Empirical Likelihood Estimation based on Inequality Moment Constraints

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
University of Pennsylvania

April 9, 2004

Abstract

This paper extends moment based estimation procedures to models in which some of the moment conditions take the form of weak inequalities, e.g., Kuhn-Tucker conditions, rather than equalities. We consider maximum empirical likelihood estimators (MELE), in which the parametric moment conditions appear as inequality constraints in the extremum estimation and develop a large sample distribution theory for the MELEs and the empirical likelihood ratios.

Preliminary and Incomplete

JEL CLASSIFICATION: C32

KEY WORDS: Empirical Likelihood, Inequality Moment Restrictions.

*Correspondence: H.R. Moon: Department of Economics, KAP 300, University Park Campus University of Southern California, Los Angeles, CA 90089. Email: moonr@usc.edu. F. Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104. Email: schorf@ssc.upenn.edu. We thank Shakeeb Khan, Yuichi Kitamura, Masao Ogaki, Nick Souleles, and seminar participants at the 2003 Econometrics Society Winter Meetings, Michigan State, Penn State, University of Pennsylvania, and Virginia for helpful comments.

1 Introduction

This paper extends empirical likelihood (EL) estimation techniques to models in which a subset of moment conditions take the form of weak inequalities rather than equalities, that is, eq_mom

$$\mathbb{E}[g_1(X_i,\theta)] = 0 \quad \text{and} \quad \mathbb{E}[g_2(X_i,\theta)] \geq 0 \tag{1}$$

if $\theta = \theta_0$. We are interested in estimating θ and testing the hypothesis that $\mathbb{E}[H'g_2(X_i,\theta)] = 0$, where the matrix H' selects a subset of elements from the vector-valued function $g_2(X_i,\theta)$. Throughout the paper we assume that θ_0 is identifiable based on the moment condition $\mathbb{E}[g_1(X_i,\theta_0)] = 0$.[1] Inequality moment conditions are quite common in economic models. For instance, they arise in environments in which agents face borrowing, regulatory, or incentive compatibility constraints.

Based on our identification assumption, the parameter θ could in principle be estimated with the first moment condition and the resulting estimate could be plugged into $g_2(X_i,\theta)$ to test the hypothesis that the second moment condition holds with equality. However, such a procedure ignores information if in fact $\mathbb{E}[H'g_2(X_i,\theta)] = 0$. Hence, it is preferable to use both moment conditions for estimation. If it is true that $\mathbb{E}[H'g_2(X_i,\theta_0)] = 0$ then the second moment condition will provide information in addition to the condition $\mathbb{E}[g_1(X_i,\theta_0)] = 0$. If, however, $\mathbb{E}[g_2(X_i,\theta_0)] > 0$ then the second moment condition is asymptotically uninformative. In this case the estimator proposed in this paper is asymptotically equivalent to the estimator obtained by using only $g_1(X_i,\theta)$. In addition to the estimator, this paper studies tests of the hypothesis $\mathbb{E}[g_2(X_i,\theta_0)] = 0$, a test for overidentifying moment restrictions, and a likelihood-ratio based coefficient test for θ_0.

To conduct estimation and inference we use the empirical likelihood framework. EL inference is attractive for two reasons. First, since moment conditions are imposed as parametric constraints on the empirical likelihood function, it is conceptually straightforward to extend the analysis from equality moment conditions to inequality moment conditions. Second, information-theoretic estimators such as EL have emerged as an attractive alternative to generalized method of moments (GMM) estimators. For instance, Kitamura (2001) showed that the empirical likelihood ratio test for moment restrictions is asymptotically optimal under the Generalized Neyman-Pearson criterion. Newey and Smith (2004) find that

[1] The estimation of models in which the moment conditions only enable the identification of (non-singleton) subsets of Θ has been explored, for instance, by Tamer (2002) and Chernozhukov, Hong, and Tamer (2002). It will not be pursued in this paper.

the asymptotic bias of EL estimators does not grow with the number of moment conditions and that bias-corrected EL estimators have higher-order efficiency properties. A detailed discussion of empirical likelihood methods in econometrics and statistics is provided in the monograph by Owen (2001).

One can introduce an additional parameter vector $\vartheta = \mathbb{E}[g_2(X_i, \theta)]$ and express the second moment condition as $\mathbb{E}[g_2(X_i, \theta_0) - \vartheta_0] = 0$, where $\vartheta_0 \geq 0$. Thus, rather than using the inequality moment condition directly, it could be translated into an inequality restriction on a set of parameters. There exists an extensive literature on estimation and inference in the presence of inequality parameter constraints of the form $\psi(\theta, \vartheta) \geq 0$, where $\psi(\cdot)$ is a deterministic function of the model parameters, e.g., Chernoff (1954), Kudo (1963), Perlman (1969), Gourieroux, Holly and Monfort (1982), Shapiro (1985), Kodde and Palm (1986), and Wolak (1991). Extensive literature surveys are provided in in Gourieroux and Monfort (1995) and Sen and Silvapulle (2002). EL inference subject to a constraint of the form $\psi(\theta, \vartheta) \geq 0$ has been considered by El Barmi (1995), El Barmi and Dykstra (1995), and Owen (2001). However, neither of them provides a complete limit distribution theory and considers the case in which the inequalities stem directly from the moment conditions.

Notice that the special case of $\mathbb{E}[H' g_2(X_i, \theta_0)] = 0$ translates into $H'\vartheta_0 = 0$, which means that ϑ_0 lies on the boundary of its domain. Hence, our asymptotic analysis is closely related to Andrews' (1999, 2001) work on estimation and testing when a parameter is on the boundary of the parameter space. Andrews (1999) considers estimators that are defined defined as extremum of an objective function. He constructs a stochastic quadratic approximation of this objective function that is valid in large samples and shows that the asymptotic distribution of interest is given by the distribution of the possibly constrained extremum of the quadratic limit objective function. EL estimators, however, are more conveniently expressed as the solution to a saddlepoint problem. Unlike the previous literature, e.g., Kitamura and Stutzer (1997) and Newey and Smith (2004), that develops the EL limit theory from an expansion of the first-order conditions associated with the saddlepoint, we follow Chernoff (1954) and Andrews (1999) by deriving a quadratic approximation of the EL objective function and analyzing the distribution of its saddlepoint. Rather than introducing the parameter ϑ directly, we restrict the Kuhn-Tucker parameter associated with $\mathbb{E}[g_2(X_i, \theta_0)]$ in the saddlepoint formulation of the EL problem to be non-positive.

The plan of the paper is as follows. Section 2 presents the assumptions underlying our analysis and the definition of the EL objective function and estimator. To motivate our setup we provide two examples: an instrumental variable estimation problem and a model of consumption in the presence of borrowing constraints in the spirit of Zeldes (1989). Section 3 develops the asymptotic distribution theory for the EL estimator and its objective function in the presence of inequality moment conditions. Section 4 presents tests of the hypothesis $\mathbb{E}[H'g_2(X_i,\theta_0)] = 0$. Results from a small Monte Carlo study are reported in Section 5. The data generating process is given by the economic model presented in Section 2. Section 6 concludes and the Appendix contains all proofs and technical Lemmas.

We use the following notation throughout the paper: "$\xrightarrow{p}$" and "$\Longrightarrow$" denote convergence in probability and distribution, respectively. "$\equiv$" signifies distributional equivalence. If A is an $n \times m$ matrix then $\|A\| = (tr[A'A])^{1/2}$. $\mathcal{I}\{x \geq a\}$ is the indicator function that is one if $x \geq a$ and zero otherwise. We abbreviate the "weak law of large numbers" by WLLN, the "uniform WLLN" by ULLN, and use w.p.a. 1 instead of "with probability approaching one." We denote $\mathbb{R}^{n-} = \{x \in \mathbb{R}^n \mid x \leq 0\}$ and $\mathbb{R}^{n+} = \{x \in \mathbb{R}^n \mid x \geq 0\}$.

2 Notation and Setup

The moment conditions that we are exploiting for estimation are given in Equation (1). Let Θ be the domain of the parameter vector θ. The functions g_1 and g_2 are of dimension $h_1 \times 1$ and $h_2 \times 1$, respectively. Let $h = h_1 + h_2$ and $g(X_i,\theta) = [g_1(X_i,\theta)', g_2(X_i,\theta)]'$. We use $g_j^{(1)}(X_i,\theta)$ and $g_j^{(2)}(X_i,\theta)$ to denote the first and the second order partial derivatives of $g_j(X_i,\theta)$, the j'th element of the vector $g(X_i,\theta)$, with respect to θ. Moreover, we collect the first-order derivatives in the matrix $g^{(1)}(X_i,\theta) = [g_1^{(1)}(X_i,\theta), \ldots, g_h^{(1)}(X_i,\theta)]$. We begin by stating some fundamental assumptions.

Assumption 1 *The random vectors X_i, $i = 1,\ldots,n$ are* i.i.d. *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$. a_iid

Assumption 2 *The parameter space Θ for θ is an m-dimensional compact subset of* $\mathbb{R}^m$. a_theta

Assumption 3 *The function $g(x,\theta)$ is continuous at each $\theta \in \Theta$ with probability one.* a_gcontinuity

Assumption 4 $\mathbb{E}[g_1(X_i,\theta_0)] = 0$*, and* $\mathbb{E}[g_1(X_i,\theta)] \neq 0$ *for* $\theta \neq \theta_0$*. Moreover,* $\nu_0 = \mathbb{E}[g_2(X_i,\theta_0)] \geq 0$ *and* $J = \mathbb{E}[g(X_i,\theta_0)g(X_i,\theta_0)']$ *is non-singular.* a_Eg

Assumption 5 $E\left[\sup_{\theta\in\Theta} \|g(X,\theta)\|^{\alpha}\right] < \infty$ *for some* $\alpha > 2$. a_Egalp

Assumption 6 *The matrix* $\mathbb{E}[g_1^{(1)}(X_i,\theta_0)']$ *has full column rank.* $\mathbb{E}\left[\sup_{\theta\in\Theta}\|g_k^{(1)}(X,\theta)\|\right]<\infty$, $\mathbb{E}\left[\sup_{\theta\in\Theta}\|g_{j,k}^{(2)}(X,\theta)\|\right]<\infty$ *for* $j=1,...,h$. a_g1g2

Most importantly, we assume in 4 that the parameter θ_0 is identifiable based on the equality moment condition $\mathbb{E}[g_1(X_i,\theta_0)]=0$. The expected value of $g_2(X_i,\theta_0)$ is denoted by $\nu_0\geq 0$.

2.1 Two Examples

Example 1: Suppose a researcher is interested in estimating the following regression model

$$X_{Y,i}=X'_{X,i}\theta_0+U_i, \tag{2}$$

where $X_{X,i}$ is an endogenous regressor that is correlated with the error term U_i. Moreover, the researcher has two sets of instrumental variables, denoted by $X_{1,i}$ and $X_{2,i}$. She is confident that the first set of instruments, $X_{1,i}$, is orthogonal to the error term U_i, but is concerned that the second set of instruments is potentially invalid. However, if $X_{2,i}$ is not orthogonal to U_i, then economic intuition suggests that the correlation is, say, positive. In this context two questions arise: how can one efficiently test whether the second set of instrument is valid? Second, how can we incorporate information from the second set of instruments in the estimation of θ_0?

In the returns-to-schooling literature $X_{Y,i}$ is a measure of income and $X_{X,i}$ is a measure of educational attainment, such as years of schooling. The error term U_i typically captures unobserved ability which is likely to be positively correlated with educational attainment. Hence, to account for the endogeneity one has to find instrumental variables that are orthogonal to innate ability, e.g., quarter of birth as in Angrist and Krueger (1991). Our framework allows the incorporation of additional instruments for which the researcher has some beliefs about the sign of their potential correlation with unobserved ability.

Example 2: Inequality moment restrictions arise, for instance, in environments in which agents face liquidity or regulatory constraints. Zeldes (1989) studies whether the presence of borrowing constraints can explain households' violation of consumption Euler equations. To motivate the econometric problem considered subsequently we present a two period version of Zeldes' infinite horizon model. Households choose consumption C in period 1 to maximize expected discounted utility:

$$\begin{aligned}\max_{C_1}\quad & U(C_1)+\beta\mathbb{E}_1\Big[U\Big((1+r)(A_1+Y_1-C_1)+Y_2\Big)\Big]\\ \text{s.t.}\quad & C_1\leq Y_1+A_1.\end{aligned}$$

In period t the household receives the income Y_t and can invest at rate r. The wealth in the initial period is A_1, whereas the wealth at the beginning of period 2 is given by $(1+r)(A_1 + Y_1 - C_1)$. The households face the borrowing constraint that period 1 consumption cannot exceed $Y_1 + A_1$. The Kuhn-Tucker condition for this constrained optimization problem is of the form

$$\mu = U^{(1)}(C_1) - \beta(1+r)\mathbb{E}_1[U^{(2)}(C_2)] \geq 0,$$

where $\mu = 0$ if $C_1 < Y_1 + A_1$. If the borrowing constraint is binding then the marginal utility of consumption at $t = 1$ exceeds the discounted expected marginal utility for $t = 2$ and $\mu > 0$.

Suppose one has a two-period panel-data set with observations on consumption and non-negative instruments. Define $\tilde{X}_i = [C_{i,1}, C_{i,2}, Z_{i,1}]'$, where the subscript i refers to the i'th household in the panel and assume that $Z_{1,i}$ is part of agent i'th information set at time $t = 1$. Moreover, let

$$\tilde{g}(\tilde{X}_i, \theta) = Z_{i,1}[U'(C_{i,1}) - \beta(1+r)U'(C_{i,2})].$$

Let θ be a parameter vector comprised of β, r, and the parameters of the utility function $U(C)$. In the absence of a borrowing constraint a popular approach to the estimation of the parameter vector θ is to exploit the moment condition

$$\mathbb{E}[\tilde{g}(\tilde{X}_i, \theta)] = 0 \quad \text{if and only if} \quad \theta = \theta_0.$$

Moment-based estimation does not require the econometrician to specify a probability distribution for income in the second period.

If a fraction of households is liquidity constrained then the moment condition becomes more complicated. Define the indicator variable S_i to be one if household i is not constrained. Zeldes (1989) replaces the indicator function S_i by an observable proxy $\tilde{S}_i$, constructed from the households' wealth-income ratio. He argues that if the wealth threshold that is used for the classification of households is sufficiently large, then it is reasonable to assume that $\tilde{S}_i \leq S_i$. Thus, some households may be classified as constrained although they are not constrained, but not vice versa. This assumption can be justified as follows. Households with a high wealth-to-income ratio are most likely not borrowing constrained. On the other hand, if a household holds no assets it could either face a binding constraint on current consumption, or it expects its future income to be constant and has no incentive to increase its current consumption.

Let $X_i = [\tilde{X}_i', \tilde{S}_i]'$. As long as $Z_{i,1}$ is chosen to be non-negative, the moment conditions that can be used to estimate θ are:

$$\begin{aligned} I\!\!E[g_1(X_i,\theta_0)] &= I\!\!E[\tilde{g}(\tilde{X}_i,\theta_0)\tilde{S}_i] = 0 \\ I\!\!E[g_2(X_i,\theta_0)] &= I\!\!E[\tilde{g}(\tilde{X}_i,\theta_0)(1-\tilde{S}_i)] \geq 0. \end{aligned}$$

Zeldes (1989) ignores the inequality moment condition when estimating θ and then evaluates the empirical moment of $g_2(X_i, \theta)$ at his estimate in order to test whether the group of households is indeed borrowing constrained. However, under the null hypothesis $I\!\!E[g_2(X_i, \theta_0)] = 0$ this approach ignores information from a large portion of the sample. Our subsequent analysis of the paper we show how the sample information can be used more efficiently by incorporating the inequality moment condition into the estimation and the construction of test statistics.

2.2 Empirical Likelihood Estimation

Among the various methods that could be used to estimate θ_0 based on the moment restrictions (1) we consider the method of maximum empirical likelihood. The notion of empirical likelihood was introduced by Owen (1988) and extended to incorporate moment restrictions by Qin and Lawless (1994). The (constrained) empirical likelihood function is eq_elobj

$$\begin{aligned} & L_{EL}(\theta,p) \\ &= \left\{ \prod_{i=1}^n p_i \;\middle|\; p_i > 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i g_1(X_i,\theta) = 0,\ \sum_{i=1}^n p_i g_2(X_i,\theta) \geq 0 \right\}, \end{aligned} \tag{3}$$

where p_i is a probability mass on X_i and $p = [p_1, \ldots, p_n]'$. The maximum empirical likelihood estimator (MELE) of θ and p is defined as

$$\{\hat{\theta}_{n,EL}, \hat{p}_{n,EL}\} = \underset{\theta\in\Theta,\ p}{\text{argmax}}\ L_{EL}(\theta,p). \tag{4}$$

Let eq_psiobj

$$\Psi_{EL}(\theta,p,\lambda_1,\lambda_2) = -\frac{1}{n}\sum_{i=1}^n \ln p_i + \lambda_1' \sum_{i=1}^n p_i g_1(X_i,\theta) + \lambda_2' \sum_{i=1}^n p_i g_2(X_i,\theta). \tag{5}$$

According to the Kuhn-Tucker Theorem there exist $\hat{\lambda}_{n,1} \in \mathbb{R}^{h_1}$ and $\hat{\lambda}_{n,2} \in \mathbb{R}^{h_2 -}$ such that $(\hat{\theta}_{n,EL}, \hat{p}_{n,EL}, \hat{\lambda}_{n,1}, \hat{\lambda}_{n,2})$ is a saddlepoint of Ψ_{EL}. Since the expected value of $g_2(X_i, \theta)$ is only required to be non-negative, $\hat{\lambda}_2$ is restricted to be less than or equal to zero. Based on the first-order conditions associated with the saddlepoint of Ψ_{EL} it is possible to express the probabilities $\hat{p}_{n,EL}$ as a function of $\hat{\lambda}_{n,1}$ and $\hat{\lambda}_{n,2}$. It is common in the empirical likelihood literature to exploit this relationship and modify the function Ψ_{EL} to eliminate the n-dimensional vector p. Let eq_gnobj

$$G_n(\theta, \lambda_1, \lambda_2) = \frac{1}{n}\sum_{i=1}^{n} \ln\left(1 + \lambda_1' g_1(X_i, \theta) + \lambda_2' g_2(X_i, \theta)\right) \tag{6}$$

and

$$\begin{aligned} \hat{\Lambda}_{n,1}(\theta) &= \{\lambda \in \mathbb{R}^{h_1} \mid \lambda' g_1(X_i, \theta) \geq -1 + \kappa, i = 1, \ldots, n\}, \\ \hat{\Lambda}_{n,2}^{-}(\theta) &= \{\lambda \in \mathbb{R}^{h_2 -} \mid \lambda' g_2(X_i, \theta) \geq -1 + \kappa, i = 1, \ldots, n\} \end{aligned}$$

for some $\kappa > 0$, and define the estimator $\hat{\theta}_n$ based on the following saddlepoint problem eq_gel.saddle

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmin}} \max_{\lambda_1 \in \hat{\Lambda}_{n,1}(\theta),\ \lambda_2 \in \hat{\Lambda}_{n,2}^{-}(\theta)} G_n(\theta, \lambda_1, \lambda_2). \tag{7}$$

The domains of λ_1 and λ_2 are chosen to ensure that the argument of the logarithm in (6) is strictly positive.

The (Kuhn-Tucker) first-order conditions associated with Ψ_{EL} are of the form eq_elfoc1-3

$$\begin{aligned} p_i &= \frac{1}{n\left(1 + \lambda_1' g_1(X_i, \theta) + \lambda_2 g_2(X_i, \theta)\right)}, & (8) \\ 0 &= \sum_{i=1}^{n} p_i g_1(X_i, \theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{g_1(X_i, \theta)}{1 + \lambda_1' g_1(X_i, \theta) + \lambda_2 g_2(X_i, \theta)}, & (9) \\ 0 &\leq \sum_{i=1}^{n} p_i g_2(X_i, \theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{g_2(X_i, \theta)}{1 + \lambda_1' g_1(X_i, \theta) + \lambda_2 g_2(X_i, \theta)}, & (10) \end{aligned}$$

where $\lambda_{2,j} = 0$ if the j'th element of (10) is strictly positive and $\lambda_{2,j} \leq 0$ otherwise. The objective function (6) is obtained by replacing the probabilities p_i in the the function Ψ_{EL} with (8). It is straightforward to verify that the first-order conditions for the modified saddle-point problem (7) are given by (9) and (10). Hence, as long as the constraints $\lambda_k' g_k(X_i, \theta) \geq -1 + \kappa$ that appear in the definitions of $\hat{\Lambda}_{n,1}(\theta)$ and $\hat{\Lambda}_{n,2}^{-}(\theta)$ are not binding, $\hat{\theta}_n$ and the associated $\hat{\lambda}_{n,1}$ and $\hat{\lambda}_{n,2}$ satisfy the first-order conditions for a saddlepoint of Ψ_{EL}.

It turns out that the large sample behavior of the saddlepoint of the function $G_n(\theta, \lambda_1, \lambda_2)$ is difficult to analyze directly, since the minimization with respect to λ_2 is restricted to non-positive values. We therefore define the function eq_gnstarobj

$$G_n^*(\theta, \nu, \lambda_1, \lambda_2) = G_n(\theta, \lambda_1, \lambda_2) - \nu' \lambda_2, \tag{11}$$

where ν is a $h_2 \times 1$ vector. In order to develop an asymptotic distribution theory for the estimator $\hat{\theta}_n$ it is more convenient to study the following problem eq_gelstarsaddle

$$\min_{\theta\in\Theta,\ \nu\geq 0}\ \max_{\lambda_1\in\hat\Lambda_{n,1}(\theta),\ \lambda_2\in\hat\Lambda_{n,2}(\theta)} G_n^*(\theta,\nu,\lambda_1,\lambda_2). \tag{12}$$

In the G_n^* formulation the vector λ_2 in the interior maximization problem is not restricted to be negative, that is,

$$\lambda_2\in\hat\Lambda_{n,2}(\theta)=\{\lambda\in\mathbb{R}^{h_2}\mid \lambda' g_2(X_i,\theta)\geq -1+\kappa, i=1,\ldots,n\}.$$

This will make it easier to approximate the profile of G_n^* that is obtained by maximization with respect to λ_1 and λ_2 for each value of θ and ν.

As mentioned in the Introduction, one could also rewrite the second moment condition as

$$I\!\!E[g_2(X_i,\theta_0)-\vartheta_0]=I\!\!E[\tilde g_2(X_i,\theta_0,\vartheta_0)]=0$$

and restrict the auxiliary parameter ϑ_0 to be nonnegative. The estimators $\hat\theta_n$ and $\hat\vartheta_n$ can be defined as the saddlepoint eq_geltilde.saddle

$$\min_{\theta\in\Theta,\ \vartheta\geq 0}\ \max_{\lambda_1\in\hat\Lambda_{n,1}(\theta),\ \lambda_2\in\hat\Lambda_{n,2}(\theta)} G_n^*(\theta,\nu,\lambda_1,\lambda_2), \tag{13}$$

where eq_geltilde

$$\tilde G_n(\theta,\vartheta,\lambda_1,\lambda_2)=\frac{1}{n}\sum_{i=1}^n \ln(1+\lambda_1' g_1(X_i,\theta)+\lambda_2'[g_2(X_i,\theta)-\vartheta]). \tag{14}$$

As in the G_n^* formulation the vector λ_2 is not constrained to be less than or equal to zero. The following lemma states that the three functions G_n, G_n^*, and $\tilde G_n$ have the same saddlepoints.

Lemma 1 *$\hat\theta$, $\hat\lambda_1$, $\hat\lambda_2$ are a solution to the saddlepoint problem (7)*

(i) if and only if $\hat\theta$, $\hat\lambda_1$, $\hat\lambda_2$, and $\hat\nu$ are a solution to the saddlepoint problem (12);

(ii) if and only if $\hat\theta$, $\hat\lambda_1$, $\hat\lambda_2$, and $\hat\vartheta$ are a solution to the saddlepoint problem (13).

The elements of the $h_2\times 1$ vector $\hat\nu$ are defined as

$$\hat\nu_j=\hat\vartheta_j=\begin{cases}\left.\frac{\partial G_n(\theta,\lambda_1,\lambda_2)}{\partial\lambda_{2,j}}\right|_{\hat\theta,\hat\lambda_1,\hat\lambda_2} & \textit{if}\quad \hat\lambda_{2,j}=0\\ 0 & \textit{if}\quad \hat\lambda_{2,j}<0,\quad j=1,\ldots,h_2.\end{cases}$$

l_gelequiv

From the definition of the function G_n in (6) and the first-order condition (10) it can be deduced that eq_nuhat

$$\hat\nu=\hat\vartheta=\sum_{i=1}^n \hat p_i g_2(X_i,\hat\theta), \tag{15}$$

that is, the $h_2 \times 1$ vector $\hat{\nu}$ in the G_n^* formulation of the saddlepoint problem provides an estimate of the expected value of g_2. To obtain a more compact notation we let

$$\lambda = [\lambda_1', \lambda_2']', \quad \text{and} \quad \hat{\Lambda}_n(\theta) = \hat{\Lambda}_{n,1}(\theta) \otimes \hat{\Lambda}_{n,2}(\theta).$$

$G_n(\theta, \lambda)$ is used to abbreviate $G_n(\theta, \lambda_1, \lambda_2)$. We define the $h_2 \times h$ matrix $M = [0\ I]$ such that

$$G_n^*(\theta, \nu, \lambda) = G_n(\theta, \lambda) - \nu' M \lambda. \tag{16}$$

We will subsequently study the saddlepoint of $G_n^*(\theta, \nu, \lambda)$ given by

$$\begin{aligned} \{\hat{\theta}_n,\ \hat{\nu}_n\} &= \underset{\theta \in \Theta,\ \nu \geq 0}{\operatorname{argmin}} \ \max_{\lambda \in \hat{\Lambda}_n(\theta)} G_n^*(\theta, \nu, \lambda) \\ \hat{\lambda}(\theta, \nu) &= \max_{\lambda \in \hat{\Lambda}_n(\theta)} G_n^*(\theta, \nu, \lambda). \end{aligned}$$

The introduction of the vector ν will make it easier to approximate the profile objective function

$$\bar{G}_n^*(\theta, \nu) = G_n^*(\theta, \nu, \hat{\lambda}(\theta, \nu)) \tag{17}$$

and will ultimately to a simplification of the asymptotic analysis.

3 Large Sample Analysis of the MELE

The large sample analysis proceeds in three steps. First, we establish the consistency of the MELE. Second we construct a quadratic approximation, denoted by $G_{nq}^*(\theta, \nu, \lambda)$ of the objective function $G_n^*(\theta, \nu, \lambda)$ in the neighborhood of $\theta = \theta_0$, $\nu = \nu_0$, and $\lambda = 0$ and show that the saddlepoint estimators defined on $G_n^*(\theta, \nu, \lambda)$ and $G_{nq}^*(\theta, \nu, \lambda)$ are $\sqrt{n}$-consistent. The third step consists of proving that the estimators obtained from G_n^* and and its quadratic approximation G_{nq}^* are distributionally equivalent in large samples.

3.1 Consistency

It is well known that the MELE with equality moment conditions is consistent. Since Assumption 4 guarantees that θ_0 is identifiable from $\mathbb{E}[g_1(X_i, \theta_0)] = 0$ it is not surprising that $\hat{\theta}_n$ is also consistent in our framework. However, we can also show that $\hat{\nu}_n$, characterized in Lemma 1 as derivative of $G_n(\theta, \lambda_1, \lambda_2)$ with respect to λ_2, converges to $\nu_0 = \mathbb{E}[g_2(X_i, \theta_0)]$. The vector of estimated Kuhn-Tucker parameters $\hat{\lambda}$ converges to zero. The consistency result is formally stated in the following theorem.

Theorem 1 *Suppose that Assumptions 1 to 5 are satisfied. Then* $\hat{\theta}_n \xrightarrow{p} \theta_0$ *and* $\hat{\nu}_n \xrightarrow{p} \nu_0$*. Moreover,* $\hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{p} 0$*.* t_consist

3.2 Quadratic Approximation of Objective Function

We proceed with a second-order Taylor approximation of the objective function G_n^*. Let $\beta = [\theta', \nu', \lambda']'$, $\beta_0 = [\theta_0', \nu_0', 0]'$, and abbreviate $G_n^*(\theta, \nu, \lambda)$ as $G_n^*(\beta)$. Define $G_n^{*(1)}(\beta)$ and $G_n^{*(2)}(\beta)$ to be the first and the second order partial derivatives of $G_n^*(\beta)$, respectively, and write the objective function as eq_gnapprox

$$G_n^*(\beta) = G_{nq}^*(\beta) + \frac{1}{n}\mathcal{R}_n(\beta), \tag{18}$$

where eq_gnq

$$G_{nq}^*(\beta) = G_n^*(\beta_0) + G_n^{*(1)}(\beta_0)'(\beta - \beta_0) + \frac{1}{2}(\beta - \beta_0)' G_n^{*(2)}(\beta_0)(\beta - \beta_0). \tag{19}$$

$\frac{1}{n}\mathcal{R}_n(\beta)$ is the remainder term of the Taylor approximation. The domain of β is given by

$$\mathcal{B}_n = \left\{\beta = [\theta', \nu', \lambda']' \mid \theta \in \Theta,\ \nu \in \mathbb{R}^{h_2+}, \lambda \in \hat{\Lambda}_n(\theta)\right\}$$

and a bound for the remainder is provided in the following lemma.

Lemma 2 *Suppose Assumptions 1 to 6 are satisfied, then for all* $\gamma_n \longrightarrow 0$ l_remain

$$\sup_{\beta \in \mathcal{B}_n : \|\beta - \beta_0\| \le \gamma_n} \frac{|\mathcal{R}_n(\beta)|}{(1 + \|\sqrt{n}(\beta - \beta_0)\|^2)} = o_p(1), \tag{20}$$

where $\mathcal{R}_n(\beta)$ *is the remainder term in (18).*

The first and second derivatives of G_n^* evaluated at β_0 are of the form eq_g1g2

$$G_n^{*(1)}(\beta_0) = [0, 0, n^{-1/2} Z_n'], \quad G_n^{*(2)}(\beta_0) = \begin{bmatrix} 0 & 0 & Q_n \\ 0 & 0 & -M \\ Q_n' & -M' & -J_n \end{bmatrix}, \tag{21}$$

where

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n g(X_i, \theta_0) - M'\nu_0,\ Q_n = \frac{1}{n}\sum_{i=1}^n g^{(1)}(X_i, \theta_0),\ J_n = \frac{1}{n}\sum_{i=1}^n g(X_i, \theta_0) g(X_i, \theta_0)'.$$

We proceed by transforming the parameter vector β. Let Let $b = [s', u', l']' = \sqrt{n}(\beta - \beta_0)$. The domain of b will be denoted by B_n, where B_n is defined such that

$$s \in S_n = \sqrt{n}(\Theta - \theta_0),\ u \in U_n = \sqrt{n}\left(\mathbb{R}^{h_2+} - \nu_0\right),\ l \in L_n(s) = \{l \mid l/\sqrt{n} \in \Lambda_n(\theta_0 + s/\sqrt{n})\}.$$

Notice that S_n expands to $\mathbb{R}^m$ and the j'th ordinate of U_n expands to $\mathbb{R}$ if $\mathbb{E}[g_{2,j}(X_i,\theta_0)] > 0$. The objective function G_n^* can be expressed in terms of the "local" deviations b from β_0 as eq_glocal

$$\mathcal{G}_n^*(s,u,l) = nG_n^*(\theta_0 + n^{-1/2}s, \nu_0 + n^{-1/2}u, n^{-1/2}l) = \mathcal{G}_{nq}^*(s,u,l) + \mathcal{R}. \tag{22}$$

We deduce from (19) and (21) that the quadratic approximation of the objective function is of the form eq_gnqlocal

$$\begin{aligned}\mathcal{G}_{nq}^*(s,u,l) &= -\frac{1}{2}(l - J_n^{-1}[Z_n + Q_n's - M'u])'J_n(l - J_n^{-1}[Z_n + Q_n's - M'u]) \\ &\quad +\frac{1}{2}(Z_n + Q_n's - M'u)'J_n^{-1}(Z_n + Q_n's - M'u). \end{aligned}\tag{23}$$

For notational convenience we will stack the parameters s and u into the vector $\phi = [s', u']'$ with domain $\Phi_n = S_n \otimes U_n$. Let $R_n = [-Q_n', M']'$ then we define eq_gnqphil

$$\begin{aligned}\mathcal{G}_{nq}^*(\phi,l) &= -\frac{1}{2}(l - J_n^{-1}[Z_n - R_n'\phi])'J_n(l - J_n^{-1}[Z_n - R_n'\phi]) \\ &\quad +\frac{1}{2}(Z_n - R_n'\phi)'J_n^{-1}(Z_n - R_n'\phi).\end{aligned}\tag{24}$$

The coefficient matrices of the function $\mathcal{G}_{nq}^*$ have the following limit distribution. Notice that the limit covariance matrix of Z_n depends not just θ_0 but also on ν_0.

t_jrz

Theorem 2 *Suppose Assumptions 1 to 6 are satisfied. Then*

$$(J_n, R_n, Z_n) \Longrightarrow (J, R, Z),$$

where $J = \mathbb{E}[g(X_i,\theta_0)g(X_i,\theta_0)']$, $R = \left[-\mathbb{E}[g^{(1)}(X_i,\theta_0)]', M'\right]'$ *and* $Z \sim \mathcal{N}(0, J - M'\nu_0\nu_0'M)$.

We now define two estimators: $\hat{b}$ is the standardized version of the actual empirical likelihood estimator. The second estimator, $\tilde{b}_q$ is obtained by solving a saddlepoint problem based on the objective $\mathcal{G}_{nq}^*(\phi,l)$ without restricting b to lie in B_n. Formally,

$$\begin{aligned}\hat{l}(\phi) &= \text{argmax}_{l\in L_n(\phi)}\, \mathcal{G}_n^*(\phi,l), \quad \hat{\phi} = \text{argmin}_{\phi\in\Phi_n}\, \mathcal{G}_n^*(\phi,\hat{l}(\phi)) \\ \tilde{l}_q(\phi) &= \text{argmax}_{l\in\mathbb{R}^h}\, \mathcal{G}_{nq}^*(\phi,l), \quad \tilde{\phi}_q = \text{argmin}_{\phi\in\Phi}\, \mathcal{G}_{nq}^*(\phi,\tilde{l}_q(\phi)),\end{aligned}$$

where $L_n(\phi)$ corresponds to $L_n(s)$ defined above and eq_phidomain

$$\Phi = \left\{\phi = [s', u'] \in \mathbb{R}^m \otimes \mathbb{R}^{h_2} \mid u_j \geq 0 \text{ if } \mathbb{E}[g_{2,j}(X_i,\theta_0)] = 0\right\}. \tag{25}$$

The vectors $\tilde{b}_q$ and $\tilde{\beta}_{nq}$ are defined by stacking and transforming the elements of $\tilde{\phi}_q$ and $\tilde{l}_q(\tilde{\phi}_q)$ appropriately.

Theorem 3 *Suppose Assumptions 1 to 6 are satisfied, then*

(i) $\sqrt{n}(\tilde{\beta}_{nq} - \beta_0) = O_p(1)$

(ii) $\sqrt{n}(\hat{\beta}_n - \beta_0) = O_p(1)$,

(iii) $nG^*_n(\hat{\beta}_n) = nG^*_{nq}(\hat{\beta}_n) + o_p(1)$,

(iv) $nG^*_{nq}(\hat{\beta}_n) = nG^*_{nq}(\tilde{\beta}_{nq}) + o_p(1)$,

(v) $nG^*_n(\hat{\beta}_n) = nG^*_{nq}(\tilde{\beta}_{nq}) + o_p(1)$.

t_op1

Theorem 3 establishes that $\hat{\beta}_n$ and $\tilde{\beta}_{nq}$ are $\sqrt{n}$-consistent. Moreover, the theorem states that the discrepancy between $\mathcal{G}^*_n(\beta)$ evaluated at $\hat{\beta}_n$ and $\mathcal{G}^*_{nq}(\beta)$ evaluated at $\tilde{\beta}_{nq}$ vanishes. Thus, the large-sample behavior of likelihood ratios can be approximated by the behavior of $\mathcal{G}^*_{nq}(\tilde{\beta}_{nq})$.

3.3 Limit Distribution of MELE

We begin by studying the limit distribution of $\tilde{b}_q$. From (24) it follows immediately that $\mathcal{G}^*_{nq}(\phi, l)$ is maximized with respect to $l \in \mathbb{R}^h$ by eq_lqtilde

$$\tilde{l}_q(\phi) = J_n^{-1}(Z_n - R'_n\phi). \quad (26)$$

According to Assumption 4 the limit of J_n is non-singular. Moreover, the function $g(x, \theta)$ is continuous at each $\theta \in \Theta$ (Assumption 3). Hence, $\tilde{l}_q(\phi)$ is well defined w.p.a. 1 and the concentrated objective function is of the form eq_gnqbar

$$\bar{\mathcal{G}}^*_{nq}(\phi) = \mathcal{G}^*_{nq}(\phi, \tilde{l}_q(\phi)) = \frac{1}{2}(Z_n - R'_n\phi)'J_n^{-1}(Z_n - R'_n\phi). \quad (27)$$

The limit distribution of $\tilde{\phi}_q$ can be determined from $\bar{\mathcal{G}}^*_{nq}(\phi)$. We then use (26) to obtain the distribution of $\tilde{l}_q(\tilde{\phi}_q)$. The results are summarized in the following theorem.

t_cltbq

Theorem 4 *Suppose Assumptions 1 to 6 are satisfied. Then*

$$(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) \Longrightarrow (\mathcal{P}, \mathcal{L}), \quad \textit{and} \quad \mathcal{G}^*_{nq}(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) \Longrightarrow \mathcal{G}^*_q(\mathcal{P}, \mathcal{L}),$$

where

$$\begin{aligned} \mathcal{P} &= \underset{\phi \in \Phi}{\textit{argmin}} \, \frac{1}{2}(Z - R'\phi)'J^{-1}(Z - R'\phi), \\ \mathcal{L} &= J^{-1}(Z - R'\mathcal{P}), \\ \mathcal{G}^*_q(\mathcal{P}, \mathcal{L}) &= \frac{1}{2}(Z - R'\mathcal{P})'J^{-1}(Z - R'\mathcal{P}). \end{aligned}$$

The final step in obtaining the limit distribution for $\hat{\beta}_n$ is to show that $\hat{b}$ and $\tilde{b}_q$ are asymptotically equivalent.

Theorem 5 *Suppose Assumptions 1 to 6 are satisfied, then* $\hat{b} = \tilde{b}_q + o_p(1)$. t_limdis

We will now explore the limit distribution in more detail. Let us assume that $g_2(x, \theta)$ can be partitioned into $[g_{2,1}'(x,\theta), g_{2,2}'(x,\theta)]'$ where $E[g_{2,1}(X_i, \theta_0)] = 0$ and $E[g_{2,1}(X_i, \theta_0)] > 0$. We partition the random vector Z and the matrices R and J

$$Z = \begin{bmatrix} Z_1 \\ Z_{2,1} \\ Z_{2,2} \end{bmatrix}, \quad R' = \begin{bmatrix} -Q_1' & 0 & 0 \\ -Q_{2,1}' & I & 0 \\ -Q_{2,2}' & 0 & I \end{bmatrix}, \quad J = \begin{bmatrix} J_{11} & J_{12,1} & J_{12,2} \\ J_{21,1} & J_{22,11} & J_{22,12} \\ J_{21,2} & J_{22,21} & J_{22,22} \end{bmatrix}.$$

The partitions conform with $g(x, \theta) = [g_1'(x,\theta), g_{2,1}'(x,\theta), g_{2,2}'(x,\theta)]'$. Moreover, we use $u = [u_1', u_2']'$. Using the formulas for marginal and conditional means and variances of a multivariate normal distribution it is straightforward to verify that eq_limobjphi

$$\begin{aligned}
&(Z - R'\phi)' J^{-1} (Z - R'\phi) \\
&= \begin{bmatrix} Z_1 + Q_1's \\ Z_{2,1} + Q_{2,1}'s - u_1 \end{bmatrix}' \begin{bmatrix} J_{11} & J_{12,1} \\ J_{21,1} & J_{22,11} \end{bmatrix}^{-1} \begin{bmatrix} Z_1 + Q_1's \\ Z_{2,1} + Q_{2,1}'s - u_1 \end{bmatrix} \qquad (28) \\
&\quad + \left(Z_{2,2} + Q_{2,2}'s - u_2 - \begin{bmatrix} J_{21,2} \\ J_{22,21} \end{bmatrix}' \begin{bmatrix} J_{11} & J_{12,1} \\ J_{21,1} & J_{22,11} \end{bmatrix}^{-1} \begin{bmatrix} Z_1 + Q_1's \\ Z_{2,1} + Q_{2,1}'s - u_1 \end{bmatrix} \right)' \\
&\quad \times \left(J_{22,22} - \begin{bmatrix} J_{21,2} \\ J_{22,21} \end{bmatrix}' \begin{bmatrix} J_{11} & J_{12,1} \\ J_{21,1} & J_{22,11} \end{bmatrix}^{-1} \begin{bmatrix} J_{12,2} \\ J_{22,12} \end{bmatrix} \right)^{-1} \\
&\quad \times \left(Z_{2,2} + Q_{2,2}'s - u_2 - \begin{bmatrix} J_{21,2} \\ J_{22,21} \end{bmatrix}' \begin{bmatrix} J_{11} & J_{12,1} \\ J_{21,1} & J_{22,11} \end{bmatrix}^{-1} \begin{pmatrix} Z_1 + Q_1's \\ Z_{2,1} + Q_{2,1}'s - u_1 \end{pmatrix} \right).
\end{aligned}$$

Now we decompose the limit random variable $\mathcal{P} = [\mathcal{S}', \mathcal{U}_1', \mathcal{U}_2']'$. The limit distribution of $\hat{u}_2$ is obtained by minimizing (28) over $\mathbb{R}^{h_2}$. Hence,

$$\mathcal{U}_2 = Z_{2,2} + Q_{2,2}'\mathcal{S} - \begin{bmatrix} J_{21,2} \\ J_{22,21} \end{bmatrix}' \begin{bmatrix} J_{11} & J_{12,1} \\ J_{21,1} & J_{22,11} \end{bmatrix}^{-1} \begin{bmatrix} Z_1 + Q_1'\mathcal{S} \\ Z_{2,1} + Q_{2,1}'\mathcal{S} - \mathcal{U}_1 \end{bmatrix},$$

which implies that the second summand in (28) is zero. We can draw two important conclusions from this algebraic manipulation. First, since the first summand does not depend on any partition of Z, Q, and J associated with $g_{2,2}(x, \theta)$ we deduce that the subset of inequality moment conditions that holds with strict inequality does not influence the distribution of $\mathcal{S}$ and $\mathcal{U}_1$ and, therefore, asymptotically does not provide any additional information on θ. Second, although the distribution of the random vector Z depends on ν_0, notice that

$$\begin{bmatrix} Z_1 \\ Z_{2,1} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} J_{11} & J_{12,1} \\ J_{21,1} & J_{22,11} \end{bmatrix}\right).$$

Thus, neither the distribution of $\mathcal{S}$, nor the distribution of $\mathcal{G}_q^*(\mathcal{P}, \mathcal{L})$ depends on the specific values of $\nu_{0,j}$ if $\nu_{0,j} > 0$.

Suppose that all the $g_2(x, \theta)$-moment conditions hold with strict inequality, that is, $\mathbb{E}[g_2(X_i, \theta_0)] > 0$, then it is not necessary to partition $g_2(x, \theta)$ and the notation simplifies considerably:

$$\mathcal{S} = -(Q_1 J_{11}^{-1} Q_1')^{-1} Q_1 J_{11}^{-1} Z_1 \equiv \mathcal{N}\Big(0, (Q_1 J_{11}^{-1} Q_1')^{-1}\Big).$$

Using the formula for the inverse of a partitioned matrix it can be verified that

$$\mathcal{L}_1 = J_{11}^{-1}(Z_1 + Q_1'\mathcal{S}), \quad \mathcal{L}_2 = 0.$$

Finally,

$$2\mathcal{G}_q^*(\mathcal{P}, \mathcal{L}) = Z_1'[J_{11}^{-1} - J_{11}^{-1} Q_1' (Q_1 J_{11}^{-1} Q_1')^{-1} Q_1 J_{11}^{-1}] Z_1, \tag{29}$$

which corresponds to a χ^2 random variable with $m - h_1$ degrees of freedom. Thus, the limit distributions reduce to the well-known case in which estimation and inference is based only on $\mathbb{E}[g_1(X_i, \theta_0)] = 0$.

4 Testing

In this section we develop test procedures for the hypothesis that the second moment condition holds with equality, a general specification test based on overidentifying moment conditions, and a coefficient test. The limit distributions of the proposed test statistics are generally non-standard and depend on nuisance parameters.

4.1 Tests for the Inequality Moment Condition

We begin by considering tests for the following null and alternative hypotheses:

$$\begin{aligned} H_0 &: \mathbb{E}[g_2(X_i,\theta_0)] = 0 \\ H_1 &: \mathbb{E}[g_2(X_i,\theta_0)] \geq 0,\ \mathbb{E}[g_{2,j}(X_i,\theta_0)] > 0 \\ &\quad \text{for at least one } 0 \leq j \leq h_2 \end{aligned}$$

Under both H_0 and H_1 we maintain that $\mathbb{E}[g_1(X_i,\theta_0)] = 0$. We consider four tests that are asymptotically equivalent: an empirical likelihood ratio test, a Hausman test, a Wald test, and a directed score (or Lagrange Multiplier) test.

1. Under the null hypothesis we can define the following constrained empirical likelihood function: eq_elobj0

$$\begin{aligned} &L^0_{EL}(\theta,p) \\ &= \left\{ \prod_{i=1}^n p_i \,\middle|\, p_i > 0,\ \sum_{i=1}^n p_i = 1,\ \sum_{i=1}^n p_i g_1(X_i,\theta) = 0,\ \sum_{i=1}^n p_i g_2(X_i,\theta) = 0 \right\}. \end{aligned} \tag{30}$$

   The empirical likelihood estimator is a saddlepoint of the function $G_n(\theta,\lambda_1,\lambda_2)$ defined in (6). However, unlike in the case of $\mathbb{E}[g_2(X_i,\theta_0)] \geq 0$ the Lagrange multiplier λ_2 is not constrained to be negative. Since $G^*_n(\theta,0,\lambda) = G_n(\theta,\lambda)$ we can define

$$\begin{aligned} \hat{\theta}^0_n &= \underset{\theta\in\Theta}{\text{argmin}} \max_{\lambda\in\hat{\Lambda}_n(\theta)} G^*_n(\theta,0,\lambda) \\ \hat{\lambda}^0(\theta,0) &= \underset{\lambda\in\hat{\Lambda}_n(\theta)}{\text{argmax}}\, G^*_n(\theta,0,\lambda). \end{aligned}$$

   The empirical likelihood ratio is based on the ratio of L^0_{EL} and L_{EL}. In terms of the objective function G^*_n it can be expressed as eq_lrtest

$$\mathcal{LR}^\nu_n = 2n\Big(G^*_n(\hat{\theta}^0_n, 0, \hat{\lambda}^0(\hat{\theta}^0_n, 0)) - G^*_n(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n))\Big). \tag{31}$$

2. A Hausman test statistic is constructed based on the difference between the stacked parameters $\hat{\psi}_n = [\hat{\theta}_n', \hat{\nu}_n']'$ and $\hat{\psi}_n^0 = [\hat{\theta}_n^{0\prime}, 0]'$. Under H_0 both $\hat{\psi}_n^0$ and $\hat{\psi}_n$ provide consistent estimates of θ_0 and ν_0. The estimator $\hat{\psi}_n^0$ is more efficient under the null hypothesis, whereas it is inconsistent under the alternative. Large discrepancies between $\hat{\psi}_n^0$ and $\hat{\psi}_n$ are evidence against the null hypothesis. Define eq_htest

$$\mathcal{H}_n = n(\hat{\psi}_n - \hat{\psi}_n^0)'\hat{R}_n \hat{J}_n^{-1} \hat{R}_n' (\hat{\psi}_n - \hat{\psi}_n^0). \tag{32}$$

3. The Wald test examines whether $\mathbb{E}[g_2(X_i, \theta_0)]$ is equal to zero based on $\hat{\nu}_n$: eq_wtest

$$\mathcal{W}_n = n\hat{\nu}_n'(M\hat{J}_n^{-1}M' - M\hat{J}_n^{-1}\hat{Q}_n'(\hat{Q}_n\hat{J}_n^{-1}\hat{Q}_n')^{-1}\hat{Q}_n\hat{J}_n^{-1}M')\hat{\nu}_n \tag{33}$$

4. A score test can be obtained as follows. Notice from (11) that the slope $\partial G_n^*/\partial\nu$ equals $-\lambda_2$, where λ_2 is the Lagrange multiplier associated with the second moment constraint in the empirical likelihood function. Under the null hypothesis one expects the constrained minimum of G_n^* with respect to $\nu \geq 0$ to be close to zero, which means that a large negative slope (positive value of $\hat{\lambda}_{n,2}^0$) is evidence against the null hypothesis. Following Andrews (2001), we define the directed score $\hat{\lambda}_{n,2}^+$ that minimizes

$$\left(\kappa - [G_n^{*(2)}(\hat{\beta}_n^0)^{-1}]_{\nu\nu}\hat{\lambda}_{n,2}(\hat{\theta}_n^0, 0)\right)' \left([G_n^{*(2)}(\hat{\beta}_n^0)^{-1}]_{\nu\nu}\right)^{-1} \left(\kappa - [G_n^{*(2)}(\hat{\beta}_n^0)^{-1}]_{\nu\nu}\hat{\lambda}_{n,2}(\hat{\theta}_n^0, 0)\right)$$

with respect to $\kappa \in \mathbb{R}^{h_2+}$. The score statistic is of the form eq_dtest

$$\mathcal{D}_n = n\hat{\lambda}_{n,2}^{+\prime}[G_n^{*(2)}(\hat{\beta}_n^0)^{-1}]_{\nu\nu}\hat{\lambda}_{n,2}^+. \tag{34}$$

Notice that $\mathcal{D}_n$ is large whenever $\hat{\lambda}_{n,2}^0$ takes a large positive value. If $\hat{\lambda}_{n,2}^0 \leq 0$ then $\mathcal{D}_n = 0$. Here $\hat{\beta}_n^0 = [\hat{\theta}_n^{0\prime}, 0, \hat{\lambda}_n^{0\prime}]'$ and $[G_n^{*(2)}(\hat{\beta}_n^0)^{-1}]_{\nu\nu}$ is the 2-2 block of the partitioned inverse of $G_n^{*(2)}$. The partitions conform with the partition of $\beta = [\theta', \nu', \lambda']'$.

The asymptotics of $\hat{\theta}_n^0$ and $\hat{\lambda}^0(\hat{\theta}_n^0, 0)$ are well known (e.g., Newey and Smith (2004)) and follow from straightforward modifications of the proofs of Theorems 3, 4, and 5. Let

$$\begin{aligned} \tilde{l}_q^0(\phi) &= \operatorname*{argmax}_{l \in \mathbb{R}^h} \mathcal{G}_{nq}^*(\phi, l) \\ \tilde{\phi}_q^0 &= \operatorname*{argmin}_{\phi \in \mathbb{R}^m \otimes \{0\}^{h_2}} \mathcal{G}_{nq}^*(\phi, \tilde{l}_q^0(\phi)), \end{aligned}$$

and define $\hat{b}^0$, $\tilde{b}_q^0$, and $\tilde{\beta}_{nq}^0$ in the same manner as $\hat{b}$, $\tilde{b}_q$, and $\tilde{\beta}_{nq}$ have been defined above. The main results are summarized in the following corollary.

Corollary 1 *Suppose Assumptions 1 to 6 are satisfied, then* c_limdis0 *(i)* $nG_n^*(\hat{\beta}_n^0) = nG_{nq}^*(\tilde{\beta}_{nq}^0) + o_p(1)$,

(ii) $(\tilde{\phi}_q^0, \tilde{l}_q^0(\tilde{\phi}_q^0)) \Longrightarrow (\mathcal{P}^0, \mathcal{L}^0)$,

(iii) $\mathcal{G}_{nq}^*(\tilde{\phi}_q^0, \tilde{l}_q^0(\tilde{\phi}_q^0)) \Longrightarrow \mathcal{G}_q(\mathcal{P}^0, \mathcal{L}^0)$,

(iv) $\hat{b}^0 = \tilde{b}_q^0 + o_p(1)$,

where $\mathcal{P}^0 = [\mathcal{S}^{0'}, 0]'$, $\mathcal{S}^0 = -(QJ^{-1}Q')^{-1}QJ^{-1}Z$, *and* $\mathcal{L}^0 = J^{-1}(Z + Q'\mathcal{S}^0)$.

The limit distribution of the test-statistics is given in the following theorem.

t_limittest

Theorem 6 *Suppose Assumptions 1 to 6 are satisfied. Then, under the null hypothesis,*

$$\mathcal{LR}_n^{\nu},\ \mathcal{H}_n,\ \mathcal{W}_n,\ \mathcal{D}_n \Longrightarrow \mathcal{U}'\Lambda^{-1}\mathcal{U},$$

where

$$\mathcal{U} = \underset{u \in \mathbb{R}^{h_2+}}{argmin}\, (u - Z_u)'\Lambda^{-1}(u - Z_u),$$

$\Lambda = (M[J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}]M')^{-1}$, *and* $Z_u \sim \mathcal{N}(0, \Lambda)$.

The test-statistics have a so-called $\bar{\chi}^2$ limit distribution. Kudo (1963) shows that the limit distribution can be characterized as follows:

$$P\left\{\mathcal{U}'\Lambda^{-1}\mathcal{U} \geq z\right\} = \sum_{\emptyset \subseteq K \subseteq \mathcal{K}} P\{\chi^2_{d(K)} \geq z\}\mathbb{P}\{(\Lambda_{K^c})^{-1}\}\mathbb{P}\{\Lambda_{K|K^c}\}.$$

The summation runs over all subsets K of $\mathcal{K} = \{1, \ldots, h_2\}$, $d(K)$ is the number of elements in K, K^c is the complement of K, Λ_K is the covariance matrix of $Z_{u,i}$, $i \in K$, $\Lambda_{K|K^c}$ is the same conditional on $Z_{u,j} = 0$, $j \in K^c$, and $\mathbb{P}\{\Sigma\}$ is the probability that the variables distributed in a multivariate normal distribution with mean zero and covariance matrix Σ are all positive. The random variable $\chi^2_{d(K)}$ has the χ^2 distribution with $d(K)$ degrees of freedom, where $\chi^2_{d(\emptyset)} = 0$, $\mathbb{P}\{\Lambda_{\emptyset|\mathcal{K}}\} = 1$, and $\mathbb{P}\{(\Lambda_{\mathcal{K}^c})^{-1}\} = \mathbb{P}\{(\Lambda_{\emptyset})^{-1}\} = 1$.

4.2 Testing Overidentifying Restrictions

We now study a specification test for the moment conditions. The null and alternative hypotheses are of the form

$$\begin{aligned} H_0 &: \mathbb{E}[g_1(X_i, \theta_0)] = 0 \text{ and } \mathbb{E}[g_2(X_i, \theta_0)] \geq 0 \\ H_1 &: \mathbb{E}[g_{1,j}(X_i, \theta_0)] \neq 0 \text{ or } \mathbb{E}[g_{2,k}(X_i, \theta_0)] < 0 \\ &\quad \text{for at least one } 0 \leq j \leq h_1 \text{ or } 0 \leq k \leq h_2. \end{aligned}$$

We consider the test statistic eq_jtest

$$\mathcal{J}_n = 2n\mathcal{G}_n^*(\hat{\beta}_n), \tag{35}$$

which can be interpreted as empirical likelihood ratio statistic since in the absence of moment restrictions the log empirical likelihood function is simply a function of the sample size: $n\ln(n)$. It can also be interpreted as J-statistic in the spirit of Hansen's (1982) test of overidentifying moment conditions in the GMM framework. Define the concentrated limit objective function

$$\bar{\mathcal{G}}_q^*(\phi) = \frac{1}{2}(Z - R'\phi)'J^{-1}(Z - R'\phi). \tag{36}$$

The limit distribution under H_0 can be directly obtained as a corollary from Theorems 4 and 5.

Corollary 2 *Suppose Assumptions 1 to 6 are satisfied, then* c_limitj

$$\mathcal{J}_n \Longrightarrow \min_{\phi\in\Phi(\nu_0)} 2\bar{\mathcal{G}}_q^*(\phi).$$

We now use the notation $\Phi(\nu_0)$ to signify the dependence of the domain of ϕ, defined in (25), on the nuisance parameter ν_0.

4.3 Coefficient Tests

Finally, we consider a hypothesis test for the parameter vector θ. We partition θ into $\theta = [\theta_1', \theta_2']$, where $\theta_j \in \Theta_j$ is $m_j \times 1$, $j = 1, 2$. The null and alternative hypotheses are of the form

$$\begin{aligned} H_0 &: \theta_1 = \theta_{1,0} \\ H_1 &: \theta_1 \neq \theta_{1,0}. \end{aligned}$$

The test statistic is given by the ratio of the unrestricted maximum of the empirical likelihood function $L_{EL}(\theta, p)$ and the constrained maximum subject to the restriction $\theta_1 = \theta_{1,0}$. We will express the test in terms of the function $G_n^*(\theta, \nu, \lambda)$.[2] Let

$$\begin{aligned} \{\hat{\theta}_{2,n},\ \hat{\nu}_n^0\} &= \operatorname*{argmin}_{\theta_2\in\Theta_2,\ \nu\geq 0}\ \max_{\lambda\in\hat{\Lambda}_n(\theta_{1,0},\theta_2)} G_n^*(\theta_{1,0}, \theta_2, \nu, \lambda) \\ \hat{\lambda}(\theta_{1,0}, \theta_2, \nu) &= \max_{\lambda\in\hat{\Lambda}_n(\theta_{1,0},\theta_2)} G_n^*(\theta, \nu, \lambda). \end{aligned}$$

Our test statistic is given by

$$\mathcal{LR}_n^\theta = 2n\Big(G_n^*(\theta_{1,0}, \hat{\theta}_{2,n}^0, \hat{\nu}_n^0, \hat{\lambda}^0(\theta_{1,0}, \hat{\theta}_{2,n}^0, \hat{\nu}_n^0)) - G_n^*(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n))\Big). \tag{37}$$

[2] In slight abuse of notation we write $G_n^*(\theta_1, \theta_2, \nu, \lambda)$ instead of $G_n^*([\theta_1', \theta_2']', \nu, \lambda)$.

As before, let

$$\bar{\mathcal{G}}_q^*(\phi) = \frac{1}{2}(Z - R'\phi)' J^{-1}(Z - R'\phi).$$

Define the set eq_phi0domain

$$\Phi_0(\nu_0) = \{\phi = [s_1', s_2', u']' \in \{0\}^{m_1} \otimes \mathbb{R}^{m_2+h_2} \mid u_j \geq 0 \text{ if } I\!\!E[g_{2,j}(X_i, \theta_0)] = 0\}. \quad (38)$$

The limit distribution under H_0 can be easily obtained as a corollary from Theorems 4 and 5.

Corollary 3 *Suppose Assumptions 1 to 6 are satisfied, then* c_limitlrt

$$\mathcal{LR}_n^\theta \Longrightarrow \left(\min_{\phi \in \Phi_0(\nu_0)} 2\bar{\mathcal{G}}_q^*(\phi) \right) - \left(\min_{\phi \in \Phi(\nu_0)} 2\bar{\mathcal{G}}_q^*(\phi) \right).$$

As in the specification test, the limit distribution of the likelihood ratio statistic for the coefficient test depends on the nuisance parameter ν_0 through the domain of ϕ.

4.4 Implementation

- The limit distribution of all the test statistics depends on the matrices Q and J, which can be consistently estimated as follows: eq_qjhat

$$\hat{J}_n = \frac{1}{n}\sum_{i=1}^n g(X_i, \hat{\theta}_n) g(X_i, \hat{\theta}_n)', \quad \hat{Q}_n = \frac{1}{n}\sum_{i=1}^n g^{(1)}(X_i, \hat{\theta}_n), \quad \hat{R}_n' = [-\hat{Q}_n', M']. \quad (39)$$

  Asymptotic critical values for the various test statistics can be obtained by computer simulation as, for instance, proposed by Andrews (2001).

- The distribution of the test statistics for the specification and the coefficient tests also depends on the "true" value of ν_0 through the domain Φ, defined in (25). For both tests the null hypothesis is composite in the sense that it only requires $\nu_0 \geq 0$. Since the dependence of the limit distribution on ν_0 is discontinuous whenever $\nu_{0,j} = 0$ for some j, the standard approach of replacing ν_0 by an consistent estimate is not applicable.

- Alternatives: (i) derive a critical value for the least favorable configuration (LFC) of the null hypothesis. By construction the LFC test is very conservative. (ii) Construct a confidence interval for the nuisance parameter ν_0 and take the sup-critical value for values of ν_0 in the confidence set. Since ν_0 can be consistently estimated it is possible to extract information about the nuisance parameter from the sample and to construct a less conservative critical value. The size of the test is then adjusted for the use of the confidence interval. References: Dufour (1992), Berger and Boos (1994), Silvapulle (1996), Hansen (2003).

- Details on the implementation are provided in the context of a specific example in the next section.

5 Example

In the remainder of the paper we apply the results obtained in the previous sections to a simultaneous equations model of the form

$$
\begin{aligned}
X_{Y,i} &= X'_{X,i}\theta + U_i && (40)\\
X_{X,i} &= X'_{1,i}\gamma_1 + X'_{2,i}\gamma_2 + \epsilon_i && (41)\\
X'_{2,i} &= X'_{1,i}\rho_{1,2} + U_i\rho_{u,2} + \eta'_i, && (42)
\end{aligned}
$$

where $X_{X,i}$ is an $m \times 1$ vector of endogenous regressors, and $X_{1,i}$ $(h_1 \times 1)$ and $X_{2,i}$ $(h_2 \times 1)$ are two vectors of instruments. While $X_{1,i}$ is assumed to be uncorrelated with the error term U_i, the second instrument, $X_{2,i}$ is potentially positively correlated with U_i. We subsequently study the limit distribution of the proposed EL estimator and likelihood ratio test statistics as well as the local power of the tests in the context of a specific example. We make the following assumptions.

Assumption 7 *(i) The random vector $V_i = [U_i, \epsilon'_i, \eta'_i, X'_{1,i}]'$ is independently and identically distributed. Its distribution is denoted by $\mathcal{P}$.*

(ii) Under the distribution $\mathcal{P}$ the following moment conditions hold: $\mathbb{E}_{\mathcal{P}}[X'_{1,i}U_i] = 0$, $\mathbb{E}_{\mathcal{P}}[X'_{1,i}\eta_i] = 0$, and $\mathbb{E}_{\mathcal{P}}[\epsilon'_i U_i] = 0$.

(iii) Moment-bound assumption for V_i under $\mathcal{P}$.

(iv) Define $\psi = [\theta', \rho'_{u,2}, \gamma'_1, \gamma'_2, vec(\rho_{1,2})]' \in \Psi$, where Ψ is a compact subset of $\mathbb{R}^{m+h_1+2h_2+h_1h_2}$ and $vec(\cdot)$ vectorizes the elements of a matrix. Ψ excludes values of ψ for which $\gamma_{1,j}$ or $\gamma_{2,j}$ is equal to zero.

(v) $\mathcal{P}$ has density with respect to Lebesgue measure that is continuous in its arguments. a_mcexp

- Assumption 7(ii) ensures that $X_{1,i}$ is a valid instrument for the estimation of θ and that the correlation between $X_{2,i}$ and U_i is due to $\rho_{u,2}$. Moreover, it is implicitly assumed that the endogenous regressor $X_{X,i}$ and the error term U_i are correlated.

- Show: Moment-bounds for V_i and domain restriction on Ψ will ensure that moment bounds in Assumptions 1 to 6 are satisfied.

- Let $X_i = [X_{Y,i}, X_{X,i}, X'_{1,i}, X_{2,i}]'$ and define

$$
\begin{aligned}
g_1(X_i, \theta) &= X_{1,i}(X_{Y,i} - X_{X,i}\theta) && (43)\\
g_2(X_i, \theta) &= X_{2,i}(X_{Y,i} - X_{X,i}\theta). && (44)
\end{aligned}
$$

Estimation and inference will be based on the moment conditions

$$\mathbb{E}[g_1(X_i,\theta)] = 0 \qquad \mathbb{E}[g_2(X_i,\theta)] = \rho_{u,2}\mathbb{E}[U_i^2] \geq 0$$

for $\theta = \theta_0$. Using the notation of Sections 2 to 4, $\nu_0 = \rho_{u,2}\mathbb{E}[U_i^2]$

- Moreover, it is straightforward to verify that

$$Z_{1,n} = \frac{1}{\sqrt{n}}\sum X_{1,i}U_i, \quad Z_{2,n} = \frac{1}{\sqrt{n}}\sum (X_{2,i}U_i - \rho_{u,2}\mathbb{E}[U_i^2]), \quad Z_n = [Z_{1,n}', Z_{2,n}]'$$

and

$$Q_n = -\frac{1}{n}\sum [X_{1,i}'X_{X,i}, X_{2,i}X_{X,i}], \quad J_n = \frac{1}{n}\sum \begin{bmatrix} X_{1,i}U_i^2X_{1,i}' & X_{1,i}U_i^2X_{2,i}' \\ & X_{2,i}U_i^2X_{2,i}' \end{bmatrix}.$$

- Remainder of this section:
  - Refinement of asymptotic results so that we can study power of tests and implement tests for composite null hypotheses.
  - Implementation of specification and coefficient tests
  - Numerical illustration that includes simulation of the limit distribution and a small Monte Carlo experiment.

5.1 Refinement of Asymptotic Results

Local Alternatives

In order to study power functions for the test statistics the limit distributions derived in Sections 3 and 4 have to be generalized to the case of a drifting DGP. While details are available from the authors on request, we briefly summarize the effect of the drift terms on the limit distribution. The objective function $G_n^*(\beta)$, see Equations (11) and (18) is approximated around

$$\beta_{n0} = [\theta_0' + n^{-1/2}s_0', \nu_0' + n^{-1/2}u_0', 0]',$$

where $\theta_0 = 0$ and $\nu_0 = 0$ in the context of the simultaneous equations model (40) - (42). The second-order approximation of the objective function leads to[3]

$$\begin{aligned} Z_n^a &= \frac{1}{n^{1/2}}\sum_{i=1}^n [g(X_i, n^{-1/2}s_0) - n^{-1/2}M'u_0] \\ Q_n^a &= \frac{1}{n}\sum g^{(1)}(X_i, n^{-1/2}s_0) \\ J_n^a &= \frac{1}{n}\sum g(X_i, n^{-1/2}s_0)g(X_i, n^{-1/2}s_0)'. \end{aligned}$$

[3]The superscript a signifies that the derivatives are evaluated at β_{n0} (drifting DGP / local alternative).

Let $R_n^a = [-Q_n^{a'}, M']$, then the quadratic approximation of the objective function is given by

$$\begin{aligned} \mathcal{G}_{nq}^{*a}(\phi, l) &= -\frac{1}{2}\left(l - (J_n^a)^{-1}\left[\widetilde{Z}_n^a - R_n^{a'}\phi\right]\right)' J_n^a \left(l - (J_n^a)^{-1}\left[\widetilde{Z}_n^a - R_n^{a'}\phi\right]\right) \\ &\quad + \frac{1}{2}\left(\widetilde{Z}_n^a - R_n^{a'}\phi\right)' (J_n^a)^{-1} \left(\widetilde{Z}_n^a - R_n^{a'}\phi\right), \end{aligned} \tag{45}$$

where

$$\widetilde{Z}_n^a = Z_n^a + R_n^{a'}\phi_0$$

is a shifted version of Z_n^a and $\phi_0 = [s_0', u_0']'$. We maintain the definitions

$$s = \sqrt{n}(\theta - \theta_0), \quad u = \sqrt{n}(\nu - \nu_0), \quad l = \sqrt{n}\lambda.$$

As in Theorem 2, we obtain that J_n^a, R_n^a, $\widetilde{Z}_n^a$ converge jointly in distribution:

$$(J_n^a, R_n^a, Z_n^a) \Longrightarrow (J, R, \widetilde{Z}). \tag{46}$$

The matrices J and R are the same as in Section 3. The random vector $\widetilde{Z}$ absorbs the drift terms and is defined as

$$\widetilde{Z} = Z + R'\phi_0. \tag{47}$$

Hence, the limit distribution of estimators and test statistics under the local alternatives can be derived from the quadratic limit objective function

$$\begin{aligned} \mathcal{G}_q^*(\phi, l) &= -\frac{1}{2}(l - J^{-1}[\widetilde{Z} - R'\phi])' J (l - J^{-1}[\widetilde{Z} - R'\phi]) \\ &\quad + \frac{1}{2}(\widetilde{Z} - R'\phi)' J^{-1} (\widetilde{Z} - R'\phi). \end{aligned} \tag{48}$$

As before, the constraint $\nu \geq 0$ translates into $u \geq 0$ if $\nu_0 = 0$ as in our example.

Uniform Convergence in Distribution

- Show that all weak convergence statements of Sections 3 and 4 hold uniformly in Ψ. We exploit that (i) Ψ is compact, (ii) the mapping from V_i into X_i is linear with a Jacobian that is non-singular for $\psi \in \Psi$, (iii) based on an assumption on the distribution of V_i, such as continuous density wrt. Lebesgue measure to show that the family of probability measures for X is compact and then use a shorter version of the argument in one of the previous notes on uniform convergence.

Additional Estimators and Tests

- We can construct an estimate of ν_0 based on $I\!E[g_1(X_i,\theta_0)] = 0$: Define

$$\hat{\nu}_{n(1)} = \frac{1}{n}\sum X_{2,i}(X_{Y,i} - X_{X,i}\hat{\theta}_{(1)}). \tag{49}$$

It can be verified that

$$\sqrt{n}\hat{\nu}_{n(1)} = u_0 + Z_{2,n} - Q'_{2,n}(Q_{1,n}J_{11,n}^{-1}Q'_{1,n})^{-1}Q_{1,n}J_{11,n}^{-1}Z_{1,n} + o_p(1). \tag{50}$$

The asymptotic variance of this estimator is

$$\Omega_{(1)} = J_{22} + Q'_2(Q_1 J_{11}^{-1} Q'_1)^{-1} Q_2 - 2Q'_2(Q_1 J_{11}^{-1} Q'_1) Q_1 J_{11}^{-1} J_{12}. \tag{51}$$

5.2 Testing Composite Null Hypotheses

- For ease of exposition assume $h_2 = 1$. Generalizations are straightforward.
- Relevant for specification test and the coefficient test, since distribution under null hypothesis depends on ν_0.
- Use $\mathcal{T}_n$ to denote test statistic: either $\mathcal{J}$ test of Section 4.2 or $\mathcal{LR}_n^\theta$ test of Section 4.3. Based on the limit results we can find asymptotic critical values $c(\nu_0)$ such that

$$\lim_{n\longrightarrow\infty} \mathcal{P}_{\nu_0}\{\mathcal{T}_n \geq c_\alpha(\nu_0)\} = \alpha.$$

To implement the test we need to know ν_0.
- Least favorable configuration (LFC) approach: use

$$\sup_{\nu_0} c_\alpha(\nu_0).$$

Example: specification test. Use K to denote subsets of $\mathcal{K} = \{1,\ldots,h_2\}$ and define eq_phikdomain

$$\Phi^{(K)} = \{\phi = [s', u']' \in I\!R^{m+h_2} \mid u_j \geq 0 \ \text{ if } \ j \in K\} \tag{52}$$

To account in our probability statements for the dependence of Z on ν we subsequently index P with the subscript ν. While the probability distributions of the limit random variables also depend on θ, we do not use a subscript to signify this dependence. We deduce from Corollary 2 that the rejection probability can be bounded by the maximum over all possible values of $I\!E[g_2(X_i,\theta_0)]$:

$$P_{\nu_0}\{\mathcal{J}_n > c\} \leq \max_{\emptyset\subseteq K\subseteq\mathcal{K}} \max_{\nu|\nu_j=0, j\in K} \left\{ P_\nu \left\{ \min_{\phi\in\Phi^{(K)}} 2\bar{\mathcal{G}}_q^*(\phi) > c \right\} \right\} + o(1) \tag{53}$$

The bound depends on the distribution of the minimum $\min_{\phi\in\Phi^{(K)}} 2\bar{\mathcal{G}}_q^*(\phi)$. While the distribution of the objective function $2\bar{\mathcal{G}}_q^*(\phi)$ depends through Z on ν, we demonstrated in Section 3.3 that the distribution of the constrained minimum does not. More precisely, it only matters whether $\nu_{0,j} > 0$. The specific value of $\nu_{0,j}$ is not important. Hence, we can simplify the bound on the rejection probability as follows:

$$P_{\nu_0}\{\mathcal{J}_n > c\} \leq \max_{\emptyset \subseteq K \subseteq \mathcal{K}} \left\{ P_{\nu=0}\left\{ \min_{\phi \in \Phi^{(K)}} 2\bar{\mathcal{G}}_q^*(\phi) > c \right\} \right\} + o(1). \tag{54}$$

Since

$$\min_{\phi \in \Phi^{(K)}} 2\bar{\mathcal{G}}_q^*(\phi) \leq \min_{\phi \in \Phi^{(\mathcal{K})}} 2\bar{\mathcal{G}}_q^*(\phi)$$

for all $\emptyset \subseteq K \subseteq \mathcal{K}$ we can deduce that

$$P_{\nu=0}\left\{ \min_{\phi \in \Phi^{(K)}} 2\bar{\mathcal{G}}_q^*(\phi) > c \right\} \leq P_{\nu=0}\left\{ \min_{\phi \in \Phi^{(\mathcal{K})}} 2\bar{\mathcal{G}}_q^*(\phi) > c \right\}.$$

Therefore, eq_boundj

$$P_{\nu_0}\{\mathcal{J}_n > c\} \leq P_{\nu=0}\left\{ \min_{\phi \in \Phi^{(\mathcal{K})}} 2\bar{\mathcal{G}}_q^*(\phi) > c \right\} + o(1) \tag{55}$$

and we can choose the critical value for the specification test based on the limit distribution obtained under the constrained minimization.

LFC approach is very conservative which potentially leads to low power. Coefficient testing has no straightforward LFC solution.

- Dufour (1992), Berger and Boos (1994), and Silvapulle (1996) proposed to construct a confidence set for the nuisance parameter of a composite hypothesis and to construct a sup-critical value subject to the restriction that the nuisance parameter has to lie in the confidence interval. Idea: Let $CI^{\nu}_{1-\alpha_n}(X_1, \ldots, X_n)$ be a confidence interval for ν with coverage probability greater than $1 - \alpha_n$ and $c_{n,\alpha}(\nu)$ a critical value for a size α test based on n observations, then the test that rejects H_0 if

$$\mathcal{T}_n \geq \sup_{\nu \in CI^{\nu}_{1-\alpha_n}} c_{n,\alpha}(\nu)$$

has size bounded by $\alpha + \alpha_n$. If the confidence interval extends only over a small subset of the nuisance parameter domain, this test is potentially more powerful than the LFC test.

- Hansen (2003) proposes an asymptotic version of this confidence-interval based test in which he lets α_n tend to zero as the sample size increases.

- A confidence interval for ν can be constructed as follows:

$$CI^{\nu}_{1-\alpha_n} = \left[\max\{0, \hat{\nu}_{(1)} - n^{-1/2} z^{crit}_{\alpha_n/2} \hat{\Omega}_{(1)}\}\ ,\ \hat{\nu}_{(1)} + n^{-1/2} z^{crit}_{\alpha_n/2} \hat{\Omega}_{(1)} \right], \tag{56}$$

- Given the uniform convergence in distribution the confidence interval has the property that

$$\lim \sup_{\nu} \mathcal{P}_{\nu}\{\nu \in CI^{\nu}_{1-\alpha_n}\} - (1-\alpha_n) = 0. \tag{57}$$

- Let $\nu_{n0} = \nu_0 + n^{-1/2}u_0$.

- According to our large sample analysis $\sqrt{n}(\hat{\nu}_{(1)} - \nu_0) = u_0 + O_p(1)$.

- If $\nu_0 > 0$ then the confidence interval will eventually be bounded away from 0 and we select the critical value that corresponds to $I\!E[g_2(X_i, \theta_0)] > 0$.

- If $\nu = 0$ then the confidence interval for $u = \sqrt{n}\nu$ will eventually cover the interval $[0, M]$ for any $M > 0$ if we let $\alpha_n \longrightarrow 0$.

- Use results in Hansen (2003) and in previous note on composite hypothesis testing to show that (i) if ν_0 is near zero then the confidence interval approach is asymptotically equivalent to searching for the least favorable configuration; (ii) if ν_0 is different from zero then the confidence interval approach is asymptotically equivalent to using critical values based on $\nu_0 > 0$.

5.3 Numerical Illustration

Parameterization:

- Make assumption about matrix dimensions: $m = 1$, $h_1 = 2$, and $h_2 = 1$.
- Error terms: the random variables U_i, η_i, and $X_{1,i}$ have zero mean and are independent of each other; ϵ_i has mean zero, is independent of $X_{X,i}$, and η_i, but is correlated with U_i. All errors are normally distributed.
- Rather than parameterizing the DGP directly in terms of γ_1, γ_2, σ_ϵ, and σ_η, we will make assumptions about the covariance matrix of the instrumental variables $X_{Z,i} = [X'_{X,i}, X_{2,i}]'$ and the correlation between instruments and regressors, denoted by $\rho_{1,X}$ (2×1) and $\rho_{2,X}$ (1×1). Specifically, we restrict the instruments $X_{1,i}$ to have unit covariance matrix, we assume that $\rho'_{1,2}\rho_{1,2} < 1$ and let $\sigma^2_\eta = 1 - \rho'_{1,2}\rho_{1,2}$, such that

  $$I\!E[X_{Z,i}X'_{Z,i}] = \begin{bmatrix} I & \rho_{1,2} \\ & 1 + \rho^2_{u,2} \end{bmatrix}.$$

  This covariance matrix determines the matrix J that appears in the characterizations of the limit distributions.

  Next, we assume that the endogenous regressor has unit variance[4] and calculate γ_1, γ_2, and σ^2_ϵ based on $\rho_{1,X}$, $\rho_{2,X}$, and $\rho_{1,2}$. Finally, it is assumed that the variance of the error term U_i is one. Since we are interested in the behavior of our proposed estimator and test statistics under local alternatives we assume that both θ and $\rho_{U,2}$ are local to zero in the sense

  $$\theta = n^{-1/2}s_0 \quad \text{and} \quad \rho_{u,2} = \nu = n^{-1/2}u_0. \tag{58}$$

  Hence, the design parameters of the DGP are

  $$s_0,\ u_0,\ \rho_{1,2},\ \rho_{1,X},\ \rho_{2,X},\ \rho_{u,\epsilon} \tag{59}$$

  and the matrices Q and J are given by

  $$Q = -\begin{bmatrix} \rho'_{1,X} & \rho_{2,X} \end{bmatrix}, \quad J = \begin{bmatrix} I & \rho_{1,2} \\ & 1 \end{bmatrix}.$$

  The vectors γ_1 and γ_2 are implicitly determined by the design parameters and the above auxiliary assumptions.

[4] While not all choices of the correlation parameters are consistent with $\sigma^2_\epsilon > 0$, the ones reported in the paper lead to a positive variance.

- Table 1 summarizes the parameter values used in the Monte Carlo simulation. All numerical results reported subsequently are based on 50,000 draws from the limit distribution and 10,000 Monte Carlo samples, respectively. The critical values for the test statistic that are used to calculate the finite sample and asymptotic power function are obtained from the simulated limit distribution.

- We will use the following Wald statistic as an alternative test for $\mathbb{E}[g_2(X_i, \theta_0)] = 0$:

$$\mathcal{W}^{\nu}_{n,(1)} = \frac{\left(\max\left\{0, \sqrt{n}\hat{n}\hat{u}_{n(1)}\right\}\right)^2}{\hat{\Omega}_{(1)}}, \tag{60}$$

  where $\hat{\Omega}_{(1)}$ is a consistent estimate of $\Omega_{(1)}$. This test is used in pretest procedures and as a benchmark for the inequality test proposed in this paper.

Estimation

We consider four different estimators of θ in the simultaneous equations model (40) - (42):

(i) $\hat{\theta}_{(0)}$ is MELE subject to $\frac{1}{n}\sum g_1(X_i, \theta) = 0$ and $\frac{1}{n}\sum g_2(X_i, \theta) \geq 0$.
(ii) $\hat{\theta}_{(1)}$ is MELE subject to $\frac{1}{n}\sum g_1(X_i, \theta) = 0$.
(iii) $\hat{\theta}_{(12)}$ is MELE subject to $\frac{1}{n}\sum g_1(X_i, \theta) = 0$ and $\frac{1}{n}\sum g_2(X_i, \theta) = 0$.
(iv) $\hat{\theta}_{(p)}$ is $\hat{\theta}_{(1)}$ if $\mathcal{W}^{\nu}_{n,(1)}$ rejects the null hypothesis that $\nu_0 = 0$, and $\hat{\theta}_{(12)}$ otherwise.

Insert discussion of pretest estimator

The estimator $\hat{\theta}_{(1)}$ does not use the second moment condition and is not affected by the parameter u_0. As discussed in Section 3, its limit distribution is given by $-(Q_1 J_{11}^{-1} Q_1')^{-1} Q_1 J_{11}^{-1} Z_1$, where the partitions of J, Q, and Z conform with the partitioning of $g(X_i, \theta)$ into g_1 and g_2. To obtain the distribution under the drifting DGP, Z has to be replaced with $\widetilde{Z}$ defined in (47). Thus,

$$\sqrt{n}(\hat{\theta}_{(1)} - \theta_0) \Longrightarrow \mathcal{N}\left(s_0, (Q_1 J_{11}^{-1} Q_1')^{-1}\right).$$

The estimator $\hat{\theta}_{(12)}$ is based on the assumption that the second moment condition is satisfied with equality. Its limit distribution is given by $-(QJ^{-1}Q')^{-1}QJ^{-1}\widetilde{Z}$, that is,

$$\sqrt{n}(\hat{\theta}_{(12)} - \theta_0) \Longrightarrow \mathcal{N}\left(s_0 - (QJ^{-1}Q')^{-1}QJ^{-1}M'u_0, (QJ^{-1}Q')^{-1}\right).$$

The top portion of Table 2 reports the bias and mean squared error (MSE) for the three empirical likelihood estimators, calculated based on the asymptotic distribution ($T = \infty$)

under DGP 1. As we previously showed, the limit distribution of $\hat{\theta}_{(1)}$ is not affected by u_0. The estimator is asymptotically unbiased and its MLE is equal to 2. For $u_0 = 0$ the estimator $\hat{\theta}_{(12)}$ which assumes that $\mathbb{E}[g_2(X_i, \theta_0)] = 0$ is more efficient than $\hat{\theta}_{(1)}$ since it uses an additional valid instrument. Its MSE equals 1.6. However, as u_0 increases the performance of $\hat{\theta}_{(12)}$ quickly deteriorates due to the bias introduced by imposing an invalid moment condition. This deterioration can be avoided by treating the second moment condition as inequality. If $u_0 = 0$ the MSE of our proposed estimator is 1.8 and lies between $MSE(\hat{\theta}_{(12)})$ and $MSE(\hat{\theta}_{(1)})$. Not surprisingly, $\hat{\theta}_{(0)}$ is asymptotically biased. As u_0 increases the inequality becomes less informative, the bias vanishes, and $\hat{\theta}_{(0)}$ becomes more and more similar $\hat{\theta}_{(1)}$.

The top part of Table 2 contains summary statistics for the finite sample distribution ($T = 100$) of the three estimators. Even in the absence of correlation between $X_{2,i}$ and U_i the estimators $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(12)}$ are slightly biased. Moreover, the finite sample variance of the estimators appears to be larger than the asymptotic variance. The gains from using the inequality moment condition for $T = 100$ are not as large as for $T = 100$. Finite sample bias and variance of $\hat{\theta}_{(0)}$ exceed the asymptotic bias and variance. As one lets u_0 increase, the behavior of $\hat{\theta}_{(0)}$ becomes very similar to that of $\hat{\theta}_{(1)}$. The larger u_0, the less likely it is that the inequality constraint $\nu \geq 0$ is binding and the less informative the inequality moment condition.

Tests for the Inequality Moment Condition

In Section 4.1 we proposed four asymptotically equivalent tests for the null and alternative hypotheses:

$$\begin{aligned} H_0 &: \mathbb{E}[g_2(X_i, \theta_0)] = 0 \\ H_1 &: \mathbb{E}[g_2(X_i, \theta_0)] \geq 0,\ \mathbb{E}[g_{2,j}(X_i, \theta_0)] > 0 \\ &\quad \text{for at least one } 0 \leq j \leq h_2 \end{aligned}$$

We subsequently simulate the limit distribution and calculate asymptotic power functions for $u_0 > 0$. In the Monte Carlo experiment we study the small sample performance of the empirical likelihood ratio test $\mathcal{LR}_n^{\nu}$.

In order to assess the benefit from using the inequality moment condition in the construction of the test statistic, we compare $\mathcal{LR}_n^{\nu}$ to the Wald statistic constructed from an estimate of θ based on $\mathbb{E}[g_1(X_i, \theta_0)] = 0$ specified above.

Figure 1 depicts the local asymptotic power for the empirical likelihood ratio test $\mathcal{LR}_n^{\nu}$ and the simple Wald test $\mathcal{W}_{n(1)}^{\nu}$. Critical values for both tests are chosen to obtain a 10 percent significance level. Under the parameterization of the DGP, our proposed empirical likelihood ratio test uniformly dominates the simple Wald test. Figure 2 shows the finite sample power of our likelihood ratio test $\mathcal{LR}_n^\nu$ and the simple Wald test. Except for a minor size distortion of the $\mathcal{LR}_n^\nu$ test, the finite sample power functions are similar to the asymptotic power functions.

Let $Z_\mathcal{N}$ be a $\mathcal{N}(0,1)$ random variable and

$$\Omega_{(0)} = \left(M[J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}]M' \right)^{-1}. \tag{61}$$

Tedious calculations involving inverses of partitioned matrices yield the following representation of the limit distribution of the test statistics in the context of our example

$$\mathcal{LR}_n^\nu \implies \left(Z_\mathcal{N} + \frac{u_0}{\Omega_{(0)}^{1/2}} \right)^2 \mathcal{I}\left\{ Z_\mathcal{N} + \frac{u_0}{\Omega_{(0)}^{1/2}} > 0 \right\} \tag{62}$$

$$\mathcal{W}_{n(1)}^\nu \implies \left(Z_\mathcal{N} + \frac{u_0}{\Omega_{(1)}^{1/2}} \right)^2 \mathcal{I}\left\{ Z_\mathcal{N} + \frac{u_0}{\Omega_{(1)}^{1/2}} > 0 \right\}, \tag{63}$$

where

$$\Omega_{(1)} - \Omega_{(0)} = \rho_{1,2}' \left[I - \frac{\rho_{1,X}\rho_{1,X}'}{\rho_{1,X}'\rho_{1,X}} \right] \rho_{1,2} = \frac{(\rho_{11,2}\rho_{12,X} - \rho_{12,2}\rho_{11,X})^2}{\rho_{1,X}'\rho_{1,X}} \geq 0$$

Thus, the magnitude of the power gain depends on the correlation between the two sets of instruments and the correlation between X_1 and X_X.

Testing Overidentifying Restrictions

We now consider testing the null and alternative hypotheses:

$$\begin{aligned} H_0 &: \mathbb{E}[g_1(X_i,\theta_0)] = 0 \text{ and } \mathbb{E}[g_2(X_i,\theta_0)] \geq 0 \\ H_1 &: \mathbb{E}[g_{1,j}(X_i,\theta_0)] \neq 0 \text{ or } \mathbb{E}[g_{2,k}(X_i,\theta_0)] < 0 \\ &\quad \text{for at least one } 0 \leq j \leq h_1 \text{ or } 0 \leq k \leq h_2. \end{aligned}$$

based on the test statistic J_n, define in (35). Figure 3 depicts the asymptotic rejection probabilities of the $\mathcal{J}_n$ test. By construction the critical value is chosen such that the rejection probability is $\alpha = 0.1$ for $u_0 = 0$. For values of $u_0 > 0$ the actual size of the test is slightly less than the nominal size as expected from the theoretical arguments. The test has power against $u_0 < 0$. For values $u_0 < -4$ the power is essentially one. Figure 4 graphs the small sample rejection probabilities. Except for a small size distortion the finite sample power function is well approximated by the large sample results.

5.4 Coefficient Tests

This subsection considers empirical likelihood ratio tests for the hypothesis $H_0 : \theta = 0$ versus the alternative $H_1 : \theta \neq 0$. We compare the performance of three tests:

(i) $\mathcal{LR}^{\theta}_{n(0)}$ is the empirical likelihood ratio test based on $\mathbb{E}[g_1(X_i, \theta_0)] = 0$ and $\mathbb{E}[g_2(X_i, \theta_0)] \geq 0$.

(ii) $\mathcal{LR}^{\theta}_{n(1)}$ is based on $\mathbb{E}[g_1(X_i, \theta_0)] = 0$.

(iii) $\mathcal{LR}^{\theta}_{n(12)}$ is based on $\mathbb{E}[g_1(X_i, \theta_0)] = 0$ and $\mathbb{E}[g_2(X_i, \theta_0)] = 0$.

(iv) $\mathcal{LR}^{\theta}_{n(pre)}$: Use $\mathcal{LR}^{\theta}_{n(1)}$ if $\mathcal{W}^{\nu}_{n,(1)}$ rejects the null hypothesis that $\nu_0 = 0$, and $\mathcal{LR}^{\theta}_{n(12)}$ otherwise.

insert discussion of pre-test procedure:

As the test for overidentifying restrictions, the $\mathcal{LR}^{\theta}_{n(0)}$ test is a test for a composite null hypothesis. The limit distribution of the test statistic depends on the nuisance parameter ν. Unlike for the $\mathcal{J}_n$ test, it is not straightforward to derive the least favorable configuration analytically. We recommend to use the confidence-interval approach outlined in the previous subsection to implement the coefficient test. In the context of our simultaneous equations model with drifting coefficients this is asymptotically equivalent to use sup-critical values over $u_0 \geq 0$.

Figure 5 depicts the asymptotic power functions of the three tests for $u_0 = 0$, $u_0 = 1$, and $u_0 = 4$. For $u_0 = 4$ the inequality moment condition is uninformative and the tests $\mathcal{LR}^{\theta}_{n(0)}$ and $\mathcal{LR}^{\theta}_{n(1)}$ are indistinguishable. The third test, $\mathcal{LR}^{\theta}_{n(12)}$ is severely biased and size distorted, because it imposes an invalid moment condition. As u_0 is reduced to 1 a discrepancy between the $\mathcal{LR}^{\theta}_{n(0)}$ and $\mathcal{LR}^{\theta}_{n(1)}$ tests emerges. For values $s_0 < 0$ there is a slight power gain due to the use of the inequality constraint. If $u_0 = 0$ the power differential is more pronounced. For values of $s_0 < 0$ $\mathcal{LR}^{\theta}_{n(0)}$ dominates the two other tests. If, on the other hand, $s_0 > 0$ the rejection frequency of $\mathcal{LR}^{\theta}_{n(0)}$ is slightly less than the rejection frequency of $\mathcal{LR}^{\theta}_{n(1)}$.

Figure 6 displays finite sample power functions. The small sample ranking of the tests is the same as the large sample ranking. For all three tests, the actual size is slightly larger than the nominal size. Unlike the asymptotic power functions of $\mathcal{LR}^{\theta}_{n(1)}$ and $\mathcal{LR}^{\theta}_{n(12)}$, the finite sample power functions are slightly asymmetric. The tests are more powerful against negative than positive alternatives.

6 Conclusion

(to be written)

A Proofs and Derivations

A.1 Empirical Likelihood Estimation

Proof of Lemma 1: We will verify the saddlepoint properties directly. (i) Suppose $\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2$ is a saddlepoint of G_n^*. If $\hat{\lambda}_{2,j} = 0$ it lies in the interior of $\hat{\Lambda}_2(\theta)$ and satisfies the first-order condition

$$\hat{\nu}_j = \left. \frac{\partial G_n(\theta, \lambda_1, \lambda_2)}{\partial \lambda_{2,j}} \right|_{\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2}.$$

If $\hat{\lambda}_{2,j} = 0$ then $\hat{\nu}_j$ minimizes G_n^* with respect to $\nu_j \geq 0$. Moreover, it is straightforward to verify that $\hat{\lambda}_2$ cannot be strictly positive. Hence, $\hat{\nu}' \hat{\lambda}_2 = 0$ and

$$G_n(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) = G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) \leq G_n^*(\theta, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) = G_n(\theta, \hat{\lambda}_1, \hat{\lambda}_2)$$

for all $\theta \in \Theta$. Moreover,

$$G_n(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) = G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) \geq G_n^*(\hat{\theta}, \hat{\nu}, \lambda_1, \hat{\lambda}_2) = G_n(\hat{\theta}, \lambda_1, \hat{\lambda}_2)$$

for all $\lambda_1 \in \hat{\Lambda}_{n,1}(\hat{\theta})$. Using the same argument as above it follows for $\hat{\lambda}_{2,j} < 0$ and $\hat{\nu}_j = 0$ that

$$G_n(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) \geq G_n(\hat{\theta}, \hat{\lambda}_1, \lambda_{2,(j)}),$$

where $\lambda_{2,(j)} \in \hat{\Lambda}_{n,2}(\hat{\theta})$ is obtained by replacing the j'th element of $\hat{\lambda}_2$ by $\lambda_{2,j} \leq 0$. Finally, if $\hat{\lambda}_{2,j} = 0$ then

$$\left. \frac{\partial G_n(\theta, \lambda_1, \lambda_2)}{\partial \lambda_{2,j}} \right|_{\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2} = \hat{\nu}_j \geq 0.$$

Since the function $G_n(\theta, \lambda_1, \lambda_2)$ is globally concave in λ_2 we deduce that

$$G_n(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) \geq G_n(\theta, \hat{\lambda}_1, \lambda_{2,(j)}).$$

As before, $\lambda_{2,(j)} \in \hat{\Lambda}_{n,2}(\hat{\theta})$ is obtained by replacing the j'th element of $\hat{\lambda}_2$ by $\lambda_{2,j} \leq \hat{\lambda}_{2,j} = 0$. Hence, we have established that $\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2$ is a saddlepoint of G_n.

Now suppose $\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2$ is a saddlepoint of G_n. The following inequalities are straightforward to verify:

$$\begin{aligned} G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) &\leq G_n^*(\theta, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) \\ G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) &\geq G_n^*(\hat{\theta}, \hat{\nu}, \lambda_1, \hat{\lambda}_2). \end{aligned}$$

Recall that $\hat{\nu}'\hat{\lambda}_2 = 0$ and $\nu'\lambda_2 \leq 0$. Therefore,

$$\begin{aligned} G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) &= G_n(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) - \hat{\nu}'\hat{\lambda}_2 \\ &\leq G_n(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) - \nu'\hat{\lambda}_2 \\ &= G_n^*(\hat{\theta}, \nu, \hat{\lambda}_1, \hat{\lambda}_2). \end{aligned}$$

If $\hat{\lambda}_{2,j} < 0$ then $\hat{\nu}_j = 0$ and

$$\begin{aligned} G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) &= G_n(\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2) - \hat{\nu}'\hat{\lambda}_2 \\ &\geq G_n(\hat{\theta}, \hat{\lambda}_1, \lambda_{2,(j)}) - \hat{\nu}'\lambda_{2,(j)} \\ &= G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \lambda_{2,(j)}), \end{aligned}$$

where $\lambda_{2,(j)}$ is defined as above. Now suppose that $\hat{\lambda}_{2,j} = 0$. Then

$$\left.\frac{\partial G_n^*(\theta, \nu, \lambda_1, \lambda_2)}{\partial \lambda_{2,j}}\right|_{\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2} = \left.\frac{\partial G_n(\theta, \lambda_1, \lambda_2)}{\partial \lambda_{2,j}}\right|_{\hat{\theta}, \hat{\lambda}_1, \hat{\lambda}_2} - \hat{\nu}_{2,j} = 0$$

Since G_n^* is globally concave in $\lambda_{2,j}$ we deduce that

$$G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \hat{\lambda}_2) \geq G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}_1, \lambda_{2,(j)}),$$

because G_n attains at $\hat{\lambda}_{2,j}$ its maximum with respect to $\lambda_{2,j}$.

The proof of (ii) is very similar to (i) and therefore omitted. ■

A.2 Consistency

In addition to the domain $\hat{\Lambda}_n(\theta)$ defined in the main text we also define the shrinking domain $\Lambda_n^\zeta = \{\lambda \in I\!R^h : \|\lambda\| \leq n^{-\zeta}\}$, where $\frac{1}{\alpha} < \zeta < \frac{1}{2}$. According to Assumption 5 the constant $\alpha > 2$ is such that $I\!E[\sup_{\theta \in \Theta} \|g(X, \theta)\|^\alpha] < \infty$. The relationship between $\hat{\Lambda}_n(\theta)$ and Λ_n^ζ is characterized in Lemma A.1 below.

A.2.1 Main Result

Proof of Theorem 1: We have to show that for any $\delta > 0$

$$\lim_{n \longrightarrow \infty} P\left\{\hat{\theta}_n \in \mathbb{B}(\theta_0, \delta), \hat{\nu}_n \in \mathbb{B}(\nu_0, \delta)\right\} = 1,$$

where

$$\mathbb{B}(\theta, \delta) = \{\tilde{\theta} \in \Theta \mid \|\theta - \tilde{\theta}\| < \delta\}, \quad \mathbb{B}(\nu, \delta) = \{\tilde{\nu} \in \mathbb{R}^{h_2+} \mid \|\nu - \tilde{\nu}\| < \delta\}.$$

Define

$$\Theta_0^c = \Theta \cap \mathbb{B}\left(\theta_0, \delta\right)^c \quad \text{and} \quad N_0^c = \mathbb{R}^{h_2+} \cap \mathbb{B}\left(\nu_0, \delta\right)^c.$$

We show the following two statements are true: (i) For a given $\varepsilon, \delta > 0$ and ζ such that $\frac{1}{\alpha} < \zeta < \frac{1}{2}$, there exist positive constants η and κ and $\bar{n}$ such that for $n \geq \bar{n}$ consist.s1

$$P\left\{\bar{G}_n^*\left(\theta_0, \nu_0\right) \geq n^{-\zeta-\kappa}\eta\right\} < \frac{\varepsilon}{2} \tag{A.1}$$

and (ii) consist.s2

$$P\left\{\min_{\theta\in\Theta_0^c,\ \nu\in N_0^c} \bar{G}_n^*\left(\theta, \nu\right) \leq n^{-\zeta}\eta\right\} < \frac{\varepsilon}{2}. \tag{A.2}$$

Then, from (A.1) and (A.2) we deduce that there exists an $\eta > 0$ such that for $n \geq \bar{n}$:

$$\begin{aligned} & P\left\{\hat{\theta}_n \in \mathbb{B}\left(\theta_0, \delta\right),\ \hat{\nu}_n \in \mathbb{B}(\nu_0, \delta)\right\} \\ \geq \quad & P\left\{\bar{G}_n^*\left(\theta_0, \nu_0\right) < n^{-\zeta-\kappa}\eta,\ \min_{\theta\in\Theta_0^c,\ \nu\in N_0^c} \bar{G}_n^*\left(\theta, \nu\right) > n^{-\zeta}\eta\right\} \geq 1 - \epsilon. \end{aligned}$$

Proof of (i). By Lemma A.2

$$\bar{G}_n^*(\theta_0, \nu_0) = \max_{\lambda\in\hat{\Lambda}_n(\theta_0)} G_n^*(\theta_0, \nu_0, \lambda) \leq O_p(1/n).$$

Choose $\kappa > 0$ such that $\zeta + \kappa < 1$. Then

$$n^{\zeta+\kappa}\bar{G}_n^*(\theta_0) \leq O_p(n^{\zeta+\kappa-1}) = o_p(1)$$

as required.

Proof of (ii). To obtain a lower bound for $\bar{G}_n^*(\theta, \nu)$ we will evaluate the function $G_n^*(\theta, \nu, \lambda)$ at $\lambda = n^{-\zeta}u(\theta, \nu)$, where the function $u(\theta, \nu)$ is defined as

$$u(\theta, \nu) = \begin{cases} 0 & \text{if } \theta = \theta_0,\ \nu = \nu_0 \\ \frac{\boldsymbol{E}[g(X,\theta)] - M'\nu}{\|\boldsymbol{E}[g(X,\theta)] - M'\nu\|} & \text{otherwise} \end{cases}$$

such that $\|u(\theta, \nu)\| \leq 1$.

Moreover, we truncate the function $g(x, \theta)$ as follows. Choose a positive constant $\xi < \zeta - \frac{1}{\alpha}$. Let

$$\mathcal{X}_n = \left\{x : \sup_{\theta\in\Theta} \|g\left(x, \theta\right)\| \leq n^{\zeta-\xi}\right\} \quad \text{and} \quad g_n\left(x, \theta\right) = I\left\{x \in \mathcal{X}_n\right\} g\left(x, \theta\right).$$

We then replace the terms

$$\ln(1 + \lambda' g(x, \theta)) - \lambda' M\nu$$

in the definition of the objective function $G_n^*(\theta, \nu, \lambda)$ by

$$q_n(x, \theta, \nu) = \ln\left(1 + n^{-\zeta}u'(\theta, \nu)g_n(x, \theta)\right) - n^{-\zeta}u(\theta, \nu)'M\nu.$$

and will later argue that the approximation error is negligible.

A second-order Taylor expansion of q_n around $u = 0$ yields consist.mv

$$n^{\zeta} q_n(x,\theta,\nu) = u(\theta,\nu)'(g_n(x,\theta) - M'\nu) - \frac{1}{2}\frac{n^{-2\zeta}u'(\theta,\nu)g_n(x,\theta)g_n(x,\theta)'u(\theta,\nu)}{\left(1+n^{-\zeta}u_*'(\theta,\nu)g_n(x,\theta)\right)^2}, \quad (A.3)$$

where $u_*'(\theta,\nu)$ lies between zero and $u(\theta,\nu)$. The second-order term of the Taylor approximation (A.3) can be bounded as follows. For given x, θ, and ν

$$\left|n^{-\zeta} u^{*'}(\theta,\nu)\, g_n(x,\theta)\right| \le n^{-\zeta}\left\|g_n(x,\theta)\right\| \le n^{-\xi}.$$

Therefore, consist.term2bnd

$$n^{-2\zeta}\frac{u'(\theta,\nu)g_n(x,\theta)g_n(x,\theta)'u(\theta,\nu)}{\left(1+n^{-\zeta}u_*'(\theta,\nu)g_n(x,\theta)\right)^2} \le n^{-2\zeta}\frac{\|g_n(x,\theta)\|^2\|u(\theta,\nu)\|^2}{(1-n^{-\xi})^2} \le n^{-2\xi} \longrightarrow 0 \quad (A.4)$$

Now consider the expected value of $q_n(x,\theta,\nu)$. For large enough n consist.lbnd

$$\begin{aligned} n^{\zeta}\mathbb{E}\left[q_n(X,\theta,\nu)\right] &= u(\theta,\nu)'(\mathbb{E}[g_n(X,\theta)] - M'\nu) + o(1) \quad (A.5)\\ &= \begin{cases} o(1) & \text{if } \theta=\theta_0,\ \nu=\nu_0 \\ \|\mathbb{E}[g(X,\theta)] - M'\nu\| + o(1) > 0 & \text{otherwise} \end{cases}. \end{aligned}$$

The $o(1)$ terms absorb the second-order term of the Taylor approximation and the discrepancy between $\mathbb{E}[g_n(X,\theta)]$ and $\mathbb{E}[g(X,\theta)]$, which vanishes as $\mathcal{X}_n$ expands. From (A.5) and the monotone convergence theorem we can deduce that

$$\lim_{n\to\infty} n^{\zeta}\lim_{\delta\downarrow 0}\mathbb{E}\left[\inf_{\theta^*\in\mathbb{B}(\theta,\delta),\ \nu^*\in\mathbb{B}(\nu,\delta)} q_n(X,\theta^*,\nu^*)\right] \begin{cases} =0 & \text{if } \theta=\theta_0 \\ >0 & \text{otherwise} \end{cases}.$$

According to Assumption 5 there exists a finite K such that

$$\sup_{\theta\in\Theta}\|\mathbb{E}[g_2(X,\theta)]\| < K < \infty.$$

Hence, for $\|\nu\| > 2K$ we can deduce that

$$\lim_{n\to\infty} n^{\zeta}\lim_{\delta\downarrow 0}\mathbb{E}\left[\inf_{\theta^*\in\mathbb{B}(\theta,\delta),\ \|\nu^*\|\ge 2K} q_n(X,\theta^*,\nu^*)\right] > K.$$

Since Θ is compact by assumption the set $\Theta\cap\mathbb{B}(\theta_0,\delta)^c$ is compact. Moreover, define the compact set $\mathbb{R}_K^{h_2+} = \{x \in \mathbb{R}^{h_2+}, \|x\| \le 2K\}$. We can cover both $\Theta\cap\mathbb{B}(\theta_0,\delta)^c$ and $\mathbb{R}_K^{h_2+}\cap\mathbb{B}(\nu_0,\delta)^c$ with $\Theta_j = \mathbb{B}(\theta_j,\delta_j)$ and $N_j = \mathbb{B}(\nu_j,\delta_j)$'s, $j = 1,\ldots,J$ taking each δ_j small enough such there exist η_j's such that

$$n^{\zeta}\mathbb{E}\left[\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n(X,\theta,\nu)\right] \ge 2\eta_j,\ n\ge n_j \quad (A.6)$$

for some positive numbers $\eta_j = \eta_j(\delta)$, $j = 1, \ldots, J$. By the WLLN[5] for a given $\varepsilon > 0$, we can find $\bar{n}_j's$ such that $n \geq \bar{n}_j$ implies that

$$\begin{aligned} \frac{\varepsilon}{2(J+1)} &\geq P\left\{\frac{1}{n}\sum_{i=1}^{n}\inf_{\theta\in\Theta_j,\,\nu\in N_j} q_n(X_i,\theta,\nu) < \mathbb{E}\left[\inf_{\theta\in\Theta_j,\,\nu\in N_j} q_n(X_i,\theta,\nu)\right] - n^{-\zeta}\eta_j\right\} \\ &\geq P\left\{\frac{1}{n}\sum_{i=1}^{n}\inf_{\theta\in\Theta_j,\,\nu\in N_j} q_n(X_i,\theta,\nu) < n^{-\zeta}\eta_j\right\}, \end{aligned}$$

for $j = 1, \ldots, J$. Moreover, there is an $\bar{n}_{J+1}$ such that

$$\frac{\varepsilon}{2(J+1)} \geq P\left\{\frac{1}{n}\sum_{i=1}^{n}\inf_{\theta\in\Theta,\,\|\nu\|\geq 2K} q_n(X_i,\theta,\nu) < n^{-\zeta}K\right\} \tag{A.8}$$

Now let letting $\eta = \min\{\eta_1, \ldots, \eta_J, K\}$ and $\bar{n} = \max_{j=1,\ldots,J+1} \bar{n}_j$, we have for $n \geq \bar{n}$

$$\begin{aligned} &P\left\{\min_{\theta\in\Theta_0^c,\,\nu\in N_0^c}\frac{1}{n}\sum_{i=1}^{n} q_n(X_i,\theta,\nu) < n^{-\zeta}\eta\right\} \\ &\leq P\left\{\min\left\{\min_{j=1,\ldots,J}\left\{\inf_{\theta\in\Theta_j,\,\nu\in N_j}\frac{1}{n}\sum_{i=1}^{n} q_n(X_i,\theta,\nu)\right\},\ \inf_{\theta\in\Theta,\,\|\nu\|\geq 2K}\frac{1}{n}\sum_{i=1}^{n} q_n(X_i,\theta,\nu)\right\} < n^{-\zeta}\eta\right\} \\ &\leq \sum_{j=1}^{J} P\left\{\inf_{\theta\in\Theta_j,\,\nu\in N_j}\frac{1}{n}\sum_{i=1}^{n} q_n(X_i,\theta,\nu) < n^{-\zeta}\eta_j\right\} + P\left\{\inf_{\theta\in\Theta,\,\|\nu\|>2K}\frac{1}{n}\sum_{i=1}^{n} q_n(X_i,\theta,\nu) < n^{-\zeta}\eta_j\right\} \\ &\leq \frac{\varepsilon}{2}. \end{aligned}$$

Notice that $n^{-\zeta}u(\theta,\nu) \in \Lambda_n^{\zeta} \subseteq \hat{\Lambda}_n(\theta)$ for all $\theta \in \Theta$ w.p.a. 1 by Lemma A.1 . Then, by Lemma A.4 and by the definition of $\hat{\lambda}_n(\theta)$,

$$\begin{aligned} &\min_{\theta\in\Theta_0^c,\,\nu\in N_0^c}\frac{1}{n}\sum_{i=1}^{n} q_n(X_i,\theta,\nu) \\ &= \min_{\theta\in\Theta_0^c,\,\nu\in N_0^c}\left[\frac{1}{n}\sum_{i=1}^{n}\ln\left(1+n^{-\zeta}u'(\theta,\nu)\,g(X_i,\theta)\right)\right] - n^{-\zeta}\nu' M u(\theta,\nu) + o_p\left(n^{-\zeta}\right) \\ &\leq \min_{\theta\in\Theta_0^c,\,\nu\in N_0^c}\left[\frac{1}{n}\sum_{i=1}^{n}\ln\left(1+\hat{\lambda}_n'(\theta,\nu)\,g(X_i,\theta)\right)\right] - \nu' M\lambda(\theta,\nu) + o_p\left(n^{-\zeta}\right). \end{aligned}$$

Therefore,

$$P\left\{\min_{\theta\in\Theta_0^c,\nu\in N_0^c}\bar{G}_n^*(\theta,\nu) < n^{-\zeta}\eta\right\} \leq \frac{\varepsilon}{2},$$

as required for (ii).

Since $\hat{\theta}_n \xrightarrow{p} \theta_0$ and $\hat{\nu}_n \xrightarrow{p} \nu_0$ we can deduce from Lemma A.2 that $\hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{p} 0$. ■

[5] Notice that

$$I\!\!E\; n^{\zeta}\inf_{\theta\in\Theta_j,\,\nu\in N_j} q_n(X,\theta,\nu) \;\leq I\!\!E\;\sup_{\theta\in\Theta}\|g(X,\theta)\| \;+2K + n^{-\zeta}\frac{I\!\!E\;\sup_{\theta\in\Theta}\|g(X,\theta)\|^{2}}{1-n^{-\xi}} < \infty. \tag{A.7}$$

A.2.2 Technical Lemmas

Lemma A.1 *Suppose that Assumptions 1 to 5 are satisfied. Then,*

(i) $\sup_{\theta\in\Theta,\lambda\in\Lambda_n^\zeta,1\le i\le n} \left|\lambda' g\left(X_i,\theta\right)\right| \longrightarrow_p 0,$

(ii) $\Lambda_n^\zeta \subseteq \bigcap_{\theta\in\Theta}\hat{\Lambda}_n(\theta)$ *w.p.a. 1.*

l_suplg

Proof of Lemma A.1: See proof of Lemma A1 in Newey and Smith (2004). ■

l_l1

Lemma A.2 *Suppose that Assumptions 1 to 5 are satisfied. Let* $\bar\theta\in\Theta$ *and* $\bar\nu\ge 0$ *be sequences such that* $\bar\theta \xrightarrow{p} \theta_0$, *and* $\bar\nu \xrightarrow{p} \nu_0$, *where* $\nu_0 = \mathbb{E}[g_2(X_i,\theta_0)]$. *Moreover,* $\frac{1}{\sqrt{n}}\sum_{i=1}^n g_1(X_i,\bar\theta) = O_p(1)$ *and* $\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(g_2(X_i,\bar\theta)-\bar\nu\right) = O_p(1)$. *Then,*

(i) $\hat\lambda(\bar\theta,\bar\nu)$ *exists w.p.a. 1,*

(ii) $\hat\lambda(\bar\theta,\bar\nu) = O_p(n^{-1/2})$,

(iii) $G_n^*\left(\bar\theta,\bar\nu,\hat\lambda(\bar\theta,\bar\nu)\right) \le O_p\left(\frac{1}{n}\right)$.

Proof of Lemma A.2: (i) Define

$$\tilde\lambda(\bar\theta,\bar\nu) = \arg\max_{\lambda\in\Lambda_n^\zeta} G_n^*(\bar\theta,\bar\nu,\lambda)$$

Since Λ_n^ζ is compact and $\ln\left(1+\lambda' g(X_i,\bar\theta)\right) - \bar\nu' M\lambda$ is continuous and strictly concave in λ the optimal solution $\tilde\lambda(\bar\theta,\bar\nu)$ exists and is unique. Statement (i) then follows from Lemma A.1.

(ii) and (iii) Write $\bar g_i = g(X_i,\bar\theta)$. For some constant C

$$\begin{aligned} 0 = G_n^*(\bar\theta,\bar\nu,0) &\le G_n^*(\bar\theta,\bar\nu,\tilde\lambda(\bar\theta,\bar\nu)) \\ &= \frac{1}{n}\sum_{i=1}^n \ln\left(1+\tilde\lambda(\bar\theta,\bar\nu)'\bar g_i\right) - \bar\nu' M\tilde\lambda(\bar\theta,\bar\nu) \\ &= \tilde\lambda(\bar\theta,\bar\nu)'\left(\frac{1}{n}\sum_{i=1}^n \bar g_i - M'\bar\nu\right) - \frac{1}{2}\tilde\lambda(\bar\theta,\bar\nu)'\left(\frac{1}{n}\sum_{i=1}^n \frac{\bar g_i\bar g_i'}{(1+\lambda_*'\bar g_i)^2}\right)\tilde\lambda(\bar\theta,\bar\nu) \\ &\le \tilde\lambda(\bar\theta,\bar\nu)'\left(\frac{1}{n}\sum_{i=1}^n \bar g_i - M'\bar\nu\right) - \frac{C}{4}\tilde\lambda(\bar\theta,\bar\nu)'\tilde\lambda(\bar\theta,\bar\nu), \end{aligned} \tag{A.9}$$

where λ_* lies on the line joining $\tilde\lambda(\bar\theta,\bar\nu)$ and 0. The last inequality holds because

$$\max_{1\le i\le n} |\lambda_*'\bar g_i| = o_p(1)$$

according to Lemma A.1 and $\frac{1}{n}\sum_{i=1}^n \bar g_i\bar g_i'$ converges in probability to J, a positive definite matrix, by the ULLN. The remainder of the proof follows the proof of Lemma A2 in Newey and Smith (2004). ■

Lemma A.3 *Suppose that Z_i is a sequence of iid random variables such that $\mathbb{E}|Z_i|^\alpha < \infty$. Then,* $\max_{1\le i\le n}|Z_i| = O_p\left(n^{1/\alpha}\right)$. l_maxz

Proof of Lemma A.3: The result follows from

$$\max_{1\le i\le n}|Z_i| = \left[\max_{1\le i\le n}|Z_i|^\alpha\right]^{1/\alpha} \le n^{1/\alpha}\left[\frac{1}{n}\sum_{i=1}^n |Z_i|^\alpha\right]^{1/\alpha} = O_p\left(n^{1/\alpha}\right). \blacksquare$$

Lemma A.4 *Let $g_n(x,\theta) = \mathcal{I}\{x\in\mathcal{X}\}g(x,\theta)$ where*

$$\mathcal{X}_n = \left\{x : \sup_{\theta\in\Theta}\|g(x,\theta)\| \le n^{\zeta-\xi}\right\}.$$

Define

$$\begin{aligned} q_n(X_i,\theta,\nu) &= \ln\left[1+n^{-\zeta}u'(\theta,\nu)g_n(X_i,\theta)\right] - n^{-\zeta}u(\theta,\nu)'M\nu \\ \tilde{q}_n(X_i,\theta,\nu) &= \ln\left[1+n^{-\zeta}u'(\theta,\nu)g(X_i,\theta)\right] - n^{-\zeta}u(\theta,\nu)'M\nu \end{aligned}$$

and assume that $\|u(\theta,\nu)\|\le 1$. Then,

$$\sup_{\theta\in\Theta,\ \nu\ge 0}\left|\frac{1}{n}\sum_{i=1}^n\left\{q_n(X_i,\theta,\nu)-\tilde{q}_n(X_i,\theta,\nu)\right\}\right| = o_p\left(n^{-\zeta}\right).$$

l_supq

Proof of Lemma A.4: By the mean value theorem, eq_qqdiff

$$\begin{aligned} &\sup_{\theta\in\Theta,\ \nu\ge 0}\left|\frac{1}{n}\sum_{i=1}^n\left\{q_n(X_i,\theta,\nu)-\tilde{q}_n(X_i,\theta,\nu)\right\}\right| \\ &= \sup_{\theta\in\Theta,\ \nu\ge 0}\left|\frac{1}{n}\sum_{i=1}^n\left(\frac{n^{-\zeta}u'(\theta,\nu)g(X_i,\theta)}{1+n^{-\zeta}u_*'(\theta,\nu)g(X_i,\theta)}\right)\mathcal{I}\{X_i\notin\mathcal{X}_n\}\right| \qquad \text{(A.10)} \\ &\le \max_{1\le i\le n}\sup_{\theta\in\Theta,\ \nu\ge 0}\left|\frac{n^{-\zeta}u'(\theta,\nu)g(X_i,\theta)}{1+n^{-\zeta}u_*'(\theta,\nu)g(X_i,\theta)}\right|\frac{1}{n}\sum_{i=1}^n\mathcal{I}\left\{\sup_{\theta\in\Theta}\|g(X_i,\theta)\| > n^{\zeta-\xi}\right\} \\ &\le \frac{1}{n^{\zeta-\xi}}\left(\max_{1\le i\le n}\sup_{\theta\in\Theta,\ \nu\ge 0}\left|\frac{n^{-\zeta}u'(\theta,\nu)g(X_i,\theta)}{1+n^{-\zeta}u_*'(\theta,\nu)g(X_i,\theta)}\right|\right)\left(\frac{1}{n}\sum_{i=1}^n\sup_{\theta\in\Theta}\|g(X_i,\theta)\|\right) \end{aligned}$$

where $u_*(\theta,\nu)$ is located between 0 and $u(\theta,\nu)$. The second term on the right-hand side of (A.10) can be bounded as follows. According to Lemma A.3

$$n^{-\zeta}\max_{1\le i\le n}\sup_{\theta\in\Theta}\|g(X_i,\theta)\| = n^{-\zeta+1/a}O_p(1).$$

Moreover, $\|u(\theta,\nu)\|\le 1$. Therefore,

$$\begin{aligned} \max_{1\le i\le n}\sup_{\theta\in\Theta,\ \nu\ge 0}\left|\frac{n^{-\zeta}u'(\theta,\nu)g(X_i,\theta)}{1+n^{-\zeta}u^{*\prime}(\theta,\nu)g(X_i,\theta)}\right| &\le \frac{2n^{-\zeta}\max_{1\le i\le n}\sup_{\theta\in\Theta}\|g_k(X_i,\theta)\|}{1-2n^{-\zeta}\max_{1\le i\le n}\sup_{\theta\in\Theta}\|g(X_i,\theta)\|} \\ &= \frac{n^{-\zeta+1/a}O_p(1)}{1-n^{-\zeta+1/a}O_p(1)} = n^{-\zeta+1/a}O_p(1). \end{aligned}$$

By Assumption 5 and the ULLN the third term on the right-hand side of (A.10) is $O_p(1)$. Since $\xi < \zeta - \frac{1}{\alpha}$, we are able to deduce that

$$n^{\zeta} \sup_{\theta\in\Theta,\ \nu\geq 0} \left| \frac{1}{n}\sum_{i=1}^{n} \left\{ q_n\left(X_i,\theta,\nu\right) - \tilde{q}_n\left(X_i,\theta,\nu\right)\right\} \right| = n^{1/a-\zeta+\xi} O_p\left(1\right) = o_p\left(1\right),$$

as required. ■

A.3 Quadratic Approximation of the Objective Function

We begin by deriving the coefficient matrices for the quadratic approximation of the objective function (19). A direct calculation shows that eq_g1beta

$$G_n^{*(1)}\left(\beta\right) = \left[G_n^{*(1)}\left(\beta\right)'_{\theta}, G_n^{*(1)}\left(\beta\right)'_{\nu}, G_n^{*(1)}\left(\beta\right)'_{\lambda} \right]', \tag{A.11}$$

where

$$\begin{aligned} G_n^{*(1)}\left(\beta\right)_{\theta} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{g^{(1)}\left(X_i,\theta\right)\lambda}{1+\lambda' g(X_i,\theta)} \right), \\ G_n^{*(1)}\left(\beta\right)_{\nu} &= -M\lambda, \\ G_n^{*(1)}\left(\beta\right)_{\lambda} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{g(X_i,\theta)}{1+\lambda' g(X_i,\theta)} \right) - M'v. \end{aligned}$$

At β_0 the first derivatives simplify to

$$G_n^{*(1)}\left(\beta_0\right)_{\theta} = 0, \quad G_n^{*(1)}\left(\beta_0\right)_{\nu} = 0, \quad G_n^{*(1)}\left(\beta_0\right)_{\lambda} = \frac{1}{n}\sum g(X_i,\theta_0) - M'\nu_0 = n^{-1/2} Z_n,$$

which leads to the formula for $G_n^{*(1)}(\beta_0)$ that appears in Equation (21) of the main text. We proceed by partitioning the matrix of second derivative as follows eq_g2beta

$$G_n^{*(2)}\left(\beta\right) = \begin{pmatrix} G_n^{*(2)}\left(\beta\right)_{\theta\theta'} & G_n^{*(2)}\left(\beta\right)_{\theta\nu'} & G_n^{*(2)}\left(\beta\right)_{\theta\lambda'} \\ G_n^{*(2)}\left(\beta\right)_{\nu\theta'} & G_n^{*(2)}\left(\beta\right)_{\nu\nu'} & G_n^{*(2)}\left(\beta\right)_{\nu\lambda'} \\ G_n^{*(2)}\left(\beta\right)_{\lambda\theta'} & G_n^{*(2)}\left(\beta\right)_{\lambda\nu'} & G_n^{*(2)}\left(\beta\right)_{\lambda\lambda'} \end{pmatrix}, \tag{A.12}$$

where

$$\begin{aligned} G_n^{*(2)}(\beta)_{\theta\theta'} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sum_{j=1}^{h}\lambda_j g_j^{(2)}(X_i,\theta)}{1+\lambda' g(X_i,\theta)} - \frac{g^{(1)}\left(X_i,\theta\right)\lambda\lambda' g^{(1)}\left(X_i,\theta\right)'}{(1+\lambda' g(X_i,\theta))^2} \right), \\ G_n^{*(2)}\left(\beta\right)_{\theta\nu'} &= 0, \quad G_n^{*(2)}\left(\beta\right)_{\nu\nu'} = 0, \quad G_n^{*(2)}\left(\beta\right)_{\lambda\nu'} = -M', \\ G_n^{*(2)}\left(\beta\right)_{\lambda\theta'} &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{g^{(1)}\left(X_i,\theta\right)'}{1+\lambda' g(X_i,\theta)} - \frac{g\left(X_i,\theta\right)\lambda' g^{(1)}(X_i,\theta)'}{(1+\lambda' g(X_i,\theta))^2} \right), \\ G_n^{*(2)}\left(\beta\right)_{\lambda\lambda'} &= -\frac{1}{n}\sum_{i=1}^{n}\frac{g(X_i,\theta)g(X_i,\theta)'}{(1+\lambda' g(X_i,\theta))^2}. \end{aligned}$$

At β_0 the second derivatives simplify to

$$\begin{aligned} G_n^{*(2)}(\beta_0)_{\theta\theta'} &= 0, \quad G_n^{*(2)}(\beta_0)_{\theta\lambda'} = \frac{1}{n}\sum_{i=1}^{n} g^{(1)}(X_i,\theta) = Q_n, \\ G_n^{*(2)}(\beta_0)_{\lambda\lambda'} &= -\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta)g(X_i,\theta)' = -J_n, \end{aligned}$$

which leads to the formula for $G_n^{*(2)}(\beta_0)$ that appears in Equation (21) of the main text.

In addition to the estimators $\hat{b}$ and $\tilde{b}_q$ defined in the main text, we will introduce a third estimator, $\hat{b}_q$, based on the quadratic approximation $\mathcal{G}^*_{nq}(\phi, l)$ subject to the restriction that $\hat{b}_q \in B_n$. Formally,

$$\hat{l}_q(\phi) = \text{argmax}_{l\in L_n(\phi)}\, \mathcal{G}^*_{nq}(\phi, l), \quad \hat{\phi}_q = \text{argmin}_{\phi\in\Phi_n}\, \mathcal{G}^*_{nq}(\phi, \hat{l}_q(\phi)).$$

A.3.1 Main Results

Proof of Lemma 2: By Lemma 1(a) of Andrews (1999), it is sufficient to prove

$$\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n} \left\| G_n^{*(2)}(\beta) - G_n^{*(2)}(\beta_0) \right\| = o_p(1),$$

for every sequence $\gamma_n \longrightarrow 0$. $G_n^{*(2)}$ is defined in (A.12). To verify this sufficient condition we will subsequently show that

(i) $\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n} \left\| G_n^{*(2)}(\beta)_{\theta\theta'} - G_n^{*(2)}(\beta_0)_{\theta\theta'} \right\| = o_p(1)$,

(ii) $\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n} \left\| G_n^{*(2)}(\beta)_{\lambda\theta'} - G_n^{*(2)}(\beta_0)_{\lambda\theta'} \right\| = o_p(1)$,

(iii) $\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n} \left\| G_n^{*(2)}(\beta)_{\lambda\lambda'} - G_n^{*(2)}(\beta_0)_{\lambda\lambda'} \right\| = o_p(1)$.

We begin by showing that

$$\sup_{\beta\in\mathcal{B}_n} \frac{1}{|1+\lambda' g(X_i,\theta)|} = O_p(1). \tag{A.13}$$

Since

$$\sup_{\beta\in\mathcal{B}_n,\, 1\le i\le n} |\lambda' g(X_i,\theta)| = o_p(1)$$

it follows that for any given $0<\delta<\frac{1}{2}$

$$P\left\{ \sup_{\beta\in\mathcal{B}_n, 1\le i\le n} |\lambda\prime g(X_i,\theta)| > \delta \right\} \longrightarrow 0.$$

Set $K > \frac{1}{\delta} > 2$. Then,

$$\begin{aligned} P\left\{\sup_{\beta\in\mathcal{B}_n,1\le i\le n}\frac{1}{|1+\lambda' g(X_i,\theta)|}>K\right\} &\le P\left\{\sup_{\beta\in\mathcal{B}_n,1\le i\le n}|1+\lambda' g(X_i,\theta)|<\frac{1}{M}\right\} \\ &\le P\left\{\sup_{\beta\in\mathcal{B}_n,1\le i\le n}|\lambda' g(X_i,\theta)|>\delta\right\}\longrightarrow 0, \end{aligned}$$

which proves (A.13).

(i) Notice that

$$\begin{aligned} &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lambda_j g_j^{(2)}(X_i,\theta)}{1+\lambda' g(X_i,\theta)}\right)\right\| \\ \le\ & \sup_{\lambda\in\Lambda_n^{\zeta}}|\lambda_j|\left(\sup_{\beta\in\mathcal{B}_n,1\le i\le n}\frac{1}{|1+\lambda' g(X_i,\theta)|}\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum\left\|g_j^{(2)}(X_i,\theta)\right\|\right) \\ =\ & O(n^{-\zeta})O_p(1)O_p(1)=o_p(1), \end{aligned}$$

where the last inequality holds by the definition of Λ_n^{ζ}, (A.13) and the ULLN under Assumption 6. Moreover,

$$\begin{aligned} &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g^{(1)}(X_i,\theta)'\lambda\lambda' g^{(1)}(X_i,\theta)}{(1+\lambda' g(X_i,\theta))^2}\right)\right\| \\ \le\ & \sup_{\lambda\in\Lambda_n^{\zeta}}\|\lambda_k\|^2\left(\sup_{\beta\in\mathcal{B}_n,\,1\le i\le n}\frac{1}{(1+\lambda' g(X_i,\theta))^2}\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\left\|g^{(1)}(X_i,\theta)\right\|\right) \\ =\ & O\left(n^{-2\zeta}\right)O_p(1)\,O_p(1)=o_p(1). \end{aligned}$$

The last inequality holds by the definition of Λ_n^{ζ}, (A.13) and the ULLN under Assumption 6.

(ii) Apply the triangle inequality to

$$\begin{aligned} &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g^{(1)}(X_i,\theta)}{1+\lambda' g(X_i,\theta)}-g^{(1)}(X_i,\theta_0)\right)\right\| \\ \le\ & \sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g^{(1)}(X_i,\theta)}{1+\lambda' g(X_i,\theta)}-g^{(1)}(X_i,\theta)\right)\right\| \\ & +\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(g^{(1)}(X_i,\theta)-\mathbb{E}\left[g^{(1)}(X_i,\theta)\right]\right)\right\| \\ & +\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_n}\left\|\mathbb{E}\left[g^{(1)}(X_i,\theta)\right]-E\left[g^{(1)}(X_i,\theta_0)\right]\right\| \\ & +\left\|\frac{1}{n}\sum_{i=1}^{n}\left\{g^{(1)}(X_i,\theta_0)-\mathbb{E}\left[g^{(1)}(X_i,\theta_0)\right]\right\}\right\| \\ =\ & I_d+o_p(1)+o_p(1)+o_p(1), \end{aligned}$$

where the last equality holds by the ULLN under Assumption 6, the uniform continuity of $\mathbb{E}\left[g^{(1)}\left(X_i,\theta\right)\right]$ in θ, and the WLLN. Next,

$$\begin{aligned} I_d &\leq \sup_{\beta\in\mathcal{B}_n}\left|\lambda' g\left(X_i,\theta\right)\right|\left(\sup_{\beta\in\mathcal{B}_n}\frac{1}{\left|1+\lambda' g\left(X_i,\theta\right)\right|}\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\left\|g^{(1)}\left(X_i,\theta\right)\right\|\right) \\ &= o_p(1)\,O_p(1)\,O_p(1)=o_p(1) \end{aligned}$$

by Lemma A.1, (A.13), and the ULLN under Assumption 6. Moreover,

$$\begin{aligned} &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\leq\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\frac{g\left(X_i,\theta\right)}{1+\lambda' g\left(X_i,\theta\right)}\frac{\lambda\prime g\left(X_i,\theta\right)}{1+\lambda' g\left(X_i,\theta\right)}\right\| \\ &\leq \sup_{\lambda\in\Lambda_n^{\zeta}}\|\lambda\|\left(\sup_{\beta\in\mathcal{B}_n,1\leq i\leq n}\frac{1}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum\left\|g\left(X_i,\theta\right)\right\|^2\right) \\ &= O\left(n^{-\zeta}\right)O_p(1)\,O_p(1)=o_p(1)\,. \end{aligned}$$

(iii) Similar as before, we have

$$\begin{aligned} &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\leq\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'\right)\right\| \\ \leq\; &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\leq\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right)\right\| \\ &+\sup_{\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X_i,\theta\right)g\left(X_i,\theta\right)'-\mathbb{E}\left[g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right]\right)\right\| \\ &+\sup_{\Theta}\left\|\mathbb{E}\left[g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right]-\mathbb{E}\left[g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'\right]\right\| \\ &+\sup_{\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'-\mathbb{E}\left[g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'\right]\right)\right\| \\ =\; &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\leq\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right)\right\|+o_p(1)\,. \end{aligned}$$

Next,

$$\begin{aligned} &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_0\|\leq\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right)\right\| \\ &\leq \sup_{\beta\in\mathcal{B}_n,1\leq i\leq n}\left|\lambda' g\left(X_i,\theta\right)\right|\left(\sup_{\beta\in\mathcal{B}_n,1\leq i\leq n}\frac{1}{\left|1+\lambda' g\left(X_i,\theta\right)\right|}\right) \\ &\quad\times\left(\sup_{\beta\in\mathcal{B}_n,1\leq i\leq n}\frac{1}{\left|1+\lambda' g\left(X_i,\theta\right)\right|}+1\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\left\|g\left(X_i,\theta\right)\right\|^2\right) \\ &= o_p(1)\,O_p(1)\,O_p(1)\,O_p(1)=o_p(1)\,.\quad\blacksquare \end{aligned}$$

Proof of Theorem 2 is omitted. ■

Proof of Theorem 3: (i) Follows from Lemma A.6.

(ii) According to Lemma A.2, $\hat{\lambda}(\hat{\theta},\hat{\nu}) = O_p(n^{-1/2})$. It remains to show that $\hat{\phi} = \sqrt{n}[(\hat{\theta} - \theta_0)', (\hat{\nu} - \nu_0)']'$ is stochastically bounded. The saddlepoint property implies that eq.ap.saddle

$$0 = \mathcal{G}_n^*(\hat{\phi}, 0) \leq \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) \leq \mathcal{G}_n^*(0, \hat{l}(0)). \tag{A.14}$$

Then using the quadratic approximation (18), the bound for the remainder term given in Lemma 2 and the definition of $\hat{l}$ and $\hat{\phi}$ we obtain eq.ap.saddle1

$$\begin{aligned} \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) &= \mathcal{G}_{nq}^*(\hat{\phi}, \hat{l}(\hat{\phi})) + (1 + \|\hat{\phi}\|^2 + \|\hat{l}(\hat{\phi})\|^2) o_p(1) \\ &= \frac{1}{2}(Z_n - R_n'\hat{\phi})' J_n^{-1} (Z_n - R_n'\hat{\phi}) \\ &\quad -\frac{1}{2}(\hat{l}(\hat{\phi}) - J_n^{-1}[Z_n - R_n'\hat{\phi}])' J_n (\hat{l}(\hat{\phi}) - J_n^{-1}[Z_n - R_n'\hat{\phi}]) \\ &\quad +(1 + \|\hat{\phi}\|^2 + \|\hat{l}(\hat{\phi})\|^2) o_p(1) \\ &= \frac{1}{2}(Z_n - R_n'\hat{\phi})' J_n^{-1} (Z_n - R_n'\hat{\phi}) + (1 + \|\hat{\phi}\|^2 + \|\hat{l}(\hat{\phi})\|^2) o_p(1). \end{aligned} \tag{A.15}$$

The last equality is a consequence of Lemma A.8. Similarly, we can deduce from Lemmas A.2, 2, and Theorem 2 that eq.ap.saddle2

$$\mathcal{G}_n^*(0, \hat{l}(0)) = -\frac{1}{2}\hat{l}\,(0)'\, J_n \hat{l}\,(0) + Z_n'\hat{l}\,(0) + (1 + \|\hat{l}(0)\|^2) o_p(1) = O_p\,(1)\,. \tag{A.16}$$

Hence, from (A.14), (A.15), and (A.16) we obtain the inequality eq.ap.doublebound

$$0 \leq \frac{1}{2}(Z_n + o_p(1) - R_n'\hat{\phi})' J_n^{-1} (Z_n + o_p(1) - R_n'\hat{\phi}) \leq O_p(1). \tag{A.17}$$

Notice that $Z_n + o_p(1) = O_p\,(1)$. According to Assumptions 4 and 6 R_n is full rank and J_n is positive definite w.p.a. 1. Therefore, (A.17) implies that $\hat{\phi}$ is stochastically bounded.

(iii) We deduce from Lemma 2 and Part (ii) that

$$\begin{aligned} nG_n^*(\hat{\beta}_n) &= \mathcal{G}_{nq}^*(\sqrt{n}(\hat{\beta}_n - \beta_0)) + (1 + \|\sqrt{n}(\hat{\beta}_n - \beta_0)\|^2) o_p(1) \\ &= nG_{nq}^*(\hat{\beta}_n) + O_p(1) o_p(1). \end{aligned}$$

(iv) We proceed by establishing $o_p(1)$ bounds for $nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq})$.

We begin with the upper bound. Using (iii) can rewrite the differential as eq_gndiff

$$\begin{aligned} nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq}) &= \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) + o_p(1) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) \\ &\leq \mathcal{G}_n^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \hat{l}_q(\tilde{\phi}_q)) + o_p(1). \end{aligned} \tag{A.18}$$

Replacing $\hat{\phi}$ by $\hat{\phi}_q$ raises $\mathcal{G}_n^*$, whereas substituting $\tilde{l}_q$ with $\hat{l}$ lowers $\mathcal{G}_{nq}^*$. Using Lemma 2 the first term on the right-hand side of (A.18) can be rewritten as eq.ap.ineq.Gnstar1

$$\begin{aligned}\mathcal{G}_n^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) &= \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p(1)\left(1 + \|\hat{\phi}_q\|^2 + \|\hat{l}(\hat{\phi}_q)\|^2\right) \\ &= \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p(1).\end{aligned} \tag{A.19}$$

The second equality in (A.19) is a consequence of Lemmas A.2 and A.6. According to Lemma A.8

$$\hat{l}(\bar{\phi}) = (J_n + o_p(1))^{-1}[Z_n - (R_n' + o_p(1))\bar{\phi}]$$

for $\bar{\phi} = O_p(1)$. Hence,

$$\hat{l}(\tilde{\phi}_q) - \hat{l}(\hat{\phi}_q) = (J_n + o_p(1))^{-1}[Z_n - (R_n' + o_p(1))](\tilde{\phi}_q - \hat{\phi}_q) = o_p(1)$$

by Lemma A.6. Since $\mathcal{G}_{nq}^*(\phi, l)$ is continuous in its arguments we can now express the second term on the right-hand side of (A.18) as eq.ap.ineq.Gnstar2

$$\mathcal{G}_{nq}^*(\tilde{\phi}_q, \hat{l}(\tilde{\phi}_q)) = \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p\left(1\right) \tag{A.20}$$

Plugging (A.19) and (A.20) into (A.18) we obtain the upper bound

$$nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq}) \leq o_p(1).$$

Using similar arguments, we can establish a lower bound as follows:

$$\begin{aligned}nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq}) &= \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) + o_p(1) \\ &\geq \mathcal{G}_n^*(\hat{\phi}, \hat{l}_q(\hat{\phi})) - \mathcal{G}_{nq}^*(\hat{\phi}, \tilde{l}_q(\hat{\phi})) + o_p(1) \\ &= \mathcal{G}_n^*(\hat{\phi}, \hat{l}_q(\hat{\phi})) - \mathcal{G}_{nq}^*(\hat{\phi}, \hat{l}_q(\hat{\phi})) + o_p(1) \\ &= o_p(1)\end{aligned}$$

which proves (iv). ■

(v) Follows from parts (iii) and (iv).

A.3.2 Technical Lemmas

l_exist.bqtilde

Lemma A.5 *Suppose Assumptions 1 to 6 are satisfied. Then, $\tilde{b}_q$ exists uniquely w.p.a. 1.*

Proof of Lemma A.5: The subsequent statements are true w.p.a. 1. Notice that $\bar{\mathcal{G}}_{nq}^*(\phi)$, defined in (27), is strictly convex function of ϕ because $R_n' = [-Q_n', M']$ is a full rank matrix under Assumption 6 and J_n^{-1} is positive definite under Assumption 4. Hence, $R_n' J_n^{-1} R_n$ is a positive definite matrix. Moreover, the domain Φ is convex. Therefore, $\tilde{\phi}_q$ is unique. Finally, from (26) we deduce that $\tilde{l}_q$ exists uniquely. ■

l_bqop1

Lemma A.6 *Suppose Assumptions 1 to 6 are satisfied. Then*

(i) $\tilde{b}_q = O_p(1)$,

(ii) $\hat{b}_q = \tilde{b}_q + o_p(1)$.

Proof of Lemma A.6: (i) We will show that $\tilde{\phi}_q = O_p(1)$. For notational simplicity, denote

$$A_{1n} = R_n' J_n^{-1} R_n,\ A_{2n} = A_{1n}^{-1} R_n' J_n^{-1} Z_n,\ \text{and } A_{3n} = Z_n' J_n^{-1} Z_n - A_{2n}' A_{1n} A_{2n},$$

and write the concentrated quadratic objective function (27) as

$$\bar{\mathcal{G}}^*_{nq}(\phi) = \frac{1}{2}(\phi + A_{2n})' A_{1n} (\phi + A_{2n}) + \frac{1}{2} A_{3n}.$$

Observe that J_n, R_n, and Z_n converge weakly according to Theorem 2. Moreover based on Assumptions 4 and 6 A_{1n} is positive definite w.p.a. 1. Let

$$\bar{\phi}_q = \text{argmin}_{\phi \in \mathbb{R}^{m+h_2}} \bar{\mathcal{G}}^*_{nq}(\phi) = -A_{2n} = O_p(1).$$

Notice that $\tilde{\phi}_q$ is the projection of $\bar{\phi}_q$ onto the set Φ with respect to the inner product $\langle x, y \rangle = x' A_{1n} y$. Then,

$$\|\tilde{\phi}_q\| \leq \lambda_{min}^{-1}(A_{1n}) \langle \tilde{\phi}_q, \tilde{\phi}_q \rangle^{1/2} \leq \lambda_{min}^{-1}(A_{1n}) \langle \bar{\phi}_q, \bar{\phi}_q \rangle^{1/2} = O_p(1)$$

where $\lambda_{min}(A_{1n})$ denotes the smallest eigenvalue of A_{1n} and is strictly positive w.p.a. 1. Finally, from (26) we can deduce that $\tilde{l}(\tilde{\phi}) = O_p(1)$.

(ii) According to Lemma A.5 the saddlepoint problem $\min_{\phi \in \Phi} \max_{l \in \mathbb{R}^h} \mathcal{G}^*_{nq}(\phi, l)$ has a unique solution $\tilde{b}_q$ on the domain $B = \Phi \otimes \mathbb{R}^h$. Since $B_n \subset B$ for any $\epsilon > 0$

$$\begin{aligned} P\left\{\|\hat{b}_q - \tilde{b}_q\| > \epsilon\right\} &\leq P\left\{\tilde{b}_q \in B \backslash B_n\right\} \\ &\leq P\{\tilde{b}_q \in B \backslash (\Phi_n \otimes \sqrt{n}\Lambda_n^\zeta)\} + o(1), \end{aligned}$$

where the $o(1)$ term in the last line holds by Lemma A.1(ii). The set $\sqrt{n}\Lambda_n^\zeta$ consists of the elements in Λ_n^ζ multiplied by $\sqrt{n}$ and expands to $\mathbb{R}^h$ because $\zeta < 1/2$. Since the true parameter θ_0 is in the interior of Θ, Φ_n expands to $\mathbb{R}^{m+h_2}$ if $\nu_0 > 0$ and $\mathbb{R}^m \otimes \mathbb{R}^{h_2+}$ otherwise. Since $\tilde{b}_q = O_p(1)$, we deduce $P\{\tilde{b}_q \in B \backslash (\Phi_n \otimes \sqrt{n}\Lambda_n^\zeta)\} = o(1)$. Therefore $\hat{b}_q = \tilde{b}_q + o_p(1)$, as required. ■

*** Note: we have to cite the following lemma. The subsequent lemma is written in terms of $\hat{\nu}$ rather than a general sequence $\bar{\nu}$.

l_gop1

Lemma A.7 *Suppose Assumptions 1 to 6 are satisfied. Then,*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g\left(X_i,\hat{\theta}\right)-M'\hat{v}\right]=O_p\left(1\right).$$

Proof of Lemma A.7: Let $\hat{g}_i = g\left(X_i,\hat{\theta}\right) - M'\hat{\nu}$ and $\hat{g} = \frac{1}{n}\sum_{i=1}^{n}\left[g\left(X_i,\hat{\theta}\right)-M'\hat{\nu}\right]$. Define $\hat{u}\left(\hat{\theta},\hat{\nu}\right)=n^{-\zeta}\frac{\hat{g}}{\|\hat{g}\|}$. (Recall the definition of $u\left(\theta,\nu\right)$ in the proof of consistency.)

Approximation $G_n^*\left(\theta,\nu,\lambda\right)$ with respect to λ around $\lambda=0$ at $\left(\theta,\nu,\lambda\right)=\left(\hat{\theta},\hat{\nu},\hat{u}\left(\hat{\theta},\hat{\nu}\right)\right)$. Then,

$$\begin{aligned}
& G_n^*\left(\hat{\theta},\hat{\nu},\hat{u}\left(\hat{\theta},\hat{\nu}\right)\right) \\
= \;& G_n^*\left(\hat{\theta},\hat{\nu},0\right)+\frac{\partial G_n^*\left(\hat{\theta},\hat{\nu},0\right)}{\partial\lambda'}\hat{u}\left(\hat{\theta},\hat{\nu}\right)+\frac{1}{2}\hat{u}\left(\hat{\theta},\hat{\nu}\right)'\frac{\partial^2 G_n^*\left(\hat{\theta},\hat{\nu},\ddot{\lambda}\right)}{\partial\lambda\partial\lambda'}\hat{u}\left(\hat{\theta},\hat{\nu}\right) \\
= \;& \hat{g}'\hat{u}\left(\hat{\theta},\hat{\nu}\right)-\frac{1}{2}\hat{u}\left(\hat{\theta},\hat{\nu}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}_i\hat{g}_i'}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right),
\end{aligned}$$

where $\ddot{\lambda}$ is located between 0 and $\hat{u}\left(\hat{\theta},\hat{\nu}\right)$.

Notice that $\max_{1\leq i\leq n}\left|\hat{u}\left(\hat{\theta},\hat{\nu}\right)'\hat{g}_i\right|\rightarrow_p 0$ and $\hat{u}\left(\hat{\theta},\hat{\nu}\right)\in\hat{\Lambda}_n\left(\hat{\theta}\right)$ by Lemma A.1 w.p.a.1. Also, $\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'\leq\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}\|g\left(X_i,\theta\right)\|\right)I\rightarrow_p CI$. Then,

$$\begin{aligned}
& \hat{g}'\hat{u}\left(\hat{\theta},\hat{\nu}\right)-\frac{1}{2}\hat{u}\left(\hat{\theta},\hat{\nu}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}_i\hat{g}_i'}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right) \\
= \;& n^{-\zeta}\|\hat{g}\|-\frac{1}{2}\hat{u}\left(\hat{\theta},\hat{\nu}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}_i\hat{g}_i'}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right) \\
\geq \;& n^{-\zeta}\|\hat{g}\|-\frac{1}{2}\max_{1\leq i\leq n}\left(\frac{1}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right) \\
\geq \;& n^{-\zeta}\|\hat{g}\|-Cn^{-2\zeta}. \qquad \text{(A.21)}
\end{aligned}$$

Then,

$$n^{-\zeta}\|\hat{g}\|-Cn^{-2\zeta}\leq G_n^*\left(\hat{\theta},\hat{\nu},\hat{u}\left(\hat{\theta},\hat{\nu}\right)\right)\leq G_n^*\left(\hat{\theta},\hat{\nu},\hat{\lambda}\right)\leq\sup_{\lambda\in\hat{\Lambda}_n\left(\theta_0\right)}G_n^*\left(\theta_0,\nu_0,\lambda\right)\leq O_p\left(\frac{1}{n}\right), \qquad \text{(A.22)}$$

where the first inequality is from (A.21), the second and third inequalities hold because $\left(\hat{\theta},\hat{\nu},\hat{\lambda}\right)$ is a saddle point, and the last inequality is from Lemma A.2 with $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[g\left(X_i,\theta_0\right)-M'\nu_0\right]=$

$O_p(1)$ by the CLT. Also, by $\zeta < \frac{1}{2}$, $\zeta - 1 < -\frac{1}{2} < -\zeta$. Solving (A.22) for $\|\hat{g}\|$ gives

$$\|\hat{g}\| \leq O_p\left(n^{-\zeta}\right). \tag{A.23}$$

Now, for a given sequence $\varepsilon_n \to 0$, let $\bar{\lambda} = \varepsilon_n \hat{g}$. By (A.23), $\bar{\lambda} = o_p\left(n^{-\zeta}\right)$, and so $\bar{\lambda} \in \Lambda_n^\zeta$ w.p.a.1. Then, as in (A.22), we have

$$\bar{\lambda}'\hat{g} - C\left\|\bar{\lambda}\right\| = \varepsilon_n \|\hat{g}\|^2 - C\varepsilon_n^2 \|\hat{g}\|^2 \leq O_p\left(\frac{1}{n}\right).$$

Since, for n large enough, $1 - C\varepsilon_n$ is bounded away from zero, it follows that $\varepsilon_n \|\hat{g}\|^2 = O_p\left(\frac{1}{n}\right)$. Since ε_n is an arbitrary sequence that tends to zero, we deduce that

$$\|\hat{g}\| = O_p\left(\frac{1}{\sqrt{n}}\right),$$

as required. ■

1.13

Lemma A.8 *Suppose that Assumptions 1 to 6 are satisfied. Let $\bar{\theta} \in \Theta$ and $\bar{\nu} \geq 0$ be sequences such that $\bar{\theta} \xrightarrow{p} \theta_0$ and $\bar{\nu} \xrightarrow{p} \nu_0$. Let $\hat{l} = \sqrt{n}\hat{\lambda}$, and $\bar{\phi} = [\bar{s}', \bar{u}']$, where $\bar{s} = \sqrt{n}(\bar{\theta} - \theta_0)$ and $\bar{u} = \sqrt{n}(\bar{\nu} - \nu_0)$. Then*

$$0 = Z_n - (R_n' + o_p(1))\bar{\phi} - (J_n + o_p(1))\hat{l}(\bar{\phi}).$$

Proof of Lemma A.8: In view of Lemmas A.1(ii) and A.2, we deduce that $\hat{\lambda}(\bar{\theta}, \bar{v})$ is in the interior of $\hat{\Lambda}(\bar{\theta})$ w.p.a. 1. Hence, $\hat{\lambda}$ satisfies the first-order conditions associated with $\max_{\lambda \in \hat{\Lambda}(\bar{\theta})} G_n^*\left(\bar{\theta}, \bar{\nu}, \lambda\right)$:

$$0 = \frac{1}{n}\sum_{i=1}^{n} \frac{g(X_i, \bar{\theta})}{1 + \hat{\lambda}' g(X_i, \bar{\theta})} - M'\bar{\nu}.$$

Let $\beta = [\theta', \nu', \lambda]'$. We now apply the mean-value theorem and multiply by $\sqrt{n}$:

$$0 = \sqrt{n} G_n^{*(1)}(\beta_0)_\lambda + G_n^{*(2)}(\beta_*)_{\lambda\theta'}\bar{s} - M'\bar{\nu} + G_n^{*(2)}(\beta_*)_{\lambda\lambda'}\hat{l},$$

where β_* lies on the line joining β_0 and $\bar{\beta} = [\bar{\theta}', \bar{\nu}', \hat{\lambda}(\bar{\theta}, \bar{\nu})']'$. The matrices $G_n^{*(1)}(\beta)$ and $G_n^{*(2)}(\beta)$ and their partitions are defined in (A.11) and (A.12). Using the same arguments as in the proof of Lemma 2 and the definitions of J_n, Q_n, R_n, and Z_n in (21) we obtain the desired result. ■

A.4 Limit Distribution

Proof of Theorem 4: By the theorem of the maximum (e.g., see Berge, 1963) $\tilde{\phi}_q$ is a continuous function of Z_n, J_n, and R_n. Moreover, from direct inspection we know that $\tilde{l}_q$

is continuous in Z_n, J_n, R_n, and $\tilde{\phi}_n$. The statement of the theorem then follows from the continuous mapping theorem. ■

Proof of Theorem 5: According to Theorem 3(iii):

$$\mathcal{G}^*_{nq}(\hat{\phi}, \hat{l}(\hat{\phi})) = \mathcal{G}^*_{nq}(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) + o_p(1). \tag{A.24}$$

Since $\hat{\phi} = O_p(1)$ we can deduce from Lemma A.8 that

$$\hat{l}(\hat{\phi}) = \tilde{l}_q(\hat{\phi}) + o_p(1). \tag{A.25}$$

and

$$\mathcal{G}^*_{nq}(\hat{\phi}, \hat{l}(\hat{\phi})) = \mathcal{G}^*_{nq}(\hat{\phi}, \tilde{l}_q(\hat{\phi})) + o_p(1). \tag{A.26}$$

Let $\bar{\mathcal{G}}^*_{nq}(\phi) = \mathcal{G}^*_{nq}(\phi, \tilde{l}_q(\phi))$. Combining (A.24) and (A.26) then yields

$$\bar{\mathcal{G}}^*_{nq}(\hat{\phi}) = \bar{\mathcal{G}}^*_{nq}(\tilde{\phi}_q) + o_p(1). \tag{A.27}$$

Since $\bar{\mathcal{G}}^*_{nq}(\phi)$ is a strictly convex quadratic function of ϕ and $\tilde{\phi}_q$ uniquely minimizes $\bar{\mathcal{G}}^*_{nq}(\phi)$ over a convex domain Φ, we deduce from (A.27) that

$$\hat{\phi} = \tilde{\phi}_q + o_p\left(1\right).$$

Using (A.25) once more we conclude that

$$\hat{l}(\hat{\phi}) = \tilde{l}_q(\hat{\phi}) + o_p(1) = \tilde{l}_q(\tilde{\phi}_q) + o_p(1)$$

which completes the proof. ■

A.5 Testing

Proof of Corollary 1: omitted. ■

Proof of Theorem 6: The proof consists of two parts. First, we will show that the test statistics $\mathcal{LR}_n$, $\mathcal{H}_n$, $\mathcal{W}_n$, $\mathcal{D}_n$ have the same limit distribution. Second, we demonstrate that the limit distribution has the conjectured form.

(i) We begin by characterizing the limit distributions $\mathcal{P}$ and $\mathcal{P}^0$. The concentrated limit objective function is of the form

$$\begin{aligned}
\bar{\mathcal{G}}_q(\phi) &= \frac{1}{2}(Z - R'\phi)'J^{-1}(Z - R'\phi) \\
&= \frac{1}{2}[\phi - (RJ^{-1}R')^{-1}RJ^{-1}Z]'RJ^{-1}R'[\phi - (RJ^{-1}R')^{-1}RJ^{-1}Z] + g(J, R, Z),
\end{aligned}$$

where the function $g(J, R, Z)$ does not depend on ϕ. Define the matrix partitions

$$(RJ^{-1}R')^{-1}RJ^{-1}Z = \begin{bmatrix} Z_s \\ Z_u \end{bmatrix} = \begin{bmatrix} QJ^{-1}Q' & -QJ^{-1}M' \\ -MJ^{-1}Q' & MJ^{-1}M' \end{bmatrix}^{-1} \begin{bmatrix} -QJ^{-1}Z \\ MJ^{-1}Z \end{bmatrix}$$

and

$$\Omega = J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}.$$

Using the formula for the inverse of a partitioned matrix it can be verified that eq_zu

$$Z_u = (M\Omega M')^{-1}M\Omega Z. \tag{A.28}$$

We can express $\bar{\mathcal{G}}_q(\phi) = \bar{\mathcal{G}}_q(s, u)$ as

$$\begin{aligned} \bar{\mathcal{G}}_q(s, u) &= \frac{1}{2}[(s - Z_s) - (QJ^{-1}Q')^{-1}(QJ^{-1}M')(u - Z_u)]' \\ &\quad \times QJ^{-1}Q'[(s - Z_s) - (QJ^{-1}Q')^{-1}(QJ^{-1}M')(u - Z_u)] \\ &\quad + \frac{1}{2}(u - Z_u)'M\Omega M'(u - Z_u) + g(J, R, Z). \end{aligned}$$

Decompose $\mathcal{P}^0 = [\mathcal{S}^{0'}, 0']'$ and $\mathcal{P} = [\mathcal{S}', \mathcal{U}']'$. We deduce that eq_s0su

$$\begin{aligned} \mathcal{S}^0 &= Z_s - (QJ^{-1}Q')^{-1}QJ^{-1}M'Z_u \qquad \text{(A.29)} \\ \mathcal{S} &= Z_s - (QJ^{-1}Q')^{-1}QJ^{-1}M'(Z_u - \mathcal{U}) \\ \mathcal{U} &= \operatorname*{argmin}_{u \in \mathbb{R}^{h_2+}} (u - Z_u)'M\Omega M'(u - Z_u). \end{aligned}$$

First, we show that the limit distributions of $\mathcal{W}_n$ and $\mathcal{D}_n$ are identical. Notice that $\hat{J}_n$ and $\hat{R}_n$ are consistent estimators of J and R. In view of Theorems 4 and 5 we conclude that

$$\mathcal{W}_n \Longrightarrow \mathcal{U}'M\Omega M'\mathcal{U}.$$

Now consider the score statistic. Since $\sqrt{n}(\hat{\beta}_n^0 - \beta_0) = O_p(1)$ we deduce (see proof of Lemma 2)

$$G_n^{*(2)}(\hat{\beta}_n^0)^{-1} \xrightarrow{p} \begin{bmatrix} 0 & 0 & Q \\ 0 & 0 & -M \\ Q' & -M' & -J \end{bmatrix}^{-1}.$$

Using the formula for the inverse of a partitioned matrix we obtain

$$[G_n^{*(2)}(\hat{\beta}_n^0)^{-1}]_{\nu\nu} \xrightarrow{p} M\Omega M'.$$

We deduce from Corollary 1 that

$$\sqrt{n}\hat{\lambda}_2^0(\hat{\theta}_n^0, 0) \Longrightarrow MJ^{-1}(Z - R'\mathcal{P}^0) = M\Omega Z.$$

Thus,

$$\begin{bmatrix} \sqrt{n}\hat{\lambda}^{+}_{n,2} \\ \left[G^{*(2)}_{n}(\hat{\beta}^{0}_{n})^{-1}\right]_{\nu\nu} \end{bmatrix} \Longrightarrow \begin{bmatrix} \mathcal{L}^{+}_{2} \\ M\Omega M' \end{bmatrix},$$

where $\mathcal{L}^{+}_{2}$ is defined as eq_l2plus

$$\mathcal{L}^{+}_{2} = \underset{\kappa\in\mathbb{R}^{h_2+}}{\operatorname{argmin}}\,(\kappa - Z_u)'(M\Omega M')^{-1}(\kappa - Z_u). \tag{A.30}$$

Thus, we can deduce from a comparison of (A.29) and (A.30) that $\mathcal{U}$ and $\mathcal{L}^{+}_{2}$ are distributionally equivalent. Therefore,

$$\mathcal{D}_n \Longrightarrow \mathcal{L}^{+'}_{2} M\Omega M'\mathcal{L}^{+}_{2} \equiv \mathcal{U}'M\Omega M'\mathcal{U},$$

that is, $\mathcal{D}_n$ and $\mathcal{W}_n$ have the same limit distribution.

Second, we will argue that $\mathcal{H}_n$ has the same limit distribution as $\mathcal{W}_n$. In view of Theorems 4 and 5 we conclude that

$$\mathcal{H}_n \Longrightarrow (\mathcal{P} - \mathcal{P}^0)'RJ^{-1}R'(\mathcal{P} - \mathcal{P}^0).$$

The limit distribution of $\mathcal{H}_n$ can be rewritten as follows

$$\begin{aligned} &(\mathcal{P} - \mathcal{P}^0)'RJ^{-1}R'(\mathcal{P} - \mathcal{P}^0) \\ &= \begin{bmatrix} (\mathcal{S} - \mathcal{S}^0)' & \mathcal{U}' \end{bmatrix}' \begin{bmatrix} QJ^{-1}Q' & -QJ^{-1}M' \\ -MJ^{-1}Q' & MJ^{-1}M' \end{bmatrix} \begin{bmatrix} (\mathcal{S} - \mathcal{S}^0) \\ \mathcal{U} \end{bmatrix} \\ &= \mathcal{U}'M\Omega M'\mathcal{U}, \end{aligned}$$

where $\mathcal{S} - \mathcal{S}^0 = (QJ^{-1}Q')^{-1}QJ^{-1}Q'\mathcal{U}$ according to (A.29).

Third, to show that the limit distributions of $\mathcal{LR}^{\nu}_{n}$ and $\mathcal{H}_n$ are identical, notice that according to Theorems 4 and 5

$$\mathcal{LR}_n \Longrightarrow 2\big(\bar{\mathcal{G}}^{*}_{q}(\mathcal{P}^0) - \bar{\mathcal{G}}^{*}_{q}(\mathcal{P})\big).$$

Using (A.29) we can verify directly that

$$(\mathcal{P} - \mathcal{P}^0)'RJ^{-1}R'\mathcal{P}^0 = 0.$$

In consequence,

$$\begin{aligned} &2\big(\bar{\mathcal{G}}^{*}_{q}(\mathcal{P}^0) - \bar{\mathcal{G}}^{*}_{q}(\mathcal{P})\big) \\ &= \big(\mathcal{P}^0 - [Z'_s, Z'_u]'\big)' RJ^{-1}R' \big(\mathcal{P}^0 - [Z'_s, Z'_u]'\big) - (\mathcal{P} - [Z'_s, Z'_u]')' RJ^{-1}R' (\mathcal{P} - [Z'_s, Z'_u]') \\ &= \mathcal{P}^{0'}(RJ^{-1}R')\mathcal{P}^0 - \mathcal{P}^{'}(RJ^{-1}R')\mathcal{P} \\ &= (\mathcal{P} - \mathcal{P}^0)'RJ^{-1}R'(\mathcal{P} - \mathcal{P}^0) \end{aligned}$$

as required.

(ii) Recall that $\mathcal{U}$ is given by

$$\mathcal{U} = \operatorname*{argmin}_{u \in \mathbb{R}^{h_2+}} (u - Z_u)' M \Omega M' (u - Z_u).$$

Moreover, we deduce from the definition of (A.28) that

$$Z_u \sim \mathcal{N}\Big(0, (M\Omega M')^{-1}\Big).$$

The statement of the theorem follows by defining $\Lambda = (M\Omega M')^{-1}$. ■

Proof of Corollary 2: omitted. ■

Proof of Corollary 3: omitted. ■

References

Andrews, Donald W.K. (1999): "Estimation when a Parameter is on a Boundary," **Econometrica**, **67**, 1341-1383.

Andrews, Donald W.K. (2001): "Testing when a Parameter is on the Boundary of the Maintained Hypothesis," **Econometrica**, **69**, 683-734.

Angrist, Joshua and Alan Krueger (1991): "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, **106**, 979-1014.

Berge, Claude (1963): *"Topological Spaces,"* MacMillan, New York.

Berger, Roger L., and Dennis D. Boos (1994): "P Values Maximized Over a Confidence Set for the Nuisance Parameter," *Journal of the American Statistical Association*, **89**, 1012-1016.

Chernoff, Herman (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, **25**, 573-578.

Chernozhukov, Victor, Han Hong, and Elie Tamer (2002): "Parameter Set Inference in a Class of Econometric Models," *Manuscript*, Princeton University, Department of Economics.

El Barmi, Hammou, and Richard Dykstra (1995): "Testing for and against a Set of Linear Inequality Constraints in a Multinomial Setting," *Canadian Journal of Statistics*, **23**, 131-143.

El Barmi, Hammou (1995): "Empirical Likelihood Ratio Test for or against a Set of Inequality Constraints," *Journal of Statistical Planning and Inference*, **55**, 191-204.

Gourierioux, Christian, Alberto Holly, and Alan Monfort (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, **50**, 63-79.

Gourieroux, Christian, and Alan Monfort (1995): *"Statistics and Econometric Models,"* Volume 2, Cambridge University Press, Cambridge.

Hansen, Peter Reinhard (2003): "Asymptotic Tests of Composite Hypotheses," *Manuscript*, Department of Economics, Brown University.

Kitamura, Yuichi (2001): "Asymptotic Optimality of Empirical Likelihood for Testing Moment Restrictions," *Econometrica*, **69(6)**, 1661-1672.

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

Kudo, Akio (1963): "A Multivariate Analogue of the One-sided Test," *Biometrika*, **50**, 403-418.

Kodde, David A., and Franz C. Palm (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, **54**, 1243-1248.

Newey, Whitney K. and Richard J. Smith (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica*, **72**, 219-256.

Owen, Art B. (1988): "Empirical Likelihood Ratio Confidence Intervals for a Single Functional," *Biometrika*, **75**, 237-249.

Owen, Art B. (2001): *"Empirical Likelihood,"* Chapman & Hall, New York.

Perlman, Michael D. (1969): "One-Sided Testing Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, **40**, 549-567.

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

Sen, Pranab, and Mervyn J. Silvapulle (2002): "An Appraisal of Some Aspects of Statistical Inference under Inequality Constraints," *Journal of Statistical Planning and Inference*, **107**, 3-43.

Shapiro, Alexander (1985): "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures under Inequality Constraints," *Biometrika*, **72**, 133-144.

Silvapulle, Mervyn J. (1996): "A Test in the Presence of Nuisance Parameters," *Journal of the American Statistical Association*, **91**, 1690-1693.

Tamer, Elie (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, **70**, 147-165.

Wolak, Frank A. (1991): "The Local and Global Nature of Hypothesis Tests Involving Inequality Constraints in Nonlinear Models," *Econometrica*, **59**, 981-995.

Wu, C. F. J. (1990): "On the Asymptotic Properties of the Jackknife Histogram", *Annals of Statistics*, **18**, 1483-1452

Zeldes, Stephen P. (1989): "Consumption and Liquidity Constraints: An Empirical Investigation," *Journal of Political Economy*, **97(2)**, 305-346.

Table 1: Parameterization of DGP

	DGP 1	DGP 2
$\rho_{1,2}$	$[0.5, -0.1]'$	
$\rho_{1,X}$	$[0.5, 0.5]'$	
$\rho_{2,X}$	0.5	
$\rho_{\epsilon,u}$	0.8	

Table 2: Sampling Distribution of $\hat{\theta}$

	$\hat{\theta}_{EL,0}$		$\hat{\theta}_{EL,1}$		$\hat{\theta}_{EL,12}$	
u_0	Bias	MSE	Bias	MSE	Bias	MSE
			$T = \infty$			
0.00	-0.25	1.80	0.00	2.00	0.00	1.60
0.50	-0.12	1.84	0.00	2.00	0.32	1.71
1.00	-0.05	1.91	0.00	2.00	0.65	2.02
1.50	-0.02	1.96	0.00	2.00	0.97	2.55
2.00	-0.01	1.99	0.00	2.00	1.30	3.29
3.00	0.00	2.00	0.00	2.00	1.95	5.41
			$T = 100$			
0.00	-0.36	2.18	-0.10	2.21	-0.07	1.78
0.50	-0.24	2.15	-0.11	2.22	0.25	1.74
1.00	-0.17	2.19	-0.11	2.23	0.56	1.91
1.50	-0.15	2.23	-0.12	2.24	0.85	2.28
2.00	-0.15	2.26	-0.12	2.25	1.14	2.83
3.00	-0.15	2.29	-0.13	2.27	1.68	4.35

Notes: The table reports bias and mean squared error (MSE) based on the simulation of the limit distribution ($T = 100$), and a finite-sample Monte Carlo experiment ($T = 100$).

Figure 1: Asymptotic Power Function for $H_0 : \mathbb{E}[g_2(X_i, \theta_0)] = 0$

Notes: Size of test is 10 percent. Power function is computed based on 50,000 draws from limit distribution.

Figure 2: Finite Sample Power Function for $H_0: \mathbb{E}[g_2(X_i, \theta_0)] = 0$

Notes: Size of test is 10 percent, sample size is $T = 100$. Power function is computed based on 10,000 Monte Carlo replications.

Figure 3: Asymptotic Rejection Probabilities for Test of $H_0 : \mathbb{E}[g_1(X_i, \theta_0)] = 0$ and $\mathbb{E}[g_2(X_i, \theta_0)] \geq 0$

Notes: Size of test is 10 percent. Power function is computed based on 50,000 draws from limit distribution. Critical value is obtained by taking the maximum over $u_0 \in [0, 5]$.

Figure 4: FINITE SAMPLE REJECTION PROBABILITIES FOR TEST OF $H_0: \mathbb{E}[g_1(X_i, \theta_0)] = 0$ AND $\mathbb{E}[g_2(X_i, \theta_0)] \geq 0$

Notes: Size of test is 10 percent, sample size is $T = 100$. Power function is computed based on 10,000 Monte Carlo replications. Critical value is obtained by taking the maximum over $u_0 \in [0, 5]$.

Figure 5: Asymptotic Power Function for Test of $H_0 : \theta_0 = 0$

Notes: Size of test is 10 percent. Power function is computed based on 50,000 draws from limit distribution. Critical values are obtained by taking the maximum over $u_0 \in [0, 5]$.

Figure 6: Finite Sample Power Function for Test of $H_0 : \theta_0 = 0$

Notes: Size of test is 10 percent, sample size is $T = 100$. Power function is computed based on 10,000 Monte Carlo replications. Critical value is obtained by taking the maximum over $u_0 \in [0, 5]$.