Model Selection Tests for Nonnested Moment Inequality Models (Job Market Paper)

Xiaoxia Shi*
Yale University
xiaoxia.shi@yale.edu
http://pantheon.yale.edu/~xs37

First Draft: November, 2008
This Draft: Jan 8, 2010

Abstract

This paper proposes Vuong-type tests to select between two partially-identifed moment inequality models based on their Kullback-Leibler distances to the true data distribution. The candidate models can be non-overlapping or overlapping. We treat the two cases separately and develop a testing procedure for each. Both testing procedures are shown to have correct asymptotic size in a uniform sense, despite the potential lack of point identification of the model parameters. The tests are shown to be consistent and have nontrivial power against $n^{-1/2}$-local alternatives. Monte Carlo simulation of an entry game is used to demonstrate the finite-sample performance of the tests. The tests are relatively easy to implement.

Keywords: Asymptotic size, Kullback-Leibler divergence, Model selection test, Moment inequalities, Nonnested models, Overlapping models, Partial identification.

JEL Classification Numbers: C12, C52

*The author is deeply indebted to Donald Andrews for guidance and encouragement. The author benefited most from the comments of Xiaohong Chen, Edward Vytlacil, Yuichi Kitamura, Taisuke Otsu, and Peter Phillips. The author thanks Lanier Benkard, Steven Berry, Philip Haile, Yoonseok Lee, Simon Lee, Yoon-Jae Whang, Jörg Stoye and seminar participants at Yale University for helpful comments. The financial support from a Carl Arvid Anderson Fellowship is gratefully acknowledged. The usual disclaimer applies.

1 Introduction

Models defined by moment inequalities (and possibly some equalities) have gained substantial popularity over recent years as researchers try to move away from ad hoc structural assumptions in various areas of economics.[1] Model selection problems in this context arise naturally when researchers consider more than one economic theory, each generating a set of moment inequalities, or when they consider different parametrizations to form the moment functions.[2] Because moment inequality models are non-traditional in the ways discussed shortly below, existing model selection methods (e.g. Vuong (1989), Kitamura (2000), AIC, or BIC) are not readily applicable. When the candidate models are nested, the emerging literature on inference for moment inequality models sometimes can be used to choose the best model.[3] However, the procedures in this literature are inapplicable when the competing models are nonnested.

This paper provides a way to select the better model from two competing nonnested moment inequality models. We design quasi-likelihood-ratio tests for the null hypothesis: both models are equally close to the true data distribution in terms of the Kullback-Leibler (KL) divergence. When the null does not hold, the tests direct the researcher to the model that is closer to the true distribution with probability approaching one. Our tests are relatively easy to compute for two reasons. First, they use standard normal critical values. Second, although the sample criterion functions can have multiple (or even a continuum of) maximizers due to partial identification, one does not need to compute all the maximizers to implement the tests.

Moment inequality models are non-traditional in two ways. First, parameters in these models typically are not point-identified. For that reason, the maximizers of a sample criterion function do not converge to a point in the parameter space. Thus, traditional model selection methods that rely on the asymptotic normality of the maximizers do not apply. We show that the sample maximizers approach a pseudo-true set as the sample size

[1] They have been used to model discrete games with multiple equilbria (Andrews, Berry, and Jia (2004), Ciliberto and Tamer (2009)), to deal with missing or interval data (Manski and Tamer (2002)), to study dynamic games that are otherwise too complicated to analyze empirically (Pakes, Porter, Ho, and Ishii (2007), Pakes (2009)) and to increase the precision of estimators in dynamic macroeconomics models (Moon and Schorfheide (2006)).

[2] Examples are given in Section 2.

[3] A non-exhaustive list of papers on estimation and inference of moment inequality models includes Chernozhukov, Hong, and Tamer (2007), Andrews and Jia (2008), Bugni (2008), Canay (2008), Romano and Shaikh (2008) Andrews and Guggenberger (2009), Andrews and Soares (2010) and Andrews and Shi (2009).

approaches infinity. With this result, we then use a special sandwich technique to derive the asymptotic distribution of the quasi-likelihood-ratio statistic in order to develop the test for non-overlapping models.

Second, moment inequality models have slackness parameters whose (pseudo-) true values are on the boundary of the parameter space[4]. The parameter-on-the-boundary problem makes the criterion function for the original model parameters non-differentiable even in the limit.[5] The non-differentiability can occur anywhere in the original parameter space.[6] Thus, the first-order-condition method or the standard quadratic approximation method cannot be used to derive the convergence rate of the estimators. We construct quadratic bounds for the sample and population criterion functions. By combining those bounds, we show that the sample maximizers approach the pseudo-true set at $n^{-1/2}$-rate, which is the rate we need in order to develop the test for overlapping models.

A testing procedure similar to ours is first proposed in Vuong (1989) for fully parametric models that can be estimated using maximum likelihood (ML). Kitamura (2000) extends the Vuong test to models defined by moment *equalities* that cannot be estimated using ML. Kitamura (2000) employs exponential tilting estimation, which is adapted to moment inequality models in the current paper. Chen, Hong, and Shum (2007) combine ML and empirical likelihood and propose a Vuong-type procedure to select between a parametric model and a moment equality model. All three papers assume that the true parameters are point-identified and are in the interior of the parameter space − assumptions that are suitable for parametric models and moment equality models, but not suitable for the moment inequality models considered in this paper.

In addition to addressing the partial identification and parameter-on-the-boundary problems, another important feature distinguishing our tests from the other Vuong-type tests is that we choose the critical values based on uniform asymptotics which guarantee correct asymptotic sizes of the tests. Vuong-type tests with critical values chosen based on pointwise asymptotics may have size distortion when the candidate models are overlapping. The reason is that the pointwise asymptotic distributions of the test statistics are discon-

[4] The slackness parameter for a moment inequality model, $Em(X_i, \theta) \geq 0$, is $a = Em(X_i, \theta)$. One can view the moment inequality model as a moment equality model with the augmented parameter, $(\theta', a')'$: $Em(X_i, \theta) - a = 0$. The space of a is $R_+^{d_m}$. The true value of a is on the boundary of $R_+^{d_m}$ whenever a moment inequality holds as an equality under the true data distribution. In this example, $\{X_i\}$ is the data, m is a R^{d_m}-valued moment function and θ is a finite-dimensional parameter.

[5] This makes our problem unlike the quantile regression problems, where the sample criterion functions are non-differentiable but the probability limits of them are.

[6] Therefore, there is no straightforward way of writing down the left/right derivatives.

tinuous in the model parameters. When the true parameters are close to the discontinuity point, the finite sample distributions of the test statistics are not well approximated by their pointwise asymptotic distributions. The poor approximation causes size distortion in finite samples (Shi (2009)). We choose our critical values taking into consideration the distributions of the test statistics near the discontinuity and by doing so control the finite sample null rejection probabilities of the tests uniformly over the composite null.

An alternative to our Vuong-type framework is the Cox-type nonnested hypothesis testing framework. For a Cox-type test, the null hypothesis is that a model $\mathcal{P}$ is correctly specified and the alternative hypothesis is that an alternative model $\mathcal{Q}$ is correctly specified. Though frequently used to choose one model from multiple candidate models, Cox-type tests are intended as a procedure for model evaluation rather than model selection. A Cox-type test does not have a clear interpretation when both models are misspecified. For details on Cox-type tests, see the seminal paper by Cox (1961), the survey papers by Gourieroux and Monfort (1994) and Pesaran and Weeks (1999), generalizations to the encompassing principle by Mizon and Richard (1986), and the extension to moment equality models by Ramalho and Smith (2002). It is of interest to extend the moment encompassing principle to partially-identified moment inequality models possibly using some of the techniques developed in this paper. We leave this to a separate project.

The rest of the paper is organized as follows. Section 2 introduces the model selection problem for moment inequality models and gives a few examples. Section 3 presents preliminaries on the pseudo-distance measure and the solution to the distance-minimizing problem. Section 4 describes the tests, one for non-overlapping models and the other for overlapping models. Sections 5 and 6 establish the asymptotic size of the test for non-overlapping models and that for overlapping models, respectively. Section 7 determines the power properties of the tests. Section 8 presents Monte Carlo simulation results for an entry game example. Section 9 concludes. The proofs are in the appendix.

We use $N_\delta(\theta)$ to denote a closed ball centered at θ with radius δ, $||\cdot||$ to denote the Euclidean norm, and "<<" to denote "is absolutely continuous with respect to (w.r.t., hereafter)". We use X_i to denote an observation, $\mathcal{X}$ to denote the space on which X_i is defined. We use $\mathcal{P}$ and $\mathcal{Q}$ to denote the candidate models, and P and Q to denote generic distributions in $\mathcal{P}$ and $\mathcal{Q}$, respectively. We use μ to denote a generic true distribution on $\mathcal{X}$, which does not necessarily belong to either of the models. We use greek letters θ and β to denote the finite-dimensional parameters in the models, Θ and B to denote the corresponding parameter spaces, and m and g to denote the moment functions.

2 Model Selection Problems

We consider two moment inequality/equality models $\mathcal{P} = \bigcup_{\theta\in\Theta} \mathcal{P}_\theta$ and $\mathcal{Q} = \bigcup_{\beta\in B} \mathcal{Q}_\beta$, where $\mathcal{P}_\theta$ and $\mathcal{Q}_\beta$ are the set of distributions that are consistent with the moment conditions for parameters θ and β, respectively:

$$\begin{aligned}\mathcal{P}_\theta &= \left\{\begin{matrix} P : E_P m_j(X_i,\theta) = 0 \text{ for } j = 1,...,d_p, \\ E_P m_j(X_i,\theta) \geq 0 \text{ for } j = d_p+1,...,d_m \end{matrix}\right\} \\ \mathcal{Q}_\beta &= \left\{\begin{matrix} Q : E_Q g_j(X_i,\beta) = 0 \text{ for } j = 1,...,d_q, \\ E_Q g_j(X_i,\beta) \geq 0 \text{ for } j = d_q+1,...,d_g \end{matrix}\right\}. \end{aligned} \tag{2.1}$$

In the above equation, $\{X_i \in \mathcal{X}\}_{i=1}^n$ is a random sample generated from μ, $m \equiv (m_1,...,m_{d_p}, m_{p+1},...,m_{d_m})'$ and $g \equiv (g_1,...,g_{d_q},\ g_{q+1},...,g_{d_g})'$ are R^{d_m} and R^{d_g}-valued moment functions known up to the finite-dimensional parameters θ and β, respectively, $\Theta \subset R^{d_\theta}$, $B \subset R^{d_\beta}$, and E_P denotes the expectation under the distribution P. The true distribution μ may or may not belong to either model. Model $\mathcal{P}$ is called **correctly specified** if $\mu \in \mathcal{P}$ and is called **misspecifed** otherwise. The parameters θ and β may or may not be point-identified.

The goal of this paper is to compare models $\mathcal{P}$ and $\mathcal{Q}$ and select the one that is closer to the true distribution μ in terms of a pseudo-distance measure. Let $d(P,\mu)$ be a pseudo-distance between a distribution P and μ. The pseudo distance from a model $\mathcal{P}$ to μ is defined by $d(\mathcal{P},\mu) = \inf_{P\in\mathcal{P}} d(P,\mu)$. We want to construct model selection tests for the null hypothesis

$$H_0 : d(\mathcal{P},\mu) = d(\mathcal{Q},\mu). \tag{2.2}$$

The choice of d is discussed in the next section.

Now, we give a few illustrative examples of model selection problems in the context of moment inequalities. We present a Monte Carlo simulation for Example 3 in Section 8.

Example 1 (Interval Outcome in Regression Models). Consider the regression models with interval outcomes from Manski and Tamer (2002). A model selection problem of potential interest is selecting different regressors or functional forms for the regression functions. Let Y be a latent random variable (e.g. wealth) that is not perfectly observed. Only an upper bound, $\overline{Y}$, and a lower bound, $\underline{Y}$, on Y are observed. Let X be a vector of explanatory variables and $Y = r(X,\theta) + \varepsilon$, where r is a function known up to a finite-dimensional parameter θ. Let Z be a vector of potential instrument variables such that $E(\varepsilon \cdot I(Z)) = 0$ for some positive (vector-valued) function I of Z. Then, the models $\mathcal{P}$ and $\mathcal{Q}$ are:

$$\begin{aligned} \mathcal{P} &= \{P : E_P[(\overline{Y} - r_1(X,\theta))I(Z)] \geq 0 \ \&\ E_P[(r_1(X,\theta) - \underline{Y})I(Z)] \geq 0,\ \theta \in \Theta\} \\ \mathcal{Q} &= \{Q : E_Q[(\overline{Y} - r_2(X,\beta))I(Z)] \geq 0 \ \&\ E_Q[(r_2(X,\beta) - \underline{Y})I(Z)] \geq 0,\ \beta \in B\}, \end{aligned} \tag{2.3}$$

where r_1 and r_2 are two regression functions. Note that the distributions P and Q are defined on the space of the **observed** random variables ($\overline{Y}$, $\underline{Y}$, X, Z).

Another model selection problem arises when one considers a different choice of instruments. The formulation of the competing models is similar to (2.3), except that r_1 and r_2 are the same and we have I_1 instead of I in model $\mathcal{P}$ and I_2 in model $\mathcal{Q}$.

Example 2 (Interval Regressor in Regression Models). Consider the regression models with interval regressors from Manski and Tamer (2002). Let Y be a continuous dependent variable, v be a regressor that is not observed perfectly but in intervals $[\underline{v}, \overline{v}]$. Let X represent other regressors. Assume that $E(Y|X,v) = f(x,v,\theta)$, where f is a function known up to the finite-dimensional parameter θ. Manski and Tamer (2002) assume that f is weakly *increasing* in v, and obtain the following moment inequality model:

$$\begin{aligned} \mathcal{P} = \{P : & E_P[(Y - f(X,\underline{v},\theta))I(X,\underline{v},\overline{v})] \geq 0 \\ & \&\ E_P[(f(X,\overline{v},\theta) - Y)I(X,\underline{v},\overline{v})] \geq 0,\ \theta \in \Theta\}, \end{aligned} \tag{2.4}$$

where $I(X,\underline{v},\overline{v})$ is a vector of positive instrument functions.[7] On the other hand, if we assume that f is weakly *decreasing* in v, we have a different moment inequality model:

$$\begin{aligned} \mathcal{Q} = \{Q : & E_Q[(f(X,\underline{v},\beta) - Y)I(X,\underline{v},\overline{v})] \geq 0 \\ & \&\ E_Q[(Y - f(X,\overline{v},\beta))I(X,\underline{v},\overline{v})] \geq 0,\ \beta \in B\}. \end{aligned} \tag{2.5}$$

By comparing models $\mathcal{P}$ and $\mathcal{Q}$, one can determine the sign of $\partial f/\partial v$, which sometimes is the objective of investigation.

Example 3 (Entry Game – Cross-firm Effect). Consider the entry game example from Andrews, Berry, and Jia (2004) and Ciliberto and Tamer (2009). Model selection

[7] Note that the probability measure P's are defined on the space of $(Y, X, \overline{v}, \underline{v})$.

tests in this context can be used to determine the sign of the cross-firm effect. Consider a 2×2 entry game with the following payoff matrix:

		Firm II 0	Firm II 1
Firm I	0	$0, 0$	$0,\ X_2'\theta_2 - \varepsilon_2$
	1	$X_1'\theta_1 - \varepsilon_1,\ 0$	$X_1\theta_1 + a_1 - \varepsilon_1,\ X_2\theta_2 + a_2 - \varepsilon_2$

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 θ_1 and θ_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, $F_{\varepsilon_1}(\cdot;\theta_\varepsilon)$ 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 θ_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) and Ciliberto and Tamer (2009) assume $a_1 \leq 0$ and $a_2 \leq 0$ and obtain the following moment inequality model:

$$\begin{aligned} \mathcal{P} = \{P : & E_P[(p_j(X,\theta) - 1(Y = j))I(X)] = 0, \text{ for } j = (0,0) \text{ or } (1,1) \\ & E_P[(p_j(X,\theta) - 1(Y = j))I(X)] \geq 0,\ j = (0,1), \text{ or } (1,0) \\ & \theta \equiv (\theta_1', \theta_2', a_1, a_2, \theta_\varepsilon',)' \in \Theta\}, \end{aligned} \tag{2.6}$$

where $I(X)$ is a vector of positive instrument functions, and

$$\begin{aligned} p_{(0,0)}(X,\theta) &= 1 - F_{\varepsilon_1}(X_1'\theta_1;\theta_\varepsilon) - F_{\varepsilon_2}(X_2'\theta_2;\theta_\varepsilon) + F_{\varepsilon_1,\varepsilon_2}(X_1'\theta_1, X_2'\theta_2;\theta_\varepsilon) \\ p_{(0,1)}(X,\theta) &= F_{\varepsilon_2}(X_2'\theta_2;\theta_\varepsilon) - F_{\varepsilon_1,\varepsilon_2}(X_1'\theta_1 + a_1, X_2'\theta_2;\theta_\varepsilon) \\ p_{(1,0)}(X,\theta) &= F_{\varepsilon_1}(X_1'\theta_1;\theta_\varepsilon) - F_{\varepsilon_1,\varepsilon_2}(X_1'\theta_1, X_2'\theta_2 + a_2;\theta_\varepsilon) \\ p_{(1,1)}(X,\theta) &= F_{\varepsilon_1,\varepsilon_2}(X_1'\theta_1 + a_1, X_2'\theta_2 + a_2;\theta_\varepsilon). \end{aligned} \tag{2.7}$$

There are reasons to be unsure about the signs of a_1 and a_2 in some markets. If we assume $a_1 \geq 0$ and $a_2 \geq 0$, we obtain a different moment inequality model:

$$\begin{aligned}\mathcal{Q} = \{Q : & E_Q[(p_j(X,\beta) - 1(Y = j))I(X)] \geq 0, \text{ for } j = (0,0) \text{ or } (1,1) \\ & E_Q[(p_j(X,\beta) - 1(Y = j))I(X)] = 0, \ j = (0,1), \text{ or } (1,0) \\ & \beta \equiv (\theta_1', \theta_2', a_1, a_2, \theta_\varepsilon',)' \in B\},\end{aligned} \tag{2.8}$$

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 the signs of the cross-firm effects.

Example 4 (Entry Game – Choosing Information Structure) Model selection test also can be used to choose the information structure of a game-theoretical model. Berry and Tamer (2006) show that the entry game described in Example 3 can be modeled by a different set of moment inequalities, if we assume that the firms do not know their competitors' idiosyncratic profits $(\varepsilon_1, \varepsilon_2)$ but have beliefs about the distributions of $(\varepsilon_1, \varepsilon_2)$. By comparing the new moment inequality model to $\mathcal{P}$ (or $\mathcal{Q}$) in Example 3, one can determine which information structure is more appropriate.

3 Preliminaries on the Pseudo-distance Measure

There are many possible choices of pseudo-distances on the space of probability distributions. One may prefer one distance to another in a specific problem. Since we deal with a generic problem, we choose the Kullback-Leibler (KL) divergence – defined below – for convenience and for its analogy to the likelihood principle in parametric models. Our results can be extended to the more general f-divergence family which includes the KL divergence, the quadratic distance and the Hellinger distance (Csiszár and Shields (2004), Kitamura (2007)). The KL divergence from P to μ is

$$d(P,\mu) = \begin{cases} \int p_\mu \log p_\mu d\mu & \text{if } P << \mu \\ \infty & \text{otherwise} \end{cases}, \tag{3.1}$$

where p_μ is the density of P with respect to μ. The pseudo-distance above also is called the I-divergence, or the relative entropy of P to μ. For moment condition models, the KL divergence motivates the exponential tilting estimation (Kitamura and Stutzer (1997)) and the empirical likelihood estimation (Qin and Lawless (1994)).

The rest of the discussions in this section – with the exclusion of the formal assumptions and lemmas – are in terms of model $\mathcal{P}$, but they apply to model $\mathcal{Q}$ as well.

In order to measure the distance from the model to the true distribution, one needs to solve the minimization problem $\inf_{P\in\mathcal{P}} d(P,\mu)$. The problem is solved in two steps:

$$\inf_{P\in\mathcal{P}} d(P,\mu) = \inf_{\theta\in\Theta} \inf_{P\in\mathcal{P}_\theta} d(P,\mu), \tag{3.2}$$

where $\mathcal{P}_\theta$ is defined in (2.1). The first step $\inf_{P\in\mathcal{P}_\theta} d(P,\mu)$ is an infinite dimensional minimization problem and can be solved through a finite-dimensional dual problem. The second step is a finite-dimensional minimization problem which may have multiple solutions because model $\mathcal{P}$ may be partially-identified. We discuss both steps in the following subsections.

3.1 The Dual Problem

The first step minimization $\inf_{P\in\mathcal{P}_\theta} d(P,\mu)$ has a unique solution, if the solution exists. The reason is that $d(P,\mu)$ is strictly convex in P and the set $\mathcal{P}_\theta$ is defined by constraints linear in P and thus is convex. We follow Csiszár (1975) and call the solution to $\inf_{P\in\mathcal{P}_\theta} d(P,\mu)$ the ***I*-projection** of μ on $\mathcal{P}_\theta$. Denote the I-projection as $P^*_{\mu,\theta}$.

For models defined by equality constraints, Csiszár (1975) gives sufficient conditions for the existence of $P^*_{\mu,\theta}$ and shows that $\inf_{P\in\mathcal{P}_\theta} d(P,\mu)$ has a finite-dimensional dual problem under those conditions. We adapt Csiszár's (1975) approach to the context of moment inequality models. The sufficient conditions are given in Assumption 1.

For a data distribution μ, let the Lagrange multipliers for each θ and β be

$$\begin{aligned}\gamma^*_\mu(\theta) &= \arg\min_{\gamma\in R_\infty^{d_p}\times R_{+,\infty}^{d_m-d_p}} E_\mu \exp(\gamma' m(X_i,\theta)),\\ \lambda^*_\mu(\beta) &= \arg\min_{\lambda\in R_\infty^{d_q}\times R_{+,\infty}^{d_g-d_q}} E_\mu \exp(\lambda' g(X_i,\beta)),\end{aligned} \tag{3.3}$$

where $R_\infty = R\cup\{\infty,-\infty\}$ and $R_{+,\infty} = R_+\cup\{\infty\}$. For every $\theta\in\Theta$, $\gamma^*_\mu(\theta)$ is uniquely defined under Assumption 1(a) below.

Assumption 1. (a) *For all* $\theta\in\Theta$, $E_\mu||m(X_i,\theta)||^2<\infty$ *and* $E_\mu[m(X_i,\theta)m(X_i,\theta)']$ *is positive definite,*

(b) *for all* $\theta\in\Theta$, $||\gamma^*_\mu(\theta)||<\infty$, *and*

(c) *parts* (a)-(b) *hold with* g, β *and* λ *in place of* m, θ *and* γ.

Assumption 1(a) is a standard assumption in the literature of moment condition models. Assumption 1(b) requires the model not to be too misspecified. We give a sufficient condition that is easy to verify for Assumption 1(b) in Assumption 1(b)*.

Assumption 1(b)*. *For all* $\theta \in \Theta$ *and all* $\gamma \in R^{d_p} \times R_+^{d_m - d_p}$, $\Pr_\mu(\gamma' m(X_i, \theta) > 0) > 0$.

Assumption 1(b)* states that for all $\theta \in \Theta$, any linear combination of the moment functions with weights in $R^{d_p} \times R_+^{d_m-d_p}$ is positive with positive probability. Under Assumption 1(b)*, $\Pr_\mu(\gamma' m(X_i, \theta) = \infty) > 0$ for any $\gamma \in R_\infty^{d_p} \times R_{+,\infty}^{d_m-d_p}$ such that $||\gamma|| = \infty$. Thus, $E_\mu \exp(\gamma' m(X_i, \theta)) = \infty$ for these γ's. Because $E_\mu \exp(\gamma_\mu^*(\theta)' m(X_i, \theta)) \leq E_\mu \exp(0' m(X_i, \theta)) = 1$, we have $||\gamma_\mu^*(\theta)|| < \infty$, i.e., Assumption 1(b) holds. Assumption 1(b)* is violated, for example, when the model is $\mathcal{P} = \{P : E_P(X_{1,i} - \theta) \geq 0, E_P(\theta - X_{2,i}) \geq 0\}$, and $X_{1,i} < X_{2,i}$ a.s. $[\mu]$.[8]

Lemma 1 below establishes that $\inf_{P \in \mathcal{P}} d(P, \mu)$ is attained and can be solved through a finite-dimensional dual problem under Assumption 1.

For a data distribution μ, Let the dual criterion functions for the two models be

$$\mathcal{M}_\mu(\gamma, \theta) = E_\mu \exp\big(\gamma' m(X_i, \theta)\big) \text{ and } \mathcal{N}_\mu(\lambda, \beta) = E_\mu \exp\big(\lambda' g(X_i, \beta)\big). \tag{3.4}$$

Lemma 1. *Suppose Assumption* 1 *holds. Then,*

(a) *for all* $\theta \in \Theta$, *the I-projection,* $P_{\mu,\theta}^*$, *of* μ *on* $\mathcal{P}_\theta$ *exists and its density w.r.t.* μ *is*

$$p_{\theta,\mu}^*(x) = \exp\big(\gamma_\mu^*(\theta)' m(x, \theta)\big) / \mathcal{M}_\mu\big(\gamma_\mu^*(\theta), \theta\big),$$

(b) *for all* $\theta \in \Theta$, $d(\mathcal{P}_\theta, \mu) = -\log[\mathcal{M}_\mu(\gamma_\mu^*(\theta), \theta)]$,

(c) *parts* (a)-(b) *hold with* g, β, λ, Q, $\mathcal{Q}$ *and* $\mathcal{N}$ *in place of* m, θ, γ, P, $\mathcal{P}$ *and* $\mathcal{M}$.

Lemma 1(a) shows that the infimum $\inf_{P \in \mathcal{P}_\theta} d(P, \mu)$ is attained at $P_{\mu,\theta}^*$. Lemma 1(b) shows that the infimum can be obtained through solving the finite-dimensional dual problem: $\min_{\gamma \in R^{d_p} \times R_+^{d_m - d_p}} \mathcal{M}_\mu(\gamma, \theta)$. Note that the nonnegativity constraints in the dual problem reflect the inequality constraints in the primal problem.

[8] To check that, pick $a = (1, 1)'$. Then, $\Pr_\mu(a' m(X_i, \theta) > 0) = \Pr_\mu(X_{1,i} - X_{2,i} > 0) = 0$.

3.2 The Pseudo-true Set and the Pseudo-true Distribution

The second step infimum in (3.2), $\inf_{\theta\in\Theta} d(\mathcal{P}_\theta, \mu)$, is attained if $d(\mathcal{P}_\theta, \mu)$ is continuous in θ and Θ is compact. These are guaranteed by Assumption 2 below.

Assumption 2. (a) Θ *and* B *are compact, and*

(b) *for all* $\theta \in \Theta$ *and* $\beta \in B$, $m(X_i, \theta)$ *and* $g(X_i, \beta)$ *are continuous in* θ and β *a.s.* $[\mu]$, *respectively.*

Assumption 2(a) is a standard assumption on the parameter space. Assumption 2(b) requires the moment functions to be almost surely continuous in the parameters.

Lemma 2 below shows that the infimum $\inf_{\theta\in\Theta} d(\mathcal{P}_\theta, \mu)$ is attained and has a saddle-point dual representation.

Lemma 2. *Suppose Assumptions* 1 *and* 2 *hold. Then,*

(a) *there exists a* $\theta^* \in \Theta$ *such that* $\mathcal{M}_\mu(\gamma_\mu^*(\theta^*), \theta^*) = \sup_{\theta\in\Theta} \mathcal{M}_\mu(\gamma_\mu^*(\theta), \theta)$,

(b) $d(\mathcal{P}, \mu) = -\log\left[\max_{\theta\in\Theta} \min_{\gamma\in R^{d_p}\times R_+^{d_m-d_p}} \mathcal{M}_\mu(\gamma, \theta)\right]$, *and*

(c) *parts* (a)-(b) *hold with* g, β, λ, q, Q, $\mathcal{Q}$ *and* $\mathcal{N}$ *in place of* m, θ, γ, p, P, $\mathcal{P}$ *and* $\mathcal{M}$.

Remark. The function $\gamma_\mu^*(\theta)$ usually has kinks because of the nonnegativity constraints in the minimization problem that defines it. This reflects the parameter-on-the-boundary problem discussed in the introduction. At the kinks, $\gamma_\mu^*(\theta)$ is not differentiable in θ. The kinks can occur anywhere in Θ. Thus, the population criterion function, $\mathcal{M}_\mu(\gamma_\mu^*(\theta), \theta)$, for θ is non-differentiable even in the interior of Θ. The non-differentiability requires a new approach to derive the convergence rate of the estimators of the pseudo-true set (defined below).

Because model $\mathcal{P}$ may be partially-identified, $\mathcal{M}_\mu(\gamma_\mu^*(\theta), \theta)$ may have multiple maximizers. We call the set of maximizers the **pseudo-true set:**

$$\Theta_\mu^* = \arg\max_{\theta\in\Theta} \mathcal{M}_\mu(\gamma_\mu^*(\theta), \theta). \tag{3.5}$$

The concept of "pseudo-true set" is generalized from the "pseudo-true parameter" concept in the literature of point-identified misspecified models. The prefix "pseudo" signifies the possibility that the model may be misspecified.

It is easy to confuse the partial identification of the parameters in a model $\mathcal{P}$ with the non-uniqueness of the solution to $\min_{P\in\mathcal{P}} d(P, \mu)$. However, the two types of non-uniqueness come from separate sources. The former may occur when the model is incompletely specified or when there are redundant parameters. As a result, one data generating process may be consistent with multiple parameter values. The latter may occur when $\mathcal{P}$ is misspecified and non-convex. Since moment inequality models are not more likely to be non-convex than moment equality models or parametric likelihood models, we adopt the tradition in the model selection literature and assume that the solution to $\min_{P\in\mathcal{P}} d(P,\mu)$ is unique. We call the unique solution the **pseudo-true distribution** of model $\mathcal{P}$ under μ, and denote it by P^*_μ. Then,

$$d(P^*_\mu,\mu) = d(\mathcal{P},\mu). \tag{3.6}$$

We define the pseudo-true distribution of model $\mathcal{Q}$ under μ in the same way and denote it by Q^*_μ. The following assumption states the uniqueness of the pseudo-true distributions.

Assumption 3. *The pseudo-true distributions, P^*_μ and Q^*_μ, of models $\mathcal{P}$ and $\mathcal{Q}$, respectively, are unique under μ.*

Remark. The same assumption is made in Vuong (1989), Kitamura (2000), Chen, Hong, and Shum (2007), as well as in the quasi-maximum likelihood literature (e.g. White (1982)). These authors assume that there is a unique pseudo-true parameters, which implies the uniqueness of the pseudo-true distribution. Our assumption 3 is weaker, as it allows for the non-uniqueness of the pseudo-true parameters that is caused by partial identification.

Assumption 3 implies that $P^*_{\theta,\mu} = P^*_{\theta^*,\mu} = P^*_\mu$ for all θ, $\theta^* \in \Theta^*_\mu$. Thus, by Lemma 1(a)

$$\gamma^*_\mu(\theta)' m(X_i,\theta) = \gamma^*_\mu(\theta^*)' m(X_i,\theta^*) \text{ a.s. } [\mu] \text{ for all } \theta,\theta^* \in \Theta^*_\mu. \tag{3.7}$$

This property is useful in deriving the asymptotic normality of the quasi-likelihood ratio statistic (see below).

4 Model Selection Tests

In this section we introduce the test statistics first. Then, we formally define non-overlapping models and overlapping models and discuss how the relationship between candidate models affects the asymptotic distributions of the test statistics. Finally, we describe the model selection tests.

4.1 Test statistics

We define the test statistics in this section and give informal discussions on the asymptotics in order to introduce the tests.

Lemma 2(b) in the previous section implies that the null (2.2) can be written as

$$H_0 : \max_{\theta\in\Theta} \mathcal{M}_\mu(\gamma^*_\mu(\theta), \theta) = \max_{\beta\in B} \mathcal{N}_\mu(\lambda^*_\mu(\beta), \beta). \tag{4.1}$$

The test statistics are based on the sample analogues of $\max_{\theta\in\Theta} \mathcal{M}_\mu(\gamma^*_\mu(\theta), \theta)$ and $\max_{\beta\in B}$ $\mathcal{N}_\mu(\lambda^*_\mu(\beta), \beta)$.

Let

$$\widehat{\mathcal{M}}_n(\gamma, \theta) = n^{-1}\sum_{i=1}^{n} \exp(\gamma' m(X_i, \theta)) \text{ and } \widehat{\mathcal{N}}_n(\lambda, \beta) = n^{-1}\sum_{i=1}^{n} \exp(\lambda' g(X_i, \beta)). \tag{4.2}$$

Let

$$\begin{aligned} \hat{\gamma}_n(\theta) &= \arg \min_{\gamma\in R^{d_p}\times R_+^{d_m-d_p}} \widehat{\mathcal{M}}_n(\gamma, \theta), \;\; \hat{\lambda}_n(\beta) = \arg \min_{\lambda\in R^{d_q}\times R_+^{d_g-d_q}} \widehat{\mathcal{N}}_n(\lambda, \beta), \\ \hat{\Theta}_n &= \arg \max_{\theta\in\Theta} \widehat{\mathcal{M}}_n(\hat{\gamma}_n(\theta), \theta), \text{ and } \hat{B}_n = \arg \max_{\beta\in B} \widehat{\mathcal{N}}_n(\hat{\lambda}_n(\beta), \beta), \end{aligned} \tag{4.3}$$

where $\hat{\Theta}_n$ and $\hat{B}_n$ are not necessarily singletons.

We use the quasi-likelihood ratio statistic:

$$\widehat{QLR}_n = \max_{\theta\in\Theta} \widehat{\mathcal{M}}_n(\hat{\gamma}_n(\theta), \theta) - \max_{\beta\in B} \widehat{\mathcal{N}}_n(\hat{\lambda}_n(\beta), \beta). \tag{4.4}$$

As we show in later sections, under H_0 and appropriate conditions,

$$\begin{aligned} n^{1/2}\widehat{QLR}_n &\to_d N(0, \omega^2_\mu), \text{ where} \\ \omega^2_\mu &= E_\mu\big[\exp\big(\gamma^*_\mu(\theta^*)' m(X_i, \theta^*)\big) - \exp\big(\lambda^*_\mu(\beta^*)' g(X_i, \beta^*)\big)\big]^2, \end{aligned} \tag{4.5}$$

with $\theta^* \in \Theta^*_\mu$ and $\beta^* \in B^*_\mu$.[9]

[9] By (3.7), ω^2_μ is invariant to the choice of $\theta^* \in \Theta^*_\mu$ and $\beta^* \in B^*_\mu$.

To form the tests, we also use a variance statistic: $\widehat{\omega}_n^2 = \sup_{\theta \in \hat{\Theta}_n, \beta \in \hat{B}_n} \widehat{\omega}_n^2(\theta, \beta)$, where

$$\widehat{\omega}_n^2(\theta, \beta) = n^{-1} \sum_{i=1}^{n} \left[\exp\left(\hat{\gamma}_n(\theta)' m(X_i, \theta)\right) - \exp\left(\hat{\lambda}_n(\beta)' g(X_i, \beta)\right) \right]^2. \tag{4.6}$$

The "sup" in the definition of $\widehat{\omega}_n^2$ is not important either in the asymptotic theory or in practice. In theory, all the results in this paper go through with $\widehat{\omega}_n^2$ defined to be $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n)$ for any arbitrary $\hat{\theta}_n \in \hat{\Theta}_n$ and $\hat{\beta}_n \in \hat{B}_n$. In practice, different choices of $\hat{\theta}_n$ and $\hat{\beta}_n$ in $\hat{\Theta}_n$ and $\hat{B}_n$ typically produce the same $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n)$.

Under H_0 and appropriate conditions

$$\omega_\mu^{-2} \widehat{\omega}_n^2 \to_p 1 \text{ if } \omega_\mu^2 > 0 \text{ and } n\hat{\omega}_n^2 = O_p(1) \text{ if } \omega_\mu^2 = 0. \tag{4.7}$$

The possibility that ω_μ^2 may equal zero or be arbitrarily close to zero complicates the testing problem. The reason is that the asymptotic distribution of $n^{1/2}\widehat{QLR}_n$ is degenerate or nearly degenerate in that case. Whether the degeneracy can occur depends on the relationship between the candidate models. We distinguish two types of relationships in Definition NO below.

Definition NO. *The models $\mathcal{P}$ and $\mathcal{Q}$ are* **non-overlapping** *if* $\inf_{P \in \mathcal{P},\, Q \in \mathcal{Q}} \int |dP/d\nu_{P,Q} - dQ/d\nu_{P,Q}| d\nu_{P,Q} > 0$*, where* $\nu_{P,Q} = (P+Q)/2$*. The models $\mathcal{P}$ and $\mathcal{Q}$ are* **overlapping** *otherwise.*

Remarks. (a) Our categorization of the model relationships is similar to but different from that in Vuong (1989). We distinguish the two types based on uniform asymptotics – whether $N(0,1)$ can uniformly approximate the finite sample distribution of the studentized quasi-likelihood ratio statistic. Vuong (1989) distinguishes the two types – "strictly nonnested" and "overlapping" – based on pointwise asymptotics. Thus, we treat models $\mathcal{P}$ and $\mathcal{Q}$ as overlapping if it is possible for P_μ^* and Q_μ^* to get arbitrarily close to each other, while Vuong (1989) does not treat them as overlapping as long as $P_\mu^* \neq Q_\mu^*$ under each null distribution μ.

(b) The connection between our categorization and Vuong's (1989) is as follows. Models $\mathcal{P}$ and $\mathcal{Q}$ are non-overlapping if and only if they are strictly nonnested in Vuong's (1989) sense, i.e. $\mathcal{P} \cap \mathcal{Q} \neq \phi$, if both models are variation-closed[10]. A sufficient condition for

[10] A set $\mathcal{P}$ is variation-closed if it is closed under the pseudo-metric $d_v(P,Q) = \int |dP/d\nu_{P,Q} - dQ/d\nu_{P,Q}| d\nu_{P,Q}$. See e.g. Csiszár (1975).

a moment inequality model $\mathcal{P}$ to be variation-closed is that the moment functions are bounded.

(c) The overlapping case includes the nested case, i.e. $\mathcal{P} \subset \mathcal{Q}$ or $\mathcal{Q} \subset \mathcal{P}$. The results in this paper for overlapping models hold for nested models except for Theorem 2(b). The test for overlapping models is asymptotically conservative if the two models are nested.

With non-overlapping models, ω_μ^2 is bounded away from zero (shown in Lemma 4 below). Thus, the studentized quasi-likelihood ratio test statistic $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n$ converges in distribution to $N(0,1)$ uniformly under H_0. The test statistic can be combined with the standard normal critical value to form a model selection test. With overlapping models, ω_μ^2 may be arbitrarily close to zero under some null distribution μ. For such μ's, $N(0,1)$ approximates the finite sample distribution of $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n$ poorly. In the extreme case that $\omega_\mu^2 = 0$, $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n$ may not converge in distribution at all due to partial identification. Thus, one needs to choose the critical value with care to take into account the lack of uniformity in asymptotics.

4.2 Tests

Let $\alpha \in (0,1)$. Let $z_{\alpha/2}$ denote the $(1-\alpha/2)$ quantile of the standard normal distribution. We propose tests for non-overlapping models and overlapping models. The test for non-overlapping models does not require a tuning parameter and needs slightly weaker assumptions to be justified than the test for overlapping models. However, for the test to be valid, the candidate models should be non-overlapping according to Definition NO. The test for overlapping models is more general and can be applied to non-overlapping models as well.

The Test for non-overlapping models. The test of nominal size α rejects H_0 in favor of model $\mathcal{P}$ if $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n > z_{\alpha/2}$, rejects H_0 in favor of model $\mathcal{Q}$ if $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n < -z_{\alpha/2}$, and accepts H_0 otherwise.

The Test for overlapping models. Let b_n be a sequence of positive numbers such that $b_n^{-1} + n^{-1}b_n \to 0$. The test of nominal size α rejects H_0 in favor of model $\mathcal{P}$ if $n\hat{\omega}_n^2 > b_n$ and $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n > z_{\alpha/2}$, rejects H_0 in favor of model $\mathcal{Q}$ if $n\hat{\omega}_n^2 > b_n$ and $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n < -z_{\alpha/2}$, and accepts H_0 otherwise.

The test for overlapping models is a two-step test. In the first step, one tests $H_{00} : \omega_\mu^2 = 0$ by comparing $n\hat{\omega}_n^2$ to a critical value b_n that increases to infinity with the sample size. If the first step rejects, one proceeds to the second step where one compares $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n$ to the standard normal critical value. Such a two step structure is the same as the test proposed for overlapping models in Vuong (1989). However, the critical values chosen in Vuong (1989) and in this paper are of different nature. The critical value used in Vuong (1989) is a quantile of the asymptotic distribution of $n\hat{\omega}_n^2$, while we let the critical value increase with n. The reason for such a difference is two-fold. First, using our critical value guarantees that the test has correct asymptotic size while using a finite critical value does not. Shi (2009) shows that the original Vuong test has size distortion because the pointwise asymptotic distributions of the test statistics do not approximate the finite sample distribution uniformly well. The same lack of uniformity issue exists for the test statistics in this paper, as discussed in Section 4.1. Second, the asymptotic distribution of $n\hat{\omega}_n^2$ does not always exist with moment inequality models due to partially identification.

In the first step of the test for overlapping models, b_n is called a tuning parameter. The tuning parameter is analogous to the moment selection criteria in Andrews and Soares (2010). We show below that our test for overlapping models has correct asymptotic size as long as $b_n^{-1} + n^{-1}b_n \to 0$. In practice, suitable choices of b_n are $2\log n$ and $2\log\log n$, which are analogous to the BIC and the Hannan-Quinn information criterion, respectively, as noted in Cheng (2008). The trade-off one should have in mind when choosing b_n is the trade-off between the size and the power of the test in finite samples. A bigger b_n better controls the size but may result in lower power, and vice versa.

5 Asymptotic Size of the Test for Non-overlapping Models

In this section, we show that the asymptotic size of the test for non-overlapping models defined in the previous section is correct.

Let $\mathcal{H}_0^{no}$ denote the set of null distributions in the case of non-overlapping models. We define $\mathcal{H}_0^{no}$ below. The size of the test for non-overlapping models of nominal size α over $\mathcal{H}_0^{no}$ is

$$SZ_n^{no}(\alpha) = \sup_{\mu\in\mathcal{H}_0^{no}} \Pr_\mu\left(n^{1/2}|\widehat{QLR}_n|/\widehat{\omega}_n > z_{\alpha/2}\right). \tag{5.1}$$

We approximate it using the asymptotic size:

$$AsySZ^{no}(\alpha) = \limsup_{n\to\infty} SZ_n^{no}(\alpha). \tag{5.2}$$

The following assumption is imposed on the parameter space and the moment functions.

Assumption 4. (a) *Θ and B are compact, and*

(b) *the moment functions $m(x,\theta)$ and $g(x,\beta)$ are continuously differentiable in θ and β over Θ and B, respectively, for all $x \in \mathcal{X}$.*

Let $m_i(\theta) \equiv m(X_i,\theta)$ and $g_i(\beta) \equiv g(X_i,\beta)$. Let $m_{j,i}$ and $g_{j,i}$ denote the jth elements of m_i and g_i, respectively.

For a data distribution μ, and parameters $\theta \in \Theta$ and $\beta \in B$, let

$$\begin{aligned} S^m_\mu(\gamma,\theta) &= E_\mu e^{\gamma' m_i(\theta)} m_i(\theta) m_i(\theta)' \\ S^g_\mu(\lambda,\beta) &= E_\mu e^{\lambda' g_i(\beta)} g_i(\beta) g_i(\beta)'. \end{aligned} \tag{5.3}$$

For $\theta^* \in \Theta^*_\mu$ and $\beta^* \in B^*_\mu$, let

$$\begin{aligned} \Lambda^*_{\mu,i} &= e^{\gamma^*_\mu(\theta^*)' m_i(\theta^*)} - e^{\lambda^*_\mu(\beta^*)' g_i(\beta^*)}, \\ \mathcal{M}^*_\mu &= \mathcal{M}_\mu(\gamma^*_\mu(\theta^*),\theta^*), \text{ and } \mathcal{N}^*_\mu = \mathcal{N}_\mu(\lambda^*_\mu(\beta^*),\beta^*). \end{aligned} \tag{5.4}$$

Note that $\omega^2_\mu = E_\mu(\Lambda^*_{\mu,i})^2$. By (3.7), $\Lambda^*_{\mu,i}$, $\mathcal{M}^*_\mu$ and $\mathcal{N}^*_\mu$ do not vary with θ^* and β^* in Θ^*_μ and B^*_μ.

Let $eig_{\min}(A)$ denote the smallest eigenvalue of a matrix A. For a positive number M, let Γ^m_M denote $N_M(0_{d_m}) \cap (R^{d_p} \times R^{d_m-d_p}_+)$, where $N_M(0_{d_m})$ is a closed ball in R^{d_m} centered at the origin with radius M. Let Γ^g_M denote $N_M(0_{d_g}) \cap (R^{d_q} \times R^{d_g-d_q}_+)$. Let $\phi = (\gamma',\theta')'$ and $\psi = (\lambda',\beta')'$. Let " $\wedge$ " and " $\vee$ " denote the minimum operator and the maximum operator, respectively. Let $N_\varepsilon(\Theta^*_\mu) = \bigcup_{\theta\in\Theta^*_\mu} N_\varepsilon(\theta)$ and $N_\varepsilon(B^*_\mu) = \bigcup_{\beta\in B^*_\mu} N_\varepsilon(\beta)$.

Define $\mathcal{H}^{no}_0$ to be the set of μ such that

(i) $\{X_i\}^n_{i=1}$ is an i.i.d. sample from μ,

(ii) $d(\mathcal{P},\mu) = d(\mathcal{Q},\mu) \le M_1$,

(iii) Assumption 3 holds,

(iv) for all $\varepsilon > 0$, there exists $\delta_\varepsilon > 0$ not dependent on μ such that

$$\sup_{\theta\in\Theta\backslash N_\varepsilon(\Theta^*_\mu)} \mathcal{M}_\mu(\gamma^*_\mu(\theta),\theta) < \mathcal{M}^*_\mu - \delta_\varepsilon \text{ and } \sup_{\beta\in B\backslash N_\varepsilon(B^*_\mu)} \mathcal{N}_\mu(\lambda^*_\mu(\beta),\beta) < \mathcal{N}^*_\mu - \delta_\varepsilon\ ,$$

(v) $\sup_{\theta\in\Theta} ||\gamma^*_\mu(\theta)|| \vee \sup_{\beta\in B} ||\lambda^*_\mu(\beta)|| \le M - \delta$,

$$\text{(vi)} \quad \inf_{\phi \in \Gamma_M^m \times \Theta} eig_{\min}(S_\mu^m(\phi)) \wedge \inf_{\psi \in \Gamma_M^g \times B} eig_{\min}(S_\mu^g(\psi)) > \delta, \text{ and} \tag{5.5}$$

$$\text{(vii)} \quad E_\mu \sup_{\phi \in \Gamma_M^m \times \Theta} \big[e^{(2+\delta)\gamma' m_i(\theta)} + || \frac{\partial e^{\gamma' m_i(\theta)}}{\partial \phi} ||^{2+\delta} + || \frac{\partial^2 e^{\gamma' m_i(\theta)}}{\partial \gamma \partial \phi} ||^{2+\delta} + \sum_{j=1}^{d_m} || \frac{\partial^3 e^{\gamma' m_i(\theta)}}{\partial \gamma_j \partial \gamma \partial \phi'} || \big] +$$

$$E_\mu \sup_{\psi \in \Gamma_M^g \times B} \big[e^{(2+\delta)\lambda' g_i(\beta)} + || \frac{\partial e^{\lambda' g_i(\beta)}}{\partial \psi} ||^{2+\delta} + || \frac{\partial^2 e^{\lambda' g_i(\beta)}}{\partial \lambda \partial \psi'} ||^{2+\delta} + \sum_{j=1}^{d_g} || \frac{\partial^3 e^{\lambda' g_i(\beta)}}{\partial \lambda_j \partial \lambda \partial \psi'} || \big] < M,$$

where M_1, M and δ are positive constants. The set $\mathcal{H}_0^{no}$ depends on those constants, but for notational simplicity, we suppress their dependence.

Remarks. (a) Condition (i) in (5.5) is the i.i.d. assumption. The results in this paper can be extended to the stationary case. The equality part of Condition (ii) requires μ to be a null distribution. The inequality part of (ii) excludes the distributions that are further than M_1 away from the models according to the KL distance. This part is needed to guarantee that ω_μ^2 is bounded away from zero. It is implied by condition (v) if the moment functions are bounded, but not necessarily so otherwise. Condition (iii) requires the pseudo-true distribution of both models to be unique. Condition (iv) prevents the population dual criterion functions $\mathcal{M}_\mu(\gamma_\mu^*(\theta), \theta)$ and $\mathcal{N}_\mu(\lambda_\mu^*(\beta), \beta)$ from being flat near the pseudo-true sets. Condition (v) restricts $\gamma_\mu^*(\theta)$ and $\lambda_\mu^*(\beta)$ to compact sets. Condition (vi) guarantees that $\mathcal{M}_\mu(\gamma, \theta)$ and $\mathcal{N}_\mu(\lambda, \beta)$ are convex in γ and λ, respectively, so that the estimators of $\gamma_\mu^*(\theta)$ and $\lambda_\mu^*(\beta)$ are $n^{-1/2}$-consistent. Condition (vii) imposes moment restrictions. The exponential moment restrictions may exclude some interesting cases in practice, but are satisfied in many other cases. For example, they are satisfied by models in Examples 3 and 4 above and by models in Examples 1 and 2 if the variables do not have heavy tails.

(b) Assumption 4 implies Assumption 2 and Conditions (v)-(vii) imply Assumption 1. Therefore, the duality results in Lemma 2 hold for $\mu \in \mathcal{H}_0^{no}$ under Assumption 4.

In order to derive the asymptotic size of the test, we show the consistency of the set estimators $\hat{\Theta}_n$ and $\hat{B}_n$ first. Lemma 3 below establishes the consistency of $\hat{\Theta}_n$ and $\hat{B}_n$ w.r.t. the *left* Hausdorff distance. The *left* Hausdorff distance between two subsets, A_1, A_2, of a Euclidean space is the maximum distance of any point in A_1 to A_2:

$$\rho_{lh}(A_1, A_2) = \sup_{a \in A_1} \inf_{a' \in A_2} ||a - a'||. \tag{5.6}$$

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

Lemma 3. *Suppose Assumption* 4 *holds. Then, under all sequences* $\{\mu_n\}_{n=1}^{\infty}$ *such that each* μ_n *satisfies conditions* (i) *and* (iii)-(vii) *in* (5.5), $\rho_{lh}(\hat{\Theta}_n, \Theta^*_{\mu_n}) + \rho_{lh}(\hat{B}_n, B^*_{\mu_n}) \to_p 0$.

Remark. Lemma 3 shows that all points in $\hat{\Theta}_n$ approach $\Theta^*_{\mu_n}$. It does not imply that the neighborhoods of all points in $\Theta^*_{\mu_n}$ are visited by $\hat{\Theta}_n$ even in the limit. Thus, $\hat{\Theta}_n$ is not necessarily consistent w.r.t. the more standard Hausdorff distance. Consistency w.r.t. ρ_{lh} is sufficient for our purpose.

The following lemma guarantees that the asymptotic variance of $n^{1/2}\widehat{QLR}_n$ is bounded away from zero with non-overlapping models.

Lemma 4. *If the models* $\mathcal{P}$ *and* $\mathcal{Q}$ *are non-overlapping, then* $\underline{\omega} \equiv \inf_{\mu \in \mathcal{H}_0^{no}} \omega_\mu^2 > 0$.

Now, we give the theorem describing the asymptotic distribution of $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n$ and the asymptotic size of the test for non-overlapping models.

Theorem 1. *Suppose Assumption* 4 *holds and the models* $\mathcal{P}$ *and* $\mathcal{Q}$ *are non-overlapping. Then,*

(a) *under all sequences* $\{\mu_n \in \mathcal{H}_0^{no}\}_{n=1}^{\infty}$, $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n \to_d N(0,1)$, *and*

(b) *for* $\alpha \in (0,1)$, $AsySZ^{no}(\alpha) = \alpha$.

Remark. Part (a) of Theorem 1 shows that the studentized quasi-likelihood ratio test statistic has a parameter-free asymptotic distribution, which is an attractive feature of the Vuong-type tests. Part (b) shows that the test proposed has correct asymptotic size over a rich set of null distributions. Part (b) follows easily from part (a) by the procedure for deriving asymptotic sizes developed in Andrews and Guggenberger (2009).

6 Asymptotic Size of the Test for Overlapping Models

Let $\mathcal{H}_0^{ol}$ denote the set of null distributions in the case of overlapping models. We define $\mathcal{H}_0^{ol}$ below. The size of the test for overlapping models of nominal size α over $\mathcal{H}_0^{ol}$, is

$$SZ_n^{ol}(\alpha) = \sup_{\mu \in \mathcal{H}_0^{ol}} \Pr_\mu \left(n\hat{\omega}_n^2 > b_n \ \& \ n^{1/2}|\widehat{QLR}_n|/\widehat{\omega}_n > z_{\alpha/2} \right). \tag{6.1}$$

We approximate it using the asymptotic size:

$$AsySZ^{ol}(\alpha) = \limsup_{n\to\infty} SZ_n^{ol}(\alpha). \tag{6.2}$$

In the definition of the asymptotic size, the limsup is taken after the $\sup_{\mu\in\mathcal{H}_0^{ol}}$. Thus, in order to obtain $AsySZ^{ol}(\alpha)$, we need to approximate the distribution of the test statistics uniformly well over $\mathcal{H}_0^{ol}$. This is harder to achieve with overlapping models because the asymptotic distributions of $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n$ and $n\hat{\omega}_n^2$ under H_0 are discontinuous in the true distribution, specifically, in ω_μ^2, as discussed in Section 4.1. We seek to approximate the finite sample distributions of the test statistics in all areas of ω_μ^2 by deriving the asymptotic distributions under drifting sequences of null distributions $\{\mu_n\}_{n=1}^\infty$. Under the drifting sequences, $n\omega_{\mu_n}^2$ can drift to a finite number or infinity, each case approximating the finite sample situation in which ω_μ^2 is close or equal to zero, or ω_μ^2 is bounded away from zero. The idea of using drifting sequences is adopted from Andrews and Guggenberger (2009).

A stronger assumption on the smoothness of the moment functions than Assumption 4 is needed.

Assumption 5. (a) Θ *and* B *are compact, and*

(b) *the moment functions* $m(x,\theta)$ *and* $g(x,\beta)$ *are three times continuously differentiable in* θ *and* β *over* Θ *and* B*, respectively, for all* $x\in\mathcal{X}$.

Let the notation be the same as in the previous section. Define the set of null distributions $\mathcal{H}_0^{ol}$ to be the set of μ such that

(i) Conditions (i), (iii), (v) and (vi) in (5.5) hold,

(ii) $d(\mathcal{P},\mu) = d(\mathcal{Q},\mu)$,

(iii) $\mathcal{M}_\mu^* - \mathcal{M}_\mu(\gamma_\mu^*(\theta),\theta) > C\cdot(\rho_{lh}^2(\theta,\Theta_\mu^*)\wedge\delta)$,
$\mathcal{N}_\mu^* - \mathcal{N}_\mu(\lambda_\mu^*(\beta),\beta) > C\cdot(\rho_{lh}^2(\beta,B_\mu^*)\wedge\delta)$,

(iv) $$E_\mu \sup_{\phi\in\Gamma_M^m\times\Theta}\Big[e^{(2+\delta)\gamma' m_i(\theta)} + ||\frac{\partial e^{\gamma' m_i(\theta)}}{\partial\phi}||^{2+\delta} + ||\frac{\partial^2 e^{\gamma' m_i(\theta)}}{\partial\phi\partial\phi}||^{1+\delta} + \sum_{j=1}^{d_m+d_\theta}||\frac{\partial^3 e^{\gamma' m_i(\theta)}}{\partial\phi_j\partial\phi\partial\phi'}||\Big] +$$
$$E_\mu \sup_{\psi\in\Gamma_M^g\times B}\Big[e^{(2+\delta)\lambda' g_i(\beta)} + ||\frac{\partial e^{\lambda' g_i(\beta)}}{\partial\psi}||^{2+\delta} + ||\frac{\partial^2 e^{\lambda' g_i(\beta)}}{\partial\psi\partial\psi'}||^{1+\delta} + \sum_{j=1}^{d_g+d_\beta}||\frac{\partial^3 e^{\lambda' g_i(\beta)}}{\partial\psi_j\partial\psi\partial\psi'}||\Big]$$
$< M$, and

(v) $E_\mu(\omega_\mu^{-1}\Lambda_{\mu,i}^*)^{2+\delta} < M$ if $\omega_\mu^2 > 0$. (6.3) where M, C and δ are positive constants. The set $\mathcal{H}_0^{ol}$ depends on M, C and δ, but for notational simplicity, we suppress these arguments. Condition (ii) states that μ is a null distribution. We do not need the distances from the models to the null distribution to be bounded from above because ω_μ^2 can be close or equal to zero for overlapping models. Condition (iii) requires the population criterion functions to be locally quadratic in a neighborhood of the identified set. This condition is analogous to the identification conditions in the literature of point-identified models.[11] It gives us the $n^{-1/2}$-consistency of the set estimators. Condition (iv) is similar to condition (vii) in (5.5). Condition (iv) in (6.3) requires the third derivatives of $e^{\gamma' m_i(\theta)}$ with respect to $\phi \equiv (\gamma', \theta')'$ to be dominated by a μ-integrable function, while in condition (vii) in (5.5), the second and third derivatives are with respect to γ only. The additional condition (v) is needed because we want to derive the asymptotic distribution of the studentized quasi-likelihood ratio statistic when the standard deviation of $\widehat{QLR}_n$ converges to zero in probability.

Let ω_n^2 abbreviate $\omega_{\mu_n}^2$. The following definition specifies the drifting sequences of μ's that we consider.

Definition SEQ. *For $\sigma \in [0, \infty]$, let Seq_σ be the set of sequences $\{\mu_{u_n} \in \mathcal{H}_0^{ol}\}_{n=1}^\infty$, such that $\{u_n\}_{n=1}^\infty$ is a subsequence of $\{n\}$, and*

$$u_n \omega_{u_n}^2 \to \sigma^2. \tag{6.4}$$

Let $Seq = \bigcup_{\sigma \in [0,\infty]} Seq_\sigma$.

Notice that we allow σ to take values in the extended real space.

The following Lemma is concerned with the convergence rate of the set estimators under drifting sequences of null distributions. In this lemma, the drifting sequences $\{\mu_n\}_{n=1}^\infty$ are not necessarily in Seq. Lemma 5 is obtained using the quadratic bounding approach described in the introduction. This approach takes into account the non-differentiability of the population and the sample criterion functions.

Lemma 5. *Suppose Assumption* 5 *holds. Then, under any drifting sequence $\{\mu_n\}_{n=1}^\infty$ such that each μ_n satisfies conditions* (i) *and* (iii)-(v) *of* (6.3)*, we have $\rho_{lh}(\hat{\Theta}_n, \Theta_n^*) + \rho_{lh}(\hat{B}_n, B_n^*) = O_p(n^{-1/2})$.*

Lemma 6 below establishes the asymptotic distributions of the test statistics under

[11] A similar condition is used in Chernozhukov, Hong, and Tamer (2007) in the context of partially identified models.

drifting sequences in Seq.

Lemma 6. *Suppose Assumption* 5 *holds. Then for* $\sigma \in [0, \infty]$ *and a subsequence* $\{u_n\}_{n=1}^{\infty}$ *of* $\{n\}$*, under a drifting sequence* $\{\mu_{u_n}\}_{n=1}^{\infty} \in Seq_{\sigma}$,

(a) *if* $\sigma \in [0, \infty)$, $u_n \widehat{\omega}_{u_n}^2 = O_p(1)$, *and*

(b) *if* $\sigma = \infty$, $u_n^{1/2} \widehat{QLR}_{u_n} / \omega_{u_n} \to_d N(0, 1)$ *and* $\widehat{\omega}_{u_n}^2 / \omega_{u_n}^2 \to_p 1$.

It follows easily from Lemma 6 that $AsySZ^{ol}(\alpha) \leq \alpha$. An extra condition is needed for the test not to be asymptotically conservative and is stated as Assumption 6 below. Assumption 6 requires the existence of at least one $\mu \in \mathcal{H}_0^{ol}$ under which the pseudo-true distributions from the two models are not the same. Assumption 6 is not restrictive for nonnested models because for a $\mu \in \mathcal{H}_0^{ol}$ that belongs to neither $\mathcal{P}$ or $\mathcal{Q}$, the pseudo-true distributions typically are different except in some pathological cases. Assumption 6 is violated when $\mathcal{P}$ and $\mathcal{Q}$ are nested, which is the reason that we restrict our attention to nonnested models.

Assumption 6. *There exists* $\mu \in \mathcal{H}_0^{ol}$, *such that* $P_{\mu}^* \neq Q_{\mu}^*$.

Theorem 2 below summarizes the null properties for the model selection test for overlapping models.

Theorem 2. *Suppose Assumption* 5 *holds. Then, for all* $\alpha \in (0, 1)$,

(a) $AsySZ^{ol}(\alpha) \leq \alpha$, *and*

(b) *if Assumption* 6 *also holds, then* $AsySZ^{ol}(\alpha) = \alpha$.

Remark. Part (a) of Theorem 2 shows that the test proposed has correct asymptotic size over a rich set of null distributions. Part (b) shows that the test is not asymptotically conservative.

7 Power Properties of the Tests

We now show that our model selection tests are consistent against fixed alternatives and have power against $n^{-1/2}$-local alternatives. The consistency of the tests implies that if the two candidate models are not equally close to the true distribution (in terms of the KL distance), the tests select the closer model with probability approaching one. The local power result implies that in finite samples, the tests select the model closer to the true distribution with high probabilities if the models are not equally close to the truth.

Assumption FA below specifies the properties of a fixed alternative μ under which model $\mathcal{P}$ is closer to μ than model $\mathcal{Q}$.

Assumption FA. *The true distribution* μ *satisfies that*

(a) $d(\mathcal{P}, \mu) < d(\mathcal{Q}, \mu)$,

(b) $\sup_{\theta\in\Theta} ||\gamma^*_\mu(\theta)|| \vee \sup_{\beta\in B} ||\lambda^*_\mu(\beta)|| < \infty$

(c) *for some* $\delta > 0$ *and* $M = \delta + \sup_{\theta\in\Theta} ||\gamma^*_\mu(\theta)|| \vee \sup_{\beta\in B} ||\lambda^*_\mu(\beta)||$, *conditions* (i), (iii) *and* (vi) *in* (5.5) *hold, and*

(d) $E_\mu \sup_{\phi\in\Gamma^m_M\times\Theta} \left[e^{2\gamma' m_i(\theta)} + ||\partial^2 e^{\gamma' m_i(\theta)}/\partial\gamma\partial\gamma'||\right] + E_\mu \sup_{\psi\in\Gamma^g_M\times B} \left[e^{2\lambda' g_i(\beta)} + ||\partial^2 e^{\lambda' g_i(\beta)} /\partial\lambda\partial\lambda'||\right] < \infty$.

Assumption FA(a) implies that μ is a fixed alternative distribution. Assumption FA(b) is essentially the same condition as (v) in (5.5). Assumption FA(d) is a weaker version of condition (vii) in (5.5).

Theorem 3 below shows that our tests for both non-overlapping models and overlapping models are consistent against all fixed alternatives.

Theorem 3. *Suppose Assumptions* 4 *and* FA *hold. Then,*

(a) $\lim_{n\to\infty} \Pr_\mu \left(n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n > z_{\alpha/2}\right) = 1,$ *and*

(b) $\lim_{n\to\infty} \Pr_\mu \left(n\widehat{\omega}^2_n > b_n \;\&\; n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n > z_{\alpha/2}\right) = 1.$

Remark. Part (a) shows that the model selection test for non-overlapping models has power approaching one against fixed alternatives. Part (b) shows that the test for overlapping models has the same property. The theorem implies that, if model $\mathcal{P}$ is closer to the true data generating process than $\mathcal{Q}$, the tests select $\mathcal{P}$ with probability approaching one as the sample size grows large.

Assumption LA1 below specifies the properties of a sequence of $n^{-1/2}$-local alternatives against which the test for non-overlapping models has nontrivial power.

Assumption LA1. *The sequence of true distributions* $\{\mu_n\}^\infty_{n=1}$ *satisfies that*

(a) $n^{1/2}\big(d(\mathcal{P}, \mu_n) - d(\mathcal{Q}, \mu_n)\big) \to h_1 \in R\backslash\{0\}$,

(b) $d(\mathcal{P}, \mu_n) \to h_2 \in R_+$ *and* $\omega^2_{\mu_n} \to \sigma^2 \in R_+$, *and*

(c) *each* μ_n *satisfies conditions* (i) *and* (iii)-(vii) *in* (5.5).

Assumption LA1(a) simply states that μ_n belongs to the $n^{-1/2}$-local alternative hypothesis $H_1 : n^{1/2}(d(\mathcal{P}, \mu_n) - d(\mathcal{Q}, \mu_n)) \to h_1 \neq 0$. When $h_1 > 0$, model $\mathcal{Q}$ is closer to μ_n than is $\mathcal{P}$, and when $h_1 < 0$, model $\mathcal{P}$ is closer. Assumption LA1(b) requires the true distributions to be within finite KL distance from the models. Assumption LA1(c) states the maintained hypothesis.

Theorem 4 below shows that our model selection test for non-overlapping models has nontrivial power against the $n^{-1/2}$-local alternatives specified in Assumption LA1.

Theorem 4. *Suppose Assumptions* 4 *and* LA1 *hold and the models* $\mathcal{P}$ *and* $\mathcal{Q}$ *are non-overlapping. Then,*

(a) $\lim_{n\to\infty} \Pr_{\mu_n}\left(n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n > z_{\alpha/2}\right) = \Phi\left(-z_{\alpha/2} - \sigma^{-1}e^{-h_2}h_1\right)$, *where* Φ *is the c.d.f. of the standard normal distribution, and*

(b) $\lim_{n\to\infty} \Pr_{\mu_n}\left(n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n < -z_{\alpha/2}\right) = \Phi\left(-z_{\alpha/2} + \sigma^{-1}e^{-h_2}h_1\right)$.

Remark. Part (a) provides the asymptotic probability of selecting model $\mathcal{P}$ under the local alternatives defined in Assumption LA1, while Part (b) provides the asymptotic probability of selecting model $\mathcal{Q}$ under the same local alternatives. The theorem implies that the asymptotic power function of the test is $\Phi\left(-z_{\alpha/2} - \sigma^{-1}e^{-h_2}h_1\right) + \Phi\left(-z_{\alpha/2} + \sigma^{-1}e^{-h_2}h_1\right)$. The minimum value of the asymptotic power is the asymptotic size, α, and is reached only at $h_1 = 0$ (i.e. under the null). This shows that the test has nontrivial power against $n^{-1/2}$-local alternatives. The asymptotic power increases to one as $\sigma^{-1}e^{-h_2}|h_1|$ increases to infinity – as the local alternatives approach the global alternatives. Theorem 4 also shows that the test is directional. That is, when $h_1 > 0$, the asymptotic probability of selecting model $\mathcal{Q}$ is bigger than $\alpha/2$, while that of selecting model $\mathcal{P}$ is smaller than $\alpha/2$.

Assumption LA2 below specifies the properties a sequence of $n^{-1/2}$-local alternatives against which the test for overlapping models has nontrivial power.

Assumption LA2. *The sequence of true distributions* $\{\mu_n\}_{n=1}^{\infty}$ *satisfies that*

(a) $n^{1/2}\omega_{\mu_n}^{-1}\big(d(\mathcal{P},\mu_n) - d(\mathcal{Q},\mu_n)\big) \to h_1^{ol} \in R\backslash\{0\}$,

(b) $d(\mathcal{P},\mu_n) \to h_2^{ol} \in R_+$ *and* $b_n^{1/2}n^{-1/2}\omega_{\mu_n}^{-1} \to 0$, *and*

(c) *each* μ_n *satisfies conditions* (i) *and* (iii)-(v) *in* (6.3).

Assumption LA2(a) states that μ_n belongs to the $n^{-1/2}$-local alternative hypothesis $H_1 : n^{1/2}\omega_{\mu_n}^{-1}(d(\mathcal{P},\mu_n) - d(\mathcal{Q},\mu_n)) \to h_1^{ol} \neq 0$. The difference in distances is normalized by ω_{μ_n} to deal with the potential situation in which $\omega_{\mu_n} \to 0$. Assumption LA2(b) requires the true distribution to be within finite KL distance from the models and ω_{μ_n} not to vanish too fast. Assumption LA2(c) states the maintained hypothesis.

Theorem 5 below shows that our model selection test for overlapping models has nontrivial power against the $n^{-1/2}$-local alternatives specified in Assumption LA2.

Theorem 5. *Suppose Assumptions* 5 *and* LA2 *hold. Then,*

(a) $\lim_{n\to\infty} \Pr_{\mu_n}\left(n\hat{\omega}_n > b_n \ \&\ n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n > z_{\alpha/2}\right) = \Phi(-z_{\alpha/2} - e^{-h_2}h_1^{ol})$, *and*

(b) $\lim_{n\to\infty} \Pr_{\mu_n}\left(n\hat{\omega}_n > b_n \ \&\ n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n < -z_{\alpha/2}\right) = \Phi(-z_{\alpha/2} + e^{-h_2}h_1^{ol})$.

The interpretation of Theorem 5 is similar to that of Theorem 4.

8 Simulation

This section reports some Monte Carlo results for the model selection problem in Example 3. We demonstrate that the null rejection probability of the proposed test for overlapping models is close to the nominal size. The test selects the better model with high probabilities and selects the worse model very rarely at sample sizes that are relevant in practice.

We consider the entry game in Example 3 with no covariates (hence $X = 1$) for simplicity. Since there are no covariates, the data include only observations on the equilibrium outcome Y. The variable Y can take only four values. Thus, each distribution of Y is a multinomial distribution which can be written as a vector in the three-dimensional simplex:

$$\Delta^3 = \left\{(p_0, p_1, p_2, p_3) \in [0,1]^4 : \sum_{j=0}^{3} p_j = 1\right\}. \tag{8.1}$$

We assume symmetric firms, so that $\theta_1 = \theta_2 = c$, $a_1 = a_2 = a$ and $F_{\varepsilon_1} = F_{\varepsilon_2} = F$. We also assume that there is no correlation between ε_1 and ε_2. In this particular entry game model without covariates, allowing for asymmetry between the firms makes the models in Example 3 untestable in the sense that $\mathcal{P} = \mathcal{Q} = int(\Delta^3)$. Then, no statistical method is useful in distinguishing the two models. Allowing for correlation between the error terms makes model $\mathcal{P}$ untestable.[12]

We normalize $F(\cdot;\theta_\varepsilon)$ to $\Phi(\cdot)$, the standard normal c.d.f. This is a normalization rather than a restriction on the models for the following reasons. Before the normalization, the predicted probability functions in (2.7) are

$$\begin{aligned} p_{(0,0)}(c,a,\theta_\varepsilon) &= (1 - F(c;\theta_\varepsilon))^2 \\ p_{(0,1)}(c,a,\theta_\varepsilon) &= p_{(1,0)}(c,a,\theta_\varepsilon) = F(c;\theta_\varepsilon)(1 - F(c+a;\theta_\varepsilon)) \\ p_{(1,1)}(c,a,\theta_\varepsilon) &= F^2(c+a;\theta_\varepsilon). \end{aligned} \tag{8.2}$$

[12] Adding covariates can make the models testable even when asymmetry and error correlation are allowed. We will investigate models with covariates using Monte Carlo simulation in later versions of this paper.

In both model $\mathcal{P}$ and model $\mathcal{Q}$, the parameters are (partially) identified through $p_{(0,0)}$, $p_{(0,1)}$, $p_{(1,0)}$ and $p_{(1,1)}$, which are in turn (partially) identified by the observed probabilities $\Pr_\mu(Y = j)$, $j = (0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$. From (8.2), we see that only the two functions, $F(c;\theta_\varepsilon)$ and $F(c+a;\theta_\varepsilon)$ can be (partially) identified. For each ($F(\cdot,\theta_\varepsilon)$, c, a), let $c' = \Phi^{-1}(F(c;\theta_\varepsilon))$, and $a' = \Phi^{-1}(F(c+a;\theta_\varepsilon)) - c'$. Then, $\Phi(c') = F(c;\theta_\varepsilon)$ and $\Phi(c'+a') = F(c+a;\theta_\varepsilon)$. That is, ($F(\cdot,\theta_\varepsilon)$, c, a) is observationally equivalent to (Φ, c', a'). Thus, we can normalize $F(\cdot,\theta_\varepsilon)$ to Φ.

After the normalization, model $\mathcal{P}$ becomes the set of P such that for some $c \in R$, $a \in R_-$

$$\begin{aligned} E_P\left[(1-\Phi(c))^2 - 1(Y=(0,0))\right] &= 0, \\ E_P\left[\Phi^2(c+a)) - 1(Y=(1,1)\right] &= 0, \\ E_P\left[\Phi(c)(1-\Phi(c+a)) - 1(Y=j)\right] &\geq 0 \text{ for } j=(0,1),\ (1,0). \end{aligned} \tag{8.3}$$

Model $\mathcal{Q}$ becomes the set of Q such that for some $c \in R$, $a \in R_+$,

$$\begin{aligned} E_Q\left[(1-\Phi(c))^2 - 1(Y=(0,0))\right] &\geq 0, \\ E_Q\left[\Phi^2(c+a)) - 1(Y=(1,1)\right] &\geq 0, \\ E_Q\left[\Phi(c)(1-\Phi(c+a)) - 1(Y=j)\right] &= 0 \text{ for } j=(0,1),\ (1,0). \end{aligned} \tag{8.4}$$

We generate the data according to three data generating processes (DGPs). DGP NL below is a null DGP. DGPs 1 and 2 below are alternative DGPs, each generated from one candidate model.

DGP NL: $\{Y_i\}_{i=1}^n$ is generated from a multinomial distribution with probability vector: $\mu_0 = (0.21267,\ 0.21267,\ 0.01,\ 0.56466)$. The distribution μ_0 is calibrated so that $d(\mathcal{P}, \mu_0) = d(\mathcal{Q}, \mu_0)$.

DGP 1: $a = -0.25$, $c = 0.5$. The firms play a complete information simultaneous-move Nash game. In the case of multiple equilibria, we set the equilibrium selection mechanism to always select the equilibrium $(1,0)$.[13] Under DGP 1, model $\mathcal{P}$ is correctly specified and model $\mathcal{Q}$ is not. Thus, $\mathcal{P}$ is better than $\mathcal{Q}$.

DGP 2: $a = 0.25$, $c = 0.5$. The firms play a complete information simultaneous-move

[13] When $a < 0$, the game has multiple equilibria only when $c + a < \varepsilon_1 < c$ and $c + a < \varepsilon_2 < c$. The multiple equilibria include $(0,1)$ and $(1,0)$: either player 1 enters or player 2 enters. The equilibrium selection mechanism is used only to generate the data and is not used when carrying out the test.

Nash game. In the case of multiple equilibria, we set the equilibrium selection mechanism to always select the equilibrium $(1, 1)$.[14] Under DGP 2, model $\mathcal{Q}$ is correctly specified and model $\mathcal{P}$ is not. Thus, $\mathcal{Q}$ is better than $\mathcal{P}$.

The two models are overlapping because they describe the same distributions when $a = 0$. Therefore, we use the test for overlapping models. For the first step of the test, we choose $b_n = 2 \log n$, where n is the sample size. We consider four sample sizes, $n = 500$, 750, 1000, 1500. These sample sizes are relevant in practice. For example, in Ciliberto and Tamer (2009), where an airline entry game is considered, the sample size (the number of flight routes) is over 2000. The number of simulation draws is set at 2000.

In Table 1 below, the three numbers in each parenthesis are the probabilities that the test selects model $\mathcal{P}$, accepts the null, or selects model $\mathcal{Q}$. The second column shows the probabilities under DGP NL at different sample sizes. The third and the fourth column show the same except under DGPs 1 and 2.

The most important observation from the table is that the test selects the better model with high probabilities, and the probabilities increase to one quickly as the sample size gets large. This is consistent with our theoretical power results. The second observation is that the probabilities that the test makes wrong selection are controlled well by the nominal size.

Table 1. Prob. of choosing $\mathcal{P}$, accepting H_0, and choosing $\mathcal{Q}$

n	DGP Null	DGP 1	DGP 2
500	(.034, .945, .021)	(.519, .481, 0)	(0, .527, .473)
750	(.036, .940, .024)	(.753, .247, 0)	(0, .308, .692)
1000	(.035, .944, .021)	(.864, .136, 0)	(0, .175, .825)
1500	(.037, .938, .025)	(.961, .039, 0)	(0, .060, .940)

The computation cost of the test is relatively low. In the simulation example described

[14] When $a > 0$, the game has multiple equilibria only when $c < \varepsilon_1 < c + a$ and $c < \varepsilon_2 < c + a$. The multiple equilibria include $(0, 0)$ and $(1, 1)$: either both enter or neither does.

above, it takes around 12 seconds to run one simulation iteration with sample size 1500.[15] For models with covariates and more than two parameters, computation time is longer, but we expect it to be in a reasonable range for the reasons discussed in the introduction of this paper.

9 Concluding Remarks

In this paper, we propose Vuong-type model selection tests for moment inequality models. We allow the candidate models to be overlapping or non-overlapping and propose one test for each. The tests are robust to partial identification of the model parameters. We show that the tests have correct asymptotic size in a uniform sense and have good power properties. The simulation results are consistent with our theoretical conclusions.

[15] The simulation is run on a desktop PC with a 2.4 Ghz cpu.

APPENDIX

Throughout the appendix, we replace μ_n with n when μ_n is in a subscript and it does not cause confusion to do so. For example, we write $\gamma^*_{\mu_n}(\theta)$ as $\gamma^*_n(\theta)$. Let $\hat{\phi}_n(\theta) = (\hat{\gamma}_n(\theta)', \theta')'$ and $\phi^*_n(\theta) = (\gamma^*_n(\theta)', \theta')'$, and $\hat{\psi}_n(\beta)$ and $\psi^*_n(\beta)$ be defined analogously. We let "r.h.s." denote "right-hand-side" and "l.h.s." denote "left-hand-side".

Let "LLN" denote the weak law of large number for row-wise i.i.d. triangular arrays. The weak law of large number we use here is Theorem 2 in Andrews (1988). Theorem 2 in Andrews (1988) is a law of large numbers for L^1-mixingale triangular arrays. Row-wise i.i.d. triangular arrays are trivially L^1-mixingales. The uniform integrability condition required in that theorem is guaranteed by higher-order-moment-finite conditions in this paper.

We first present a few auxiliary lemmas, the proofs of which are given in Appendix C. Lemma A1 is an instrumental result for the uniform stochastic boundedness of empirical processes, which is useful for establishing Lemmas A2-A4. Lemma A2 establishes the uniform convergence and rate of convergence of various stochastic processes, which is useful for proving the main lemmas and theorems. Lemma A3 establishes the uniform consistency of $\hat{\gamma}_n(\theta)$, the rate of convergence of $\hat{\gamma}_n(\theta)$, and the continuity of $\gamma^*_n(\theta)$. Lemmas A2 and A3 are stated under Assumption 4 and conditions (i) and (iii)-(vii) in (5.5). Their results hold under Assumption 5 and conditions (i) and (iii)-(v) in (6.3) because Assumption 5 implies Assumption 4 and conditions (i) and (iii)-(v) in (6.3) imply conditions (i), (iii)-(vii) in (5.5).

Lemmas 3, 5 and A2-A4 are stated in terms of $\{n\}$, but because they only impose termwise restrictions on the sequence $\{\mu_n\}_{n=1}^\infty$, their conclusions hold with $\{n\}$ replaced with any subsequence of $\{n\}$.

Lemma A1. *Consider the triangular array of empirical processes* $\{\nu_n(\phi) : \phi \in \Phi\}_{n=1}^\infty$. *If* (i) (Φ, ρ) *is a totally bounded pseudo-metric space,* (ii) $\nu_n(\phi)$ *is stochastically equicontinuous w.r.t.* ρ *and* (iii) *for every* $\phi \in \Phi$, $||\nu_n(\phi)|| = O_p(1)$, *then* $\sup_{\phi\in\Phi} ||\nu_n(\phi)|| = O_p(1)$.[16]

[16] Note that here, Φ denotes the space of ϕ. In the main sections of this paper, Φ stands for the c.d.f. of the standard normal distribution. Hopefully, there is no confusion caused by this abuse of notation.

Let ρ_0, ρ_1 be pseudo-metrics on $\Gamma_M^m \times \Theta$ defined by:

$$\rho_0(\phi_1, \phi_2) = \sup_{N \geq 1} \left[E_n (e^{\gamma_1' m_i(\theta_1)} - e^{\gamma_2' m_i(\theta_2)})^2 \right]^{1/2}$$
$$\rho_1(\phi_1, \phi_2) = \sup_{N \geq 1} \left[E_n || \partial e^{\gamma_1' m_i(\theta_1)} / \partial \phi - \partial e^{\gamma_2' m_i(\theta_2)} / \partial \phi ||^2 \right]^{1/2},$$

where $\phi_1, \phi_2 \in \Gamma_M^m \times \Theta$ and $\phi_j = (\gamma_j', \theta_j')'$, $j = 1, 2$.

Lemma A2. *Suppose Assumption* 4 *holds. Under any sequence* $\{\mu_n\}_{n=1}^\infty$ *such that each* μ_n *satisfies conditions* (i) *and* (iii)-(vii) *in* (5.5), *we have*

(a) *the triangular array of empirical processes* $\{\nu_n^0(\phi) \equiv n^{1/2}(\widehat{\mathcal{M}}_n(\phi) - \mathcal{M}_{\mu_n}(\phi))$: $\phi \in \Gamma_M^m \times \Theta\}$ *is stochastically equicontinuous w.r.t.* ρ_0,

(b) $\sup_{\phi \in \Gamma_M^m \times \Theta} |n^{1/2}(\widehat{\mathcal{M}}_n(\phi) - \mathcal{M}_{\mu_n}(\phi))| = O_p(1)$,

(c) *the triangular array of empirical processes* $\{\nu_n^1(\phi) \equiv n^{1/2}(\partial \widehat{\mathcal{M}}_n(\phi)/\partial\gamma - \partial \mathcal{M}_{\mu_n}(\phi)/\partial\gamma)$: $\phi \in \Gamma_M^m \times \Theta\}$ *is stochastically equicontinuous w.r.t.* ρ_1,

(d) $\sup_{\phi \in \Gamma_M^m \times \Theta} ||n^{1/2}(\partial \widehat{\mathcal{M}}_n(\phi)/\partial\gamma - \partial \mathcal{M}_{\mu_n}(\phi)/\partial\gamma)|| = O_p(1)$,

(e) *for all random sequences* $\{\phi_{1,n} \in \Gamma_M^m \times \Theta\}_{n=1}^\infty$ *and* $\{\phi_{2,n} \in \Gamma_M^m \times \Theta\}_{n=1}^\infty$ *such that* $||\phi_{1,n} - \phi_{2,n}|| \to_p 0$, *we have*

$$||\partial^2 \widehat{\mathcal{M}}_n(\phi_{1,n})/\partial\gamma\partial\gamma' - \partial^2 \mathcal{M}_{\mu_n}(\phi_{2,n})/\partial\gamma\partial\gamma'|| \to_p 0$$
$$|\widehat{\mathcal{M}}_n(\phi_{1,n}) - \mathcal{M}_{\mu_n}(\phi_{2,n})| \to_p 0, \text{ and}$$

(f) *parts* (a)-(e) *hold with* Θ, γ, ϕ, $\mathcal{M}$ *and* m *replaced with* B, λ, ψ, $\mathcal{N}$ *and* g, *and with* ρ_0 *and* ρ_1 *redefined for model* $\mathcal{Q}$ *accordingly.*

Lemma A3. *Suppose Assumption* 4 *holds. Under any sequence* $\{\mu_n\}_{n=1}^\infty$ *such that each* μ_n *satisfies conditions* (i) *and* (iii)-(vii) *in* (5.5), *we have*

(a) *for any two random sequences* $\{\theta_{1,n} \in \Theta\}_{n=1}^\infty$ *and* $\{\theta_{2,n} \in \Theta\}_{n=1}^\infty$ *such that* $||\theta_{1,n} - \theta_{2,n}|| \to_p 0$,

$$||\hat{\gamma}_n(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})|| \to_p 0,$$

(b) $\sup_{\theta \in \Theta} ||\hat{\gamma}_n(\theta) - \gamma_n^*(\theta)|| = O_p(n^{-1/2})$,

(c) *for any two random sequences* $\{\theta_{1,n} \in \Theta\}_{n=1}^\infty$ *and* $\{\theta_{2,n} \in \Theta\}_{n=1}^\infty$ *such that* $||\theta_{1,n} - \theta_{2,n}|| \to_p 0$,

$$||\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})|| = O_p(||\theta_{1,n} - \theta_{2,n}||), \text{ and}$$

(d) *parts* (a)-(c) *hold with* θ, Θ, γ, ϕ, $\mathcal{M}$, m *replaced with* β, B, λ, ψ, $\mathcal{N}$, g.

Lemma A4. *Suppose Assumption* 5 *holds. Then, under any sequence* $\{\mu_n\}_{n=1}^{\infty}$ *such that each* μ_n *satisfies conditions* (i) *and* (iii)-(v) *in* (6.3),

(a) *for any two random sequences* $\{\phi_{1,n} \in \Gamma_M^m \times \Theta\}_{n=1}^{\infty}$ *and* $\{\phi_{2,n} \in \Gamma_M^m \times \Theta\}_{n=1}^{\infty}$ *such that* $||\phi_{1,n} - \phi_{2,n}|| \to_p 0$,

$$||\partial^2 \widehat{\mathcal{M}}_n(\phi_{1,n})/\partial\phi\partial\phi' - \partial^2 \mathcal{M}_{\mu_n}(\phi_{2,n})/\partial\phi\partial\phi'|| \to_p 0, \text{ and}$$

(b) *part* (a) *hold with* Θ, ϕ, $\mathcal{M}$ *and* m *replaced with* B, ψ, $\mathcal{N}$ *and* g.

A Proof of the Theorems

Proof of Theorem 1. (a) Let $\hat{\theta}_n \in \hat{\Theta}_n$ and $\hat{\beta}_n \in \hat{B}_n$ be such that $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n) \geq \widehat{\omega}_n^2 - 2^{-n}$. Then, part (a) is implied by:

$$n^{1/2}\widehat{QLR}_n/\omega_n \to_d N(0,1), \text{ and} \tag{A.1}$$

$$\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n)/\omega_n^2 \to_p 1. \tag{A.2}$$

Next, we show (A.1) and (A.2).

Let $\theta_n^* \in \Theta_n^*$ and $\beta_n^* \in B_n^*$ satisfy $||\hat{\theta}_n - \theta_n^*|| \leq \rho_{lh}(\hat{\theta}_n, \Theta_n^*) + 2^{-n}$ and $||\hat{\beta}_n - \beta_n^*|| \leq \rho_{lh}(\hat{\beta}_n, B_n^*) + 2^{-n}$. Then, Lemmas 3 and A3(a) imply that

$$||\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)|| \to_p 0 \text{ and } ||\hat{\psi}_n(\hat{\beta}_n) - \psi_n^*(\beta_n^*)|| \to_p 0. \tag{A.3}$$

First, we show (A.1). Observe that

$$\begin{aligned} \omega_n^{-1} n^{1/2}\widehat{QLR}_n &= \omega_n^{-1} n^{-1/2} \sum_{i=1}^n \left[\exp(\hat{\gamma}_n(\hat{\theta}_n)' m_i(\hat{\theta}_n)) - \exp(\hat{\lambda}_n(\hat{\beta}_n)' g_i(\hat{\beta}_n)\right] \\ &= \omega_n^{-1} n^{-1/2} \sum_{i=1}^n \Lambda_{n,i}^* + A_{n,1} + A_{n,2}, \end{aligned} \tag{A.4}$$

where $A_{n,1} = \omega_n^{-1} n^{-1/2} \sum_{i=1}^n \left[\exp(\hat{\gamma}_n(\hat{\theta}_n)' m_i(\hat{\theta}_n)) - \exp(\gamma_n^*(\theta_n^*)' m_i(\theta_n^*))\right]$ and $A_{n,2} = \omega_n^{-1} n^{-1/2} \sum_{i=1}^n \left[\exp(\lambda_n^*(\beta_n^*)' g_i(\beta_n^*)) - \exp(\hat{\lambda}_n(\hat{\beta}_n)' g_i(\hat{\beta}_n))\right]$.

By the Lyapounov CLT for triangular arrays,

$$\omega_n^{-1} n^{-1/2} \sum_{i=1}^{n} \Lambda_{n,i}^* \rightarrow_d N(0,1). \tag{A.5}$$

The CLT applies because (a) $E_n \Lambda_{n,i}^* = 0$ by condition (ii) in (5.5) and Lemma 2(b), (b) $\omega_n^{-2} E_n(\Lambda_{n,i}^*)^2 = 1$ and (c) $E_n(\omega_n^{-1}\Lambda_{n,i}^*)^{2+\delta} \leq \underline{\omega}^{-2-\delta} E_n(\Lambda_{n,i}^*)^{2+\delta} < \infty$ by Lemma 4 and condition (vii) in (5.5).

It is left to show $A_{n,1} = o_p(1)$ and $A_{n,2} = o_p(1)$ before we can conclude that (A.1) holds. It suffices to show $A_{n,1} = o_p(1)$ since the arguments for $A_{n,2} = o_p(1)$ are analogous. Because we do not have convergence rates for $\hat{\Theta}_n$ and $\hat{B}_n$ under the conditions of Theorem 1, the usual approach of expanding $\exp(\hat{\gamma}_n(\hat{\theta}_n)' m_i(\hat{\theta}_n))$ around $\phi_n^*(\theta_n^*)$ does not give us the desired result. Instead, we show $A_{n,1} = o_p(1)$ by bounding $A_{n,1}$ from both above and below by $o_p(1)$. The lower bound of $A_{n,1}$ is obtained by replacing $\hat{\theta}_n$ with θ_n^* in the expression of $A_{n,1}$ and using the convergence rate result for $\hat{\gamma}_n(\cdot)$ (Lemma A3(b)):

$$\begin{aligned}
A_{n,1} &\geq \omega_n^{-1} n^{-1/2} \sum_{i=1}^{n} \left[\exp(\hat{\gamma}_n(\theta_n^*)' m_i(\theta_n^*)) - \exp(\gamma_n^*(\theta_n^*)' m_i(\theta_n^*)) \right] \\
&= \omega_n^{-1} \left[\partial \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))/\partial \gamma' \right] \left[n^{1/2} (\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*)) \right] \\
&\quad + \omega_n^{-1} n^{1/2} \big(\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*) \big)' \left[\partial^2 \widehat{\mathcal{M}}_n(\tilde{\phi}_n)/\partial \gamma \partial \gamma' \right] \big(\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*) \big) \\
&\geq \omega_n^{-1} \left[\partial \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))/\partial \gamma' - \partial \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))/\partial \gamma' \right] \left[n^{1/2} (\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*)) \right] \\
&\quad + \omega_n^{-1} n^{1/2} \big(\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*) \big)' \left[\partial^2 \widehat{\mathcal{M}}_n(\tilde{\phi}_n)/\partial \gamma \partial \gamma' \right] \big(\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*) \big) \\
&= o_p(1),
\end{aligned} \tag{A.6}$$

where $\tilde{\phi}_n$ lies on the line segment joining $\phi_n^*(\theta_n^*)$ and $\hat{\phi}_n(\theta_n^*)$, the first inequality holds because $\hat{\theta}_n$ is a maximizer of $\widehat{\mathcal{M}}_n(\hat{\gamma}_n(\cdot), \cdot)$, the first equality holds by a Taylor expansion of $\exp(\hat{\gamma}_n(\theta_n^*)' m_i(\theta_n^*))$ around $\gamma_n^*(\theta_n^*)$, and the second inequality holds by the Kuhn-Tucker conditions of the minimization problem $\min_{\gamma \in R^{d_p} \times R_+^{d_m - d_p}} \mathcal{M}_{\mu_n}(\gamma, \theta_n^*)$, and the second equality holds by Lemmas 3, 4, A2(d)-(e) and A3(b).

The upper bound of $A_{n,1}$ is obtained by replacing $\hat{\gamma}_n$ with γ_n^* in the expression of $A_{n,1}$ and applying Lemma A2(a):

$$A_{n,1} \leq \omega_n^{-1} n^{-1/2} \sum_{i=1}^{n} \left[\exp(\gamma_n^*(\hat{\theta}_n)' m_i(\hat{\theta}_n)) - \exp(\gamma_n^*(\theta_n^*)' m_i(\theta_n^*)) \right]$$

$$= \omega_n^{-1}\left[\nu_n^0(\phi_n^*(\hat\theta_n)) + n^{1/2}\big(\mathcal{M}_{\mu_n}(\phi_n^*(\hat\theta_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))\big) - \nu_n^0(\phi_n^*(\theta_n^*))\right]$$
$$\leq \omega_n^{-1}[\nu_n^0(\phi_n^*(\hat\theta_n)) - \nu_n^0(\phi_n^*(\theta_n^*))] = o_p(1), \tag{A.7}$$

where the first inequality holds because $\hat\gamma_n(\hat\theta_n)$ is the minimizer of $\widehat{\mathcal{M}}_n(\cdot,\hat\theta_n)$, the first equality holds for the empirical process $\{\nu_n^0(\phi) : \phi \in \Gamma_M^m \times \Theta\}_{n=1}^\infty$ defined in Lemma A2(a), the second inequality holds because θ_n^* is a maximizer of $\mathcal{M}_{\mu_n}(\phi_n^*(\theta))$, and the second equality holds by Lemmas 3, 4, A2(a) and A3(c).

Therefore, $A_{n,1} = o_p(1)$.

Next, we show (A.2). By a mean-value expansion of $\exp(\hat\gamma_n(\hat\theta_n)'m_i(\hat\theta_n))$ around $\phi_n^*(\theta_n^*)$, we have

$$\begin{aligned}
&\omega_n^{-2}\widehat\omega_n^2(\hat\theta_n,\hat\beta_n)\\
&= \omega_n^{-2}n^{-1}\sum_{i=1}^n(\Lambda_{n,i}^*)^2 + \\
&2\omega_n^{-2}n^{-1}\sum_{i=1}^n\left[\frac{\partial e^{\tilde\gamma_n' m_i(\tilde\theta_n)}}{\partial\phi'}(\hat\phi_n(\hat\theta_n)-\phi_n^*(\theta_n^*)) - \frac{\partial e^{\tilde\lambda_n' g_i(\tilde\beta_n)}}{\partial\psi'}(\hat\psi_n(\hat\beta_n)-\psi_n^*(\beta_n^*))\right]\Lambda_{n,i}^* + \\
&\omega_n^{-2}n^{-1}\sum_{i=1}^n\left[\frac{\partial e^{\tilde\gamma_n' m_i(\tilde\theta_n)}}{\partial\phi'}(\hat\phi_n(\hat\theta_n)-\phi_n^*(\theta_n^*)) - \frac{\partial e^{\tilde\lambda_n' g_i(\tilde\beta_n)}}{\partial\psi'}(\hat\psi_n(\hat\beta_n)-\psi_n^*(\beta_n^*))\right]^2\\
&\equiv W_{n,1} + W_{n,2} + W_{n,3},
\end{aligned} \tag{A.8}$$

where $(\tilde\gamma_n', \tilde\theta_n')'$ lies on the line segment joining $\hat\phi_n(\hat\theta_n)$ and $\phi_n^*(\theta_n^*)$ and $(\tilde\lambda_n', \tilde\beta_n')'$ lies on the line segment joining $\hat\psi_n(\hat\beta_n)$ and $\psi_n^*(\beta_n^*)$.

Applying the LLN to $W_{n,1}$, we have $W_{n,1} \equiv \omega_n^{-2}n^{-1}\sum_{i=1}^n(\Lambda_{n,i}^*)^2 \to_p 1$. The LLN applies because (a) $E_n\omega_n^{-2}(\Lambda_{n,i}^*)^2 = 1$, and (b) by condition (vii) in (5.5) and Lemma 4, $\sup_{n\geq1}E_n(\omega_n^{-1}\Lambda_{n,i}^*)^{2+\delta} \leq \underline{\omega}^{-2-\delta}\sup_{n\geq1}E_n(\Lambda_{n,i}^*)^{2+\delta} < \infty$.

The third summand, $W_{n,3}$, in the r.h.s. of (A.8) is $o_p(1)$ because

$$\begin{aligned}
0 &\leq W_{n,3}\\
&\leq 2\underline{\omega}^{-2}(\hat\phi_n(\hat\theta_n)-\phi_n^*(\theta_n^*))\left(n^{-1}\sum_{i=1}^n\frac{\partial e^{\tilde\gamma_n' m_i(\tilde\theta_n)}}{\partial\phi}\frac{\partial e^{\tilde\gamma_n' m_i(\tilde\theta_n)}}{\partial\phi'}\right)(\hat\phi_n(\hat\theta_n)-\phi_n^*(\theta_n^*))\\
&\quad+2\underline{\omega}^{-2}(\hat\psi_n(\hat\beta_n)-\psi_n^*(\beta_n^*))\left(n^{-1}\sum_{i=1}^n\frac{\partial e^{\tilde\lambda_n' g_i(\tilde\beta_n)}}{\partial\psi}\frac{\partial e^{\tilde\lambda_n' g_i(\tilde\beta_n)}}{\partial\psi'}\right)(\hat\psi_n(\hat\beta_n)-\psi_n^*(\beta_n^*))\\
&= o_p(1),
\end{aligned} \tag{A.9}$$

where the second inequality holds by the inequality, $(a+b)^2 \leq 2a^2 + 2b^2$ and Lemma 4, and the equality holds by (A.3) and

$$E_n \left\| n^{-1} \sum_{i=1}^{n} \frac{\partial e^{\tilde{\gamma}_n' m_i(\tilde{\theta}_n)}}{\partial \phi} \frac{\partial e^{\tilde{\gamma}_n' m_i(\tilde{\theta}_n)}}{\partial \phi'} \right\| \leq E_n || \partial e^{\tilde{\gamma}_n' m_i(\tilde{\theta}_n)} / \partial \phi ||^2 \leq M$$
$$E_n \left\| n^{-1} \sum_{i=1}^{n} \frac{\partial e^{\tilde{\lambda}_n' g_i(\tilde{\beta}_n)}}{\partial \psi} \frac{\partial e^{\tilde{\lambda}_n' g_i(\tilde{\beta}_n)}}{\partial \psi'} \right\| \leq E_n || \partial e^{\tilde{\lambda}_n' g_i(\tilde{\beta}_n)} / \partial \psi' ||^2 \leq M, \qquad \text{(A.10)}$$

which holds by the triangular inequality, the inequality $||aa'|| \leq ||a||^2$ and condition (vii) in (5.5).

The second summand, $W_{n,2}$, in the r.h.s. of (A.8) is $o_p(1)$ because, by the Cauchy-Schwartz inequality, $0 \leq W_{n,2} \leq 2[W_{n,1} \cdot W_{n,3}]^{1/2}$.

Therefore, (A.2) holds.

(b) Let $\{\mu_n \in \mathcal{H}_0^{no}\}_{n=1}^{\infty}$ satisfy $\Pr_n \left(n^{1/2} |\widehat{QLR}_n| / \widehat{\omega}_n > z_{\alpha/2} \right) \geq SZ_n^{no}(\alpha) - 2^{-n}$. Such a sequence always exists. Then,

$$\limsup_{n \to \infty} \Pr_n \left(n^{1/2} |\widehat{QLR}_n| / \widehat{\omega}_n > z_{\alpha/2} \right) = AsySZ^{no}(\alpha). \qquad \text{(A.11)}$$

By part (a), the l.h.s. of the equation above equals α. Therefore, $AsySZ^{no}(\alpha) = \alpha$. ■

Proof of Theorem 2. (a) Let $\{a_n\}$ be a subsequence of $\{n\}$ such that $AsySZ^{ol}(\alpha) = \lim_{n\to\infty} SZ_{a_n}^{ol}(\alpha)$. Such a sequence always exists. Let $\{\mu_n \in \mathcal{H}_0^{ol}\}_{n=1}^{\infty}$ be a sequence such that for each n,

$$\Pr_n \left(n\widehat{\omega}_n^2 > b_n \ \& \ n^{1/2} |\widehat{QLR}_n| / \widehat{\omega}_n > z_{\alpha/2} \right) \geq SZ_n^{ol}(\alpha) - 2^{-n}. \qquad \text{(A.12)}$$

Let $\{u_n\}$ be a subsequence of $\{a_n\}$ such that $u_n \omega_{u_n}^2 \to \sigma$, $\sigma \in [0, \infty]$. Such subsequences always exist because we allow σ to take values in the extended real space. Then,

$$AsySZ^{ol}(\alpha) = \lim_{n\to\infty} \Pr_{u_n} \left(u_n \widehat{\omega}_{u_n}^2 > b_{u_n} \ \& \ u_n^{1/2} |\widehat{QLR}_{u_n}| / \widehat{\omega}_{u_n} > z_{\alpha/2} \right). \qquad \text{(A.13)}$$

If $\sigma < \infty$, then by Lemma 6(a),

$$\lim_{n\to\infty} \Pr_{u_n} \left(u_n \widehat{\omega}_{u_n}^2 > b_{u_n} \ \& \ u_n^{1/2} |\widehat{QLR}_{u_n}| / \widehat{\omega}_{u_n} > z_{\alpha/2} \right) \leq \lim_{n\to\infty} \Pr_{u_n} \left(u_n \widehat{\omega}_{u_n}^2 > b_{u_n} \right)$$
$$= 0 < \alpha. \qquad \text{(A.14)}$$

If $\sigma = \infty$, then by Lemma 6(b)

$$\begin{aligned}
&\lim_{n\to\infty} \Pr_{u_n}\left(u_n\widehat{\omega}^2_{u_n} > b_{u_n} \ \&\ u_n^{1/2}|\widehat{QLR}_{u_n}|/\widehat{\omega}_{u_n} > z_{\alpha/2}\right) \\
\leq\ &\lim_{n\to\infty} \Pr_{u_n}\left(u_n^{1/2}|\widehat{QLR}_{u_n}|/\widehat{\omega}_{u_n} > z_{\alpha/2}\right) = \alpha. \qquad \text{(A.15)}
\end{aligned}$$

Therefore, by (A.13)-(A.15), $AsySZ^{ol}(\alpha) \leq \alpha$.

(b) Let $\mu \in \mathcal{H}_0^{ol}$ satisfy $P_\mu^* \neq Q_\mu^*$. By Assumption 6, such a μ exists. Then, $\omega_\mu^2 > 0$. By Lemma 6(b), under μ, $\widehat{\omega}_n^2 \to_p \omega_\mu^2 > 0$. Also, by Lemma 6(b), under μ, $n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n \to_d N(0,1)$. Because $n^{-1}b_n \to 0$, we have

$$\lim_{n\to\infty} \Pr_\mu\left(n\widehat{\omega}_n^2 > b_n \ \&\ n^{1/2}|\widehat{QLR}_n|/\widehat{\omega}_n > z_{\alpha/2}\right) = \alpha. \qquad \text{(A.16)}$$

By definition, $AsySZ^{ol}(\alpha) \geq \lim_{n\to\infty} \Pr_\mu\left(n\widehat{\omega}_n^2 > b_n \ \&\ n^{1/2}|\widehat{QLR}_n|/\widehat{\omega}_n > z_{\alpha/2}\right)$. Thus, we have $AsySZ^{ol}(\alpha) \geq \alpha$. Combining this with part (a), we obtain the desired result. ■

Proof of Therem 3. Let $\hat{\theta}_n \in \hat{\Theta}_n$ and $\hat{\beta}_n \in \hat{B}_n$ be such that $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n) \geq \widehat{\omega}_n^2 - 2^{-n}$. Parts (a) and (b) are implied by

$$\widehat{QLR}_n \to_p E_\mu \Lambda_{\mu,i}^* > 0 \text{ and } \widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n) \to_p \omega_\mu^2 > 0. \qquad \text{(A.17)}$$

Next, we show (A.17). First we show $E_\mu\Lambda_{\mu,i}^* > 0$ and $\omega_\mu^2 > 0$. Under Assumptions 4 and FA(b)-(d), Lemma 2(b) applies and gives

$$E_\mu\Lambda_{\mu,i}^* = \exp(-d(\mathcal{P},\mu)) - \exp(-d(\mathcal{Q},\mu)) > 0, \qquad \text{(A.18)}$$

where the inequality holds by Assumption FA(a). By the Cauchy-Schwartz inequality,

$$\omega_\mu^2 = E_\mu(\Lambda_{\mu,i}^*)^2 \geq (E_\mu\Lambda_{\mu,i}^*)^2 > 0, \qquad \text{(A.19)}$$

where the strict inequality holds by (A.18).

Now, we show the convergence results in (A.17). Let $\{\theta_n^* \in \Theta_\mu^*\}_{n=1}^\infty$ and $\{\beta_n^* \in B_\mu^*\}_{n=1}^\infty$ be sequences that satisfy: $||\hat{\theta}_n - \theta_n^*|| \leq \rho_{lh}(\hat{\theta}_n, \Theta_\mu^*) + 2^{-n}$ and $||\hat{\beta}_n - \beta_n^*|| \leq \rho_{lh}(\hat{\beta}_n, B_\mu^*) + 2^{-n}$. Such sequences always exist. Because $\mathcal{M}_\mu(\phi)$, $\mathcal{N}_\mu(\phi)$ and $E_\mu(e^{\gamma' m_i(\theta)} - e^{\lambda' g_i(\beta)})^2$ are uniformly continuous in $\phi \in \Gamma_M^m \times \Theta$ and $\psi \in \Gamma_M^g \times B$ by Assumption 4, the convergence results in (A.17) are implied by

$$\text{(i)} \sup_{\phi\in\Gamma_M^m\times\Theta}|\widehat{\mathcal{M}}_n(\phi)-\mathcal{M}_\mu(\phi)|+\sup_{\psi\in\Gamma_M^g\times B}|\mathcal{N}_n(\phi)-\mathcal{N}_\mu(\phi)|\to_p 0,$$

$$\text{(ii)} \sup_{\phi\in\Gamma_M^m\times\Theta}\sup_{\psi\in\Gamma_M^g\times B}\left|n^{-1}\sum_{i=1}^n(e^{\gamma' m_i(\theta)}-e^{\lambda' g_i(\beta)})^2-E_\mu(e^{\gamma' m_i(\theta)}-e^{\lambda' g_i(\beta)})^2\right|\to_p 0,$$

$$\text{(iii)} ||\hat{\phi}_n(\hat{\theta}_n)-\phi_\mu^*(\theta_n^*)||+||\hat{\psi}_n(\hat{\theta}_n)-\psi_\mu^*(\theta_n^*)||\to_p 0. \tag{A.20}$$

Conditions (i) and (ii) in (A.20) are implied by Lemma 2.4 in Newey and McFadden (1986). The conditions in that lemma are guaranteed by Assumptions 4, FA(b) and (d).

Condition (iii) in (A.20) is implied by Lemma 3 and Lemma A3(a) with $\{\mu_n\in\mathcal{H}_0^{no}\}_{n=1}^\infty$ replaced by $\{\mu_n=\mu\}_{n=1}^\infty$. Under Assumptions 4 and FA, with some modifications, the proofs of Lemma 3 and Lemma A3(a) still go through. The modifications are as follows: (a) the first inequality in (B.5) holds for some $\delta_\varepsilon>0$ because $\mathcal{M}_\mu(\gamma_\mu^*(\theta),\theta)$ is continuous in θ, $\Theta_\mu^*=\arg\min_{\theta\in\Theta}\mathcal{M}_\mu(\gamma_\mu^*(\theta),\theta)$ and Θ is compact; (b) the second equality in (B.5) holds by (i) in (A.20), the uniform continuity of $\mathcal{M}_\mu(\gamma,\theta)$ in $\gamma\in\Gamma_M^m$, and Lemma A3(a); (c) in the first paragraph of the proof of Lemma A3, use Assumption FA(b) instead of condition (v) of (5.5); (d) δ in (C.13) is the constant in Assumption FA(c); and (e) the convergence in (C.14) holds by (i) in (A.20) and the uniform continuity of $\mathcal{M}_\mu(\gamma,\theta)$ in $\theta\in\Theta$. ■

Proof of Theorem 4. Let $\hat{\theta}_n\in\hat{\Theta}_n$ and $\hat{\beta}_n\in\hat{B}_n$ be such that $\widehat{\omega}_n^2(\hat{\theta}_n,\hat{\beta}_n)\geq\widehat{\omega}_n^2-2^{-n}$. Then, Theorem 4 is implied by

$$n^{1/2}\widehat{QLR}_n/\widehat{\omega}_n(\hat{\theta}_n,\hat{\beta}_n)\to_d Z-\sigma^{-1}e^{-h_2}h_1 \text{ where } Z\sim N(0,1). \tag{A.21}$$

The proof of (A.21) is the same as that of Theorem 1(a) except that $\{\mu_n\in\mathcal{H}_0^{no}\}_{n=1}^\infty$ is replaced with $\{\mu_n\}_{n=1}^\infty$ defined in Assumption LA1 and that the asymptotic distribution of $\omega_n^{-1}n^{-1/2}\sum_{i=1}^n\Lambda_{n,i}^*$ (the first summand in the r.h.s. of (A.4)) is now different. Under $\{\mu_n\}_{n=1}^\infty$, the term $\omega_n^{-1}n^{-1/2}\sum_{i=1}^n\Lambda_{n,i}^*$ has a nonzero mean. Below we show that

$$\omega_n^{-1}n^{-1/2}\sum_{i=1}^n\Lambda_{n,i}^*\to_d Z-\sigma^{-1}e^{-h_2}h_1. \tag{A.22}$$

The proof of Theorem 1(a) goes through except the part that shows the asymptotic distribution of $\omega_n^{-1}n^{-1/2}\sum_{i=1}^n\Lambda_{n,i}^*$. Thus, (A.21) holds.

Now, we show (A.22). First, we have

$$\begin{aligned}\omega_n^{-1}n^{1/2}E_n\Lambda_{n,i}^* &= \omega_n^{-1}n^{1/2}\big(\max_{\theta\in\Theta}\mathcal{M}_{\mu_n}(\gamma_n^*(\theta),\theta)-\max_{\beta\in B}\mathcal{N}_n(\lambda_n^*(\beta),\beta)\big)\\ &= \omega_n^{-1}n^{1/2}\big(\exp(-d(\mathcal{P},\mu_n))-\exp(-d(\mathcal{Q},\mu_n))\big)\\ &= -\omega_n^{-1}\exp(\tilde{d}_n)\cdot n^{1/2}(d(\mathcal{P},\mu_n)-d(\mathcal{Q},\mu_n))\\ &\to -\sigma^{-1}\exp(-h_2)h_1, \end{aligned} \tag{A.23}$$

where the second equality holds by Lemma 2(b)-(c), the third inequality holds by a mean-value expansion with $\tilde{d}_n$ lying on the line segment joining $d(\mathcal{P},\mu_n)$ and $d(\mathcal{Q},\mu_n)$, and the convergence holds by Assumption LA1(a)-(b).

Applying the Lyapounov CLT to the triangular array $\{\omega_n^{-1}n^{-1/2}\sum_{i=1}^n[\Lambda_{n,i}^*-E_n\Lambda_{n,i}^*]\}_{n=1}^\infty$, we have

$$\omega_n^{-1}n^{-1/2}\sum_{i=1}^n[\Lambda_{n,i}^*-E_n\Lambda_{n,i}^*]\to_d N(0,1). \tag{A.24}$$

The CLT applies because (i) $E_n[\Lambda_{n,i}^*-E_n\Lambda_{n,i}^*]=0$, (ii) $\omega_n^{-2}E_n[\Lambda_{n,i}^*-E_n\Lambda_{n,i}^*]^2=1-\omega_n^{-2}[E_n\Lambda_{n,i}^*]^2\to 1$, by Assumption LA1(b) and (A.23), and (iii) $\sup_{n\geq 1}E_n(\Lambda_{n,i}^*/\omega_n)^{2+\delta}<\infty$ by condition (vii) in (5.5) and Lemma 4. ■

Proof of Theorem 5. Let $\hat{\theta}_n\in\hat{\Theta}_n$ and $\hat{\beta}_n\in\hat{B}_n$ be such that $\widehat{\omega}_n^2(\hat{\theta}_n,\hat{\beta}_n)\geq\widehat{\omega}_n^2-2^{-n}$. Theorem 5 is implied by,

$$\omega_n^{-2}\widehat{\omega}_n^2(\hat{\theta}_n,\hat{\beta}_n)\to_p 1,\text{ and }\omega_n^{-1}n^{1/2}\widehat{QLR}_n\to_d Z-e^{-h_2}h_1^{ol},\text{ where }Z\sim N(0,1). \tag{A.25}$$

The proof of $\omega_n^{-2}\widehat{\omega}_n^2(\hat{\theta}_n,\hat{\beta}_n)\to_p 1$ is the same as that of $\omega_n^{-2}\widehat{\omega}_n^2(\hat{\theta}_n,\hat{\beta}_n)\to_p 1$ in Lemma 6(b). The proof of $\omega_n^{-1}n^{1/2}\widehat{QLR}_n\to_d Z-e^{-h_2}h_1^{ol}$ is the same as that of $\omega_n^{-1}n^{1/2}\widehat{QLR}_n\to_d N(0,1)$ in Lemma 6(b) except with the same changes performed as in the proof of Theorem 4. A detailed proof is omitted for brevity. ■

B Proof of the Main Lemmas

Proof of Lemma 1. We only need to show parts(a)-(b) because part (c) is analogous to parts (a)-(b).

(a) By Assumption 1(a)-(b), $p_{\theta,\mu}^*$ is a well defined density function. The proof here is similar to that of the second part of Theorem 3.1 in Csiszár (1975). Let P be a distribution in $\mathcal{P}_\theta$ such that $P << \mu$ and let p_μ denote the density of P with respect to μ, then

$$\begin{aligned} d(P,\mu) - d(P, P^*_{\theta,\mu}) &= \int \log p_\mu dP - \int \log(p_\mu / p^*_{\theta,\mu}) dP \\ &= \int \log p^*_{\theta,\mu} dP \\ &= -\log E_\mu \exp(\gamma^*_\mu(\theta)' m(X_i,\theta)) + \gamma^*_\mu(\theta)' E_P m(X_i,\theta) \\ &\geq -\log E_\mu \exp(\gamma^*_\mu(\theta)' m(X_i,\theta)), \qquad \text{(B.1)} \end{aligned}$$

where the inequality holds because for $j \leq d_p$, $E_P m_j(X_i,\theta) = 0$, and for $j \geq d_p + 1$, $E_P m_j(X_i,\theta) \geq 0$ and $\gamma^*_{\mu,j}(\theta) \geq 0$. Equation (B.1) implies that

$$d(P,\mu) \geq -\log E_\mu \exp(\gamma^*_\mu(\theta)' m(X_i,\theta)). \qquad \text{(B.2)}$$

By definition,

$$\begin{aligned} d(P^*_{\theta,\mu},\mu) &= \int \log p^*_{\theta,\mu} dP^*_\theta \\ &= -\log E_\mu \exp(\gamma^*_\mu(\theta)' m(X_i,\theta)) + \gamma^*_\mu(\theta)' E_{P^*_\theta} m(X_i,\theta) \\ &= -\log E_\mu \exp(\gamma^*_\mu(\theta)' m(X_i,\theta)), \qquad \text{(B.3)} \end{aligned}$$

where the last equality holds by the Kuhn-Tucker conditions from the minimization problem: $\min_{\gamma \in R^{d_p} \times R_+^{d_m - d_p}} \mathcal{M}_\mu(\gamma,\theta)$. The Kuhn-Tucker conditions are

$$\begin{aligned} 0 &= \partial \mathcal{M}_\mu(\gamma^*_\mu(\theta),\theta)/\partial \gamma_j \equiv E_{P^*_\theta} m_j(X_i,\theta) && \text{for } j \leq d_p \\ 0 &= \gamma^*_{\mu,j}(\theta)\big(\partial \mathcal{M}_\mu(\gamma^*_\mu(\theta),\theta)/\partial \gamma_j\big) \equiv \gamma^*_{\mu,j}(\theta) E_{P^*_\theta} m_j(X_i,\theta) && \text{for } j \geq d_p + 1. \qquad \text{(B.4)} \end{aligned}$$

By (B.2) and (B.3), we have $d(P^*_{\theta,\mu},\mu) = \min_{P \in \mathcal{P}_\theta} d(P,\mu)$, i.e. $P^*_{\theta,\mu}$ is the I-projection of μ on P.

(b) Part (b) is implied by (B.3). ■

Proof of Lemma 2. (a) By Assumptions 1(a)-(b) and 2(b), $\gamma^*_\mu(\theta)$ is continuous in θ. Thus, $\mathcal{M}_\mu(\gamma^*_\mu(\theta),\theta)$ is continuous in θ. The continuity of $\mathcal{M}_\mu(\gamma^*_\mu(\theta),\theta)$ combined with Assumption 2(a) implies part (a)

(b) By part (a), $\sup_{\theta \in \Theta} \mathcal{M}_\mu(\gamma^*_\mu(\theta),\theta) = \max_{\theta \in \Theta} \mathcal{M}_\mu(\gamma^*_\mu(\theta),\theta)$. By Lemma 1(b) and the definition of $\gamma^*_\mu(\cdot)$, we have part (b).

(c) The arguments for part (c) are analogous to those for parts (a)-(b). ■

Proof of Lemma 3. It suffices to show $\rho_{lh}(\hat{\Theta}_n, \Theta_n^*) \to_p 0$ because $\rho_{lh}(\hat{B}_n, B_n^*) \to_p 0$ can be obtained by analogous arguments.

For an arbitrary $\varepsilon > 0$ and an arbitrary sequence $\{\hat{\theta}_n \in \hat{\Theta}_n\}_{n=1}^{\infty}$,

$$\begin{aligned}
&\Pr_n\left(\rho_{lh}(\hat{\theta}_n, \Theta_n^*) > \varepsilon\right) \\
\leq\ &\Pr_n\left(\mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*)) - \mathcal{M}_{\mu_n}(\phi_n^*(\hat{\theta}_n)) > \delta_\varepsilon\right) \\
=\ &\Pr_n\left([\mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*)) - \widehat{\mathcal{M}}_n(\hat{\phi}_n(\theta_n^*))] + [\widehat{\mathcal{M}}_n(\hat{\phi}_n(\theta_n^*)) - \widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n))]\right. \\
&\left.+[\widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\hat{\theta}_n))] > \varepsilon\right) \\
=\ &\Pr_n\left(o_p(1) + \widehat{\mathcal{M}}_n(\hat{\phi}_n(\theta_n^*)) - \widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) + o_p(1) > \varepsilon\right) \\
\leq\ &\Pr_n(o_p(1) + o_p(1) > \varepsilon) \to 0, \qquad \text{(B.5)}
\end{aligned}$$

where $\theta_n^* \in \Theta_n^*$, the first inequality holds by condition (iv) in (5.5), the second equality holds by Lemmas A2(e) and A3(a), and the second inequality holds because $\hat{\theta}_n$ maximizes $\widehat{\mathcal{M}}_n(\hat{\gamma}_n(\theta), \theta)$. Therefore, $\rho_{lh}(\hat{\theta}_n, \Theta_n^*) \to_p 0$. Because the choice of $\hat{\theta}_n \in \hat{\Theta}_n$ is arbitrary, (B.5) implies $\rho_{lh}(\hat{\Theta}_n, \Theta_n^*) \to_p 0$. ■

Proof of Lemma 4. For any $\mu \in \mathcal{H}_0^{no}$ and $\theta^* \in \Theta_\mu^*$, we have

$$\begin{aligned}
\omega_\mu^2 &= \mathcal{M}_\mu^2(\gamma_\mu^*(\theta^*), \theta^*) E_\mu[dP_\mu^*/d\mu - dQ_\mu^*/d\mu]^2 \\
&\geq E_\mu[dP_\mu^*/d\mu - dQ_\mu^*/d\mu]^2 \cdot \exp(-2M_1) \\
&\geq \left[E_\mu|dP_\mu^*/d\mu - dQ_\mu^*/d\mu|\right]^2 \cdot \exp(-2M_1) \\
&= \left[\int |dP_\mu^*/d\nu_{P_\mu^*,Q_\mu^*} - dQ_\mu^*/d\nu_{P_\mu^*,Q_\mu^*}| d\nu_{P_\mu^*,Q_\mu^*}\right]^2 \cdot \exp(-2M_1) \\
&\geq \left[\inf_{P\in\mathcal{P},\ Q\in\mathcal{Q}} \int |dP/d\nu_{P,Q} - dQ/d\nu_{P,Q}| d\nu_{P,Q}^2\right]^2 \cdot \exp(-2M_1) > 0, \qquad \text{(B.6)}
\end{aligned}$$

where the first equality holds by Lemma 1(a) and condition (ii) in (5.5), the first inequality holds because $\mathcal{M}_\mu(\gamma_\mu^*(\theta^*), \theta^*) = \exp(-d(\mathcal{P}, \mu)) \geq \exp(-M_1)$ by Lemma 2(b) and condition (ii) in (5.5), the second inequality holds by the convexity of $f(x) = x^2$, the second equality holds because P_μ^* and Q_μ^* are absolutely continuous w.r.t. $\nu_{P_\mu^*,Q_\mu^*}$, the third inequality holds because $P_\mu^* \in \mathcal{P}$ and $Q_\mu^* \in \mathcal{Q}$, and the last inequality holds by Definition NO. Lemma 4 is implied by (B.6). ■

Proof of Lemma 5. It suffices to show that $\rho_{lh}(\hat{\Theta}_n, \Theta_n^*) = O_p(n^{-1/2})$ because the remainder is analogous. First, by Lemma 3, we have $\rho_{lh}(\hat{\Theta}_n, \Theta_n^*) \to_p 0$.

Now, we show the rate of convergence: $\rho_{lh}(\hat{\Theta}_n, \Theta_n^*) = O_p(n^{-1/2})$. Take an arbitrary sequence $\{\hat{\theta}_n \in \hat{\Theta}_n\}_{n=1}^{\infty}$. Let $\{\theta_n^* \in \Theta_n^*\}_{n=1}^{\infty}$ be a sequence such that $||\theta_n^* - \hat{\theta}_n||^2 \leq \rho_{lh}^2(\hat{\theta}_n, \Theta_n^*) + 2^{-n}$. The proof is based on the quadratic approximation of $\widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) - \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))$ and that of $\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))$. The basic idea is from Andrews (1999), but the procedure is more involved here because (a) we deal with a saddle-point estimation problem instead of a extremum estimation problem, (b) after profiling out the first step minimization parameter γ, the criterion functions $\widehat{\mathcal{M}}_n(\hat{\phi}_n(\theta))$ and $\mathcal{M}_{\mu_n}(\phi_n^*(\theta))$ are non-differentiable in θ, and (c) there is no straightforward way of writing down the left/right derivatives w.r.t. θ. We construct quadratic bounds for the centralized population and sample criterion functions. Specifically, we show below that

$$\begin{aligned}
&\text{(i)} \;\; \big[\widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) - \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))\big] - \big[\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))\big] \\
&\qquad = O_p(n^{-1/2}) \cdot ||\hat{\theta}_n - \theta_n^*|| + o_p(1) \cdot ||\hat{\theta}_n - \theta_n^*||^2, \\
&\text{(ii)} \;\; \widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) - \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*)) \geq O_p(n^{-1}) \text{ , and} \qquad \text{(B.7)} \\
&\text{(iii)} \;\; \mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*)) \leq O_p(n^{-1}) - C \cdot \big((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta\big) \text{ ,}
\end{aligned}$$

where C and δ are the positive constants in condition (iii) of (6.3). Conditions (i)-(iii) in (B.7) imply that

$$\begin{aligned}
O_p(n^{-1}) &\leq O_p(n^{-1/2}) \cdot ||\hat{\theta}_n - \theta_n^*|| + o_p(1) \cdot ||\hat{\theta}_n - \theta_n^*||^2 - C \cdot \big((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta\big) \\
&= -C||\hat{\theta}_n - \theta_n^*||^2 + O_p(n^{-1/2}) \cdot ||\hat{\theta}_n - \theta_n^*|| + o_p(1) \cdot ||\hat{\theta}_n - \theta_n^*||^2 + o_p(1), \qquad \text{(B.8)}
\end{aligned}$$

which implies that $||\hat{\theta}_n - \theta_n^*|| = O_p(n^{-1/2})$. Therefore, the desired result, $\rho_{lh}(\hat{\Theta}_n, \Theta_n^*) = O_p(n^{-1/2})$, holds since $\hat{\theta}_n$ is arbitrarily chosen from $\hat{\Theta}_n$.

Now, we show condition (i) in (B.7). We have

$$\begin{aligned}
&\big[\widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) - \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))\big] - \big[\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))\big] \\
&= \big[\partial\widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))/\partial\phi' - \partial\mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))/\partial\phi'\big]\big[\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)\big] + \\
&\quad 2^{-1}\big[\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)\big]'\big[\partial^2\widehat{\mathcal{M}}_n(\tilde{\phi}_n)/\partial\phi\partial\phi' - \partial^2\mathcal{M}_{\mu_n}(\bar{\phi}_n)/\partial\phi\partial\phi'\big]\big[\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)\big] \\
&= O_p(n^{-1/2}) \cdot ||\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)|| + o_p(1) \cdot ||\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*||^2 \\
&= O_p(n^{-1/2}) \cdot ||\hat{\theta}_n - \theta_n^*|| + o_p(1) \cdot ||\hat{\theta}_n - \theta_n^*||^2, \qquad \text{(B.9)}
\end{aligned}$$

where both $\tilde{\phi}_n$ and $\bar{\phi}_n$ lie on the line segment joining $\hat{\phi}_n(\hat{\theta}_n)$ and $\phi_n^*(\theta_n^*)$ and they are not necessarily the same, the first equality holds by second-order Taylor expansions of $\widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n))$ and $\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n))$ around $\phi_n^*(\theta_n^*)$, the second equality holds by Lemmas A2(d) and A4(a) and the third equality holds by Lemma A3(c). Equation (B.9) implies condition (i) in (B.7).

Condition (ii) in (B.7) is implied by

$$\begin{aligned}
&\widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) - \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*)) \\
\geq\ & \widehat{\mathcal{M}}_n(\hat{\phi}_n(\theta_n^*)) - \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*)) \\
=\ & [\partial\widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))/\partial\gamma']\big[\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*)\big] + \\
& 2^{-1}\big[\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*)\big]'[\partial^2\widehat{\mathcal{M}}_n(\tilde{\phi}_n)\partial\gamma\partial\gamma']\big[\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*)\big] \\
=\ & [\partial\widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))/\partial\gamma']\big[\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*)\big] + O_p(n^{-1}) \\
\geq\ & \big[\partial\widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))/\partial\gamma' - \partial\mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))/\partial\gamma'\big]\big[\hat{\gamma}_n(\theta_n^*) - \gamma_n^*(\theta_n^*)\big] + O_p(n^{-1}) \\
=\ & O_p(n^{-1}),
\end{aligned} \tag{B.10}$$

where $\tilde{\phi}_n$ lies on the line segment joining $\hat{\phi}_n(\theta_n^*)$ and $\phi_n^*(\theta_n^*)$, the first inequality holds because $\hat{\theta}_n$ is a maximizer of $\widehat{\mathcal{M}}_n(\hat{\phi}_n(\theta))$, the first equality holds by a Taylor expansion, the second equality holds by Lemmas A2(e) and A3(b), the second inequality holds by the Kuhn-Tucker conditions of the minimization problem: $\min_{\gamma \in R^{d_p} \times R_+^{d_m - d_p}} \mathcal{M}_{\mu_n}(\gamma, \theta_n^*)$ and the last equality holds by Lemmas A2(d) and A3(b).

Condition (iii) in (B.7) is implied by

$$\begin{aligned}
& \mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*)) \\
=\ & \big[\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\hat{\theta}_n))\big] + \big[\mathcal{M}_{\mu_n}(\phi_n^*(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\theta_n^*))\big] \\
\leq\ & \big[\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n)) - \mathcal{M}_{\mu_n}(\phi_n^*(\hat{\theta}_n))\big] - C \cdot \big((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta\big) \\
=\ & -C \cdot \big((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta\big) + [\partial\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n))/\partial\gamma'][\hat{\gamma}_n(\hat{\theta}_n) - \gamma_n(\hat{\theta}_n)] - \\
& 2^{-1}\big[\hat{\gamma}_n(\hat{\theta}_n) - \gamma_n^*(\hat{\theta}_n)\big]'[\partial^2\mathcal{M}_{\mu_n}(\tilde{\phi}_n)\partial\gamma\partial\gamma']\big[\hat{\gamma}_n(\hat{\theta}_n) - \gamma_n^*(\hat{\theta}_n)\big] \\
=\ & O_p(n^{-1}) - C \cdot \big((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta\big) + [\partial\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n))/\partial\gamma']\big[\hat{\gamma}_n(\hat{\theta}_n) - \gamma_n^*(\hat{\theta}_n)\big] \\
\leq\ & O_p(n^{-1}) - C \cdot \big((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta\big) \\
& + \big[\partial\mathcal{M}_{\mu_n}(\hat{\phi}_n(\hat{\theta}_n))/\partial\gamma' - \partial\widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n))/\partial\gamma'\big]\big[\hat{\gamma}_n(\hat{\theta}_n) - \gamma_n^*(\hat{\theta}_n)\big] \\
=\ & O_p(n^{-1}) - C \cdot \big((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta\big),
\end{aligned} \tag{B.11}$$

where $\tilde{\phi}_n$ lies on the line segment joining $\phi_n^*(\hat{\theta}_n)$ and $\hat{\phi}_n(\hat{\theta}_n)$, the first inequality holds by condition (iii) in (6.3) and $||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n} \leq \rho_{lh}^2(\hat{\theta}_n, \Theta_n^*)$ by design, the second equality holds by a Taylor expansion of $\mathcal{M}_{\mu_n}(\phi_n^*(\hat{\theta}_n))$ around $\hat{\phi}_n(\hat{\theta}_n)$, the third equality holds by Lemmas A2(e) and A3(b), the second inequality holds by the Kuhn-Tucker conditions of the minimization problem: $\min_{\gamma \in R^{d_p} \times R_+^{d_m - d_p}} \widehat{\mathcal{M}}_n(\gamma, \hat{\theta}_n)$, and the last equality holds by Lemmas A2(d) and A3(b). ■

Proof of Lemma 6. Lemma 6 is stated in terms of subsequences $\{u_n\}_{n=1}^\infty$. For notational simplicity, we prove it for the sequence $\{n\}$. All of the arguments go through with $\{u_n\}$ in place of $\{n\}$.

Let $\hat{\theta}_n \in \hat{B}_n$ and $\hat{\beta}_n \in \hat{B}_n$ be such that $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n) \geq \widehat{\omega}_n^2 - 2^{-n}$.

(a) Let $\{\theta_n^* \in \Theta_n^*\}_{n=1}^\infty$ be a sequence such that $||\theta_n^* - \hat{\theta}_n||^2 \leq \rho_{lh}^2(\hat{\theta}_n, \Theta_n^*) + 2^{-n}$. We have

$$\begin{aligned}
& n\hat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n) \\
\leq\ & 3\sum_{i=1}^n (\Lambda_{n,i}^*)^2 \\
& +3n(\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)) \left(n^{-1} \sum_{i=1}^n \frac{\partial e^{\tilde{\gamma}_n' m_i(\tilde{\theta}_n)}}{\partial \phi} \frac{\partial e^{\tilde{\gamma}_n' m_i(\tilde{\theta}_n)}}{\partial \phi'} \right) (\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)) \\
& +3n(\hat{\psi}_n(\hat{\beta}_n) - \psi_n^*(\beta_n^*)) \left(n^{-1} \sum_{i=1}^n \frac{\partial e^{\tilde{\lambda}_n' g_i(\tilde{\beta}_n)}}{\partial \psi} \frac{\partial e^{\tilde{\lambda}_n' g_i(\tilde{\beta}_n)}}{\partial \psi'} \right) (\hat{\psi}_n(\hat{\beta}_n) - \psi_n^*(\beta_n^*)) \\
\equiv\ & W_{n,1} + W_{n,2} + W_{n,3},
\end{aligned} \tag{B.12}$$

where $\tilde{\phi}_n$ and $\tilde{\psi}_n$ lie on the lie segment joining $\hat{\phi}_n(\hat{\theta}_n)$ and $\phi_n^*(\theta_n^*)$ and the one joining $\hat{\psi}_n(\hat{\beta}_n)$ and $\psi_n^*(\beta_n^*)$, respectively, the inequality holds by a mean-value expansion and the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$. In (B.12), $W_{n,1} = O_p(1)$ because $E_n W_{n,1} = n\omega_n^2 \to \sigma^2 < \infty$, $W_{n,2} = O_p(1)$ by (A.10), Lemmas 5, A3(b) and A3(c), and $W_{n,3} = O_p(1)$ for analogous reasons.

Therefore, $n\hat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n) = O_p(1)$ when $\sigma < \infty$.

(b) The proof here is of the same structure as, but slightly different from, the proof of Theorem 1(a). The difference is caused by (i) ω_n^2 is not bounded away from zero under the conditions of this lemma while it is under the conditions of Theorem 1(a), and (ii) the set estimators are $n^{-1/2}$-consistent under the conditions of this lemma while they are not under the conditions of Theorem 1(a).

First, we show $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n)/\omega_n^2 \to_p 1$. Decompose $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n)/\omega_n^2$ in the same way as in (A.8). We show below that $W_{n,1} \to_p 1$, $W_{n,2} = o_p(1)$ and $W_{n,3} = o_p(1)$ when $\sigma = \infty$. These results together imply $\widehat{\omega}_n^2(\hat{\theta}_n, \hat{\beta}_n)/\omega_n^2 \to_p 1$.

Applying the LLN to $W_{n,1}$, we have $W_{n,1} \equiv \omega_n^{-2} n^{-1} \sum_{i=1}^n [\Lambda_{n,i}^*]^2 \to_p 1$. The LLN applies because (a) $E_n \omega_n^{-2} [\Lambda_{n,i}^*]^2 = 1$ and (b) $\sup_{n \geq 1} E_n [\omega_n^{-1} \Lambda_{n,i}^*]^{2+\delta} < \infty$ by condition (v) in (6.3).

The term $W_{n,3}$ is $o_p(1)$ because

$$\begin{aligned} 0 &\leq W_{n,3} \\ &\leq 2\omega_n^{-2}(\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)) \left(n^{-1} \sum_{i=1}^n \frac{\partial e^{\tilde{\gamma}_n' m_i(\tilde{\theta}_n)}}{\partial \phi} \frac{\partial e^{\tilde{\gamma}_n' m_i(\tilde{\theta}_n)}}{\partial \phi'} \right) (\hat{\phi}_n(\hat{\theta}_n) - \phi_n^*(\theta_n^*)) \\ &\quad + 2\omega_n^{-2}(\hat{\psi}_n(\hat{\beta}_n) - \psi_n^*(\beta_n^*)) \left(n^{-1} \sum_{i=1}^n \frac{\partial e^{\tilde{\lambda}_n' g_i(\tilde{\beta}_n)}}{\partial \psi} \frac{\partial e^{\tilde{\lambda}_n' g_i(\tilde{\beta}_n)}}{\partial \psi'} \right) (\hat{\psi}_n(\hat{\beta}_n) - \psi_n^*(\beta_n^*)) \\ &= o_p(1), \end{aligned} \tag{B.13}$$

where the inequality holds by the inequality, $(a+b)^2 \leq 2a^2 + 2b^2$ and the equality holds by $n\omega_n^2 \to \infty$, (A.10), Lemmas 5, A3(b) and A3(c).

The term $W_{n,2}$ is $o_p(1)$ because $0 \leq W_{n,2} \leq 2[W_{n,1} \cdot W_{n,3}]^{1/2}$ by the Cauchy-Schwartz inequality.

Next, we show $n^{1/2}\widehat{QLR}_n/\omega_n \to_d N(0,1)$. Let $A_{n,1}$ and $A_{n,2}$ be the same as in (A.4). Then, by (A.4), the desired result is implied by (i) $\omega_n^{-1} n^{-1/2} \sum_{i=1}^n \Lambda_{n,i}^* \to_d N(0,1)$, (ii) $A_{n,1} = o_p(1)$ and (iii) $A_{n,2} = o_p(1)$.

By the Lyapounov CLT, (i) holds. The CLT applies because (a) $E_n \Lambda_{n,i}^* = 0$ by condition (ii) in (6.3) and Lemma 2(b), (b) $\omega_n^{-2} E_n (\Lambda_{n,i}^*)^2 = 1$, and (c) $E_n (\omega_n^{-1} \Lambda_{n,i}^*)^{2+\delta} < \infty$ by condition (v) in (6.3).

Now we show (ii) $A_{n,1} = o_p(1)$. Because $n^{-1/2}\omega_n A_{n,1} = \widehat{\mathcal{M}}_n(\hat{\phi}_n(\hat{\theta}_n)) - \widehat{\mathcal{M}}_n(\phi_n^*(\theta_n^*))$, (ii) is implied by $n\omega_n^2 \to \infty$ and

$$\begin{aligned} O_p(n^{-1}) &\leq n^{-1/2}\omega_n A_{n,1} \\ &\leq O_p(n^{-1/2}) \cdot ||\hat{\theta}_n - \theta_n^*|| + o_p(1) \cdot ||\hat{\theta}_n - \theta_n^*||^2 + \\ &\qquad O_p(n^{-1}) - C \cdot ((||\hat{\theta}_n - \theta_n^*||^2 - 2^{-n}) \wedge \delta) \\ &= O_p(n^{-1}), \end{aligned} \tag{B.14}$$

where the first inequality holds by condition (ii) in (B.7) in the proof of Lemma 5, the second inequality holds by conditions (i) and (iii) in (B.7), and the equality holds by Lemma 5. ■

C Proof of the Auxiliary Lemmas

Proof of Lemma A1. Consider an $\varepsilon > 0$. It suffices to show there exists C_ε large enough such that

$$\limsup_{n\to\infty} \Pr_n \big(\sup_{\phi\in\Phi} ||\nu_n(\phi)|| > C_\varepsilon \big) < \varepsilon. \tag{C.1}$$

Because Φ is totally bounded w.r.t. ρ, for all $\eta > 0$, there exists a finite subset of Φ, $\{\phi_1, ..., \phi_{J_\eta}\}$ such that $\sup_{\phi\in\Phi} \min_{j\le J_\eta} \rho(\phi, \phi_j) < \eta$. Choose η such that

$$\limsup_{n\to\infty} \Pr_n \big(\sup_{\phi,\phi'\in\Phi:\rho(\phi,\phi')\le\eta} ||\nu_n(\phi) - \nu_n(\phi')|| > 1 \big) < \varepsilon/2. \tag{C.2}$$

Such an η exists because $\nu_n(\phi)$ is stochastically equicontinuous w.r.t. ρ.

Because $||\nu_n(\phi)|| = O_p(1)$, for every $\phi \in \Phi$, $\max_{j\le J_\eta} ||\nu_n(\phi_j)|| = O_p(1)$. Then, we can to choose C_ε large enough such that

$$\limsup_{n\to\infty} \Pr_n \big(\max_{j\le J_\eta} ||\nu_n(\phi_j)|| > C_\varepsilon - 1 \big) < \varepsilon/2. \tag{C.3}$$

Therefore, we have

$$\begin{aligned}
&\limsup_{n\to\infty} \Pr_n \big(\sup_{\phi\in\Phi} ||\nu_n(\phi)|| > C_\varepsilon \big) \\
\le\ &\limsup_{n\to\infty} \Pr_n \big(\sup_{\phi\in\Phi} \min_{j\le J_\eta} ||\nu_n(\phi) - \nu_n(\phi_j)|| + \max_{j\le J_\eta} ||\nu_n(\phi_j)|| > C_\varepsilon \big) \\
\le\ &\limsup_{n\to\infty} \Pr_n \big(\sup_{\phi,\phi'\in\Phi:\rho(\phi,\phi')\le\eta} ||\nu_n(\phi) - \nu_n(\phi\prime)|| + \max_{j\le J_\eta} ||\nu_n(\phi_j)|| > C_\varepsilon \big) \\
\le\ &\limsup_{n\to\infty} \Pr_n \big(\sup_{\phi,\phi'\in\Phi:\rho(\phi,\phi')\le\eta} ||\nu_n(\phi) - \nu_n(\phi\prime)|| > 1 \big) + \Pr_{\mu_n} \big(\max_{j\le J_\eta} ||\nu_n(\phi_j)|| > C_\varepsilon - 1 \big) \\
\le\ &\varepsilon,
\end{aligned} \tag{C.4}$$

where the first inequality holds by the triangle inequality, the second inequality holds by the definition of $\{\phi_1, ..., \phi_{J_\eta}\}$, the third inequality holds by $P(A\cup B) \le P(A) + P(B)$, and the last inequality holds by (C.2) and (C.3).

Equation (C.1) is implied by (C.4). ■

Proof of Lemma A2. It suffices to show parts (a)-(e) because part (f) can be obtained by analogous arguments.

(a) The proof for this part is an application of Theorem 1 in Andrews (1994). Let $\mathcal{F}$ denote the class of functions $\{e^{\gamma' m(\cdot,\theta)} : \phi \in \Gamma_M^m \times \Theta\}$. We verify the three assumptions of that theorem: (i) $\mathcal{F}$ satisfies Pollard's entropy condition with some envelope $\bar{F}$, (ii) $\limsup_{n\to\infty} E_n \bar{F}^{2+\delta}(X_i) < \infty$ for some $\delta > 0$, and (iii) $\{X_i : i \le n, n \ge 1\}$ is an m-dependent triangular array of random variables. Assumption (iii) holds trivially by condition (i) in (5.5). Assumption (i) holds because the class $\mathcal{F}$ is a type II class (i.e., a class of Lipschitz functions indexed by finite-dimensional parameters, see Andrews (1994)). It is a type II class because $\Gamma_M^m \times \Theta$ is a bounded subset of the Euclidean space and $e^{\gamma' m(\cdot,\theta)}$ is Lipschitz in ϕ:

$$|e^{\gamma_1' m(\cdot,\theta_1)} - e^{\gamma_2' m(\cdot,\theta_2)}| \le B(\cdot)||\phi_1 - \phi_2||, \tag{C.5}$$

where $B(\cdot) = \sup_{\phi\in\Gamma_M^m\times\Theta} ||\partial e^{\gamma' m(\cdot,\theta)}/\partial\phi||$. The inequality holds by a mean value expansion of $e^{\gamma_1' m(\cdot,\theta_1)}$ around ϕ_2. Then, by Theorem 2 in Andrews (1994), $\mathcal{F}$ satisfies Pollard's entropy condition with envelope $1 \vee \sup_{\phi\in\Phi} e^{\gamma' m(\cdot,\theta)} \vee B(\cdot)$.

Assumption (ii) above holds because, for some $\delta_1 > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} E_n[1 \vee \sup_{\phi\in\Gamma_M^m\times\Theta} e^{\gamma' m_i(\theta)} \vee B(X_i)]^{2+\delta_1} \\
&\le \limsup_{n\to\infty} E_n[1 + \sup_{\phi\in\Gamma_M^m\times\Theta} e^{\gamma' m_i(\theta)} + B(X_i)]^{2+\delta_1} \\
&\le C \cdot \limsup_{n\to\infty}[1 + E_n \sup_{\phi\in\Gamma_M^m\times\Theta} e^{(2+\delta_1)\gamma' m_i(\theta)} + E_n \sup_{\phi\in\Gamma_M^m\times\Theta} ||\partial e^{\gamma' m(\cdot,\theta)}/\partial\phi||^{2+\delta_1}] \\
&< \infty,
\end{aligned} \tag{C.6}$$

where C is a constant, the second inequality holds by the convexity of the function $f(x) = x^{2+\delta_1}$, and the third inequality holds by conditions and (vii) of (5.5).

Therefore, Theorem 1 in Andrews (1994) applies and $\nu_n^0(\phi)$ is stochastically equicontinuous w.r.t. ρ_0.

(b) By Lemma A1 and Lemma A2(a), it suffices to show that $(\Gamma_M^m \times \Theta, \rho_0)$ is totally bounded and $\nu_n^0(\phi) = O_p(1)$ for every $\phi \in \Gamma_M^m \times \Theta$. We show these two conditions below.

The pseudo-metric space $(\Gamma_M^m \times \Theta, \rho_0)$ is totally bounded because $\Gamma_M^m \times \Theta$ endowed with the Euclidean metric $||\cdot||$ is compact (and thus totally bounded) and for all $\phi_1, \phi_2 \in \Gamma_M^m \times \Theta$,

$$\begin{aligned}\rho(\phi_1, \phi_2) &= \sup_{N \geq 1} [E_n(\partial e^{\tilde{\gamma}' m_i(\tilde{\theta})}/\partial \phi')(\phi_1 - \phi_2))^2]^{1/2} \\ &\leq ||\phi_1 - \phi_2|| \cdot \sup_{N \geq 1} [E_n||\partial e^{\tilde{\gamma}' m_i(\tilde{\theta})}/\partial \phi'||^2]^{1/2} \leq M^{1/2}||\phi_1 - \phi_2||, \qquad \text{(C.7)}\end{aligned}$$

where $(\tilde{\gamma}', \tilde{\theta}')'$ lies on the line segment joining ϕ_1 and ϕ_2, the equality holds by a mean-value expansion, the first inequality holds by the algebraic inequalities $|a'b| \leq ||a|| \cdot ||b||$ and the second inequality holds by condition (vii) in (5.5).

For every $\phi \in \Gamma_M^m \times \Theta$, $\nu_n^0(\phi) = O_p(1)$ because

$$\begin{aligned}E_n(\nu_n^0(\phi))^2 &= E_n\Big[n^{-1/2}\sum_{i=1}^n (e^{\gamma' m_i(\theta)} - E_n e^{\gamma' m_i(\theta)})\Big]^2 \\ &= E_n(e^{\gamma' m_i(\theta)} - E_n e^{\gamma' m_i(\theta)})^2 \\ &\leq E_n e^{2\gamma' m_i(\theta)} < \infty, \qquad \text{(C.8)}\end{aligned}$$

where the second equality holds by condition (i) in (5.5), and the second inequality holds by condition (vii) in (5.5).

(c) & (d) The proof of parts (c) and (d) is essentially the same as that of parts (a) and (b) and is omitted for brevity.

(e) The proofs for the two convergence results of part (e) are similar. For brevity, we only present the proof for the first convergence result:

$$||\partial^2 \widehat{\mathcal{M}}_n(\phi_{1,n})/\partial\gamma\partial\gamma' - \partial^2 \mathcal{M}_{\mu_n}(\phi_{2,n})/\partial\gamma\partial\gamma'|| \to_p 0. \qquad \text{(C.9)}$$

Equation (C.9) is implied by the following two results:

$$\begin{aligned}&\text{(i)} \quad \sup_{\phi \in \Gamma_M^m \times \Theta} ||\partial^2 \widehat{\mathcal{M}}_n(\phi)/\partial\gamma\partial\gamma' - \partial^2 \mathcal{M}_{\mu_n}(\phi)/\partial\gamma\partial\gamma'|| \to_p 0 \text{ and} \\ &\text{(ii)} \quad ||\partial^2 \mathcal{M}_{\mu_n}(\phi_{1,n})/\partial\gamma\partial\gamma' - \partial^2 \mathcal{M}_{\mu_n}(\phi_{2,n})/\partial\gamma\partial\gamma'|| \to_p 0. \qquad \text{(C.10)}\end{aligned}$$

Thus, it suffices to show results (i) and (ii).

Result (i) in (C.10) is shown using Theorem 4 in Andrews (1992). This theorem requires four conditions: BD (boundedness), P-WLLN (pointwise weak law of large number), DM (domination) and TSE (termwise stochastic equicontinuity). We verify these conditions one by one. Let $\Gamma_M^m \times \Theta$ be endowed with the usual Euclidean metric, $||\cdot||$. Then, $\left(\Gamma_M^m \times \Theta, ||\cdot||\right)$ is totally bounded because Γ_M^m and Θ are compact subsets of the Euclidean space. Thus, the BD condition holds. The P-WLLN condition holds by the LLN. The DM condition holds because, by condition (vii) in (5.5),

$$\limsup_{n\to\infty} E_n \sup_{\phi\in\Gamma_M^m\times\Theta} ||\partial^2 e^{\gamma' m_i(\theta)}/\partial\gamma\partial\gamma'||^{1+\delta} < \infty. \tag{C.11}$$

The TSE condition holds because, for every $\varepsilon > 0$ and every j, $j' \leq d_m$,

$$\begin{aligned}
&\lim_{\zeta\to 0}\limsup_{n\to\infty} \Pr_n\big(\sup_{\phi\in\Gamma_M^m\times\Theta}\ \sup_{\phi^*:||\phi-\phi^*||<\zeta} |\partial^2 e^{\gamma^{*\prime} m_i(\theta^*)}/\partial\gamma_j\partial\gamma_{j'} - \partial^2 e^{\gamma' m_i(\theta)}/\partial\gamma_j\partial\gamma_{j'}| > \varepsilon\big) \\
&\leq \lim_{\zeta\to 0}\limsup_{n\to\infty} \Pr_n\big(\sup_{\phi\in\Gamma_M^m\times\Theta} ||\partial^3 e^{\gamma' m_i(\theta)}/\partial\gamma_j\partial\gamma_{j'}\partial\phi|| \cdot \sup_{\phi^*:||\phi-\phi^*||<\zeta} ||\phi-\phi^*|| > \varepsilon\big) \\
&\leq \lim_{\zeta\to 0}\limsup_{n\to\infty} E_n \sup_{\phi\in\Gamma_M^m\times\Theta} ||\partial^3 e^{\gamma' m_i(\theta)}/\partial\gamma_j\partial\gamma_{j'}\partial\phi|| \cdot \zeta/\varepsilon \\
&= 0,
\end{aligned} \tag{C.12}$$

where the first inequality holds by a mean-value expansion and $||a'b|| \leq ||a|| \cdot ||b||$, the second inequality holds by the Markov inequality and the equality holds by condition (vii) in (5.5). Therefore, all four conditions of Theorem 4 in Andrews (1992) hold and (i) is shown.

Result (ii) in (C.10) is shown by a mean-value expansion similar to the one used to show the TSE condition above. We omit the details for brevity. ∎

Proof of Lemma A3. (a) Let $\hat{\gamma}_n^M(\theta) = \min_{\gamma\in\Gamma_M^m} \widehat{\mathcal{M}}_n(\gamma,\theta)$. It suffices to show that $||\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})|| \to_p 0$ because $\widehat{\mathcal{M}}_n(\gamma,\theta)$ is strictly convex in γ by condition (vi) in (5.5) and $||\gamma_n^*(\theta_{2,n})|| \leq M - \delta$ by condition (v) of (5.5).

Below we show

$$\mathcal{M}_{\mu_n}(\hat{\gamma}_n^M(\theta_{1,n}),\theta_{2,n}) - \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}),\theta_{2,n}) \geq 2^{-1}\delta||\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})||^2, \tag{C.13}$$

where δ is the constant in condition (vi) of (5.5). Then, $||\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})|| \to_p 0$ is implied by the following equation: for all $\varepsilon > 0$,

$$\begin{aligned}
&\Pr_n\left(||\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})|| \geq \varepsilon\right) \\
&\leq \Pr_n\left(\mathcal{M}_{\mu_n}(\hat{\gamma}_n^M(\theta_{1,n}),\theta_{2,n}) - \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}),\theta_{2,n}) \geq 2^{-1}\delta\varepsilon^2\right)
\end{aligned}$$

$$
\begin{aligned}
&= \Pr_n \Big(\mathcal{M}_{\mu_n}(\hat{\gamma}_n^M(\theta_{1,n}), \theta_{2,n}) - \widehat{\mathcal{M}}_n(\hat{\gamma}_n^M(\theta_{1,n}), \theta_{1,n}) \\
&\quad + \widehat{\mathcal{M}}_n(\hat{\gamma}_n^M(\theta_{1,n}), \theta_{1,n}) - \widehat{\mathcal{M}}_n(\gamma_{\mu_n}^*(\theta_{2,n}), \theta_{1,n}) \\
&\quad + \widehat{\mathcal{M}}_n(\gamma_n^*(\theta_{2,n}), \theta_{1,n}) - \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}), \theta_{2,n}) \geq 2^{-1}\delta\varepsilon^2 \Big) \\
&\leq \Pr_n \Big(\mathcal{M}_{\mu_n}(\hat{\gamma}_n^M(\theta_{1,n}), \theta_{2,n}) - \widehat{\mathcal{M}}_n(\hat{\gamma}_n^M(\theta_{1,n}), \theta_{1,n}) \\
&\quad + \widehat{\mathcal{M}}_n(\gamma_n^*(\theta_{2,n}), \theta_{1,n}) - \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}), \theta_{2,n}) \geq 2^{-1}\delta\varepsilon^2 \Big) \\
&\to 0, \qquad \text{(C.14)}
\end{aligned}
$$

where the first inequality holds by (C.13), the second inequality holds by the definition of $\hat{\gamma}_n^M(\theta_{1,n})$, and the convergence holds by Lemma A2(e).

Now, it is left to show (C.13). A Taylor expansion of $\mathcal{M}_{\mu_n}(\hat{\gamma}_n^M(\theta_{1,n}), \theta_{2,n})$ around $\gamma_n^*(\theta_{2,n})$ gives

$$
\begin{aligned}
&\mathcal{M}_{\mu_n}(\hat{\gamma}_n^M(\theta_{1,n}), \theta_{2,n}) - \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}), \theta_{2,n}) \\
&= \left(\partial \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}), \theta_{2,n})/\partial\gamma'\right) \left(\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right) \\
&\quad + 2^{-1} \left(\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right)' S_{\mu_n}^m(\tilde{\gamma}_n, \theta_{2,n}) \left(\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right) \\
&\geq 2^{-1} \left(\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right)' S_{\mu_n}^m(\tilde{\gamma}_n, \theta_{2,n}) \left(\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right) \\
&\geq 2^{-1}\delta ||\hat{\gamma}_n^M(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})||^2, \qquad \text{(C.15)}
\end{aligned}
$$

where $\tilde{\gamma}_n$ lies on the line segment joining $\hat{\gamma}_n^M(\theta_{1,n})$ and $\gamma_n^*(\theta_{2,n})$, the first inequality holds by the Kuhn-Tucker conditions and the second inequality holds by condition (vi) of (5.5). Thus, (C.13) holds.

(b) Let $\{\theta_n \in \Theta\}_{n=1}^\infty$ be a random sequence such that we have $||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)|| \geq \sup_{\theta\in\Theta} ||\hat{\gamma}_n(\theta) - \gamma_n^*(\theta)|| - 2^{-n}$. Then, by part (a), $||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)|| \to_p 0$. Part (b) holds if $||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)|| = O_p(n^{-1/2})$, which is immediately implied by the following equation:

$$
\begin{aligned}
0 &\geq \widehat{\mathcal{M}}_n(\hat{\gamma}_n(\theta_n), \theta_n) - \widehat{\mathcal{M}}_n(\gamma_n^*(\theta_n), \theta_n) \\
&= \left[\partial \widehat{\mathcal{M}}_n(\gamma_n^*(\theta_n), \theta_n)/\partial\gamma'\right] (\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)) \\
&\quad + 2^{-1} (\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n))' \left[\partial^2 \widehat{\mathcal{M}}_n(\tilde{\gamma}_n, \theta_n)/\partial\gamma\partial\gamma'\right] (\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)) \\
&\geq \left[\partial \widehat{\mathcal{M}}_n(\gamma_n^*(\theta_n), \theta_n)/\partial\gamma' - \partial \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_n), \theta_n)/\partial\gamma'\right] (\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)) \\
&\quad + 2^{-1} (\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n))' \left[\partial^2 \widehat{\mathcal{M}}_n(\tilde{\gamma}_n, \theta_n)/\partial\gamma\partial\gamma'\right] (\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n))
\end{aligned}
$$

$$
\begin{aligned}
&= O_p\big(n^{-1/2} \cdot ||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)||\big) + o_p\big(||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)||^2\big) \\
&\quad +2^{-1} \left(\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)\right)' S^m_{\mu_n}(\gamma_n^*(\theta_n), \theta_n) \left(\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)\right) \\
&\geq O_p\big(n^{-1/2} \cdot ||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)||\big) + o_p\left(||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)||^2\right) \\
&\quad +2^{-1}\delta||\hat{\gamma}_n(\theta_n) - \gamma_n^*(\theta_n)||^2, \qquad \text{(C.16)}
\end{aligned}
$$

where the first inequality holds because $\hat{\gamma}_n(\theta_n)$ minimizes $\widehat{\mathcal{M}}_n(\gamma, \theta_n)$, the first equality holds by a Taylor expansion with $\tilde{\gamma}_n$ lying on the line segment joining $\hat{\gamma}_n(\theta_n)$ and $\gamma_n^*(\theta_n)$, the second inequality holds by the Kuhn-Tucker conditions for the minimization problem $\min_{\gamma \in R^{d_p} \times R_+^{d_m - d_p}} \mathcal{M}_{\mu_n}(\gamma, \theta_n)$, the second equality holds by Lemmas A2(b) and (e), and the last inequality holds by condition (vi) in (5.5).

(c) By the Kuhn-Tucker conditions from the problem $\min_{\gamma \in R^{d_p} \times R_+^{d_m - d_p}} \mathcal{M}_{\mu_n}(\gamma, \theta)$, we have

$$
\left[\partial \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{1,n}), \theta_{1,n})/\partial \gamma' - \partial \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}), \theta_{2,n})/\partial \gamma'\right] \left(\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right) \leq 0. \qquad \text{(C.17)}
$$

A mean-value expansion of $\partial \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{1,n}), \theta_{1,n})/\partial \gamma'$ around $\left(\gamma_n^*(\theta_{2,n})', \theta_{2,n}'\right)'$ gives

$$
\begin{aligned}
&\partial \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{1,n}), \theta_{1,n})/\partial \gamma - \partial \mathcal{M}_{\mu_n}(\gamma_n^*(\theta_{2,n}), \theta_{2,n})/\partial \gamma \qquad \text{(C.18)} \\
&= S^m_{\mu_n}(\tilde{\gamma}_n, \tilde{\theta}_n) \left(\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right) + \left[\partial^2 \mathcal{M}_{\mu_n}(\tilde{\gamma}_n, \tilde{\theta}_n)/\partial \gamma \partial \theta'\right] (\theta_{1,n} - \theta_{2,n}),
\end{aligned}
$$

where $(\tilde{\gamma}_n', \tilde{\theta}_n')'$ lies on the line segment joining $\phi_n^*(\theta_{1,n})$ and $\phi_n^*(\theta_{2,n})$.

By (C.17) and (C.18), we have

$$
\begin{aligned}
0 &\geq \left(\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right)' S^m_{\mu_n}(\tilde{\gamma}_n, \tilde{\theta}_n) \left(\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right) \\
&\quad + \left(\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})\right)' [\partial \mathcal{M}_{\mu_n}(\tilde{\gamma}_n, \tilde{\theta}_n)/\partial \gamma \partial \theta'](\theta_{1,n} - \theta_{2,n}) \qquad \text{(C.19)} \\
&\geq \delta||\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})||^2 + O_p\left(||\gamma_n^*(\theta_{1,n}) - \gamma_n^*(\theta_{2,n})|| \cdot ||\theta_{1,n} - \theta_{2,n}||\right),
\end{aligned}
$$

where the second inequality holds by conditions (vi) and (vii) in (5.5). The desired result is implied by (C.19). ■

Proof of Lemma A4. The proof is similar to that for Lemma A2(e) and is omitted for brevity. ■

References

Andrews, D. W. K. (1988): "Laws of Large Numbers for Dependent Non-identically Distributed Random Variables," *Econometric Theory*, 4, 458–467.

——— (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 37, pp. 2247–2294. Elsevier.

——— (1999): "Estimation when a Parameter is on a Boundary," *Econometrica*, 67, 1341–1383.

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," unpublished manuscript, Department of Economics, Yale University.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and Plug-In Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709.

Andrews, D. W. K., and P. Jia (2008): "Inference for Parameters Defined by Moment Inequalities: a Recommended Moment Selection Procedure," Cowles Foundation Discussion Paper No. 1676, Cowles Foundation, Yale University.

Andrews, D. W. K., and X. Shi (2009): "Inference Based on Conditional Moment Inequality Models," unpublished manuscript, Department of Economics, Yale University.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, forthcoming.

Berry, S., and E. Tamer (2006): "Identification in Models of Oligopoly Entry," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress*, ed. by N. Blundel, and Persson, vol. 2, pp. 46–85. Cambridge University Press.

Bugni, F. A. (2008): "Bootstrap Inference in Partially Identified Models," unpublished manuscript, Department of Economics, Northwestern University.

Canay, I. A. (2008): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," unpublished manuscript, Department of Economics, Northwestern University.

Chen, X., H. Hong, and M. Shum (2007): "Nonparametric Likelihood Ratio Model Selection Tests between Parametric Likelihood and Moment Condition Models," *Journal of Econometrics*, 141, 109–140.

Cheng, X. (2008): "Robust Confidence Intervals in Nonlinear Regression under Weak Identification," unpublished manuscript, Department of Economics, Yale University.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75, 1243–1284.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in the Airline Industry," *Econometrica*, forthcoming.

Cox, D. R. (1961): "Tests of Separate Families of Hypotheses," Proceedings of the Fourth Berkeley Symposium in Mathematical Statistics and Probability, University of California Press: Berkeley.

Csiszár, I. (1975): "I-Divergence Geometry of Probability Distributions and Minimization Problems," *The Annals of Probability*, 3, 146–158.

Csiszár, I., and P. C. Shields (2004): *Information Theory and Statistics: A Tutorial*, Foundations and Trends in Communications and Information Theory. Now Publisers Inc.

Gourieroux, C., and A. Monfort (1994): "Testing Non-nested Hypotheses," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 44, pp. 2583–2637. Elsevier.

Kitamura, Y. (2000): "Comparing Misspecified Dynamic Econometric Models Using Nonparametric Likelihood," unpublished manuscript, Department of Economics, University of Pennsylvania.

——— (2007): "Nonparametric Likelihood: Efficiency and Robustness," *The Japanese Economic Review*, 58, 26–46.

Kitamura, Y., and M. Stutzer (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation," *Econometrica*, 65, 861–874.

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

Mizon, G. E., and J.-F. Richard (1986): "The Encompassing Principle and Its Application to Testing Non-nested Hypotheses," *Econometrica*, 54, 657–78.

Moon, H. R., and F. Schorfheide (2006): "Boosting Your Instruments: Estimation with Overidentifying Inequality Moment Conditions," IEPR Working Paper No. 06.56, Institute of Economic Policy Research (IEPR).

Newey, W. K., and D. McFadden (1986): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. McFadden, vol. 4 of *Handbook of Econometrics*, chap. 36, pp. 2111–2245. Elsevier.

Pakes, A. (2009): "Models for Moment Inequalities," unpublished manuscript, Harvard University and NBER.

Pakes, A., J. Porter, K. Ho, and J. Ishii (2007): "Moment Inequalities and Their Application," CeMMAP Working Paper No. CWP16/07, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Pesaran, M. H., and M. Weeks (1999): "Non-nested Hypothesis Testing: an Overview," Cambridge Working Paper in Economics No. 9918, Faculty of Economics, University of Cambridge.

Qin, J., and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *The Annals of Statistics*, 22, 300–325.

Ramalho, J. J. S., and R. J. Smith (2002): "Generalized empirical likelihood non-nested tests," *Journal of Econometrics*, 107, 99–125.

Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138, 2786–2807.

Shi, X. (2009): "The Asymptotic Size of the Classical Vuong Test for Overlapping Models," unpublished manuscript, Department of Economics, Yale University.

Vuong, Q. H. (1989): "Likelihood Ratio Tests for Model Selection and Non-nested Hypotheses," *Econometrica*, 57, 307–33.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.