,P_n})\big[\kappa\big(\hat\gamma_{1,\ell_2,n}(\hat\theta_{1,n})'m_1(W_i,\hat\theta_{1,n})g_{\ell_2}(X_i)\big) - \kappa\big(\gamma^*_{1,\ell_2,P_n}(\theta^*_{1,n})'m_1(W_i,\theta^*_1)g_{\ell_2}(X_i)\big)\big]\Big\}\Big| \\
&= t_n^{-2}\Big|\Psi'(\mathcal{M}^*_{1,\ell_1,P_n})\big((\hat\phi_{1,\ell_1,n}(\hat\theta_{1,n}) - \phi^*_{1,\ell_1,P_n}(\theta^*_1))'\big)\Big\{\frac{1}{n}\sum_{i=1}^n\big[\frac{\partial\kappa\big(\ddot\gamma_1'm_1(W_i,\ddot\theta_1)g_{\ell_1}(X_i)\big)}{\partial\phi_1}\big]' \\
&\quad \times \big[\frac{\partial\kappa\big(\tilde\gamma_1'm_1(W_i,\tilde\theta_1)g_{\ell_2}(X_i)\big)}{\partial\phi_1}\big]\Big\}\big((\hat\phi_{1,\ell_2,n}(\hat\theta_{1,n}) - \phi^*_{1,\ell_2,P_n}(\theta^*_{1,n}))'\big) \\
&= t_n^{-2}\cdot O_p(n^{-1}r_n^{-d_x}) = o_p(1),
\end{aligned}
\tag{B.11}
$$

where the second equality holds by a mean-value expansion and the last line holds by similar arguments in the proof of Lemma A.7. These results hold uniformly over $\ell_1,\ell_2\in\mathcal{L}_{r_n}$ and they imply that $t_n^{-2}\frac{1}{n}\sum_{i=1}^n C_{2,i}^2 = o_p(1)$. Similarly, $t_n^{-2}\frac{1}{n}\sum_{i=1}^n C_{3,i}^2 = o_p(1)$. Part (a) follows.

For part (b), note that under Assumption 5.2(vii) and by Lyapounov central limit theorem, we have $\omega_{P_n}^{-1}n^{1/2}(\widehat{LR}_n - LR_{P_n}) = n^{-1/2}\sum_{i=1}^n \omega_{P_n}^{-1}\Lambda^*_{P_n,i} + o_p(1) \to_d N(0,1)$. Next, by similar argument for part (a) with ω_n^{-2} in place of t_n^{-2}, we have $\omega_{P_n}^{-2}\hat\omega_n^2 = \omega_{P_n}^{-2}n^{-1}\sum_{i=1}^n(\Lambda^*_{P_n,i})^2 + o_p(1)$. And by the fact that $t_n^{-2}\omega^2_{P_n}\to\infty$. Also, $\omega_{P_n}^{-2}n^{-1}\sum_{i=1}^n(\Lambda^*_{P_n,i})^2 \to 1$ by the law of large number. Therefore, (b) follows.■

Proof of Theorem 5.1: To show both part (a) and part (b), it suffices to show that for any subsequence $\{a_n\}$ of $\{n\}$, and any $\{P_{a_n}\in\mathcal{F}\}_{n\ge1}$, there exists a further subsequence $\{u_n\}$ of $\{a_n\}$ such that

$$
\lim_{n\to\infty} E_{P_{u_n}}\varphi_n(\alpha) = \alpha, \tag{B.12}
$$

for $\varphi_n(\alpha) = \varphi_n^{\text{2-sided}}(\alpha)$ or $\varphi_n^{\text{1-sided}}(\alpha)$. By the completeness of the real line, there is always a subsequence $\{h_n\}$ of $\{u_n\}$ such that $t_{h_n}^{-2}\omega_{P_{h_n}}^2 \to v_\infty$ for some $v_\infty \in [0,\infty]$. We discuss two cases below.

<u>Case 1:</u> $v_\infty \in [0,\infty)$. In this case, we have

$$\begin{aligned}\widehat{T}_{h_n} &= \frac{\sqrt{h_n}\widehat{LR}_{h_n} + \hat{\sigma}_{h_n}U}{\sqrt{\hat{\omega}_{P_{h_n}}^2 + \hat{\sigma}_{h_n}^2}} = \frac{\hat{\sigma}_{h_n}^{-1}\sqrt{h_n}\widehat{LR}_{h_n} + U}{\sqrt{\hat{\sigma}_{h_n}^{-2}\hat{\omega}_{h_n}^2 + 1}} \\ &\to_d U \sim N(0,1),\end{aligned} \tag{B.13}$$

where the convergence holds by Assumption 5.5(ii) and Lemma 5.1(a). Thus, in this case, (B.12) holds.

<u>Case 2:</u> $v_\infty = \infty$. In this case, by Lemma 5.1(b) and Assumption 5.5(iii), we have $\Pr_{P_{h_n}}(\sqrt{h_n}(\widehat{LR} - LR_{P_{h_n}}) \leq x|\hat{\sigma}_{h_n}) \to \Phi(x)\ \forall x \in R,\ a.s.$ and $\Pr_{P_{h_n}}(|\hat{\omega}_{h_n}/\omega_{h_n} - 1| > \varepsilon|\hat{\sigma}_{h_n}) \to 0, \forall\varepsilon > 0\ a.s.$. Define the subset Ω_0 of the underlying probability space as

$$\begin{aligned}\Omega_0 := \Big\{o \in \Omega : &\Pr_{P_{h_n}}(\sqrt{h_n}(\widehat{LR} - LR_{P_{h_n}}) \leq x|\hat{\sigma}_{h_n} = \hat{\sigma}_{h_n}(o)) \to \Phi(x)\ \forall x \in R, \text{ and} \\ &\Pr_{P_{h_n}}(|\hat{\omega}_{h_n}/\omega_{h_n} - 1| > \varepsilon|\hat{\sigma}_{h_n} = \hat{\sigma}_{h_n}(o)) \to 0, \forall\varepsilon > 0.\Big\}\end{aligned} \tag{B.14}$$

and

$$\Sigma_0 \equiv \Big\{\{\sigma_{h_n}\}_{n\geq 1} : \{\sigma_{h_n}\}_{n\geq 1} = \{\hat{\sigma}_{h_n}(o)\}_{n\geq 1} \text{ for some } o \in \Omega_0\Big\}. \tag{B.15}$$

Clearly, $P(\Omega_0) = 1$. Thus, it suffices to show that for any point $\{\sigma_{h_n}\}_{n\geq 1} \in \Sigma_0$ and any real number x,

$$\lim_{n\to\infty} \Pr_{P_{h_n}}(\widehat{T}_{h_n} \leq x|\hat{\sigma}_{h_n} = \sigma_{h_n}) = \Phi(x). \tag{B.16}$$

By the property of limits, it suffices to show that for any subsequence of $\{h_n\}$, there exists a further subsequence $\{\zeta_n\}$ such that the above equality holds with $\{h_n\}$ replaced by $\{\zeta_n\}$. We prove this sufficient condition next.

Note that for any subsequence of $\{h_n\}$, there exists a further subsequence $\{\zeta_n\}$ such that $\sigma_{\zeta_n}/\omega_{P_{\zeta_n}} \to c \in [0,\infty]$. By definition of Ω_0 and Σ_0, $\Pr_{P_{\zeta_n}}(\sqrt{\zeta_n}(\widehat{LR}_{\zeta_n} - LR_{P_{\zeta_n}})/\omega_{P_{\zeta_n}} \leq x|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \to \Phi(x)\ \forall x \in R$ and $\Pr_{P_{\zeta_n}}(|\hat{\omega}_{\zeta_n}/\omega_{P_{\zeta_n}} - 1| > \varepsilon|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \to 0, \forall\varepsilon > 0$.

If $c \in [0,\infty)$, then by the continuous mapping theorem

$$\begin{aligned}\Pr_{P_{\zeta_n}}(\widehat{T}_{\zeta_n} \leq x|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) &= \Pr_{P_{\zeta_n}}\left(\frac{\sqrt{\zeta_n}(\widehat{LR}_{\zeta_n} - LR_{P_{\zeta_n}})/\omega_{P_{\zeta_n}} + \sigma_{\zeta_n}U/\omega_{P_{\zeta_n}}}{\sqrt{\hat{\omega}_{\zeta_n}^2/\omega_{P_{\zeta_n}}^2 + \sigma_{\zeta_n}^2/\omega_{P_{\zeta_n}}^2}} \leq x\right) \\ &\to \Pr\left(\frac{Z + cU}{\sqrt{1+c^2}} \leq x\right) = \Phi(x),\end{aligned} \tag{B.17}$$

where $Z \sim N(0,1)$ and Z is independent of U. Then the sufficient condition mentioned above holds.

Similarly, if $c = \infty$,

$$\Pr_{P_{\zeta_n}}(\widehat{T}_{\zeta_n} \leq x|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \to \Pr(U \leq x) = \Phi(x). \tag{B.18}$$

Then the sufficient condition also holds.

Therefore, in case 2, (B.12) holds as well.■

Proof of Lemma 5.2: It is trivial to see that $\hat{\sigma}_{a_n}$ is independent of U and $\hat{\sigma}_{a_n} \in [0,1]$.

To check Assumption 5.5(ii), note that if $t_{a_n}^{-2}\omega_{P_{a_n}}^2 \to v_\infty \in [0,\infty)$, then $t_{b_{a_n}}^{-2}\omega_{P_{a_n}}^2 \to 0$ and by Lemma 5.1(a), we have $t_{b_{a_n}}^{-2}(\hat{\omega}_{b_{a_n}}^*)^2 \to_p 0$ and this implies that $\hat{\sigma}_{a_n} \equiv \exp(-t_{b_{a_n}}^{-2}(\hat{\omega}_{b_{a_n}}^*)^2) \to 1$. Therefore, $\hat{\sigma}_{a_n}$ satisfies Assumption 5.5(ii).

To check Assumption 5.5(iii), first from Lemma A.8,

$$
\begin{aligned}
\frac{\sqrt{a_n}(\widehat{LR}_{a_n} - LR_{P_{a_n}})}{\omega_{P_{a_n}}} &= \frac{1}{\sqrt{a_n}}\sum_{i=1}^{a_n}\frac{\Lambda_{P_{a_n},i}^*}{\omega_{P_{a_n}}} + o_p(1) \\
&= \frac{1}{\sqrt{a_n}}\sum_{i=b_{a_n}+1}^{a_n}\frac{\Lambda_{P_{a_n},i}^*}{\omega_{P_{a_n}}} + \frac{1}{\sqrt{a_n}}\sum_{i=1}^{b_{a_n}}\frac{\Lambda_{P_{a_n},i}^*}{\omega_{P_{a_n}}} + o_p(1) \\
&=: A_{1n} + A_{2n},
\end{aligned}
\tag{B.19}
$$

where $A_{1n} = \frac{1}{\sqrt{a_n}}\sum_{i=b_{a_n}+1}^{a_n}\frac{\Lambda_{P_{a_n},i}^*}{\omega_{P_{a_n}}}$ and $A_{2n} = \frac{\sqrt{a_n}(\widehat{LR}_{a_n} - LR_{P_{a_n}})}{\omega_{P_{a_n}}} - A_{1n}$.

Note that

$$
\frac{1}{\sqrt{a_n}}\sum_{i=1}^{b_{a_n}}\frac{\Lambda_{P_{a_n},i}^*}{\omega_{P_{a_n}}} = \frac{\sqrt{b_{a_n}}}{\sqrt{a_n}}\frac{1}{\sqrt{b_{a_n}}}\sum_{i=1}^{b_{a_n}}\frac{\Lambda_{P_{a_n},i}^*}{\omega_{P_{a_n}}} = o(1)\cdot O_P(1) = o_p(1), \tag{B.20}
$$

Thus, $A_{2n} \to_p 0$. This implies

$$
\Pr_{P_{a_n}}(|A_{2n}| > \varepsilon|\hat{\sigma}_{a_n}) \to_p 0 \ \forall\varepsilon > 0, \tag{B.21}
$$

by an application of the Markov inequality.

Because A_{1n} and $\hat{\sigma}_n$ are computed from separate samples, they are independent to each other. Also by Lemma A.8(i), we have

$$
A_{1n} \to N(0,1). \tag{B.22}
$$

Thus, $\Pr_{P_{a_n}}(A_{1n} \le x|\hat{\sigma}_{a_n}) = \Pr_{P_{a_n}}(A_{1n} \le x) \to \Phi(x)\ \forall x \in R$. Combining this with (B.21) and using an arguments similar to those that prove the continuous mapping theorem, we get

$$
\Pr_{P_{a_n}}\left(\frac{\sqrt{a_n}(\widehat{LR}_{a_n} - LR_{P_{a_n}})}{\omega_{P_{a_n}}} \le x|\hat{\sigma}_{a_n}\right) \to_p \Phi(x)\ \forall x \in R. \tag{B.23}
$$

Hence, the first statement of Assumption 5.5(iii) is proved. The second statement of Assumption 5.5(iii) follows from similar arguments. ■

Proof of Theorem 6.1: Let ω_* and μ_* denote ω_{P_*} and LR_{P_*} respectively. We need to consider two cases: $\omega_* > 0$ and $\omega_* = 0$. We first consider $\omega_* > 0$. When $\omega_* > 0$, we have $t_{b_n}^{-2}\omega_*^2 \to \infty$ and by Lemma 5.1, $t_{b_n}^{-2}\hat{\omega}_{b_n}^2 \to \infty$ and this implies that $\hat{\sigma}_n \xrightarrow{p} 0$ because $\hat{\sigma}_n = \exp(-t_{b_n}^{-2}\hat{\omega}_{b_n}^2)$. Also, $t_n^{-2}\omega_*^2 \to \infty$ and by Lemma 5.1, $\hat{\omega}_n/\omega_* \xrightarrow{p} 1$. Therefore,

$$
n^{-1/2}\widehat{T}_n = \frac{\widehat{LR}_n + n^{-1/2}\hat{\sigma}_n U}{\sqrt{\hat{\omega}_n^2 + \hat{\sigma}_n^2}} \xrightarrow{p} \mu_*/\omega_* > 0,
$$

so $\widehat{T}_n \to \infty$. This implies that $E_{P_*}\varphi_n(\alpha) = 1$.

Second, we consider $\omega_* = 0$. When $\omega_* = 0$, $t_n^{-2}\omega_*^2 = 0$, then by Lemma 5.1(a), $\hat{\omega}_n^2 = o_p(1)$. Also, by Assumption 5.5(ii), we have $\hat{\sigma}_n^2 \to_p 1$. Therefore, $\sqrt{n}\mu_*/\sqrt{\hat{\omega}_n^2 + \hat{\sigma}_n^2} \to_p \infty$. Also, by the same argument for the Case 1 of Theorem 5.1, we have

$$\frac{\sqrt{n}(\widehat{LR}_n - \mu_*) + \hat{\sigma}_n U}{\sqrt{\hat{\omega}_n^2 + \hat{\sigma}_n^2}} \to_d U = O_p(1).$$

Therefore,

$$\widehat{T}_n = \frac{\sqrt{n}(\widehat{LR}_n - \mu_*) + \hat{\sigma}_n U}{\sqrt{\hat{\omega}_n^2 + \hat{\sigma}_n^2}} + \frac{\sqrt{n}\mu_*}{\sqrt{\hat{\omega}_n^2 + \hat{\sigma}_n^2}} \to \infty.$$

This implies that $E_{P_*}\varphi_n(\alpha) = 1$. This completes the proof.■

Proof of Theorem 6.2: Since part (b) follows from similar arguments as part (a), for brevity, we only show part (a).

Let $\{u_n\}_{n\geq 1}$ be a subsequence of $\{n\}$ such that

$$\lim_{n\to\infty} \Pr_{P_{u_n,*}}(|\widehat{T}_{u_n}| > z_{\alpha/2}) = \liminf_{n\to\infty} \Pr_{P_{n,*}}(|\widehat{T}_n| > z_{\alpha/2}). \tag{B.24}$$

Note that such $\{u_n\}$ always exists by the definition of lim inf. By the completeness of the real line, there is always a subsequence $\{a_n\}$ of $\{u_n\}$ such that $t_{a_n}^{-2}\omega_{P_{a_n}}^2 \to v_\infty$ for some $v_\infty \in [0, \infty]$. We discuss two cases below.

Case 1: $v_\infty \in [0, \infty)$. In this case, we have

$$\begin{aligned}
\widehat{T}_{a_n} &= \frac{\sqrt{a_n}(\widehat{LR}_{a_n} - LR_{P_{a_n,*}}) + \sqrt{a_n}LR_{P_{a_n,*}} + \hat{\sigma}_{h_n} U}{\sqrt{\hat{\omega}_{P_{a_n}}^2 + \hat{\sigma}_{a_n}^2}} \\
&= \frac{\hat{\sigma}_{a_n}^{-1}\sqrt{a_n}\widehat{LR}_{a_n} + \hat{\sigma}_{a_n}^{-1}\sqrt{a_n}LR_{P_{a_n,*}} + U}{\sqrt{\hat{\sigma}_{a_n}^{-2}\hat{\omega}_{a_n}^2 + 1}} \\
&\to_p h_1 + U.
\end{aligned} \tag{B.25}$$

where the second equality holds by Assumption 5.5(ii) and Lemma 5.1(a). Thus,

$$\begin{aligned}
\lim_{n\to\infty} \Pr_{P_{a_n,*}}(|\widehat{T}_{a_n}| > z_{\alpha/2}) &= \Pr(|U + h_1| > z_{\alpha/2}) \\
&= 1 - \Phi(z_{\alpha/2} - h_1/\sqrt{\omega_\infty^2 + 1}) + \Phi(-z_{\alpha/2} - h_1/\sqrt{\omega_\infty^2 + 1}),
\end{aligned} \tag{B.26}$$

where the second equality holds because $\omega_\infty^2 = 0$ by $t_{a_n}^{-2}\omega_{P_{a_n}}^2 \to v_\infty < \infty$.

Case 2: $v_\infty = \infty$. In this case, we follow similar arguments as those in Case 2 of the proof of Theorem 5.1. First let the sets Ω_0 and Σ_0 be defined as in (B.14) and (B.15), respectively. It suffices to show that for any point $\{\sigma_{a_n}\}_{n\geq 1} \in \Sigma_0$,

$$\lim_{n\to\infty} \Pr_{P_{a_n,*}}(|\widehat{T}_{a_n}| \geq z_{\alpha/2} | \hat{\sigma}_{a_n} = \sigma_{a_n}) \geq 1 - \Phi(z_{\alpha/2} - h_1/\sqrt{\omega_\infty^2 + 1}) + \Phi(-z_{\alpha/2} - h_1/\sqrt{\omega_\infty^2 + 1}). \tag{B.27}$$

By the property of limits, it suffices to show that for any subsequence of $\{a_n\}$, there exists a further subsequence $\{\zeta_n\}$ such that the above equality holds with $\{a_n\}$ replaced by $\{\zeta_n\}$. We prove this sufficient condition next.

Note that for any subsequence of $\{a_n\}$, there exists a further subsequence $\{\zeta_n\}$ such that $\sigma_{\zeta_n} \to \sigma_\infty$ and $\sigma_{\zeta_n}/\omega_{P_{\zeta_n}} \to c$ for some $\sigma_\infty \in [0,1]$ and $c \in [0,\infty]$. By the definition of Ω_0 and Σ_0, $\Pr_{P_{\zeta_n}}(\sqrt{\zeta_n}(\widehat{LR}_{\zeta_n} - LR_{P_{\zeta_n}})/\omega_{P_{\zeta_n}} \le x|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \to \Phi(x)\ \forall x \in R$ and $\Pr_{P_{\zeta_n}}(|\hat{\omega}_{\zeta_n}/\omega_{P_{\zeta_n}} - 1| > \varepsilon|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \to 0, \forall \varepsilon > 0$.

If $c \in [0,\infty)$ and $\omega_\infty > 0$, then by the continuous mapping theorem

$$\begin{aligned}
&\Pr_{P_{\zeta_n},*}(|\widehat{T}_{\zeta_n}| \ge z_{\alpha/2}|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \\
&= \Pr_{P_{\zeta_n},*}\left(\left|\frac{\sqrt{\zeta_n}(\widehat{LR}_{\zeta_n} - LR_{P_{\zeta_n}})/\omega_{P_{\zeta_n}} + \sqrt{\zeta_n}LR_{P_{\zeta_n}}/\omega_{P_{\zeta_n}} + \sigma_{\zeta_n}U/\omega_{P_{\zeta_n}}}{\sqrt{\hat{\omega}^2_{\zeta_n}/\omega^2_{P_{\zeta_n}} + \sigma^2_{\zeta_n}/\omega^2_{P_{\zeta_n}}}}\right| \ge z_{\alpha/2}\middle|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}\right) \\
&\to \Pr\left(\left|\frac{Z + h_1/\omega_\infty + cU}{\sqrt{1+c^2}}\right| \ge z_{\alpha/2}\right) \\
&\ge 1 - \Phi(z_{\alpha/2} - h_1/\sqrt{\omega^2_\infty + 1}) + \Phi(-z_{\alpha/2} - h_1/\sqrt{\omega^2_\infty + 1}), \qquad \text{(B.28)}
\end{aligned}$$

where $Z \sim N(0,1)$ and Z is independent of U and the last equality holds because $h_1/(\omega_\infty\sqrt{1+c^2}) = h_1/\sqrt{\omega^2_\infty + c^2\omega^2_\infty}$ and $c^2\omega^2_\infty \le 1$ because it is the limit of σ_{ζ_n}. Then (B.27) holds.

If $c \in [0,\infty)$ and $\omega_\infty = 0$, then

$$\Pr_{P_{\zeta_n},*}(|\widehat{T}_{\zeta_n}| \ge z_{\alpha/2}|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \to 1, \qquad \text{(B.29)}$$

which also implies (B.27).

If $c = \infty$, we must have $\omega_\infty = 0$ because c is the limit of $\sigma_{\zeta_n}/\omega_{P_{\zeta_n}}$ and $\sigma_{\zeta_n} \le 1$. And

$$\begin{aligned}
&\Pr_{P_{\zeta_n},*}(|\widehat{T}_{\zeta_n}| \ge z_{\alpha/2}|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}) \\
&= \Pr_{P_{\zeta_n},*}\left(\left|\frac{\sqrt{\zeta_n}(\widehat{LR}_{\zeta_n} - LR_{P_{\zeta_n}})/\sigma_{\zeta_n} + \sqrt{\zeta_n}LR_{P_{\zeta_n}}/\sigma_{\zeta_n} + U}{\sqrt{\hat{\omega}^2_{\zeta_n}/\sigma^2_{\zeta_n} + 1}}\right| \ge z_{\alpha/2}\middle|\hat{\sigma}_{\zeta_n} = \sigma_{\zeta_n}\right) \\
&\to \Pr\left(|h_1/\sigma_\infty + U| \ge z_{\alpha/2}\right) \\
&\ge \Pr(|h_1 + U| > z_{\alpha/2}) \\
&= 1 - \Phi(z_{\alpha/2} - h_1/\sqrt{\omega^2_\infty + 1}) + \Phi(-z_{\alpha/2} - h_1/\sqrt{\omega^2_\infty + 1}), \qquad \text{(B.30)}
\end{aligned}$$

which also implies (B.27).

Therefore, in case 2, the desired result holds as well.■

C Proof of Auxiliary Lemmas

Proof of Lemma A.1: Lemma A.1 is the same as in Shi (2009a).■

Proof of Lemma A.2: Let E_n abbreviate E_{P_n}. For $s = 1$ and 2, define new pseudo-metrics ϱ_{s0} and ϱ_{s1} on $\Phi_s \times \mathcal{L}$ as:

$$\begin{aligned}
&\varrho_{s0}((\phi_{s1},\ell_1),(\phi_{s2},\ell_2)) \\
&= \sup_{n\ge1}[E_n\left[\kappa(\gamma'_{s1}m_{s,}(W,\theta_{s1})g_{\ell_1}(X)) - \kappa(\gamma'_{s2}m_{s,}(W,\theta_{s2})g_{\ell_2}(X))\right]^2]^{1/2}, \qquad \text{(C.1)}
\end{aligned}$$

$$
\begin{aligned}
&\varrho_{s1}((\phi_{s1},\ell_1),(\phi_{s2},\ell_2)) \\
&= \sup_{n\geq 1}[E_n\|\partial\kappa(\gamma_{s1}'m_s(W,\theta_{s1})g_{\ell_1}(X))/\partial\gamma_s - \partial\kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_2}(X)))/\partial\gamma_s\|^2]^{1/2},
\end{aligned}
\tag{C.2}
$$

where $(\phi_{s1},\ell_1),(\phi_{s2},\ell_2)\in\Phi_s\times\mathcal{L}$.

For (a), it is sufficient to show that the empirical process is stochastically equicontinuous w.r.t. ϱ_{s0} because ρ_s dominates ϱ_{s0}. Note that

$$
\begin{aligned}
&\varrho_{s0}((\phi_{s1},\ell_1),(\phi_{s2},\ell_2)) \\
&= \sup_{n\geq 1}[E_n\big(\kappa(\gamma_{s1}'m_s(W,\theta_{s1})g_{\ell_1}(X)) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_2}(X))\big)^2]^{1/2} \\
&\leq \sup_{n\geq 1}\Big\{E_n 2[\kappa(\gamma_{s1}'m_s(W,\theta_{s1})g_{\ell_1}(X)) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_1}(X))]^2 \\
&\quad + 2[\kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_1}(X)) - \kappa(\gamma_{s2}'m_{s,\ell_2}(W,\theta_{s2})g_{\ell_2}(X))]^2\Big\}^{1/2} \\
&\leq \sqrt{2}\sup_{n\geq 1}\Big\{E_n[\kappa(\gamma_{s1}'m_s(W,\theta_{s1})g_{\ell_1}(X)) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_1}(X))]^2\Big\}^{1/2} \\
&\quad + \sqrt{2}\sup_{n\geq 1}\Big\{E_n[\kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_1}(X)) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_2}(X))]^2\Big\}^{1/2}.
\end{aligned}
\tag{C.3}
$$

The first inequality holds because $(a+b)^2\leq 2(a^2+b^2)$. The second inequality follows because $(a+b)\leq(\sqrt{a}+\sqrt{b})^2$ when $a,b\geq 0$, and $\sup_{n\geq 1}(a_n+b_n)\leq\sup_{n\geq 1}a_n+\sup_{n\geq 1}b_n$ for any two sequences. Note that

$$
\begin{aligned}
&\sup_{n\geq 1}\Big\{E_n[\kappa(\gamma_{s1}'m_s(W,\theta_{s1})g_{\ell_1}(X)) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_1}(X))]^2\Big\}^{1/2} \\
&= \sup_{n\geq 1}\Big\{E_n g_{\ell_1}(X)[\kappa(\gamma_{s1}'m_s(W,\theta_{s1})) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2}))]^2\Big\}^{1/2} \\
&\leq \sup_{n\geq 1}\Big\{E_n[\kappa(\gamma_{s1}'m_s(W,\theta_{s1})) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2}))]^2\Big\}^{1/2} \\
&= \sup_{n\geq 1}\Big\{E_n\big((\phi_{s1}-\phi_{s2})(\partial\kappa(\tilde{\gamma}_s'm_s(W,\tilde{\theta}_s))/\partial\phi_s)\big)^2\Big\}^{1/2} \\
&\leq \|\phi_{s1}-\phi_{s2}\|\cdot\sup_{n\geq 1}\Big\{E_n\big\|\partial\kappa(\tilde{\gamma}_s'm_s(W,\tilde{\theta}_s))/\partial\phi_s\big\|^2\Big\}^{1/2} \leq M^{1/2}\|\phi_{s1}-\phi_{s2}\|,
\end{aligned}
\tag{C.4}
$$

where $(\tilde{\gamma}_s',\tilde{\theta}_s')$ lies on the line segment joining ϕ_{s1} and ϕ_{s2}. Note that $g_\ell(X)$ is an indicator function and $\kappa(0)=0$, so $\kappa(\gamma_s'm_s(W,\theta_s)g_\ell(X)) = g_\ell(X)\kappa(\gamma_s'm_s(W,\theta_s))$ and the first equality holds. The second equality holds by a mean-value expansion. The second inequality holds by the Cauchy-Schwartz inequality and the last inequality holds by Assumption 5.2(vi). Note that (C.4) holds uniformly over $\ell_1\in\mathcal{L}$. Also,

$$
\begin{aligned}
&\sup_{n\geq 1}\Big\{[\kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_1}(X)) - \kappa(\gamma_{s2}'m_s(W,\theta_{s2})g_{\ell_2}(X))]^2\Big\}^{1/2} \\
&= \sup_{n\geq 1}\Big\{E_n\Big[(g_{\ell_1}(X)-g_{\ell_2}(X))^2\kappa(\gamma_{s2}'m_s(W,\theta_{s2}))^2\Big]\Big\}^{1/2} \\
&\leq \sup_{n\geq 1}\Big\{E_n\Big[(g_{\ell_1}(X)-g_{\ell_2}(X))^2\sup_{\phi_s\in\Phi_s}\kappa(\gamma_s'm_s(W,\theta_s))^2\Big]\Big\}^{1/2} \\
&= \sup_{n\geq 1}\Big\{E_n\Big[(g_{\ell_1}(X)-g_{\ell_2}(X))^2 E_n\big[\sup_{\phi_s\in\Phi_s}\kappa(\gamma_s'm_s(W,\theta_s))^2\big|X]\Big]\Big\}^{1/2}
\end{aligned}
$$

$$\leq M^{1/2}\sup_{n\geq 1}\left\{E_n(g_{\ell_1}(X)-g_{\ell_2}(X))^2\right\}^{1/2}$$
$$= M^{1/2}\sup_{n\geq 1}\left\{E_n|g_{\ell_1}(X)-g_{\ell_2}(X)|\right\}^{1/2}\leq M\rho_\ell(\ell_1,\ell_2). \tag{C.5}$$

The first inequality follows by taking supremum over $\phi_s\in\Phi_s$. The third equality holds by law of iterated expectations. The second inequality holds by Assumption 5.2(vii). The last equality holds by the fact that g_{ℓ_1} and g_{ℓ_2} are indicator functions. The last inequality holds because

$$\begin{aligned} E_n[|g_{\ell_1}(X)-g_{\ell_2}(X)|] &= \int_{\mathcal{X}}|g_{\ell_1}(x)-g_{\ell_2}(x)|f_n(x)dx \\ &= \int_{C_{\ell_1}\triangle C_{\ell_2}} f_n(x)dx\leq M\lambda(C_{\ell_1}\triangle C_{\ell_2}) = M\rho_\ell(\ell_1,\ell_2)^2, \end{aligned} \tag{C.6}$$

where the first equality follows from the fact that P_x has density $f_n(x)$, the second equality holds by the fact that $|g_{\ell_1}(x)-g_{\ell_2}(x)| = 1(C_{\ell_1}\triangle C_{\ell_2})$, the first inequality holds by Assumption 5.2(viii), and the last equality holds by the definition of ρ_ℓ. Also, (C.5) holds uniformly over $\phi_s\in\Phi_s$. Therefore, (C.3), (C.4) and (C.5) together imply that

$$\begin{aligned} \varrho_{s0}((\phi_{s1},\ell_1),(\phi_{s2},\ell_2)) &\leq KM^{1/2}\|\phi_{s1}-\phi_{s2}\| + KM\rho_\ell(\ell_1,\ell_2) \\ &\leq C(\|\phi_{s1}-\phi_{s2}\|+\rho_\ell(\ell_1,\ell_2)) = C\cdot\rho_s((\phi_{s1},\ell_1),(\phi_{s2},\ell_2)), \end{aligned} \tag{C.7}$$

for some $C>0$.

To show the stochastic equicontinuity w.r.t. ϱ_{s0}, we apply the results in Andrews (1994). Recall that $\kappa(\gamma_s' m_s(W,\theta_s)g_\ell(X)) = g_\ell(X)\kappa(\gamma_s' m_s(W,\theta_s))$. Because $\mathcal{G}=\{g_\ell(\cdot):\ell\in\mathcal{L}\}$ is a class of functions of Vapnik-Červonenkis sets, then $\mathcal{G}$ is a type I classes of functions with envelope function 1. $\{\kappa(\gamma_s' m_s(\cdot,\theta_s)):\phi_s\in\Phi_s\}$ is a type II class because Φ_s is a bounded subset of the Euclidean space and $\kappa(\gamma_s' m_s(\cdot,\theta_s))$ is Lipschitz in ϕ_s:

$$|\kappa(\gamma_{s1}' m_s(\cdot,\theta_{s1}))-\kappa(\gamma_s 2' m_s(\cdot,\theta_{s2}))|\leq B(\cdot)\|\phi_{s1}-\phi_{s2}\|, \tag{C.8}$$

where $B(\cdot)=\sup_{\phi_s\in\Phi_s}\|\partial\kappa(\gamma_s' m(\cdot,\theta_s))/\partial\phi_s\|$. Hence, by Theorem 2 of Andrews (1994), $\{\kappa(\gamma_s' m_s(\cdot,\theta_s)):\phi_s\in\Phi_s\}$ satisfies Pollard's entropy condition with envelope $F(\cdot)\equiv 1\vee\sup_{\phi_s\in\Phi_s}|\kappa(\gamma_s' m_s(\cdot,\theta_s))|\vee B(\cdot)$. Hence, by Theorem 3 of Andrews (1994), $\{\kappa(\gamma_s' m_s(W,\theta_s)g_\ell(X)) = g_\ell(X_i)\cdot\kappa(m_s(W,\theta_s)):\phi_s\in\Phi_s,\ell\in\mathcal{L}\}$ satisfies Pollard's entropy conditional with envelope function $F(\cdot)$. Note that

$$\begin{aligned} &\limsup_{n\to\infty} E_n[1\vee\sup_{\phi_s\in\Phi_s}|\kappa(\gamma_s' m_s(W,\theta_s))|\vee B(X)]^{2+\delta} \\ \leq&\limsup_{n\to\infty} E_n[1+\sup_{\phi_s\in\Phi_s}|\kappa(\gamma_s' m_s(W,\theta_s))|+B(X)]^{2+\delta}, \\ \leq& C\cdot E_n[1+\sup_{\phi_s\in\Phi_s}|\kappa(\gamma_s' m_s(W,\theta_s))|^{2+\delta}] + E_n[\sup_{\phi_s\in\Phi_s}\|\partial\kappa(\gamma_s' m(W,\theta_s))/\partial\phi_s\|^{2+\delta}]<\infty \end{aligned} \tag{C.9}$$

for some $C>0$. The second inequality holds by the convexity of the function $f(x)=x^{2+\delta}$ and the last inequality holds by Assumption 5.2(vi). Therefore, by Theorem 1 in Andrews (1994), $\nu^0_{s,n}(\phi_s,\ell)$ is stochastically equicontinuous w.r.t. ϱ_{s0}.

For (b), it is sufficient to show that the metric space $(\Phi_s\times\mathcal{L},\rho_s)$ is totally bounded and $\nu^0_{s,n}(\phi_s,\ell)$ is $o_p(1)$ for all $(\phi_s,\ell)\in\Phi_s\times\mathcal{L}$. To show that $(\Phi_s\times\mathcal{L},\rho_s)$ is totally bounded, it suffices to show that both $(\Phi_s,\|\cdot\|)$ and

$(\mathcal{L}, \rho_\ell)$ are totally bounded. $(\Phi_s, \|\cdot\|)$ is totally bounded because Φ_s is compact set with Euclidean metric. To see that $(\mathcal{L}, \rho_\ell)$ is totally bounded, let $\{X_i : i = 1, \ldots\}$ be a sequence of *i.i.d.* uniform random variables over $\mathcal{X}$. Then the triangular array $\{f_{ni}(\omega, \ell) = g_\ell(X_i)/\sqrt{n}\}$ satisfies the conditions of Functional Central Limit Theorem of Pollard (1990, Theorem 10.6) and pseudo-metric on $\mathcal{L}$ is

$$[E(g_{\ell_1}(X) - g_{\ell_2}(X))^2]^{1/2} = \left[\frac{\lambda(C_{\ell_1} \triangle C_{\ell_2})}{\lambda(\mathcal{X})}\right]^{1/2} = K \cdot \rho_\ell(\ell_1, \ell_2) \tag{C.10}$$

for $K = \lambda(\mathcal{X})^{1/2}$. That is, $[E(g_{\ell_1}(X) - g_{\ell_2}(X))^2]^{1/2}$ is equivalent to $\rho_\ell(\ell_1, \ell_2)$. By Theorem 10.6 (a) of Pollard (1990), $(\mathcal{L}, [E(g_{\ell_1}(X) - g_{\ell_2}(X))^2]^{1/2})$ is totally bounded and this implies that $(\mathcal{L}, \rho_\ell)$ is totally bounded.

$\nu^0_{s,n}(\phi_s, \ell) = O_p(1)$ for each $(\phi_s, \ell) \in \Phi_s \times \mathcal{L}$ because

$$E_n \nu^0_{s,n}(\phi_s, \ell)^2 = E_n\big[\frac{1}{\sqrt{n}} \sum_{i=1}^n [\kappa(\gamma_s' m_s(W_i, \theta_s) g_\ell(X_i)) - E_n \kappa(\gamma_s' m_s(W, \theta_s) g_\ell(X))]\big] \tag{C.11}$$

$$= E_n(\kappa(\gamma_s' m_s(W, \theta_s) g_\ell(X)) - E_n \kappa(\gamma_s' m_s(W, \theta_s) g_\ell(X))^2 \tag{C.12}$$

$$\leq E_n \kappa(\gamma_s' m_s(W, \theta_s))^2 < \infty, \tag{C.13}$$

where the second equality holds by the i.i.d. assumption and the last inequality holds by Assumption 5.2(vi).

The proofs for (c) and (d) are similar to (a) and (b) respectively, and we omit them for brevity.

For (e), we show that $\sup_{\ell \in \mathcal{L}} |\widehat{\mathcal{M}}_{s,\ell,n}(\phi^{(1)}_{s,n}(\ell)) - \mathcal{M}_{s,\ell,P_n}(\phi^{(2)}_{s,n}(\ell))| \to_p 0$ and the proofs for other three convergence results are similar. The proof is done by showing that

$$\sup_{\phi_s \in \Phi_s, \ell \in \mathcal{L}} |\widehat{\mathcal{M}}_{s,\ell,n}(\phi_s) - \mathcal{M}_{s,\ell,P_n}(\phi_s)| \to_p 0 \tag{C.14}$$

$$\sup_{\ell \in \mathcal{L}} |\mathcal{M}_{s,\ell,P_n}(\phi^{(1)}_{s,n}(\ell)) - \mathcal{M}_{s,\ell,P_n}(\phi^{(2)}_{s,n}(\ell))| \to_p 0. \tag{C.15}$$

To show (C.14), we use the uniform weak law of large number in Andrews and Shi (2011, Lemma E2). Consider the triangular array of processes $\{g_\ell(X)\kappa(\gamma_s' m_s(\cdot, \theta_s)) : \phi_s \in \Phi_s, \ell \in \mathcal{L}, i \leq n, n \geq 1\}$. From part (a), it is manageable w.r.t. the envelope functions $F(\cdot) = 1 \vee \sup_{\phi_s \in \Phi_s} \kappa(\gamma_s' m_s(\cdot, \theta_s)) \vee B(\cdot)$ such that $\sup_{n \geq 1} n^{-1} \sum_{i=1}^n E_n F_{n,i}^{1+\delta} < \infty$. Therefore, by Lemma E2 of Andrews and Shi (2011), (C.14) holds.

To show (C.15), note that

$$\begin{aligned}
|\mathcal{M}_{s,\ell,P_n}(\phi_{s1}) - \mathcal{M}_{s,\ell,P_n}(\phi_{s2})| &= |E_n[g_\ell(X)\kappa(\gamma_{s1}' m_s(W, \theta_{s1})) - g_\ell(X)\kappa(\gamma_{s2}' m_s(W, \theta_{s2}))]| \\
&\leq E_n|g_\ell(X)\kappa(\gamma_{s1}' m_s(W, \theta_{s1})) - g_\ell(X)\kappa(\gamma_{s2}' m_s(W, \theta_{s2}))| \\
&= E_n|g_\ell(X)[\kappa(\gamma_{s1}' m_s(W, \theta_{s1})) - \kappa(\gamma_{s2}' m_s(W, \theta_{s2}))]| \\
&\leq E_n|\kappa(\gamma_{s1}' m_s(W, \theta_{s1})) - \kappa(\gamma_{s2}' m_s(W, \theta_{s2})| \\
&= E_n|(\partial\kappa(\tilde{\gamma}_s' m_s(W, \tilde{\theta}_s))/\partial\phi_s)(\phi_{s1} - \phi_{s2})| \\
&\leq \|\phi_{s1} - \phi_{s2}\| \cdot E_n\|(\partial\kappa(\tilde{\gamma}_s' m_s(W, \tilde{\theta}_s))/\partial\phi_s)\| \\
&\leq \|\phi_{s1} - \phi_{s2}\| \cdot E_n \sup_{\phi_s \in \Phi_s} \|(\partial\kappa(\gamma_s' m_s(W, \theta_s))/\partial\phi_s)\| \\
&\leq C \cdot \|\phi_{s1} - \phi_{s2}\|,
\end{aligned} \tag{C.16}$$

for some $0 < C < \infty$. The first equality holds by the definition of $\mathcal{M}_{s,\ell,P_n}$, the second equality holds by the fact that $|E_n Y| \le E_n|Y|$ for any random variable Y, the second inequality holds because g_ℓ is an indicator function, the third equality holds by a mean-value expansion, the third inequality holds by the Cauchy-Schwartz inequality and the last inequality holds by Assumption 5.2(vi). (C.16) holds uniformly over $\ell \in \mathcal{L}$, so (C.15) follows.■

Proof of Lemma A.3: Let $\hat{\gamma}^M_{s,\ell,n}(\theta_s) = \arg\max_{\gamma_s \in \Gamma^s_M} \widehat{\mathcal{M}}_{s,\ell,n}(\gamma_s, \theta_s)$. For Lemma A.3(a), it is sufficient to show that $\sup_{\theta_s \in \Theta_s, \ell \in \mathcal{L}_{r_n}} \|\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)\| = o_p(1)$ because by Assumption 5.2(vi) $\widehat{\mathcal{M}}_{s,\ell,n}(\gamma_s, \theta_s)$ is strictly concave in γ_s and by Assumption 5.2(v) $\|\gamma^*_{s,\ell,P_n}(\theta_s)\| \le M - \delta$. Also, define $\hat{\phi}^M_{s,\ell,n}(\theta_s) = (\hat{\gamma}^M_{s,\ell,n}(\theta_s), \theta_s)$

First, for all $\ell \in \mathcal{L}$,

$$\begin{aligned}
&\mathcal{M}_{s,\ell,P_n}(\hat{\phi}^M_{s,\ell,n}(\theta_s)) - \mathcal{M}_{s,\ell,P_n}(\phi^*_{s,\ell,P_n}(\theta_s)) \\
&= \Big(\frac{\partial \mathcal{M}_{s,\ell,P_n}(\phi^*_{s,\ell,P_n}(\theta_s))}{\partial \gamma_s}\Big)(\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)) \\
&\quad + 2^{-1}(\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s))' \Big(\frac{\partial^2 \mathcal{M}_{s,\ell,P_n}(\tilde{\gamma}_n, \theta_s)}{\partial \gamma_s \partial \gamma_s'}\Big)(\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)) \\
&\le 2^{-1}(\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s))' \Big(\frac{\partial^2 \mathcal{M}_{s,\ell,P_n}(\tilde{\gamma}_n, \theta_s)}{\partial \gamma_s \partial \gamma_s'}\Big)(\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)) \\
&\le -2^{-1} p_{P_0,\ell} \cdot \delta \cdot \|\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)\|^2,
\end{aligned} \tag{C.17}$$

where $\tilde{\gamma}_n$ lies between $\hat{\gamma}^M_{s,\ell,n}(\theta_s)$ and $\gamma^*_{s,\ell,P_n}(\theta_s)$ and δ is in condition Assumption 5.2(vi). The first inequality holds by By Lemmas A.9 and A.11, which apply because $\gamma^*_{s,\ell,P_n}(\theta_s)$ is the solution to $\max_{\gamma_s \in \Gamma_s(\theta_s)} \mathcal{M}_{s,\ell,P_n}(\gamma_s, \theta_s)$. The second inequality holds by Lemma A.5 and the fact that

$$\frac{\partial^2 \mathcal{M}_{s,\ell,P_n}(\tilde{\gamma}_n, \theta_s)}{\partial \gamma_s \partial \gamma_s'} = E_{P_0}\Big[\kappa''(\tilde{\gamma}_n' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)' g_\ell(X)\Big]. \tag{C.18}$$

For any $\varepsilon > 0$,

$$\begin{aligned}
&P_n\Big(\sup_{\theta_s \in \Theta_s, \ell \in \mathcal{L}_{r_n}} \|\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)\| > \varepsilon\Big) \\
\le &P_n\Big(\|\hat{\gamma}^M_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_s)\| > \varepsilon\Big) \\
\le &P_n\Big(\mathcal{M}_{s,\ell_n,P_n}(\hat{\phi}^M_{s,\ell_n,n}(\theta_{s,n})) - \mathcal{M}_{s,\ell_n,P_n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n})) \le -2^{-1} p_{P_{0,n},\ell_n} \delta \varepsilon^2\Big) \\
= &P_n\Big(\mathcal{M}_{s,\ell_n,P_n}(\hat{\phi}^M_{s,\ell_n,n}(\theta_{s,n})) - \widehat{\mathcal{M}}_{s,\ell_n,n}(\hat{\phi}^M_{s,\ell_n,n}(\theta_{s,n})) \\
&\quad + \widehat{\mathcal{M}}_{s,\ell_n,n}(\hat{\phi}^M_{s,\ell_n,n}(\theta_{s,n})) - \widehat{\mathcal{M}}_{s,\ell_n,n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n})) \\
&\quad + \widehat{\mathcal{M}}_{s,\ell_n,n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n})) - \mathcal{M}_{s,\ell_n,P_n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n})) \le -2^{-1} p_{P_{0,n},\ell_n} \delta \varepsilon^2\Big) \\
\le &P_n\Big(\sqrt{n}\big(\mathcal{M}_{s,\ell_n,P_n}(\hat{\phi}^M_{s,\ell_n,n}(\theta_{s,n})) - \widehat{\mathcal{M}}_{s,\ell_n,n}(\hat{\phi}^M_{s,\ell_n,n}(\theta_{s,n}))\big) \\
&\quad + \sqrt{n}\big(\widehat{\mathcal{M}}_{s,\ell_n,n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n})) - \mathcal{M}_{s,\ell_n,P_n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n}))\big) \le -2^{-1}\sqrt{n} p_{P_{0,n},\ell_n} \cdot \delta \cdot \varepsilon^2\Big) \\
\to &0.
\end{aligned} \tag{C.19}$$

The first inequality holds after we pick a sequence $\{\ell_n \in \mathcal{L}_{r_n}, \theta_{s,n} \in \Theta_s\}_{n=1}^\infty$ such that $\|\hat{\gamma}^M_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})\| \ge \sup_{\theta_s \in \Theta_s, \ell \in \mathcal{L}_{r_n}} \|\hat{\gamma}^M_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell}(\theta_s)\| - 2^{-n}$. The second inequality holds by (C.17) and the third inequality holds by the definition of $\hat{\gamma}^{M}_{s,\ell_n,n}(\theta_{s,n})$.The convergence holds by the fact that the l.h.s. term in the last $P_n(\cdot)$ is $O_p(1)$ by Lemma A.2(b) and the r.h.s. term diverges to negative infinity by Assumption 5.4 and $\sqrt{n}p_{P_{0,n},\ell_n} \geq \sqrt{n}Cr_n^{d_x}$.

For Lemma A.3(b), let $\{\ell_n \in \mathcal{L}_{r_n}, \theta_{s,n} \in \Theta_s\}_{n=1}^{\infty}$ be a random sequence such that $p_{P_{0,n},\ell_n}\|\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})\| \geq \sup_{\theta_s\in\Theta_s,\ell\in\mathcal{L}_{r_n}} p_{P_{0,n},\ell}\|\hat{\gamma}_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,n,n}(\theta_s)\| - 2^{-n}$. Then,

$$
\begin{aligned}
0 \leq & \Big(\widehat{\mathcal{M}}_{s,\ell_n,n}(\hat{\phi}_{s,\ell_n,n}(\theta_{s,n})) - \widehat{\mathcal{M}}_{s,\ell_n,n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n}))\Big) \\
\leq & \Big(\frac{\partial \widehat{\mathcal{M}}_{s,\ell_n,n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n}))}{\partial \gamma_s'}\Big)(\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})) \\
& + (\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n}))'\Big(\frac{\partial^2 \widehat{\mathcal{M}}_{s,\ell_n,n}(\tilde{\gamma}_n,\theta_{s,n})}{\partial\gamma_s\partial\gamma_s'}\Big)(\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})) \\
\leq & \left(\frac{\partial \widehat{\mathcal{M}}_{s,\ell_n,n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n}))}{\partial \gamma_s'} - \frac{\partial \mathcal{M}_{s,\ell_n,n}(\phi^*_{s,\ell_n,P_n}(\theta_{s,n}))}{\partial \gamma_s'}\right)(\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})) \\
& + (\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n}))'\Big(\frac{\partial^2 \widehat{\mathcal{M}}_{s,\ell_n,n}(\tilde{\gamma}_n,\theta_{s,n})}{\partial\gamma_s\partial\gamma_s'}\Big)(\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})) \\
\leq & O_p(n^{-1/2})\|\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})\| - p_{P_{0,n},\ell_n}\cdot\delta\cdot\|\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})\|^2 \\
& \qquad + o_p(p_{P_{0,n},\ell_n}\cdot\delta\cdot\|\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})\|^2).
\end{aligned} \tag{C.20}
$$

where the first inequality holds by the definition of $\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n})$ and the second inequality holds by a second order Taylor expansion. The third inequality holds by applying Lemmas A.9 and A.11 to the problems $\max_{\gamma_s\in\Gamma_s(\theta_{s,n})}\mathcal{M}_{s,\ell_n,n}(\gamma_s,\theta_{s,n})$ and $\max_{\gamma_s\in\Gamma_s(\theta_s)}\widehat{\mathcal{M}}_{s,\ell_n,n}(\gamma_s,\theta_{s,n})$. For the last inequality, the $O_p(n^{-1/2})$ term follows from Lemma A.2(d). Also, by Lemma A.2(e),

$$
\begin{aligned}
& \frac{1}{p_{P_0,\ell_n}}\frac{\partial^2\widehat{\mathcal{M}}_{s,\ell_n,n}(\tilde{\gamma}_n,\theta_{s,n})}{\partial\gamma_s\partial\gamma_s'} = \frac{1}{p_{P_0,\ell_n}}\Big(\frac{\partial^2\mathcal{M}_{s,\ell_n,P_n}(\tilde{\gamma}_n,\theta_{s,n})}{\partial\gamma_s\partial\gamma_s'} + O_p(n^{-1/2})\Big) \\
& = E_{P_0}\Big[\kappa''(\tilde{\gamma}_n' m_s(W,\theta_{s,n}))m_s(W,\theta_{s,n})m_s(W,\theta_{s,n})'\big|X\in C_{\ell_n}\Big] + o_p(1),
\end{aligned} \tag{C.21}
$$

where the $o_p(1)$ follows from the fact that $\sqrt{n}p_{P_0,\ell_n}$ diverges to infinity by the same argument for (C.19). Finally, by Lemma A.5 and (C.29), the last inequality follows. Then, $p_{P_{0,n},\ell_n}\|\hat{\gamma}_{s,\ell_n,n}(\theta_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta_{s,n})\| = O_p(n^{-1/2})$ and this completes the proof. For the second part, note that it is straightforward to see that $r_n^{d_x}\sup_{\ell\in\mathcal{L}_{r_n}} p^{-1}_{P_{0,n},\ell} \leq \delta_x^{-1}$ where δ_x is defined in Assumption 5.2(viii). Therefore,

$$
r_n^{d_x}\sup_{\theta_s\in\Theta_s,\ell\in\mathcal{L}_{r_n}}\|\hat{\gamma}_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)\| \tag{C.22}
$$

$$
\leq \sup_{\theta_s\in\Theta_s,\ell\in\mathcal{L}_{r_n}} p_{P_{0,n},\ell}\|\hat{\gamma}_{s,\ell,n}(\theta_s) - \gamma^*_{s,\ell,P_n}(\theta_s)\| = O_p(n^{-1/2}), \tag{C.23}
$$

and this shows the second part.

For Lemma A.3(c), let $\{\ell_n \in \mathcal{L}\}_{i=1}^{\infty}$ be such that $\big\|\gamma^*_{s,\ell_n,P_n}(\theta^{(1)}_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta^{(2)}_{s,n})\big\| \geq \sup_{\ell\in\mathcal{L}_{r_n}}\big\|\gamma^*_{s,\ell_n,P_n}(\theta^{(1)}_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta^{(2)}_{s,n})\big\| - 2^{-n}$. Then, similar to (A.19) of Shi (2009a), we have

$$
\begin{aligned}
0 \leq & - p_{P_0,\ell_n}\cdot\delta\cdot\big\|\gamma^*_{s,\ell_n,P_n}(\theta^{(1)}_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta^{(2)}_{s,n})\big\|^2 \\
& \qquad + p_{P_0,\ell_n}\cdot O_p\big(\big\|\gamma^*_{s,\ell_n,P_n}(\theta^{(1)}_{s,n}) - \gamma^*_{s,\ell_n,P_n}(\theta^{(2)}_{s,n})\big\|\cdot\big\|\theta^{(1)}_{s,n} - \theta^{(2)}_{s,n}\big\|\big).
\end{aligned} \tag{C.24}
$$

(C.24) implies that $\sup_{\ell\in\mathcal{L}} \left\|\gamma^*_{s,\ell,P_n}(\theta^{(1)}_{s,n}) - \gamma^*_{s,\ell,P_n}(\theta^{(2)}_{s,n})\right\| = O_p\left(\|\theta^{(1)}_{s,n} - \theta^{(2)}_{s,n}\|\right)$.

Lemma A.3(d) is implied by Lemma A.3(a) and Lemma A.3(b).■

Proof of Lemma A.4: The following proof holds uniformly over $P \in \mathcal{F}$ and we simplify the notation by deleting the dependence on P. For example, we have $\mathcal{M}_{s,\ell} = \mathcal{M}_{s,\ell,P}$ and $\gamma^*_{s,\ell}(\theta_s) = \gamma^*_{s,\ell,P}(\theta_s)$. Recall that $\mathcal{M}_{s,\ell}(\gamma_s, \theta_s) = E_{P_0}[g_\ell(X)\kappa(\gamma_s' m_s(W, \theta_s))]$. As a result,

$$\mathcal{M}_{s,\ell}(\phi^*_{s,\ell}(\theta_s)) = p_{P_0,\ell} E\left[\kappa(\gamma^*_{s,\ell}(\theta_s)' m_s(W, \theta_s))|X \in C_\ell\right]. \tag{C.25}$$

By a mean-value expansion, we have

$$\begin{aligned} 0 \leq \Psi\left[\mathcal{M}_{s,\ell}(\phi^*_{s,\ell}(\theta_s))\right] &= \Psi'(\ddot{M}_{s,\ell}) \cdot \mathcal{M}_{s,\ell}(\phi^*_{s,\ell}(\theta_s)) \\ &= \Psi'(\ddot{M}_{s,\ell}) \cdot p_{P_0,\ell} \cdot E[\kappa\left(\gamma^*_{s,\ell}(\theta_s)' m_s(W, \theta_s)\right)|X \in C_\ell] \leq C r^{d_x}, \end{aligned} \tag{C.26}$$

where $\ddot{M}_{s,\ell}$ is a value between 0 and $\mathcal{M}_{s,\ell}(\phi^*_{s,\ell}(\theta_s))$, and $\ddot{M}_{s,\ell}$ is bounded uniformly over $\ell \in \mathcal{L}$ by Assumptions 5.2(iv) and (vi), and the second inequality holds due to this, the continuity of $\Psi'(\cdot)$, and Assumptions (viii). This implies that

$$0 \leq r_n^{-d_x - 1} \int_{\mathcal{L}^c_{r_n}} \Psi\left[\mathcal{M}_{s,\ell}(\phi^*_{s,\ell}(\theta_s))\right] dF(\ell) \leq C, \tag{C.27}$$

where C is a positive number not dependent on θ_s and P_0. Note that the extra r_n comes from the fact that the total weight of $F(\ell)$ over $\mathcal{L}^c_{r_n}$ is of order r_n. ■

Proof of Lemma A.5: Note that

$$\begin{aligned} &E_{P_0}[\kappa''(\gamma_s' m_s(W, \theta_s) g_\ell(X)) m_s(W, \theta_s) m_s(W, \theta_s)'] \\ =&E_{P_0}[g_\ell(X)\kappa''(\gamma_s' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)'] \\ =&p_{P_0,\ell} \cdot E_{P_0}\left[\kappa''(\gamma_s' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)' \middle| X \in C_\ell\right], \end{aligned} \tag{C.28}$$

where the second equality holds by the law of iterated expectations and the last equality follows from the definition of the conditional expectation. Hence, it suffices to show that

$$eig_{\max}\left(E_{P_0}\left[\kappa''(\gamma_s' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)' \middle| X \in C_\ell\right]\right) \leq -\delta, \tag{C.29}$$

because for all $a > 0$, $eig_{\max}(a \cdot A) = a \cdot eig_{\max}(A)$.

Suppose not, i.e., $eig_{\max}(E_{P_0}[\kappa''(\gamma_s' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)'|X \in C_\ell]) > -\delta$. Let λ be the eigenvector associated with the maximum eigenvalue and $\|\lambda\| > 0$, then

$$\lambda' E_{P_0}\left[\kappa''(\gamma_s' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)' \middle| X \in C_\ell\right]\lambda > -\delta\|\lambda\|^2. \tag{C.30}$$

However, this contradicts to the following:

$$\begin{aligned} &\lambda' E_{P_0}[\kappa''(\gamma_s' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)'|X \in C_\ell]\lambda \\ =&E_{P_0}[\lambda'\kappa''(\gamma_s' m_s(W, \theta_s)) m_s(W, \theta_s) m_s(W, \theta_s)'\lambda|X \in C_\ell] \end{aligned}$$

$$
\begin{aligned}
&= E_{P_0}\Big[\lambda' E\big[\kappa''(\gamma_s' m_s(W,\theta_s)) m_s(W,\theta_s) m_s(W,\theta_s)'|X\big]\lambda \big| X \in C_\ell\Big] \\
&\leq E_{P_0}\Big[\delta\|\lambda\|^2 \big| X \in C_\ell\Big] \leq -\delta\|\lambda\|^2. \qquad \text{(C.31)}
\end{aligned}
$$

The second equality holds by law of iterated expectations and the inequity holds because of Assumption 5.2(v). Lemma A.5 follows.■

Proof of Lemma A.6: We first claim that $P_0 \in \mathcal{P}_{s,\theta_s,\ell}$ iff $d(\mathcal{P}_{s,\theta_s,\ell}, P_0)$ defined in (3.5) is equal to 0. Note that $E_{P_0}\big[\kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)\big] = 0$ when $\gamma_s = 0$. Therefore, it is straightforward to see that $d(\mathcal{P}_{s,\theta_s,\ell}, P_0) \geq 0$. Also, by Lemma A.5, $E_{P_0}\big[\kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)\big]$ is strictly concave on Γ_s^M and by definition, $\Gamma_s(\Theta_s)$ is convex, so the solution to (3.5) is unique. As a result, to show that $P_0 \in \mathcal{P}_{s,\theta_s,\ell}$ iff $d(\mathcal{P}_{s,\theta_s,\ell}, P_0) = 0$, it is equivalent to show that $P_0 \in \mathcal{P}_{s,\theta_s,\ell}$ iff $\gamma_s = 0$ is the solution to (3.5).

First, we show that if $\gamma_s = 0$ is the solution to (3.5), then $P_0 \in \mathcal{P}_{s,\theta_s,\ell}$. Note that for $j = 1\ldots, k_s$,

$$
\frac{\partial E_{P_0}\big[\kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)\big]}{\partial \gamma_{s,j}}\Big|_{\gamma_s=0} = -E_{P_0}[m_{s,j}(X,\theta_s) g_\ell(X)]. \qquad \text{(C.32)}
$$

It is obvious that $\gamma_s = 0$ is a regular point. By Lemma A.9, if $\gamma_s = 0$ is the solution to (3.5), then there exists $\mu^* \in R_+^{k_s - p_s}$ such that

$$
\frac{\partial E_{P_0}\big[\kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)\big]}{\partial \gamma_{s,j}}\Big|_{\gamma_s=0} = 0, \quad \text{for } j = 1, \ldots, p_s, \qquad \text{(C.33)}
$$

$$
\frac{\partial E_{P_0}\big[\kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)\big]}{\partial \gamma_{s,j}}\Big|_{\gamma_s=0} + \mu^*_{j-p_s} = 0, \quad \text{for } j = p_s + 1, \ldots, k_s. \qquad \text{(C.34)}
$$

This implies that

$$
E_{P_0}[m_{s,j}(X,\theta_s) g_\ell(X)] = 0, \quad \text{for } j = 1, \ldots, p_s, \qquad \text{(C.35)}
$$

$$
E_{P_0}[m_{s,j}(X,\theta_s) g_\ell(X)] = \mu^*_{j-p_s} \geq 0, \quad \text{for } j = p_s + 1, \ldots, k_s, \qquad \text{(C.36)}
$$

i.e., $P_0 \in \mathcal{P}_{s,\theta_s,\ell}$.

We show the other direction. We apply Lemma A.10 to show that if $P_0 \in \mathcal{P}_{s,\theta_s,\ell}$, then $\gamma_s = 0$ is the solution to (3.5). Let $\mu_j^* = E_{P_0}[m_{s,p_s+j}(W,\theta_s)]$ which is greater than or equal to 0 because $P_0 \in \mathcal{P}_{s,\theta_s,\ell}$. Therefore, we have

$$
-E_{P_0}[m_{s,j}(W,\theta_s)] = \frac{\partial E_{P_0}\big[\kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)\big]}{\partial \gamma_{s,j}}\Big|_{\gamma_s=0} = 0, \qquad \text{for } j = 1, \ldots, p_s, \qquad \text{(C.37)}
$$

$$
-E_{P_0}[m_{s,j}(W,\theta_s)] + \mu^*_{j-p_s} = \frac{\partial E_{P_0}\big[\kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)\big]}{\partial \gamma_{s,j}}\Big|_{\gamma_s=0} + \mu^*_{j-p_s} = 0, \quad \text{for} j \geq p_s. \qquad \text{(C.38)}
$$

On the other hand, the $L(\lambda_s, \mu)$ for our case is

$$
L(\gamma_s, \mu) = E_{P_0}\Big[\frac{\partial^2 \kappa\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big)}{\partial \gamma_s \partial \gamma_s{}'}\Big] = E_{P_0}[\kappa''\big(\gamma_s' m_s(X,\theta_s) g_\ell(X)\big) m_s(X,\theta_s) m_s(X,\theta_s)']. \qquad \text{(C.39)}
$$

Therefore, $L(0,\mu^*) = E_{P_0}[-g_\ell(X) m_s(X,\theta_s) m_s(X,\theta_s)')]$ and $eig_{\max}(L(0,\mu^*)) < 0$ by Lemma A.5. As a result, the condition 2 of Lemma A.10 holds. This implies that $\gamma_s = 0$ is the solution to (3.5) and $d(\mathcal{P}_{s,\theta_s,\ell}, P_0) = 0$.

Next we show $P_0 \in \mathcal{P}_s$ iff $d_{\mathcal{L}}(\mathcal{P}_s, P_0) = 0$. First, if $P_0 \in \mathcal{P}_s$, then there exists $\theta_s^* \in \Theta_s$ such that $P_0 \in \mathcal{P}_{s,\theta_s^*}$. This is equivalent to that $P_0 \in \mathcal{P}_{s,\theta_s^*,\ell}$ for all $\ell \in \mathcal{L}$ and it follows that $d(\mathcal{P}_{s,\theta_s^*,\ell}, P_0) = 0$. Therefore, $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s^*}, P_0) = 0$. This implies that $d_{\mathcal{L}}(\mathcal{P}_s, P_0) = \inf_{\theta_s \in \Theta_s} d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0) = 0$.

We show the other direction. Suppose $d_{\mathcal{L}}(\mathcal{P}_s, P_0) = 0$, then there exists a sequence $\{\theta_{s,n} \in \Theta_s\}_{n=1}^{\infty}$ such that $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_{s,n}}, P_0) < 1/n$. Since Θ_s is a compact set, there exists a subsequence k_n of n such that $\theta_{s,k_n} \to \theta_s^* \in \Theta_s$ as $n \to \infty$. We first claim that $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s^*}, P_0) = 0$ and it is sufficient to show that $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0)$ is continuous in θ_s. Note that $\mathcal{M}_{s,\ell,P_0}(\gamma_s, \theta_s) = E_{P_0}[\kappa(\gamma_s' m_s(W, \theta_s) g_\ell(X))]$ is uniformly continuous on $\Gamma_M^s \times \Theta_s$. By Lemma A.3(c), $\gamma_{s,\ell}^*(\theta_s)$ is uniformly continuous in $\theta_s \in \Theta_s$. These imply that for each $\ell \in \mathcal{L}$, $\mathcal{M}_{s,\ell,P_0}(\gamma_{s,\ell}^*(\theta_s), \theta_s)$ and $d(\mathcal{P}_{s,\theta_s,\ell}, P_0) = \Psi[\mathcal{M}_{s,\ell,P_0}(\gamma_{s,\ell}^*(\theta_s), \theta_s)]$ is continuous in θ_s. Hence, for any sequence $\theta_{s,n}$ that converges θ_s, $d(\mathcal{P}_{s,\theta_{s,n},\ell}, P_0)$ converges to $d(\mathcal{P}_{s,\theta_s,\ell}, P_0)$ for each $\ell \in \mathcal{L}$. By Assumptions 5.2(iv) and (vi), $d(\mathcal{P}_{s,\theta_s,\ell}, P_0)$ is bounded above uniformly in ℓ and θ_s. Finally, by the dominated convergence theorem, it follows that for any sequence $\theta_{s,n} \to \theta_s$, $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_{s,n}}, P_0) \to d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0)$. This shows the continuity of $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s}, P_0)$ in θ_s.

Next, we show that $P_0 \in \mathcal{P}_{s,\theta_s^*}$. Suppose not, then there exists ℓ^* such that $P_0 \notin \mathcal{P}_{s,\theta_s^*,\ell^*}$ and $d(\mathcal{P}_{s,\theta_s^*,\ell^*}, P_0) > 0$. Next, by the same argument for Theorem 3 of AS, there exists $\bar{\tau} > 0$ such that for all $\ell \in \mathcal{N}_{\bar{\tau}}(\ell^*)$, $P_0 \notin \mathcal{P}_{s,\theta_s^*,\ell^*}$ and $d(\mathcal{P}_{s,\theta_s^*,\ell^*}, P_0) > 0$. Finally, it follows that $d_{\mathcal{L}}(\mathcal{P}_{s,\theta_s^*}, P_0) \geq \int_{\mathcal{N}_{\bar{\tau}}(\ell^*)} dF(\ell)(\mathcal{P}_{s,\theta_s^*,\ell}, P_0) dF(\ell) > 0$. This completes Lemma A.6.■

Proof of Lemma A.7: Let $\Pr_n$ abbreviate $\Pr_{P_n}$. For simplicity, ignore the subscript s.

We first show the consistency, that is for arbitrary sequence $\{\hat{\theta}_n \in \widehat{\Theta}_n\}_{n=1}^{\infty}$ and arbitrary $\varepsilon > 0$, we have $\Pr_n(\rho_{lh}(\hat{\theta}_n, \Theta^*(P_n)) > \varepsilon) \to 0$. Note that Assumption 5.2(ii) implies that for all $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ not dependent on P_n such that

$$\inf_{\theta \in \Theta \setminus N_\varepsilon(\Theta^*(P_n))} d_{\mathcal{L}}(\mathcal{P}_\theta, P_n) > d_{\mathcal{L}}(\mathcal{P}, P_n) + \delta_\varepsilon. \tag{C.40}$$

Let θ_n^* be the point in $\Theta^*(P_n)$ that is (approximately) closest to $\hat{\theta}_n$, that is to say $\|\hat{\theta}_n - \theta_n^*\|^2 \leq \rho_{lh}^2(\hat{\theta}_n, \Theta_n^*) + 2^{-n}$. Then, the consistency is proved by the following derivation:

$$\begin{aligned}
&\Pr_n(\rho_{lh}(\hat{\theta}_n, \Theta^*(P_n)) > \varepsilon) \\
\leq &\Pr_n(d_{\mathcal{L}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - d_{\mathcal{L}}(\mathcal{P}_{\theta_n^*}, P_n) > \delta_\varepsilon) \\
= &\Pr_n\Big(\big[d_{\mathcal{L}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n)\big] + \big[d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - \hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n))\big] \\
&\quad + \big[\hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n)) - \hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta_n^*}, P_n))\big] \\
&\quad + \big[\hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta_n^*}, P_n)) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta_n^*}, P_n))\big] + \big[d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta_n^*}, P_n)) - d_{\mathcal{L}}(\mathcal{P}_{\theta_n^*}, P_n)\big] > \delta_\varepsilon \Big) \\
= &\Pr_n\Big(o_p(1) + o_p(1) + \big[\hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n)) - \hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta_n^*}, P_n))\big] + o_p(1) + o_p(1) > \delta_\varepsilon \Big) \\
\leq &\Pr_n\Big(o_p(1) + o_p(1) + o_p(1) + o_p(1) > \delta_\varepsilon \Big) \to 0,
\end{aligned} \tag{C.41}$$

where the first inequality holds by (C.40), the second equality holds by Lemma A.4 and Lemma A.2(b) and the second inequality holds by the definition of $\widehat{\Theta}_n$.

Next, we show the convergence rate. Let $\hat{\theta}_n$ and θ_n^* be the same as above. And for any measurable subset A of $\mathcal{L}$, let $\hat{d}_A(\mathcal{P}_\theta, P_n) = \int_A \Psi(\widehat{\mathcal{M}}_{\ell,n}(\hat{\gamma}_{\ell,n}(\theta), \theta)) dF(\ell)$. Let $\widehat{\mathcal{M}}_{A,n}(\phi^*(\theta)) = \int_A \Psi(\widehat{\mathcal{M}}_{\ell,n}(\gamma_{\ell,P_n}^*(\theta), \theta)) dF(\ell)$ and

$\mathcal{M}_{A,n}(\hat{\phi}(\theta)) = \int_A \Psi(\mathcal{M}_{\ell,n}(\hat{\gamma}_{\ell,n}(\theta),\theta))dF(\ell)$. Below, we show that

$$\begin{aligned}
&(1a)\ [\hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - \widehat{\mathcal{M}}_{\mathcal{L}_{r_n},n}(\phi^*(\theta_n^*))] - [\mathcal{M}_{\mathcal{L}_{r_n},n}(\hat{\phi}(\hat{\theta}_n)) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta_n^*}, P_n)] \\
&\qquad = O_p(n^{-1}r_n^{-d_x}) + O_p(n^{-1/2})\|\hat{\theta}_n - \theta_n^*\| + o_p(1)\cdot\|\hat{\theta}_n - \theta_n^*\|^2 \\
&(1b)\ [\hat{d}_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - \widehat{\mathcal{M}}_{\mathcal{L}_{r_n},n}(\phi^*(\theta_n^*))] \le O_p(n^{-1}r_n^{-d_x}) \\
&(1c)\ [\mathcal{M}_{\mathcal{L}_{r_n},n}(\hat{\phi}(\hat{\theta}_n)) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta_n^*}, P_n)] \ge O_p(n^{-1}r_n^{-d_x}) + \delta\cdot\big((\|\hat{\theta}_n - \theta_n^*\|^2 - 2^{-n})\wedge\delta\big),
\end{aligned} \tag{C.42}$$

where δ is the positive number in Assumption 5.2. The three conditions in (C.42) imply that

$$\begin{aligned}
O_p(n^{-1}r_n^{-d_x}) \ge & O_p(n^{-1/2})\cdot\|\hat{\theta}_n - \theta_n^*\| + o_p(1)(\|\hat{\theta}_n - \theta_n^*\|) \\
& + \delta\cdot\big((\|\hat{\theta}_n - \theta_n^*\|^2 - 2^{-n})\wedge\delta\big),
\end{aligned} \tag{C.43}$$

and this further implies that $\|\hat{\theta}_n - \theta_n^*\| = O_p(n^{-1/2}r_n^{-d_x/2})$.

We first show that (1a) holds. First, we have the l.h.s. of (1a) equals

$$\begin{aligned}
&\int_{\mathcal{L}_{r_n}} \Big\{\big[\Psi\big(\widehat{\mathcal{M}}_{\ell,n}(\hat{\phi}_{\ell,n}(\hat{\theta}_n))\big) - \Psi\big(\widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta_n^*))\big)\big] \\
&\qquad - \big[\Psi\big(\mathcal{M}_{\ell,n}(\hat{\phi}_{\ell,n}(\hat{\theta}_n))\big) - \Psi\big(\mathcal{M}_{\ell,P_n}(\phi^*_{\ell,P_n}(\theta_n^*))\big)\big]\Big\}d\ell \\
&= \int_{\mathcal{L}_{r_n}} \Big\{\Big(\hat{\phi}_{\ell,n}(\hat{\theta}_n) - \phi^*_{\ell,P_n}(\theta_n^*)\Big)' \times \\
&\Bigg[\Psi'\big(\widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta_n^*))\big)\left(\frac{\partial\widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta_n^*))}{\partial\phi'}\right) - \Psi'\big(\mathcal{M}_{\ell,P_n}(\phi^*_{\ell,P_n}(\theta_n^*))\big)\left(\frac{\partial\mathcal{M}_{\ell,P_n}(\phi^*_{\ell,P_n}(\theta_n^*))}{\partial\phi'}\right)\Bigg] + \\
&\Big(\hat{\phi}_{\ell,n}(\hat{\theta}_n) - \phi^*_{\ell,P_n}(\theta_n^*)\Big)'\left(\frac{\partial^2\Psi(\widehat{\mathcal{M}}_{\ell,P_n}(\bar{\phi}_{\ell,n}))}{\partial\phi\partial\phi'} - \frac{\partial^2\Psi(\mathcal{M}_{\ell,P_n}(\bar{\phi}_{\ell,n}))}{\partial\phi\partial\phi'}\right)\Big(\hat{\phi}_{\ell,n}(\hat{\theta}_n) - \phi^*_{\ell,P_n}(\theta_n^*)\Big)\Big\}dF(\ell)
\end{aligned} \tag{C.44}$$

by a second order Taylor expansion, where $\bar{\phi}_{\ell,n}$ is some values lying on the line segment joining $\hat{\phi}_{\ell,n}(\hat{\theta})$ and $\phi^*_{\ell,n}(\theta_n^*)$. Together, the first summand of the expression inside the integral in the r.h.s. of (C.44) is

$$O_p(n^{-1/2})\|\hat{\phi}_{\ell,n}(\hat{\theta}_n) - \phi^*_{\ell,P_n}(\theta_n^*)\| \tag{C.45}$$

due to a combination of Lemma A.2(b) and (d), twice-continuously differentiability of $\Psi(\cdot)$ and Assumptions 5.2(vii)-(viii). Now note that

$$\begin{aligned}
\|\hat{\phi}_{\ell,n}(\hat{\theta}_n) - \phi^*_{\ell,P_n}(\theta_n^*)\| \le & \|\hat{\gamma}_{\ell,n}(\hat{\theta}_n) - \gamma^*_{\ell,P_n}(\theta_n^*)\| + \|\hat{\theta}_n - \theta_n^*\| \\
\le & \|\hat{\gamma}_{\ell,n}(\hat{\theta}_n) - \gamma^*_{\ell,P_n}(\hat{\theta}_n)\| + \|\gamma^*_{\ell,P_n}(\hat{\theta}_n) - \gamma^*_{\ell,P_n}(\theta_n^*)\| + \|\hat{\theta}_n - \theta_n^*\| \\
= & \|\hat{\gamma}_{\ell,n}(\hat{\theta}_n) - \gamma^*_{\ell,P_n}(\hat{\theta}_n)\| + O_p(\|\hat{\theta}_n - \theta_n^*\|) + \|\hat{\theta}_n - \theta_n^*\| \\
= & O_p(n^{-1/2}r_n^{-d_x}) + O_p(\|\hat{\theta}_n - \theta_n^*\|) + \|\hat{\theta}_n - \theta_n^*\| \\
= & O_p(n^{-1/2}r_n^{-d_x}) + O_p(\|\hat{\theta}_n - \theta_n^*\|),
\end{aligned} \tag{C.46}$$

where the first two inequalities hold by triangular inequalities, the first equality follows by Lemma A.3(c) and the second equality follows by Lemma A.3(b).

Now we study the second summand of the expression inside the integral in the r.h.s. of (C.44). Using the first three lines of (C.46), we have that the absolute value of this second summand is bounded by

$$
\begin{aligned}
&\bar{K} \cdot \|\hat{\phi}_{\ell,n}(\hat{\theta}_n) - \phi^*_{\ell,P_n}(\theta^*_{\ell,n})\|^2 \cdot \left\| \frac{\partial^2 \Psi(\widehat{\mathcal{M}}_{\ell,n}(\bar{\phi}_{\ell,n}))}{\partial\phi\partial\phi'} - \frac{\partial^2 \Psi(\mathcal{M}_{\ell,P_n}(\bar{\phi}_{\ell,n}))}{\partial\phi\partial\phi'} \right\| \\
&= (O_p(n^{-1}r_n^{-2d_x}) + O_p(\|\hat{\theta}_n - \theta^*_n\|^2))O_p(n^{-1/2}) \\
&= o_p(n^{-1}r_n^{-d_x}) + O_p(n^{-1/2})O_p(\|\hat{\theta}_n - \theta^*_n\|^2),
\end{aligned} \tag{C.47}
$$

where $\bar{K}$ is a positive constant, the first equality holds by (C.46) and Lemma A.2(d) and the twice continuous differentiability of $\Psi(\cdot)$, and the second equality holds by Assumption 5.4.

The O_p and o_p terms in the above three displays are uniform over $\ell \in \mathcal{L}_{r_n}$. Thus, together, they imply that condition (1a) in C.42 holds.

For (1b) in (C.42), we have that the l.h.s. of the condition equals

$$
\begin{aligned}
&\int_{\mathcal{L}_{r_n}} \Big\{\Psi\big(\widehat{\mathcal{M}}_{\ell,n}(\hat{\phi}_{\ell,n}(\hat{\theta}_n))\big) - \Psi\big(\widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta^*_n))\big)\Big\} dF(\ell) \\
&\leq \int_{\mathcal{L}_{r_n}} \Big\{\Psi\big(\widehat{\mathcal{M}}_{\ell,n}(\hat{\phi}_{\ell,n}(\theta^*_n))\big) - \Psi\big(\widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta^*_n))\big)\Big\} dF(\ell) \\
&= \int_{\mathcal{L}_{r_n}} \Big\{\Psi'(\ddot{\mathcal{M}}_{\ell,n})\big(\widehat{\mathcal{M}}_{\ell,n}(\hat{\phi}_{\ell,n}(\theta^*_n))\big) - \widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta^*_n))\big)\Big\} dF(\ell),
\end{aligned} \tag{C.48}
$$

where the first inequality holds because $\hat{\theta}_{s,n}$ is a minimizer of the problem and the second equality holds by mean-expansions. Note that

$$
\begin{aligned}
&\widehat{\mathcal{M}}_{\ell,n}(\hat{\phi}_{\ell,n}(\theta^*_n)) - \widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta^*_n)) \\
&= \Big[\frac{\partial \widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta^*_n))}{\partial\gamma'}\Big][\hat{\gamma}_{\ell,n}(\theta^*_n) - \gamma^*_{\ell,P_n}(\theta^*_n)] \\
&\quad + [\hat{\gamma}_{\ell,n}(\theta^*_n) - \gamma^*_{\ell,P_n}(\theta^*_n)]'\Big[\frac{\partial^2 \widehat{\mathcal{M}}_{s,\ell,n}(\ddot{\phi}_\ell)}{\partial\gamma\partial\gamma'}\Big][\hat{\gamma}_{\ell,n}(\theta^*_n) - \gamma^*_{\ell,P_n}(\theta^*_n)] \\
&= \Big[\frac{\partial \widehat{\mathcal{M}}_{\ell,n}(\phi^*_\ell(\theta^*_n))}{\partial\gamma'}\Big][\hat{\gamma}_{\ell,n}(\theta^*_n) - \gamma^*_{\ell,P_n}(\theta^*_n)] + O_p(n^{-1}r_n^{-d_x}) \\
&\leq \Big[\frac{\partial \widehat{\mathcal{M}}_{\ell,n}(\phi^*_{\ell,P_n}(\theta^*_n))}{\partial\gamma'} - \frac{\partial \mathcal{M}_{\ell,P_n}(\phi^*_{\ell,P_n}(\theta^*_n))}{\partial\gamma'}\Big][\hat{\gamma}_{\ell,n}(\theta^*_n) - \gamma^*_{\ell,P_n}(\theta^*_{s,n})] + O_p(n^{-1}r_n^{-d_x}) \\
&= O_p(n^{-1}r_n^{-d_x}) + O_p(n^{-1}r_n^{-d_x}) = O_p(n^{-1}r_n^{-d_x}).
\end{aligned} \tag{C.49}
$$

Given that $\Psi'(\ddot{\mathcal{M}}_\ell) = O_p(1)$, (C.49) is sufficient for (1b).

For (1c) in (C.42), we first have that the l.h.s. of this condition equals

$$
\begin{aligned}
&[\mathcal{M}_{\mathcal{L}_{r_n},n}(\hat{\phi}(\hat{\theta}_n)) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n)] - [d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\theta^*_n}, P_n)] \\
&= [\mathcal{M}_{\mathcal{L}_{r_n},n}(\hat{\phi}(\hat{\theta}_n)) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n)] - [d_{\mathcal{L}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - d_{\mathcal{L}}(\mathcal{P}_{\theta^*_n}, P_n)] \\
&\quad + [d_{\mathcal{L}^c_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n) - d_{\mathcal{L}^c_{r_n}}(\mathcal{P}_{\theta^*_n}, P_n)] \\
&\geq [\mathcal{M}_{\mathcal{L}_{r_n},n}(\hat{\phi}(\hat{\theta}_n)) - d_{\mathcal{L}_{r_n}}(\mathcal{P}_{\hat{\theta}_n}, P_n)] + \delta \cdot [(\|\hat{\theta}_n - \theta^*_n\|^2 + 2^{-n}) \wedge \delta] \\
&\quad + [\mathcal{M}_{\mathcal{L}^c_{r_n},n}(\phi^*(\hat{\theta}_n)) - d_{\mathcal{L}^c_{r_n}}(\mathcal{P}_{\theta^*_n}, P_n)]
\end{aligned}
$$

$$= O_p(n^{-1}r_n^{-d_x}) + \delta \cdot [(\|\hat{\theta}_n - \theta_n^*\|^2 + 2^{-n}) \wedge \delta] + [\mathcal{M}_{\mathcal{L}_{r_n}^c,n}(\phi^*(\hat{\theta}_n)) - d_{\mathcal{L}_{r_n}^c}(\mathcal{P}_{\theta_n^*}, P_n)], \tag{C.50}$$

where the first inequality holds by Assumption 5.2(iii) and the second equality holds by similar arguments as those for condition (1b) in (C.42). Also,

$$\begin{aligned}
[d_{\mathcal{L}_{r_n}^c}(\mathcal{P}_{\hat{\theta}_n}, P_n) - d_{\mathcal{L}_{r_n}^c}(\mathcal{P}_{\theta_n^*}, P_n)] &= \int_{\mathcal{L}_{r_n}^c} \Big\{\Psi\big(\mathcal{M}_{\ell,P_n}(\phi_{\ell,P_n}^*(\hat{\theta}_n))\big) - \Psi\big(\mathcal{M}_{\ell,P_n}(\phi_{\ell,P_n}^*(\theta_n^*))\big)\Big\} dF(\ell) \\
&= \int_{\mathcal{L}_{r_n}^c} \Psi'\big(\ddot{\mathcal{M}}_{\ell,n}\big)[\mathcal{M}_{\ell,P_n}(\phi_{\ell,P_n}^*(\hat{\theta}_n)) - \mathcal{M}_{\ell,P_n}(\phi_{\ell,P_n}^*(\theta_n^*))] dF(\ell) \\
&= \int_{\mathcal{L}_{r_n}^c} O(r_n^{d_x}) O_p(\|\hat{\theta}_n - \theta_n^*\|) dF(\ell) \\
&= O(r_n^{d_x+1}) \cdot O_p(\|\hat{\theta}_n - \theta_n^*\|) \\
&= o(n^{-1/2}) \cdot O_p(\|\hat{\theta}_n - \theta_n^*\|),
\end{aligned} \tag{C.51}$$

where the second equality holds by a mean value expansion. The third equality in (C.51) holds because for each $\ell \in \mathcal{L}_{r_n}^c$

$$\begin{aligned}
[\mathcal{M}_{\ell,P_n}(\phi_{\ell,P_n}^*(\hat{\theta}_n)) - \mathcal{M}_{\ell,P_n}(\phi_{\ell,P_n}^*(\theta_n^*))] &= \Big[\frac{\partial \mathcal{M}_{\ell,P_n}(\ddot{\phi})}{\partial \phi'}\Big][\phi_{\ell,P_n}^*(\hat{\theta}_n) - \phi_{\ell,P_n}^*(\theta_n^*)] \\
&= p_{\ell,n}^{-1}\Big[\frac{\partial \mathcal{M}_{\ell,P_n}(\ddot{\phi}_\ell)}{\partial \phi'}\Big] p_{\ell,n}[\phi_{\ell,P_n}^*(\hat{\theta}_n) - \phi_{\ell,P_n}^*(\theta_n^*)] \\
&= O_p(1) \cdot O(r_n^{d_x}) \cdot O_p(\|\hat{\theta}_n - \theta_n^*\|),
\end{aligned} \tag{C.52}$$

where the first equality holds by a mean-value expansion, the $O_p(1)$ in the last equality holds by Assumption 5.2(viii), $p_{\ell,n} = O(r_n^{d_x})$ for all $\ell \in \mathcal{L}_{r_n}^c$ and

$$\begin{aligned}
\|\phi_{\ell,P_n}^*(\hat{\theta}_n) - \phi_{\ell,P_n}^*(\theta_n^*)\| &\le \|\gamma_{\ell,P_n}^*(\hat{\theta}_{\ell,n}) - \gamma_{\ell,P_n}^*(\theta_n^*)\| + \|\hat{\theta}_n - \theta_n^*\| \\
&= O_p(\|\hat{\theta}_n - \theta_n^*\|) + \|\hat{\theta}_n - \theta_s^*\| = O_p(\|\hat{\theta}_n - \theta_n^*\|)
\end{aligned} \tag{C.53}$$

where the last line holds by Lemma A.3(c). The fourth equality in (C.51) holds because the total weight of $F(\ell)$ over $\mathcal{L}_{r_n}^c$ is $O(r_n)$ by Assumption 5.4 and the last equality in (C.51) holds by Assumption 5.4. Therefore, (1c) holds.■

Proof of Lemma A.8: Rewrite $\sqrt{n}(\widehat{LR}_n - LR_{P_n})$ as

$$\begin{aligned}
&\sqrt{n}(\widehat{LR}_n - LR_{P_n}) \\
&= \sqrt{n}\Big(\int_{\mathcal{L}_{r_n}} \Big[\Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\hat{\phi}_{1,\ell,n}(\hat{\theta}_{1,n})\big)\Big) - \Psi\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\hat{\phi}_{2,\ell,n}(\hat{\theta}_{2,n})\big)\Big)\Big] dF(\ell) - LR_{P_n}\Big) \\
&= \sqrt{n}\int_{\mathcal{L}} \Big[\Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi_{1,\ell,P_n}^*(\theta_{1,n}^*)\big)\Big) - \Psi\big(M_{1,\ell,P_n}^*\big) - \Psi\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi_{2,\ell,P_n}^*(\theta_{2,n}^*)\big)\Big) + \Psi\big(M_{2,\ell,P_n}^*\big)\Big] dF(\ell) \\
&\quad - \sqrt{n}\int_{\mathcal{L}_{r_n}^c} \Big[\Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi_{1,\ell,P_n}^*(\theta_{1,n}^*)\big)\Big) - \Psi\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi_{2,\ell,P_n}^*(\theta_{2,n}^*)\big)\Big)\Big] dF(\ell) \\
&\quad + \sqrt{n}\int_{\mathcal{L}_{r_n}} \Big[\Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\hat{\phi}_{1,\ell,n}(\hat{\theta}_{1,n})\big)\Big) - \Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi_{1,\ell,P_n}^*(\theta_{1,n}^*)\big)\Big)\Big] dF(\ell)
\end{aligned}$$

$$
\begin{aligned}
&+\sqrt{n}\int_{\mathcal{L}_{r_n}}\Big[\Psi\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)\Big)-\Psi\Big(\widehat{\mathcal{M}}_{2.\ell,n}\big(\hat{\phi}_{2,\ell,n}(\hat{\theta}_{2,n})\big)\Big)\Big]dF(\ell)\\
&=\widetilde{\Lambda}_n-A^c_n+A_{n,1}+A_{n,2}. \qquad\qquad (\text{C.54})
\end{aligned}
$$

Re-write A^c_n as

$$
\begin{aligned}
A^c_n=&\sqrt{n}\int_{\mathcal{L}^c_{r_n}}\Big[\Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\Big)-\Psi\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)\Big)\Big]dF(\ell)\\
=&\sqrt{n}\int_{\mathcal{L}^c_{r_n}}\Big[\Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\Big)-\Psi\Big(\mathcal{M}_{1,\ell,P_n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\Big)\Big]dF(\ell)\\
&-\sqrt{n}\int_{\mathcal{L}^c_{r_n}}\Big[\Psi\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)\Big)-\Psi\Big(\mathcal{M}_{2,\ell,P_n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)\Big)\Big]dF(\ell)\\
&+\sqrt{n}\int_{\mathcal{L}^c_{r_n}}\Big[\Psi\Big(\mathcal{M}_{1,\ell,P_n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\Big)-\Psi\Big(\mathcal{M}_{2,\ell,P_n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)\Big)\Big]dF(\ell)\\
=&A^c_{n,1}-A^c_{n,2}+A^c_{n,3}. \qquad\qquad (\text{C.55})
\end{aligned}
$$

By Lemma A.4, $A^c_{n,3}=O_p(n^{1/2}r_n^{d_x+1})$. For $A^c_{n,1}$, by mean-value expansions,

$$
\begin{aligned}
|A^c_{n,1}|=&\Big|\int_{\mathcal{L}^c_{r_n}}\Psi'(\widetilde{\mathcal{M}}_{1,\ell,n})\sqrt{n}\left(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)-\mathcal{M}_{1,\ell,P_n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\right)dF(\ell)\Big|\\
\le& C\int_{\mathcal{L}^c_{r_n}}\sqrt{n}\left|\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)-\mathcal{M}_{1,\ell,P_n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\right|dF(\ell)\\
=&O_p(r_n)=o_p(n^{1/2}r_n^{d_x+1}), \qquad\qquad (\text{C.56})
\end{aligned}
$$

where $\widetilde{\mathcal{M}}_{1,\ell,n}$ is between $\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)$ and $\mathcal{M}_{1,\ell,P_n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)$. The first inequality holds by the fact that $\Psi'(\widetilde{\mathcal{M}}_{1,\ell,n})$ is uniformly bounded with probability approaching 1, the quantity $\sqrt{n}\big|\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)-\mathcal{M}_{1,\ell,P_n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\big|$ is $O_p(1)$ uniformly over ℓ by A.2(b), and the total weight of $F(\ell)$ over $\mathcal{L}^c_{r_n}$ is $O(r_n)$ by Assumption 5.4. The last equality holds because $n^{1/2}r_n^{d_x}\to\infty$. Similarly, $A^c_{n,2}=o_p(n^{1/2}r_n^{d_x+1})$. These together imply that

$$
A^c_n=O_p(n^{1/2}r_n^{d_x+1}). \qquad\qquad (\text{C.57})
$$

For $A_{n,1}$ and $A_{n,2}$, observe that they are $n^{1/2}$ times the l.h.s. of (1b) in equation (C.42) specialized to model $\mathcal{P}_1$ and model $\mathcal{P}_2$ respectively. The three conditions in (C.42) combined with Lemma A.7 show that

$$
A_{n,1}=O_p(n^{-1/2}r_n^{-d_x}),\ \text{and}\ A_{n,2}=O_p(n^{-1/2}r_n^{-d_x}). \qquad\qquad (\text{C.58})
$$

It is left to discuss $\widetilde{\Lambda}_n$. Recall that

$$
\begin{aligned}
\widetilde{\Lambda}_n=&\sqrt{n}\int_{\mathcal{L}}\Big[\Psi\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)\Big)-\Psi\big(\mathcal{M}^*_{1,\ell,P_n}\big)\Big]dF(\ell)\\
&-\sqrt{n}\int_{\mathcal{L}}\Big[\Psi\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)-\Psi\big(\mathcal{M}^*_{2,\ell,P_n}\big)\Big]dF(\ell)\\
=&\int_{\mathcal{L}}\Psi'(\mathcal{M}^*_{1,\ell,P_n})\sqrt{n}\Big(\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)-\mathcal{M}^*_{1,\ell,P_n}\Big)dF(\ell)+O_p(n^{-1/2})\\
&+\int_{\mathcal{L}}\Psi'(\mathcal{M}^*_{2,\ell,P_n})\sqrt{n}\Big(\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)-\mathcal{M}^*_{2,\ell,P_n}\Big)dF(\ell)+O_p(n^{-1/2})
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathcal{L}}\Big\{\Psi'(\mathcal{M}^*_{1,\ell,P_n})\Big[\kappa\big(\gamma^*_{1,\ell,P_n}(\theta^*_{1,n})'m_1(W_i,\theta^*_{1,n})g_\ell(X_i)\big)-\Psi\big(\mathcal{M}^*_{1,\ell,P_n}\big)\Big]\\
&\qquad-\Psi'(\mathcal{M}^*_{2,\ell,P_n})\Big[\kappa\big(\gamma^*_{2,\ell,n}(\theta^*_{2,n})'m_2(W_i,\theta^*_{2,n})g_\ell(X_i)\big)+\Psi\big(\mathcal{M}^*_{2,\ell,P_n}\big)\Big]\Big\}dF(\ell)+O_p(n^{-1/2})\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Lambda^*_{P_n,i}+O_p(n^{-1/2}).
\end{aligned}\tag{C.59}
$$

The second equality holds because $P_{0,n}\in\mathcal{F}_0$. The third equality holds by Taylor expansions and the last equality holds by expressing $\widehat{\mathcal{M}}_{1,\ell,n}\big(\phi^*_{1,\ell,P_n}(\theta^*_{1,n})\big)$ and $\widehat{\mathcal{M}}_{2,\ell,n}\big(\phi^*_{2,\ell,P_n}(\theta^*_{2,n})\big)$ and by changing the order of summation and integration.

The lemma is proved by combining (C.54), (C.57), (C.58) and (C.59).

Proof of Lemma A.9: This is identical to Theorem 20.1 of Chong and Żak (2001).■

Proof of Lemma A.10: This is identical to Theorem 20.3 of Chong and Żak (2001).■

Proof of Lemma A.11: By assumption, $\mu_j^*=0$ if $g_j(x^*)>0$ for $j=1,\ldots,m$, we just need to consider those j's such that $g_j(x^*)=0$. Note that g_j is concave and $g_j(x^*)=0$, so

$$
0\le g_j(x)\le g_j(x^*)+\frac{\partial g_j(x^*)}{\partial x}\cdot(x-x^*)=\frac{\partial g_j(x^*)}{\partial x}\cdot(x-x^*).\tag{C.60}
$$

This implies that

$$
\frac{\partial f(x^*)}{\partial x}\cdot(x-x^*)=-\sum_{j=1}^{m}\mu_j^*\cdot\frac{\partial g_j(x^*)}{\partial x}\cdot(x-x^*)\le 0.\tag{C.61}
$$

This shows Lemma A.11.■

References

Andrews, D. W. K. (1994). "Empirical process methods in econometrics," Handbook of Econometrics, in: R. F. Engle and D. McFadden (ed.), Vol. 4, 2247–2294, Elsevier.

Andrews, D. W. K., S. Berry, and P. Jia (2004). "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," *Econometrica*, forthcoming.

Andrews, D. W. K., and X. Shi (2013). "Inference Based on Conditional Moment Inequalities," *Econometrica*, Vol. 81, 509-547.

Andrews, D. W. K. and X. Shi (2011). "Supplement to 'Inference Based on Conditional Moment Inequalities,'" unpublished manuscript.

Chong, E. K. P. and S. H. Żak (2001). "An Introduction to Optimization," 2nd Ed., Wiley-Interscience.

Ciliberto, F. and E. Tamer (2009). "Market Structure and Multiple Equilibria in the Airline Industry," *Econometrica*, **77**, 1791–1828.

Donald, S. G., G. W. Imbens and W. K. Newey (2003). "Empirical Likelihood Estimation and Consistent Tests with Conditional Moment Restrictions," *Journal of Econometrics*, **117**, 55–93.

Gandhi, A., Z. Lu and X. Shi (2012). "Estimating Demand for Differentiated Products with Error in Market Shares: A Moment Inequalities Approach," Working Paper. University of Wisconsin at Madison.

Kitamura, Y., G. Tripathi and H. Ahn (2004). "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, **72**, 1667–1714.

Manski, C. F. (2005). "Partial Identification with Missing Data: Concepts and Findings," *International Journal of Approximate Reasoning*, **39**, 151–165.

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

Moon, H. R. and F. Schorfheide (2009). "Estimation with Overidentifying Inequality Moment Conditions," *Journal of Econometrics*, **153**, 136 –154.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2007): "Moment Inequalities and Their Application," Working Paper.

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

Shi, X. (2009a). "Model Selection Test for Nonnested Moment Inequality Models," Working Paper, University of Wisconsin at Madison.

Shi, X. (2009b). "A nondegenerate Vuong test," Working Paper, University of Wisconsin at Madison.

Vuong, Q. H. (1989). "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, **57**, 307–333.

Yachew, A. J. (1992): "Nonparametric Regression Tests Based on Least Squares," *Econometric Theory*, **8**, 435–451.