Empirical Likelihood Estimation with Inequality Moment Constraints

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
University of Pennsylvania

May 19, 2005

Abstract

This paper extends moment-based estimation procedures to models in which overidentifying information is provided by inequality moment conditions. We derive the large sample distribution theory for the maximum empirical likelihood estimator of the finite-dimensional parameter vector θ that indexes the moment conditions. We show that the asymptotic mean-squared error (MSE) of our estimator is smaller than the MSE of an empirical likelihood estimator that ignores the information contained in the inequality moment conditions. We propose asymptotically valid confidence sets for θ and the slackness associated with the inequality moment conditions. Based on simulations of the limit distribution of the confidence sets, we provide evidence that our approach leads to more precise inference than procedures that ignore the the inequality moment conditions. The limit distributions derived in this paper also apply to conventional GMM estimators.

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 Jin Hahn, Shakeeb Khan, Yuichi Kitamura, Masao Ogaki, Nick Souleles, and seminar participants at the 2003 Econometrics Society Winter Meetings, Columbia University, the Federal Reserve Bank of Atlanta, Michigan State University, Penn State University, UC Irvine, UC Riverside, and the Universities of Chicago, Montreal, 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,

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

if $\theta = \theta_0$. Inequality moment conditions arise from many important economic models, including intertemporal models of consumption and investment decisions that are subject to borrowing constraints, e.g., Zeldes (1989), asset pricing in the presence of financial frictions, e.g., Luttmer (1996, 1999), models of firm behavior that are based on the assumption that firms' actual choices yield higher ex ante expected profits than alternative feasible choices, see Pakes, Porter, Ho, and Ishii (2005), and instrumental variable models in which a subset of the instrumental variables is potentially correlated with the error term in the regression equation, but the direction of the potential correlation is assumed to be known, e.g., as in the case of Manski and Pepper's (2000) monotone instrumental variables.

In general, the use of inequality moment conditions may introduce identification problems, that is, there is a non-singleton subset of the parameter space that satisfies (1). Estimation and inference in the context of set identification has recently been studied by Andrews, Berry, and Jia (2004), Chernozhukov, Hong, and Tamer (2002), and Pakes, Porter, Ho, and Ishii (2005). Our paper, on the other hand, focuses on the additional information that the inequality moment condition $\mathbb{E}[g_2(X_i, \theta)] \geq 0$ can provide in a model in which θ_0 is in principle identifiable based on the equality moment condition $\mathbb{E}[g_1(X_i, \theta)] = 0$ alone. If it is the case that some elements of the vector $\mathbb{E}[g_2(X_i, \theta_0)]$ are near zero, in the sense that $\mathbb{E}[g_2(X_i, \theta_0)] = u_0/\sqrt{n}$, then the second set of moment conditions provides additional information, even asymptotically. The inequality condition constrains the limit objective function of the estimator of θ and hence reduces its variability. The larger u_0, the less informative is the second moment condition. As u_0 tends to infinity the estimation and inference procedures proposed in this paper are asymptotically equivalent to those that are based on $g_1(X_i, \theta)$ only.

A variety of approaches exist to exploit the moment conditions (1) for the estimation of θ_0. While generalized method of moments (GMM) is currently the most widely used procedure in practice, information-theoretic estimators such as empirical likelihood estimators have emerged as an attractive alternative to GMM. 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 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 review of empirical likelihood methods in econometrics and statistics is provided in the monograph by Owen (2001).

While we do not extend the above-mentioned higher-order optimality properties of empirical likelihood procedures to the class of irregular models considered in this paper, we believe that these results provide a good reason for studying empirical likelihood estimators in the context of models with moment inequality constraints. In fact, since moment conditions are imposed as parametric constraints on the empirical likelihood function, an extension to inequality conditions is quite natural.

This paper focuses on first-order asymptotic approximations and makes two contributions. First, we derive the joint limit distribution of the EL estimators of θ_0 and $\mathbb{E}[g_2(X_i, \theta_0)]$. EL estimators are 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, more recently, Andrews (1999) by deriving a quadratic approximation of the EL objective function and analyzing the distribution of its saddlepoint. The inequality moment conditions translate into sign restrictions on the corresponding Kuhn-Tucker parameters in the saddlepoint formulation of the EL problem. For the (special) case in which $g_2(X_i, \theta)$ is a scalar, we show analytically that the asymptotic mean-squared error (MSE) of our estimator is smaller than the MSE of an empirical likelihood estimator that ignores the information contained in the inequality moment conditions.

Our asymptotic analysis has a straightforward extension to the class of saddlepoint estimators that Newey and Smith (2004) refer to as generalized empirical likelihood estimators. However, the extension is not pursued in this paper. Since the concentrated limit objective function of the EL estimator has the same first-order asymptotic approximation as a GMM estimator that uses an optimal weight matrix and handles the presence of inequality moment conditions through additional slackness parameters (see below), our large sample analysis also applies to conventional GMM estimators.

Second, we invert empirical likelihood ratio test statistics to obtain confidence sets for θ_0 and $\mathbb{E}[g_2(X_i, \theta_0)]$. The near-zero slackness parameter u_0 enters the limit distributions of the EL estimator of θ and related empirical likelihood ratio statistics, which complicates statistical inference. Since u_0 cannot be consistently estimated we construct a Bonferroni type confidence set for θ_0 that takes a union of confidence sets that are valid conditional on particular values of u_0. This complication is unrelated to the saddlepoint formulation of the EL estimation problem and also arises in a more conventional GMM analysis of model (1). The nuisance parameter dependence of the limit distributions resembles the difficulties encountered in models with nearly integrated regressors, e.g., Cavanagh, Elliott, and Stock (1995). Based on simulations of the non-standard limit distribution of the empirical likelihood ratios, we show that the proposed confidence sets for θ_0 and u_0 perform well compared to the exact asymptotic confidence sets based on the $I\!E[g_1(X_i, \theta_0)] = 0$ estimator.

One can introduce an additional parameter vector $\vartheta = I\!E[g_2(X_i, \theta)]$ that captures the slackness in the inequalities and express the second moment condition as $I\!E[g_2(X_i, \theta_0) - \vartheta_0] = 0$, where $\vartheta_0 \geq 0$. Thus, rather than using the inequality moment condition directly, it could be translated into an inequality restriction on the parameter vector. 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). Detailed 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, none of the EL papers provides a complete limit distribution theory, considers the important case in which the inequalities stem directly from the moment conditions, and analyzes confidence intervals.

The special case of $I\!E[g_2(X_i, \theta_0)] = 0$ translates into $\vartheta_0 = 0$, 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. We extend some of Andrews' results to estimators that are defined as a saddlepoint rather than an extremum. Moreover, Andrews (1999) focuses on inference for ϑ_0, whereas we also discuss inference with respect to θ_0, treating ϑ_0 as a nuisance parameter.

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. We discuss several important applications of our method. Section 3 develops the asymptotic distribution theory for the EL estimator and its objective function in the presence of inequality moment conditions. Section 4 constructs interval estimators for θ_0 and $\mathbb{E}[g_2(X_i, \theta_0)]$. Since the asymptotic distributions derived in this paper are non-standard, we simulate the limit distributions of point estimators and confidence intervals in the context of a numerical example in Section 5. Moreover, we make a comparison with the asymptotic properties of simple procedures that ignore the information in the inequality moment condition. 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 $\|A\| = (tr[A'A])^{1/2}$. $I\{x \geq a\}$ 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 $(\Omega, \mathcal{F}, P)$.*

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

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

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, $\mathbb{E}[g_2(X_i, \theta_0)] = \nu_{n,0} = \nu_0 + n^{-1/2} u_0 \geq 0$ and $\mathbb{E}[g(X_i, \theta_0) g(X_i, \theta_0)'] = J_n \longrightarrow J$ is non-singular.*

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

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* $j=1,...,h$.

Most importantly, we assume in Assumption 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 $\nu_{n,0}\geq 0$. In order to be able to study the local properties of our estimation and inference procedures we allow for $n^{-1/2}$ drifts in the parameter θ and the slackness of the inequality conditions. In general, it will turn out that moment conditions for which the corresponding element of ν_0 is strictly greater than zero do not affect the limit distribution of estimators and test statistics. However, if $\nu_0=0$ and the expected value of the second set of moment conditions are close to zero in the sense that $u_0>0$ then it will influence the limit distributions that we are deriving subsequently.

2.1 Examples of Inequality Moment Conditions

We now introduce several examples for inequality moment conditions to motivate our econometric analysis. Consider the problem of estimating

$$X_Y=X_X'\theta_0+U, \tag{2}$$

where X_X is an endogenous regressor that is correlated with the error term U. Suppose two sets of instrumental variables (IV) are available: X_1 is a vector of variables that are orthogonal to U, whereas for X_2 it is economically plausible to assume that a potential violation of the orthogonality takes a particular direction, for instance, $\mathbb{E}[X_2U]\geq 0$. This inequality condition is closely related to Manski and Pepper's (2000) notion of monotone instrumental variables:

$$\mathbb{E}[U|X_2=x_2]\geq\mathbb{E}[U|X_2=\widetilde{x}_2]\quad\text{for all}\quad x_2\geq\widetilde{x}_2.$$

Suppose that X_Y is wage, X_X is schooling, and X_2 measures a person's ability. If the type of ability captured by X_2 is valued in the market, then it is reasonable to assume that X_2 is a monotone instrumental variable and hence satisfies the inequality moment condition.

In macroeconomics there is great interest in characterizing the behavior of central banks through interest rate feedback rules, e.g.

$$R=\rho R_{-1}+(1-\rho)(\psi_1\pi+\psi_2 y)+U,$$

where R is the nominal interest rate, π is inflation, and y is output (see Taylor (1999)). All variables are in deviation from their respective target levels and U is an unanticipated deviation from the policy rule, typically referred to as monetary policy shock. Hence, in our notation $X_Y = R$ and $X_X = [R_{-1}, \pi, y]'$. Many theoretical dynamic stochastic general equilibrium models (see Woodford (2003) for an extensive analysis of New Keynesian models) predict that output and inflation are correlated with the monetary policy shock U, which establishes the need for IV estimation. Natural candidates in a time series environment are lagged endogenous variables. However, many theories also predict that inflation and output fall in response to a contractionary policy shock ($U > 0$). This prediction leads to the inequality moment conditions $\mathbb{E}[-yU] \geq 0$ and $\mathbb{E}[-\pi U] \geq 0$, which can be used to sharpen the inference with respect to the policy rule coefficients.[1]

More generally, applied researchers typically pay careful attention to the validity of instrumental variables, sacrificing relevance, that is, correlation between instruments and regressors. A widely cited example is Angrist and Krueger (1991) who proposed using quarter of birth as an instrument to circumvent ability bias in estimating the returns to education. In response, the econometrics literature has developed an asymptotic theory of weak IVs, which assumes that instruments and error terms are orthogonal but the correlation between instruments and regressors vanishes as the sample size tends to infinity (see Stock, Wright, and Yogo (2002) for a survey). In this paper we take a complimentary approach. Our analysis encompasses situations in which the IV's remain asymptotically relevant, but the validity condition is weakly violated, in the sense that the correlation between instruments and regressors is potentially positive in finite samples, but converges to zero asymptotically.

Inequality moment restrictions also arise in the context of intertemporal optimization models, in which agents face liquidity or regulatory constraints. For instance, in an influential paper Zeldes (1989) studies whether the presence of borrowing constraints can explain households' violation of consumption Euler equations. For households that face a binding borrowing constraint, the marginal utility of consumption in the current period exceeds the discounted expected marginal utility in the subsequent period. Zeldes (1989) constructs an observable proxy from the wealth-to-income ratio that indicates if consumption of household i is constrained. He argues that if the wealth threshold is sufficiently large, then some households may be incorrectly classified as constrained, but it is unlikely that unconstrained households are misclassified. Hence, the optimality condition for high

[1] Such an approach is related to the methods developed by De Nicoló (2002) and Uhlig (2005) to identify impulse response functions based on sign restrictions.

wealth-to-income households translates into an equality moment condition, whereas the optimality condition for the remaining households leads to an inequality moment condition. Zeldes (1989) ignores in his empirical analysis the inequality moment condition when estimating utility function parameters and testing for the presence of borrowing constraints. However, if the marginal utility differential of the borrowing constrained households is small or the fraction of misclassified households is large then his approach potentially neglects important information.

Finally, Pakes, Porter, Ho, and Ishii (2005) provide several examples of inequality moment conditions derived from models of industrial organization, including an ordered choice problem in which banks choose the number of ATMs, and a model of buyer/seller networks, in which hospitals and Health Maintenance Organizations establish health plans and hospital networks.

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). Our analysis has a straightforward extension (not pursued in this paper) to the class of estimators that Newey and Smith (2004) refer to as generalized empirical likelihood estimators, e.g., exponential tilting and continuous updating GMM. The (constrained) empirical likelihood function is

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

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

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

Let

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

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

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

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

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

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

$$p_i = \frac{1}{n\left(1 + \lambda_1' g_1(X_i, \theta) + \lambda_2 g_2(X_i, \theta)\right)}, \tag{8}$$

$$0 = \sum_{i=1}^{n} p_i g_1(X_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{g_1(X_i, \theta)}{1 + \lambda_1' g_1(X_i, \theta) + \lambda_2 g_2(X_i, \theta)}, \tag{9}$$

$$0 \leq \sum_{i=1}^{n} p_i g_2(X_i, \theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{g_2(X_i, \theta)}{1 + \lambda_1' g_1(X_i, \theta) + \lambda_2 g_2(X_i, \theta)}, \tag{10}$$

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

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

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

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

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

$$E[g_2(X_i,\theta_0) - \vartheta_{0,n}] = E[\tilde{g}_2(X_i,\theta_0,\vartheta_{0,n})] = 0$$

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

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

where

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

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

$$\hat{\nu}_j = \hat{\vartheta}_j = \begin{cases} \left.\frac{\partial G_n(\theta,\lambda_1,\lambda_2)}{\partial \lambda_{2,j}}\right|_{\hat{\theta},\hat{\lambda}_1,\hat{\lambda}_2} & \text{if} \quad \hat{\lambda}_{2,j} = 0 \\ 0 & \text{if} \quad \hat{\lambda}_{2,j} < 0, \quad j = 1,\ldots,h_2. \end{cases}$$

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

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

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

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

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 lead 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 the difference between $\hat{\nu}_n$, characterized in Lemma 1 as derivative of $G_n(\theta, \lambda_1, \lambda_2)$ with respect to λ_2, and $\nu_{n,0} = I\!E[g_2(X_i, \theta_0)]$ converges to zero. The vector of estimated Kuhn-Tucker parameters $\hat{\lambda}$ also 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* $\hat{\nu}_n - \nu_{n,0} \xrightarrow{p} 0$. *Moreover,* $\hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{p} 0$.

3.2 Quadratic Approximation of Objective Function

We proceed with a second-order Taylor approximation of the objective function G_n^*. Let $\beta = [\theta', \nu', \lambda']'$, $\beta_{n,0} = [\theta_0', \nu_{n,0}', 0]'$, 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

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

where

$$G_{nq}^*(\beta) = G_n^*(\beta_{n,0}) + G_n^{*(1)}(\beta_{n,0})'(\beta - \beta_{n,0}) + \frac{1}{2}(\beta - \beta_{n,0})' G_n^{*(2)}(\beta_{n,0})(\beta - \beta_{n,0}). \tag{19}$$

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

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

where $\Lambda_n^{\zeta} = \left\{ \lambda \in \mathbb{R}^h : \|\lambda\| \leq n^{-\zeta} \right\}$. For technical reasons it is convenient to impose that the domain of λ shrinks at the rate $n^{-\zeta}$. We show in Lemmas A.1 and A.2 in the Appendix that this domain restriction asymptotically does not affect $\hat{\lambda}$. A bound for the remainder $\mathcal{R}_n(\beta)$ is provided in the following lemma.

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

$$\sup_{\beta \in \mathcal{B}_n : \|\beta - \beta_{n,0}\| \leq \gamma_n} \frac{|\mathcal{R}_n(\beta)|}{(1 + \|\sqrt{n}(\beta - \beta_{n,0})\|^2)} = o_p(1), \tag{20}$$

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

The first and second derivatives of G_n^* evaluated at $\beta_{n,0}$ are of the form

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

where

$$Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g(X_i, \theta_0) - M' \nu_{n,0} \right], \; Q_n = \frac{1}{n} \sum_{i=1}^{n} g^{(1)}(X_i, \theta_0), \; J_n = \frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta_0) g(X_i, \theta_0)'.$$

We proceed by transforming the parameter vector β. Let $b = [s', u', l']' = \sqrt{n}(\beta - \beta_0)$, where $\beta_0 = [\theta_0', \nu_0', 0]'$. 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 the j'th element of ν_0 is strictly positive. The objective function G_n^* can be expressed in terms of the "local" deviations b from β_0 as

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

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

$$\begin{aligned}
&\mathcal{G}_{nq}^*(s, u, l) \\
&= -\frac{1}{2}(l - J_n^{-1}[Z_n + Q_n' s - M'(u - u_0)])' J_n (l - J_n^{-1}[Z_n + Q_n' s - M'(u - u_0)]) \\
&\quad + \frac{1}{2}(Z_n + Q_n' s - M'(u - u_0))' J_n^{-1} (Z_n + Q_n' s - M'(u - u_0)).
\end{aligned} \tag{23}$$

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 $\phi_0 = [0, u_0']'$ and $R_n = [-Q_n', M']'$. Then we define

$$\begin{aligned}
\mathcal{G}_{nq}^*(\phi, l) &= -\frac{1}{2}(l - J_n^{-1}[Z_n - R_n'(\phi - \phi_0)])' J_n (l - J_n^{-1}[Z_n - R_n'(\phi - \phi_0)]) \qquad (24) \\
&\quad + \frac{1}{2}(Z_n - R_n'(\phi - \phi_0))' J_n^{-1} (Z_n - R_n'(\phi - \phi_0)).
\end{aligned}$$

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.

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

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

where $J = \lim_{n \longrightarrow \infty} \mathbb{E}[g(X_i, \theta_0) g(X_i, \theta_0)']$, $R = \lim_{n \longrightarrow \infty} \left[-\mathbb{E}[g^{(1)}(X_i, \theta_0)]', M' \right]'$ *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,

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

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

$$\Phi(\nu_0) = \left\{\phi = [s', u'] \in \mathbb{R}^m \otimes \mathbb{R}^{h_2} \mid u_j \geq 0 \text{ if } \nu_{0,j} = 0\right\}. \tag{25}$$

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

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

$$\tilde{l}_q(\phi) = J_n^{-1}\left(Z_n - R'_n(\phi - \phi_0)\right). \tag{26}$$

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

$$\bar{\mathcal{G}}^*_{nq}(\phi) = \mathcal{G}^*_{nq}(\phi, \tilde{l}_q(\phi)) = \frac{1}{2}(Z_n - R'_n(\phi - \phi_0))' J_n^{-1}(Z_n - R'_n(\phi - \phi_0)). \tag{27}$$

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.

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

$$\begin{aligned}
\mathcal{P} &= \underset{\phi \in \Phi(\nu_0)}{argmin} \frac{1}{2}(Z - R'(\phi - \phi_0))' J^{-1}(Z - R'(\phi - \phi_0)), \\
\mathcal{L} &= J^{-1}(Z - R'(\mathcal{P} - \phi_0)), \\
\mathcal{G}^*_q(\mathcal{P}, \mathcal{L}) &= \frac{1}{2}(Z - R'(\mathcal{P} - \phi_0))' J^{-1}(Z - R'(\mathcal{P} - \phi_0)).
\end{aligned}$$

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

3.4 GMM with Inequality Moment Conditions

The limit distribution derived in Theorem 4 also applies to the following GMM estimator:

$$\min_{\theta \in \Theta, \vartheta \geq 0} \quad \frac{1}{2} \left(\sum_{i=1}^{n} g(X_i, \theta) - M'\vartheta \right)' W_n \left(\sum_{i=1}^{n} g(X_i, \theta) - M'\vartheta \right), \tag{28}$$

where ϑ is an $h_2 \times 1$ vector of slackness parameters, the $h_2 \times h$ matrix $M = [0\ I]$ is defined as above, and $\{W_n\}$ is a sequence of positive-definite $h \times h$ weight matrices. Let $s = \sqrt{n}(\theta - \theta_0)$, $u = \sqrt{n}(\vartheta - \nu_0)$, and $\phi = [s', u']'$. Using definitions of Z_n, R_n, and J_n in (21) and assuming that $W_n - J_n^{-1} \xrightarrow{p} 0$ it follows from the arguments in Andrews (1999) that the objective function of the GMM estimator has a quadratic approximation of the form

$$\frac{1}{2}(Z_n - R_n'(\phi - \phi_0))' J_n^{-1}(Z_n - R_n'(\phi - \phi_0)).$$

Thus, the approximation of the GMM objective function is equivalent to the concentrated objective function $\bar{\mathcal{G}}^*_{nq}(\phi)$ of the empirical likelihood estimator in Equation (27). Therefore, the analysis in the remainder of the paper applies not only to empirical likelihood estimators but also to conventional GMM estimators.

3.5 Discussion

We will now explore the limit distribution of $\hat{b}$ in more detail. First, we will show that the limit distribution of $\hat{s}$ does not depend on the g_2-moment condition if $\nu_0 > 0$. In this case, our estimator is asymptotically equivalent to the one that only uses the g_1-moment condition. The result has a straightforward generalization: elements of the vector g_2 that have a strictly positive expected value do not affect the limit distribution of $\hat{\theta}$. Second, if $\nu_0 = 0$ and $\mathbb{E}[g_2(X_i,\theta_0)] = n^{-1/2}u_0$, then the parameter u_0 affects the shape of the limit distribution. The larger u_0 the less information about θ can be extracted from the inequality moment condition. Third, for the case $h_1 = 1$ we derive the asymptotic mean and the variance of $\hat{s}$ and compare it to the mean and variance of an estimator that only uses $\mathbb{E}[g_1(X_i,\theta_0)] = 0$ and one that potentially wrongly imposes $\mathbb{E}[g_2(X_i,\theta_0)] = 0$.

Irrelevant Inequality Moment Conditions. We partition the random vector Z and the matrices R and J as follows:

$$Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad R' = \begin{bmatrix} -Q_1' & 0 \\ -Q_2' & I \end{bmatrix}, \quad J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}.$$

The partitions conform with $g(x,\theta) = [g_1'(x,\theta), g_2'(x,\theta)]'$. Using the formulas for marginal and conditional means and variances of a multivariate normal distribution it is straightforward to verify that

$$\begin{aligned} &(Z - R'(\phi-\phi_0))'J^{-1}(Z - R'(\phi-\phi_0)) \\ &= (Z_1 + Q_1's)'J_{11}^{-1}(Z_1 + Q_1's) \\ &\quad + [Z_2 + Q_2's - (u-u_0) - J_{21}J_{11}^{-1}(Z_1+Q_1's)]' \\ &\quad \times (J_{22} - J_{21}J_{11}^{-1}J_{12})^{-1}[Z_2 + Q_2's - (u-u_0) - J_{21}J_{11}^{-1}(Z_1+Q_1's)]. \end{aligned} \tag{29}$$

If $\nu_0 > 0$ then the limit distribution of $\hat{u}$ is obtained by minimizing (29) with respect to $u \in \mathbb{R}^{h_2}$. Hence,

$$\mathcal{U} - u_0 = Z_2 + Q_2\mathcal{S} - J_{21}J_{11}^{-1}(Z_1 + Q_1'\mathcal{S}),$$

which implies that the second summand in (29) 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 $g_2(x,\theta)$ we deduce that inequality moment conditions that hold with strict inequality do not influence the distribution of $\mathcal{S}$ and, therefore, asymptotically do not provide any additional information on θ. Second, although the distribution of the random vector Z depends on ν_0, notice that $Z_1 \sim \mathcal{N}(0, J_{11})$. Thus, neither the distribution of $\mathcal{S}$, nor the distribution of $\mathcal{G}_q^*(\mathcal{P},\mathcal{L})$ depends on the specific values of ν_0 if $\nu_0 > 0$. In particular,

$$\mathcal{S} = -(Q_1J_{11}^{-1}Q_1')^{-1}Q_1J_{11}^{-1}Z_1 \equiv \mathcal{N}\Big(0, (Q_1J_{11}^{-1}Q_1')^{-1}\Big).$$

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,

$$2\mathcal{G}_q^*(\mathcal{P},\mathcal{L}) = Z_1'[J_{11}^{-1} - J_{11}^{-1}Q_1'(Q_1J_{11}^{-1}Q_1')^{-1}Q_1J_{11}^{-1}]Z_1, \tag{30}$$

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 $\mathbb{E}[g_1(X_i,\theta_0)] = 0$.

Weakly Informative Inequality Moment Conditions. Now suppose that $\mathbb{E}[g_2(X_i,\theta_0)] = n^{-1/2}u_0$, where $u_0 > 0$. Then the concentrated asymptotic objective function becomes

$$\bar{\mathcal{G}}_q^*([s',u']') = \frac{1}{2}(Z + Q's - M'(u-u_0))'J^{-1}(Z + Q's - M'(u-u_0)) \tag{31}$$

and has to be minimized subject to the constraint that $u \geq 0$. Using a change of variables and defining $\tilde{u} = u - u_0$ we obtain

$$\bar{\mathcal{G}}_q^*([s',u_0+\tilde{u}']') = \frac{1}{2}(Z + Q's - M'\tilde{u})'J^{-1}(Z + Q's - M'\tilde{u}) \tag{32}$$

where $\tilde{u} \geq -u_0$. Thus, the further $\mathbb{E}[g_2(X_i,\theta_0)]$ is apart from zero (in the local metric) the less often the constraint on $\tilde{u}$ is binding and the closer limit distribution to the one that is obtained if the second set of moment conditions is ignored.

Mean-Squared-Error Comparison. For the special case of $h_2 = 1$ we derive an analytic formula for the asymptotic mean-squared-error of the estimator $\hat{s}$. Consider the concentrated limit objective function for the estimator of ϕ:

$$\bar{\mathcal{G}}_q^*(\phi) = \frac{1}{2}(Z - R'(\phi-\phi_0))'J^{-1}(Z - R'(\phi-\phi_0)). \tag{33}$$

In the absence of a constraint on ϕ the limit covariance matrix of $\hat{\phi}$ were given by

$$\Omega = (RJ^{-1}R')^{-1} = \begin{bmatrix} \Omega_{ss} & \Omega_{su} \\ \Omega_{us} & \Omega_{uu} \end{bmatrix}.$$

The partitions of Ω conform with the partition $\phi = [s',u']'$. It can be verified that

$$\begin{aligned} \Omega_{ss} &= (Q_1J_{11}^{-1}Q_1')^{-1} \\ \Omega_{ss} - \Omega_{su}\Omega_{uu}^{-1}\Omega_{us} &= (QJ^{-1}Q')^{-1}. \end{aligned}$$

Without loss of generality we are re-normalizing the inequality moment condition such that $\Omega_{uu} = 1$. Let $\varphi(\cdot)$ denote the probability density function and $\Phi(\cdot)$ the cumulative density function of a $\mathcal{N}(0,1)$. We show in the Appendix that

$$\begin{aligned}
\mathbb{E}[\mathcal{S}] &= \Omega_{su}[\varphi(u_0) - u_0\Phi(u_0)] && (34)\\
V[\mathcal{S}] &= \Omega_{ss} + \Omega_{su}\Omega_{us}(1-\Phi(u_0))\Bigg[1 - \frac{\varphi^2(u_0)}{(1-\Phi(u_0))^2} - \frac{u_0\varphi(u_0)}{1-\Phi(u_0)} && (35)\\
&\quad + \left(u_0 + \frac{\varphi(u_0)}{1-\Phi(u_0)}\right)^2 \Phi(u_0)\Bigg] - \Omega_{su}\Omega_{us}
\end{aligned}$$

and the mean-squared-error is given by

$$MSE(\mathcal{S}) = \Omega_{ss} + \Omega_{su}\Omega_{us}[(u_0^2 - 1)(1-\Phi(u_0)) - u_0\varphi(u_0)]. \tag{36}$$

The limit distribution of the empirical likelihood estimator that is based only on $\mathbb{E}[g_1(X_i,\theta_0)] = 0$ can be expressed as $\mathcal{S}_{(1)} \sim \mathcal{N}(0,\Omega_{ss})$. Since[2]

$$(u_0^2 - 1)(1-\Phi(u_0)) - u_0\varphi(u_0) \begin{cases} = -\frac{1}{2} & \text{if } u_0 = 0 \\ < -\frac{1}{u_0}\varphi(u_0) & \text{if } u_0 > 0 \end{cases}$$

we obtain the following efficiency result:

Theorem 6 *Suppose Assumptions 1 to 6 are satisfied and $h_2 = 1$, then*

$$MSE(\mathcal{S}) \leq MSE(\mathcal{S}_{(1)}).$$

The limit distribution of the estimator $\hat{s}_{(12)}$ that imposes $\mathbb{E}[g_1(X_i,\theta_0)] = 0$ and $\mathbb{E}[g_2(X_i,\theta_0)]$ can be written as

$$\mathcal{S}_{(12)} \sim \mathcal{N}\Big((QJ^{-1}Q')^{-1}QJ^{-1}M'u_0, \Omega_{ss} - \Omega_{su}\Omega_{us}\Big)$$

Hence, for $u_0 = 0$ we obtain the ranking

$$MSE(\mathcal{S}_{(12)}) \leq MSE(\mathcal{S}) \leq MSE(\mathcal{S}_{(1)}).$$

As the slackness of the inequality constraint, u_0, increases, the performance of $\hat{s}_{(12)}$ quickly deteriorates. We provide a numerical illustration in Section 5.

4 Inference

Based on the results obtained in the previous section, we will proceed by deriving asymptotically valid confidence sets for θ and ν.

[2] See Pollard (2002, page 317).

4.1 Confidence Sets for θ

A confidence set for θ can be obtained by inverting the empirical likelihood ratio statistic for the null hypothesis $\theta_0 = \theta^H$. We will first study a joint confidence interval for all elements of the parameter vector θ. An extension to confidence regions for subsets of parameters is fairly straightforward and will be discussed at the end of this subsection. The derivation of the confidence sets is complicated by the dependence of the limit distribution of the maximized empirical likelihood function on the slackness associated with the inequality moment condition. In the subsequent analysis we will assume that the second set of moments is close to zero in the sense that $\nu_0 = 0$ and $u_0 \geq 0$.

The test statistic that is used to obtain the confidence set for θ is defined as the ratio of the unrestricted maximum of the empirical likelihood function $L_{EL}(\theta, p)$ and the constrained maximum subject to the restriction $\theta = \theta^H$. We will express the test statistic in terms of the function $G_n^*(\theta, \nu, \lambda)$. Let

$$\hat{\nu}_n^H = \text{argmin}_{\nu \geq 0} \max_{\lambda \in \hat{\Lambda}_n(\theta^H)} G_n^*(\theta^H, \nu, \lambda).$$

The test statistic is given by

$$\mathcal{LR}_n^\theta(\theta^H) = 2n\bigg(G_n^*(\theta^H, \hat{\nu}_n^H, \hat{\lambda}(\theta^H, \hat{\nu}_n^H)) - G_n^*(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n)) \bigg). \tag{37}$$

As in Section 3, let

$$\bar{\mathcal{G}}_q^*(\phi) = \frac{1}{2}(Z - R'(\phi - \phi_0))' J^{-1} (Z - R'(\phi - \phi_0)).$$

Define the set

$$\Phi_H(\nu_0) = \{\phi = [s', u']' \in \{0\}^m \otimes \mathbb{R}^{h_2} \mid u_j \geq 0 \text{ if } \nu_{0,j} = 0\}. \tag{38}$$

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

Corollary 1 *Suppose Assumptions 1 to 6 are satisfied. Moreover, $\theta^H = \theta_0$, $\nu_0 = 0$, $u_0 \geq 0$. Then*

$$\mathcal{LR}_n^\theta(\theta_0) \Longrightarrow \mathcal{LR}^\theta(u_0) \equiv \left(\min_{\phi \in \Phi_H(0)} 2\bar{\mathcal{G}}_q^*(\phi) \right) - \left(\min_{\phi \in \Phi(0)} 2\bar{\mathcal{G}}_q^*(\phi) \right).$$

The asymptotic critical value $c_\alpha^\theta(u_0)$ satisfies

$$\mathbb{P}_{u_0}\left\{ \mathcal{LR}^\theta(u_0) \leq c_\alpha^\theta(u_0) \right\} = 1 - \alpha.$$

Suppose we knew the true value u_0 of the slackness in the inequality constraint. Then a confidence set for θ with asymptotic coverage probability $1-\alpha$ can be obtained as follows:

$$\mathcal{CS}_n^\theta(u_0,\alpha) = \Big\{\theta \in \Theta \mid \mathcal{LR}_n^\theta(\theta) \leq c_\alpha^\theta(u_0)\Big\}. \tag{39}$$

We can deduce from Corollary 1 that this set has the desired coverage probability.

Corollary 2 *Suppose Assumptions 1 to 6 are satisfied. Moreover,* $\theta^H = \theta_0$, $\nu_0 = 0$, $u_0 \geq 0$. *Then*

$$P_{u_0}\Big\{\theta_0 \in \mathcal{CS}_n^\theta(u_0,\alpha)\Big\} = P_{u_0}\Big\{\mathcal{LR}_n^\theta(\theta_0) \leq c_\alpha^\theta(u_0)\Big\} \longrightarrow 1-\alpha.$$

In practice the "true" slackness parameter u_0 is, however, unknown. Since u_0 cannot be consistently estimated, we construct a Bonferroni confidence set for θ_0. Let $\mathcal{CS}_n^u(\alpha_2)$ be a confidence set for u_0 with coverage probability $1-\alpha_2$. Define,

$$\mathcal{CS}_n^\theta = \bigcup_{u \in \mathcal{CS}_n^u(\alpha_2)} \mathcal{CS}_n^\theta(u,\alpha_1). \tag{40}$$

Then,

$$\begin{aligned} P_{u_0}\Big\{\theta_0 \notin \mathcal{CS}_n^\theta\Big\} &\leq P_{u_0}\Big\{\theta_0 \notin \mathcal{CS}_n^\theta\Big\}\Big\{u_0 \in \mathcal{CS}_n^u(\alpha_2)\Big\} + P_{u_0}\Big\{u_0 \notin \mathcal{CS}_n^u(\alpha_2)\Big\} \\ &\leq P_{u_0}\Big\{\theta_0 \notin \mathcal{CS}_n^\theta(u_0,\alpha_1)\Big\} + P_{u_0}\Big\{u_0 \notin \mathcal{CS}_n^u(\alpha_2)\Big\} \longrightarrow \alpha_1 + \alpha_2. \end{aligned}$$

The Bonferroni confidence interval raises two questions. First, how should one construct the confidence set $\mathcal{CS}_n^u(\alpha_2)$, and second, how large should its coverage probability be. The next subsection discusses confidence intervals for u_0. In the numerical illustration in Section 5 we will set α_2 equal to zero.

In order to obtain a confidence set for a subset of parameters one can proceed by modifying the likelihood ratio statistic on which the confidence interval is based as follows. Without loss of generality, partition $\theta = [\theta_1', \theta_2']'$ and denote the hypothesized value of θ_1 by θ_1^H. Let

$$\{\hat{\theta}_{2,n}^H, \hat{\nu}_n^H\} = \text{argmin}_{\theta_2, \nu \geq 0} \max_{\lambda \in \hat{\Lambda}_n(\theta_1^H, \theta_2)} G_n^*(\theta_1^H, \theta_2, \nu, \lambda)$$

and redefine the test statistic as

$$\mathcal{LR}_n^\theta(\theta_1^H) = 2n\Big(G_n^*(\theta_1^H, \hat{\theta}_{2,n}^H, \hat{\nu}_n^H, \hat{\lambda}(\theta_1^H, \hat{\theta}_{2,n}^H, \hat{\nu}_n^H)) - G_n^*(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n))\Big). \tag{41}$$

The subsequent steps remain unchanged.

4.2 Confidence Sets for u

As mentioned previously, we are most interested in the case in which the second set of moment conditions is near zero, that is, $\nu_0 = 0$ and $u_0 \geq 0$. In particular, it is the local slackness parameter u_0 that affects the limit distribution of the likelihood ratios. To keep the notation simple we will focus on a joint confidence set for u. An extension to confidence sets for subsets of u is fairly straightforward. The confidence set is obtained by inverting the empirical likelihood statistic for the null hypothesis $u_0 = u^H$. Let

$$\hat{\theta}_n^H = \text{argmin}_\theta \max_{\lambda \in \hat{\Lambda}_n(\theta)} G_n^*(\theta, n^{-1/2}u^H, \lambda)$$

and define the test statistic

$$\mathcal{LR}_n^u(u^H) = 2n\bigg(G_n^*(\hat{\theta}_n^H, n^{-1/2}u^H, \hat{\lambda}(\hat{\theta}_n^H, n^{-1/2}u^H)) - G_n^*(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n))\bigg). \tag{42}$$

We summarize its limit distribution in the following theorem.

Theorem 7 *Suppose Assumptions 1 to 6 are satisfied. Moreover, $\nu_0 = 0$, $u_0 \geq 0$, and $u^H = u_0$. Then*

$$\mathcal{LR}_n^u(u_0) \Longrightarrow \mathcal{LR}^u(u_0) \equiv Z_u'\Lambda^{-1}Z_u - (\widetilde{\mathcal{U}} - Z_u)'\Lambda^{-1}(\widetilde{\mathcal{U}} - Z_u),$$

where

$$\widetilde{\mathcal{U}} = \textit{argmin}_{\tilde{u} \geq -u_0} (\tilde{u} - Z_u)'\Lambda^{-1}(\tilde{u} - Z_u),$$

$\Lambda = (M[J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}]M')^{-1}$, and $Z_u \sim \mathcal{N}(0, \Lambda)$. The asymptotic critical value $c_\alpha^u(u_0)$ satisfies

$$\mathbb{P}_{u_0}\bigg\{\mathcal{LR}^u(u_0) \leq c_\alpha^u(u_0)\bigg\} = 1 - \alpha.$$

If $u_0 = 0$ then the limit distribution simplifies to $\widetilde{\mathcal{U}}'\Lambda^{-1}\widetilde{\mathcal{U}}$ and the test-statistic has a so-called $\bar{\chi}^2$ limit distribution, e.g., Kudo (1963). As before, a confidence set for u_0 with asymptotic coverage probability $1-\alpha$ can be obtained by inverting the test statistic $\mathcal{LR}^u(u_0)$ as follows:

$$\mathcal{CS}_n^u(\alpha) = \bigg\{u \geq 0 \mid \mathcal{LR}_n^u(u) \leq c_\alpha^u(u)\bigg\}. \tag{43}$$

We can deduce from Theorem 7 that the confidence set has the desired coverage probability.

Corollary 3 *Suppose Assumptions 1 to 6 are satisfied. Moreover, $\nu_0 = 0$, $u_0 \geq 0$, and $u^H = u_0$. Then*

$$P_{u_0}\bigg\{u_0 \in \mathcal{CS}_n^u(\alpha)\bigg\} = P_{u_0}\bigg\{\mathcal{LR}_n^u(u_0) \leq c_\alpha^u(u_0)\bigg\} \longrightarrow 1 - \alpha.$$

4.3 Implementation

The asymptotic critical value functions $c_\alpha^\theta(u_0)$ and $c_\alpha^u(u_0)$ that are needed for the construction of the confidence sets depend on the matrices Q and J. First, one has to calculate the empirical likelihood estimator $\hat{\theta}_n$. Second, a consistent estimate of J and R can be computed as follows:

$$\hat{J}_n = \frac{1}{n}\sum_{i=1}^{n} g(X_i, \hat{\theta}_n) g(X_i, \hat{\theta}_n)', \quad \hat{Q}_n = \frac{1}{n}\sum_{i=1}^{n} g^{(1)}(X_i, \hat{\theta}_n), \quad \hat{R}_n' = [-\hat{Q}_n', M']. \tag{44}$$

Approximate asymptotic critical values $\hat{c}_\alpha^\theta(u_0)$ and $\hat{c}_\alpha^u(u_0)$ can be obtained by simulating $\mathcal{LR}^\theta(u_0)$ (Corollary 1) and $\mathcal{LR}^u(u_0)$ (Theorem 7) conditional on $\hat{J}_n$ and $\hat{R}_n$ for a fine grid of u_0 values (see also Andrews (2001)). Finally, the confidence sets for θ_0 and u_0 can be constructed according to Equations (39) and (43).

5 Example

In the remainder of this paper we provide a numerical example to illustrate the large sample distributions that we derived previously. Consider a simultaneous equations model of the form

$$\begin{aligned} X_{Y,i} &= X_{X,i}\theta + U_i && (45) \\ X_{X,i} &= X_{1,i}'\gamma_1 + X_{2,i}'\gamma_2 + \epsilon_i && (46) \\ X_{2,i} &= X_{1,i}'\rho_{1,2} + \frac{\rho_{u,2}}{\sqrt{n}} U_i + \eta_i, && (47) \end{aligned}$$

where $X_{X,i}$ is an endogenous regressor, and $X_{1,i}$ (2×1) and $X_{2,i}$ (1×1) are two vectors of instruments. While $X_{1,i}$ is assumed to be uncorrelated with the error term U_i, $X_{2,i}$ is potentially positively correlated with U_i, that is $\rho_{u,2} \geq 0$. We assume that the random vector $V_i = [U_i, \epsilon_i', \eta_i', X_{1,i}']'$ is independently and identically distributed and satisfies the following moment conditions: $\mathbb{E}[X_{1,i}'U_i] = 0$, $\mathbb{E}[X_{1,i}'\eta_i] = 0$, and $\mathbb{E}[\epsilon_i'U_i] \neq 0$. Let $X_i = [X_{Y,i}, X_{X,i}, X_{1,i}', X_{2,i}]'$ and define

$$\begin{aligned} g_1(X_i, \theta) &= X_{1,i}(X_{Y,i} - X_{X,i}\theta) && (48) \\ g_2(X_i, \theta) &= X_{2,i}(X_{Y,i} - X_{X,i}\theta). && (49) \end{aligned}$$

Point and interval estimation will be based on the moment conditions

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

for $\theta = \theta_0$. Using the notation of Sections 2 to 4, $\nu_0 = 0$ and $u_0 = \rho_{u,2}\mathbb{E}[U_i^2]$. 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

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

5.1 Parameterization

Since we are simulating the limit distribution of our estimators and confidence sets, we only have to parameterize the matrices J and Q. In order to make the numerical values easier to interpret we derive them from the simultaneous equations model specified above. Suppose that 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.

The matrix J is determined by the covariance matrix of the instruments $X_{Z,i} = [X_{1,i}', X_{2,i}]'$. We assume that the instruments $X_{1,i}$ have a unit covariance matrix and that $\rho_{1,2}'\rho_{1,2} < 1$. Let $\sigma_\eta^2 = 1 - \rho_{1,2}'\rho_{1,2}$ and $\mathbb{E}[U_i^2] = 1$ such that

$$J = \begin{bmatrix} I & \rho_{1,2} \\ & 1 \end{bmatrix}.$$

The vector Q is a function of the correlation between the instruments and the endogenous regressor, denoted by the 2×1 vector $\rho_{1,X}$ and the scalar $\rho_{2,X}$ (1×1). We impose that $X_{X,i}$ has unit variance and obtain[3]

$$Q = -\begin{bmatrix} \rho_{1,X}' & \rho_{2,X} \end{bmatrix}.$$

Hence, the relevant design parameters for the data generating process (DGP) are $u_0 = \rho_{u,2}$, $\rho_{1,2}$, $\rho_{1,X}$, and $\rho_{2,X}$. We consider three different parameterizations of the DGP, listed in Table 1. DGP 1 can be viewed as a benchmark. The correlations between the three instruments and the endogenous regressors are equal to 0.5. $X_{2,i}$ is positively correlated with the first element of $X_{1,i}$ and slightly negatively correlated with the second. For DGP 2 we increase the correlation between $X_{1,i}$ and $X_{2,i}$ by reducing the variance of η_i in Equation (47). This will make it easier to estimate u_0. Finally, we consider a parameterization in which we lower the correlation between the instruments $X_{1,i}$ and the endogenous regressor to 0.3.

[3]Based on $\rho_{1,X}$, $\rho_{2,X}$, and $\rho_{1,2}$ it is possible to calculate γ_1, γ_2, and σ_ϵ^2. While not all choices of the correlation parameters are consistent with $\sigma_\epsilon^2 > 0$, the ones reported in the paper lead to a positive variance.

5.2 Alternative Estimators and Confidence Sets

In order to assess the asymptotic performance of the proposed point estimator we consider two alternatives, using the following notation:

(i) $\hat{\theta}_{(0)}$ is MELE based on $\mathbb{E}[g_1(X_i, \theta)] = 0$ and $\mathbb{E}[g_2(X_i, \theta)] \geq 0$.

(ii) $\hat{\theta}_{(1)}$ is MELE based on $\mathbb{E}[g_1(X_i, \theta)] = 0$.

(iii) $\hat{\theta}_{(12)}$ is MELE based on $\mathbb{E}[g_1(X_i, \theta)] = 0$ and $\mathbb{E}[g_2(X_i, \theta)] = 0$.

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. Thus,

$$\sqrt{n}(\hat{\theta}_{(1)} - \theta_0) \Longrightarrow \mathcal{N}\Big(0, (Q_1 J_{11}^{-1} Q_1')^{-1}\Big).$$

Numerical values for the asymptotic standard deviation of the estimator can be found in Table 1. 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

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

The larger u_0, the larger the bias of the estimator that incorrectly imposes $\mathbb{E}[g_2(X_i, \theta_0)] = 0$.

In order to conduct inference with respect to θ_0 and u_0 we consider two types of confidence sets:

(i) $\mathcal{CS}_{(0)}^{\theta}$ and $\mathcal{CS}_{(0)}^{u}$ are obtained based on $\mathbb{E}[g_1(X_i, \theta)] = 0$ and $\mathbb{E}[g_2(X_i, \theta)] \geq 0$ as described in Section 4. In computing the Bonferroni interval we set $\alpha_2 = 0$ and $\alpha_1 = \alpha$ such that $\mathcal{CS}_{(0)}^{\theta} = \bigcup_{u \geq 0} \mathcal{CS}_{(0)}^{\theta}(u, \alpha)$.

(ii) $\mathcal{CS}_{(1)}^{\theta}$ is obtained based on $\mathbb{E}[g_1(X_i, \theta)] = 0$. We invert the empirical likelihood ratio test for the hypothesis $\theta_0 = \theta^H$.

(iii) $\mathcal{CS}_{(1)}^{u}$ is obtained based on $\mathbb{E}[g_1(X_i, \theta)] = 0$. The confidence set is constructed by inverting the Wald test statistic for the hypothesis $u_0 = u^H$. The test statistic is constructed as follows

$$\mathcal{W}_{(1)}^{u} = \left(\frac{\max\{-u^H, \hat{u}_{(1)} - u^H\}}{v^{1/2}(\hat{u}_{(1)})} \right),$$

where

$$\begin{aligned} \hat{u}_{(1)} &= \frac{1}{\sqrt{n}} \sum X_{2,i}(X_{Y,i} - X_{X,i}\hat{\theta}_{(1)}) \\ &= u_0 + Z_{2,n} - Q_{2,n}'(Q_{1,n} J_{11,n}^{-1} Q_{1,n}')^{-1} Q_{1,n} J_{11,n}^{-1} Z_{1,n} + o_p(1). \end{aligned}$$

The asymptotic variance of this estimator is

$$v(\hat{u}_{(1)}) = J_{22} + Q_2'(Q_1 J_{11}^{-1} Q_1')^{-1} Q_2 - 2Q_2'(Q_1 J_{11}^{-1} Q_1') Q_1 J_{11}^{-1} J_{12}.$$

Numerical values for the three DGPs are provided in Table 1.

5.3 Numerical Results

All numerical results reported subsequently are based on 100,000 draws from the limit distribution.

Table 2 reports the bias and mean squared error (MSE) for the three empirical likelihood estimators. As we previously showed, the limit distribution of $\hat{\theta}_{(1)}$ is not affected by u_0. The estimator is asymptotically unbiased and its MSE is equal to 2 under DGP 1. 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 same pattern emerges under DGP 2 and DGP 3.

Table 3 summarizes the performance of the confidence intervals for θ_0. The coverage probability is 90 percent and we report the average lengths of the confidence intervals. The simulation of the confidence intervals $\mathcal{CS}^{\theta}_{(1)}$ involves several steps. Without loss of generality we set $\theta_0 = 0$ and let $s = \sqrt{n}\theta$. First, we specify grids for u and s. For simplicity, we will denote these grids by S and U. Second, we generate draws from the asymptotic distribution of Z_n and simulate the empirical likelihood ratio statistics for each $u_0 \in U$. Based on the output of this simulation it is possible to approximate the critical values $c^{\theta}_{\alpha}(u_0)$. Third, we fix a u_0, simulate the empirical likelihood ratio statistic again, and determine for each $s \in S$ and $u \in U$ whether $\mathcal{LR}^{\theta}(\sqrt{n}s) \leq c_{\alpha}(u)$. This will lead to $\mathcal{CS}^{\theta}(u, \alpha)$. We then take the union of these confidence intervals over α to obtain $\mathcal{CS}^{(\theta)}_{(0)}$. The simulation of $\mathcal{CS}^{\theta}_{(1)}$ is considerably easier because the critical values of the empirical likelihood ratio statistic do not depend on u_0 and can simply be obtained from a χ^2 distribution.

The interval based on $\mathbb{E}[g_1(X_i, \theta_0)] = 0$ only is not affected by the local slackness parameter u_0. Its (scaled) length under DGP 1 is 4.62. The use of the inequality moment condition sharpens the inference. For $u_0 = 0$ the interval $\mathcal{CS}^{\theta}_{(0)}$ has a length of 4.39. As u_0 increases and the information in the inequality moment condition vanishes, its length expands to 4.62. A similar pattern emerges under DGP 2.

Results for the u_0 confidence intervals are reported in Table 4. Unlike $\mathcal{CS}^{\theta}_{(1)}$, the length of the interval $\mathcal{CS}^{u}_{(1)}$ varies with u_0. If u_0 is near zero, the distribution of the Wald statistic $\mathcal{W}^{u}_{(1)}$ has a point mass near zero that keeps the confidence short, since the domain of u_0 is bounded below by zero. As u_0 increases, the point mass at zero vanishes and the confidence interval becomes longer. For all values of u_0 reported in the table $\mathcal{CS}^{u}_{(0)}$ dominates $\mathcal{CS}^{u}_{(1)}$ and our procedure is able to exploit the additional information contained in the second moment condition. The percentage gain over $\mathcal{CS}^{u}_{(1)}$ is largest for DGP 2, under which $X_{2,i}$ is strongly correlated with the first element of $X_{1,i}$.

6 Conclusion

This paper developed a limit distribution theory for empirical likelihood estimators when some of the moment conditions take the form of inequalities. The inequality moment conditions provide additional information if they are close to zero. The limit distribution of the parameter estimators and empirical likelihood ratio statistics typically depend on a nuisance parameter that measures the slack in the inequality conditions. This nuisance parameter complicates statistical inference because it cannot be estimated consistently. We constructed Bonferroni type confidence sets for the parameter of interest, θ, by taking a union of sets that are valid for a particular value of the nuisance parameter. While the focus of this paper has been interval estimation, the nuisance parameter problem also arises in the context of hypothesis tests. The null distribution of an empirical likelihood ratio coefficient test is a function of u_0 and the testing problem becomes that of testing a composite hypothesis, which has been studied, for instance, by Berger and Boos (1994), Hansen (2003), and Silvapulle (1996). Finally, we have assumed throughout the paper that the parameter θ is identifiable based on the equality moment condition $\mathbb{E}[g_1(X_i, \theta_0)] = 0$. Relaxing this assumption would imply that the model parameters are likely to be only set identifiable rather than point-identifiable. We leave this interesting extension for future research.

A Proofs and Derivations

The Appendix contains detailed proofs and derivations for the results presented in the main text. Section A.1 shows the equivalence of the three formulations of the saddlepoint problem discussed in Section 2. Section A.2 contains the consistency proof. By and large, we follow the structure of the proofs in Newey and Smith (2004), making the necessary adjustments for the presence of the inequality moment conditions. In Section A.3 the quadratic approximation of the objective function is obtained. We use Lemma 1(a) of Andrews (1999) to bound the remainder term in the second-order Taylor approximation of the objective function. The proof of $\sqrt{n}$ consistency differs from Andrews (1999) because he studied an extremum estimator and we are studying a saddlepoint estimator. The proof also differs from Newey and Smith (2004), who expand the first-order condition associated with the saddlepoint, whereas we work with the quadratic approximation of the objective function. Based on the asymptotic approximation of the empirical likelihood objective function, we derive limit distributions for point and interval estimators in Sections A.4 and A.5.

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

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

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

A.2.1 Main Result

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

$$\lim_{n \longrightarrow \infty} P\left\{\hat{\theta}_n \in \mathbb{B}(\theta_0, \delta), \hat{\nu}_n \in \mathbb{B}(\nu_{n,0}, \delta)\right\} = 1,$$

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

$$\Theta_0^c = \Theta \cap \mathbb{B}\left(\theta_0, \delta\right)^c \quad \text{and} \quad N_0^c = \mathbb{R}^{h_2+} \cap \mathbb{B}\left(\nu_{n,0}, \delta\right)^c.$$

To simplify the notation we omit the subscript n from the set N_0^c. Recall that according to Assumption 5 the constant $\alpha > 2$ is such that $\mathbb{E}[\sup_{\theta\in\Theta} \|g(X, \theta)\|^\alpha] < \infty$. 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 $n \geq \bar{n}$

$$P\left\{\bar{G}_n^*\left(\theta_0, \nu_{n,0}\right) \geq n^{-\zeta-\kappa}\eta\right\} < \frac{\varepsilon}{2} \tag{A.1}$$

and (ii)

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

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

$$\begin{aligned} &P\left\{\hat{\theta}_n \in \mathbb{B}\left(\theta_0, \delta\right),\ \hat{\nu}_n \in \mathbb{B}(\nu_{n,0}, \delta)\right\} \\ \geq\ &P\left\{\bar{G}_n^*\left(\theta_0, \nu_{n,0}\right) < n^{-\zeta-\kappa}\eta,\ \min_{\theta\in\Theta_0^c,\, \nu\in N_0^c} \bar{G}_n^*\left(\theta, \nu\right) > n^{-\zeta}\eta\right\} \geq 1 - \epsilon. \end{aligned}$$

Proof of (i): By Lemma A.2

$$\bar{G}_n^*(\theta_0, \nu_{n,0}) = \max_{\lambda\in\hat{\Lambda}_n(\theta_0)} G_n^*(\theta_0, \nu_{n,0}, \lambda) \leq O_p(1/n).$$

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

$$n^{\zeta+\kappa}\bar{G}_n^*(\theta_0, \nu_{n,0}) \leq O_p(n^{\zeta+\kappa-1}) = o_p(1)$$

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

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

such that $\|u(\theta,\nu)\| \le 1$. Strictly speaking, the function $u(\theta,\nu)$ depends through $\nu_{n,0}$ on the sample size n, but for notational convenience the n subscript is omitted.

Moreover, we truncate the function $g(x,\theta)$ as follows. Since $\alpha > 2$, we can choose a positive constant ξ such that

$$\frac{1}{\alpha^2} < \xi < \frac{1}{2\alpha}.$$

Let

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

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

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

In what follows, we deduce the required result for (ii) by showing that

$$\text{(ii)-(a):} \quad \min_{\theta\in\Theta_0^c,\nu\in N_0^c} \frac{1}{n}\sum_{i=1}^n q_n(X_i,\theta,\nu) \le \min_{\theta\in\Theta_0^c,\nu\in N_0^c} \bar{G}_n^*(\theta,v) + o_p\left(n^{-\zeta}\right)$$

and

$$\text{(ii)-(b):} \; P\left\{ \min_{\theta\in\Theta_0^c,\nu\in N_0^c} \frac{1}{n}\sum_{i=1}^n q_n(X_i,\theta,\nu) < n^{-\zeta}\eta \right\} \le \frac{\varepsilon}{2}.$$

Proof of (ii)-(a): Notice that $n^{-\zeta} u'(\theta,\nu) \in \Lambda_n^{\zeta} \subset \cap_{\theta\in\Theta} \hat{\Lambda}_n(\theta)$ w.p.a.1 by Lemma A.1. Then, by Lemma A.5 and by the definition of $\hat{\lambda}_n(\theta,v)$,

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

as required.

Proof of (ii)-(b): A second-order Taylor expansion of q_n around $u(\theta,\nu) = 0$ yields

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

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

$$\sup_{\theta\in\Theta,\ \nu}\left|n^{-\zeta}u^{*\prime}(\theta,\nu)\,g_n(x,\theta)\right| \le n^{-\zeta}\sup_{\theta\in\Theta}\|g_n(x,\theta)\| \le n^{-\zeta+\xi} \le n^{-\zeta/2}$$

since $\xi < \frac{1}{2\alpha} < \frac{\zeta}{2}$. Therefore,

$$\sup_{\theta\in\Theta,\ \nu} n^{-\zeta}\frac{u(\theta,\nu)'g_n(x,\theta)g_n(x,\theta)'u(\theta,\nu)}{\left(1+n^{-\zeta}u_*(\theta,\nu)'g_n(x,\theta)\right)^2} \le \sup_{\theta\in\Theta,\ \nu} n^{-\zeta}\frac{\|g_n(x,\theta)\|^2\|u(\theta,\nu)\|^2}{(1-n^{-\zeta/2})^2} \le n^{-\zeta+2\xi} = o(1)\,. \tag{A.4}$$

Now consider the expected value of $n^{\zeta}q_n(x,\theta,\nu)$. From (A.3), (A.4), and by the dominated convergence theorem, we have

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

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

$$\lim_{n\to\infty} n^{\zeta}\lim_{\delta\downarrow 0}E\left[\inf_{\theta^*\in\mathbb{B}(\theta,\delta),\ \nu^*\in\mathbb{B}(\nu,\delta)} q_n(X,\theta^*,\nu^*)\right]\begin{cases} =0 & \text{if } \theta=\theta_0,\ \nu=\nu_{n0} \\ >0 & \text{otherwise}\end{cases}.$$

Next, according to Assumption 5 there exists a finite K such that

$$\sup_{\theta\in\Theta}\|E[g_2(X,\theta)]\| < K < \infty. \tag{A.6}$$

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 $\mathbb{R}_K^{h_2+}\cap\mathbb{B}(\nu_{n,0},\delta)^c$ 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

$$n^{\zeta}E\left[\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n(X,\theta,\nu)\right] \ge 2\eta_j,\ n\ge n_j \tag{A.7}$$

for some positive numbers $\eta_j=\eta_j(\delta)$, $j=1,\ldots,J$. By the WLLN[4] and (A.7) , for a given

[4] Notice that

$$E\left|n^{\zeta}\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n(X,\theta,\nu)\right| \le E\left[\sup_{\theta\in\Theta}\|g(X,\theta)\|\right]+2K+n^{-\zeta}\frac{E\left[\sup_{\theta\in\Theta}\|g(X,\theta)\|\right]}{\left(1-n^{-\zeta/2}\right)^2} < \infty. \tag{A.8}$$

$\varepsilon > 0$, we can find $\bar{n}_j' s$ such that $n \geq \bar{n}_j$ implies that

$$\begin{aligned}
\frac{\varepsilon}{4J} &\geq P\left\{\left|\frac{1}{n}\sum_{i=1}^{n} n^{\zeta} \inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n\left(X_i,\theta,\nu\right) - E\left[n^{\zeta}\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n\left(X_i,\theta,\nu\right)\right]\right| > \eta_j\right\} \\
&\geq P\left\{\frac{1}{n}\sum_{i=1}^{n}\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n\left(X_i,\theta,\nu\right) < E\left[\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n\left(X_i,\theta,\nu\right)\right] - n^{-\zeta}\eta_j\right\} \\
&\geq P\left\{\frac{1}{n}\sum_{i=1}^{n}\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n\left(X_i,\theta,\nu\right) < n^{-\zeta}\eta_j\right\} \\
&\geq P\left\{\inf_{\theta\in\Theta_j,\ \nu\in N_j}\frac{1}{n}\sum_{i=1}^{n} q_n\left(X_i,\theta,\nu\right) < n^{-\zeta}\eta_j\right\}
\end{aligned}$$

for $j = 1, \ldots, J$. Also, after this proof we show that w.p.a.1

$$\inf_{\theta\in\Theta,\ \|\nu\|>2K}\frac{1}{n}\sum_{i=1}^{n} n^{\zeta} q_n\left(X_i,\theta,\nu\right) \geq K. \tag{A.9}$$

For the given ε, then, we can choose an $\bar{n}_{J+1}$ such that $n \geq \bar{n}_{J+1}$ implies that

$$P\left\{\inf_{\theta\in\Theta,\ \|\nu\|>2K}\frac{1}{n}\sum_{i=1}^{n} q_n\left(X_i,\theta,\nu\right) < n^{-\zeta}K\right\} \leq \frac{\varepsilon}{4}.$$

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 $n \geq \bar{n}$

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

as required part (ii)-(b).

Combining (ii)-(a) and (ii)-(b) we have

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

as required for (ii).

Since $\hat{\theta}_n \xrightarrow{p} \theta_0$ and $\hat{\nu}_n - \nu_{n,0} \xrightarrow{p} 0$ we can deduce from Lemmas A.2 and A.3 that $\hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{p} 0$. ■

A.2.2 Technical Lemmas

Proof of (A.9): Notice from (A.3) and (A.4) that

$$n^{\zeta} q_n (X_i, \theta, \nu) \geq u(\theta, \nu)' (g_n (X_i, \theta) - M'\nu) - \frac{1}{2} n^{-\zeta+2\xi}.$$

Then we have

$$\begin{aligned}
&\inf_{\theta\in\Theta,\ \|\nu\|>2K} \frac{1}{n}\sum_{i=1}^{n} n^{\zeta} q_n (X_i, \theta, \nu) \\
\geq &\inf_{\theta\in\Theta,\ \|\nu\|>2K} \frac{1}{n}\sum_{i=1}^{n} u(\theta,\nu)' E\left[(g(X_i,\theta) - M'\nu)\right] \\
&+ \inf_{\theta\in\Theta,\ \|\nu\|>2K} \frac{1}{n}\sum_{i=1}^{n} u(\theta,\nu)' (g_n(X_i,\theta) - E\left[g(X_i,\theta)\right]) - \frac{1}{2} n^{-\zeta+2\xi}.
\end{aligned}$$

First, by the definition of $u(\theta, \nu)$, we have

$$\begin{aligned}
&\inf_{\theta\in\Theta,\ \|\nu\|>2K} \frac{1}{n}\sum_{i=1}^{n} u(\theta,\nu)' E\left[(g(X_i,\theta) - M'\nu)\right] \\
= &\inf_{\theta\in\Theta,\ \|\nu\|>2K} \|E[g(X,\theta)] - M'\nu\| \geq \inf_{\theta\in\Theta,\ \|\nu\|>2K} \|E[g_2(X,\theta)] - \nu\| \\
\geq &\inf_{\theta\in\Theta,\ \|\nu\|>2K} \left[\|\nu\| - \|E[g_2(X,\theta)]\|\right] \geq 2K - \sup_{\theta\in\Theta} \|E[g_2(X,\theta)]\| \\
\geq & K.
\end{aligned}$$

Next, by the Cauchy-Schwarz inequality and the definition of $g_n (X_i, \theta)$,

$$\begin{aligned}
&\inf_{\theta\in\Theta,\ \|\nu\|>2K} u(\theta,\nu)' \left[\frac{1}{n}\sum_{i=1}^{n}\Big(g_n(X_i,\theta) - E[g(X_i,\theta)]\Big)\right] \\
\geq & -\sup_{\theta\in\Theta,\ \|\nu\|>2K} \left\|\frac{1}{n}\sum_{i=1}^{n}\left(g(X_i,\theta) I\left\{\sup_{\theta\in\Theta}\|g(X_i,\theta)\| \leq n^{\xi}\right\} - E[g(X_i,\theta)]\right)\right\| \\
\geq & -\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\Big(g(X_i,\theta) - E[g(X_i,\theta)]\Big)\right\| - \sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta) I\left\{\sup_{\theta\in\Theta}\|g(X_i,\theta)\| > n^{\xi}\right\}\right\|.
\end{aligned}$$

The first term is

$$\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\Big(g(X_i,\theta) - E[g(X_i,\theta)]\Big)\right\| = o_p(1)$$

by the ULLN. The second term is

$$\begin{aligned}
&-\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta) I\left\{\sup_{\theta\in\Theta}\|g(X_i,\theta)\| > n^{\xi}\right\}\right\| \\
\geq & -\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}\|g(X_i,\theta)\| I\left\{\sup_{\theta\in\Theta}\|g(X_i,\theta)\| > n^{\xi}\right\} \\
\geq & -n^{-\xi(\alpha-1)}\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}\|g(X_i,\theta)\|^{\alpha} \\
= & O_p\left(n^{-\xi(\alpha-1)}\right) = o_p(1).
\end{aligned}$$

These give

$$\inf_{\theta\in\Theta,\ \|\nu\|>2K} u(\theta,\nu)' \left[\frac{1}{n}\sum_{i=1}^{n}\left\{g_n(X_i,\theta) - E\left[g(X_i,\theta)\right]\right\}\right] \geq o_p(1).$$

Therefore we have

$$\inf_{\theta\in\Theta,\ \|\nu\|>2K} \frac{1}{n}\sum_{i=1}^{n} n^{\zeta} q_n(X_i,\theta,\nu) \geq K + o_p(1),$$

as required. ■

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

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

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

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

Lemma A.2 *Suppose that Assumptions 1 to 5 are satisfied. Let* $\bar{\theta} \in \Theta$ *and* $\bar{\nu} \geq 0$ *be sequences such that* $\bar{\theta} \xrightarrow{p} \theta_0$, *and* $\bar{\nu} - \nu_{n,0} \xrightarrow{p} 0$. *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) $G_n^*\left(\bar{\theta},\bar{\nu},\hat{\lambda}(\bar{\theta},\bar{\nu})\right) \leq O_p\left(\frac{1}{n}\right)$.

Proof of Lemma A.2:

Proof of (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.

Proof of (ii) and (iii): Write $\bar{g}_i = g(X_i,\bar{\theta})$. For some constant C

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

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

$$\max_{1\leq i\leq n} |\lambda_*' \bar{g}_i| = o_p(1)$$

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 Assumptions 1 to 5 are satisfied. Then,*

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

Proof of Lemma A.3: 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 $u(\theta, \nu)$ in the proof of consistency.)

Approximation $G_n^*(\theta, \nu, \lambda)$ with respect to λ around $\lambda = 0$ at $(\theta, \nu, \lambda) = \left(\hat{\theta}, \hat{\nu}, \hat{u}\left(\hat{\theta}, \hat{\nu}\right)\right)$. 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 $\max_{1\leq i\leq n}\left|\hat{u}\left(\hat{\theta}, \hat{\nu}\right)'\hat{g}_i\right| \to_p 0$ 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, $\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i\hat{g}_i' \leq \left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\|g(X_i, \theta)\|\right) I \to_p CI$. Then,

$$\begin{aligned}
& \hat{g}'\hat{u}\left(\hat{\theta}, \hat{\nu}\right) - \frac{1}{2}\hat{u}\left(\hat{\theta}, \hat{\nu}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}_i\hat{g}_i'}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta}, \hat{\nu}\right) \\
= \;& n^{-\zeta}\|\hat{g}\| - \frac{1}{2}\hat{u}\left(\hat{\theta}, \hat{\nu}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{g}_i\hat{g}_i'}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta}, \hat{\nu}\right) \\
\geq \;& n^{-\zeta}\|\hat{g}\| - \frac{1}{2}\max_{1\leq i\leq n}\left(\frac{1}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta}, \hat{\nu}\right)'\left(\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i\hat{g}_i'\right)\hat{u}\left(\hat{\theta}, \hat{\nu}\right) \\
\geq \;& n^{-\zeta}\|\hat{g}\| - Cn^{-2\zeta}. \qquad \text{(A.10)}
\end{aligned}$$

Then,

$$n^{-\zeta}\|\hat{g}\|-Cn^{-2\zeta}\leq G_n^*\left(\hat{\theta},\hat{\nu},\hat{u}\left(\hat{\theta},\hat{\nu}\right)\right)\leq G_n^*\left(\hat{\theta},\hat{\nu},\hat{\lambda}\right)\leq\sup_{\lambda\in\hat{\Lambda}_n(\theta_0)}G_n^*\left(\theta_0,\nu_{n,0},\lambda\right)\leq O_p\left(\frac{1}{n}\right), \tag{A.11}$$

where the first inequality is from (A.10), 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 $\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[g\left(X_i,\theta_0\right)-M'\nu_{n,0}\right]=O_p\left(1\right)$ by the CLT. Also, by $\zeta<\frac{1}{2}$, $\zeta-1<-\frac{1}{2}<-\zeta$. Solving (A.11) for $\|\hat{g}\|$ gives

$$\|\hat{g}\|\leq O_p\left(n^{-\zeta}\right). \tag{A.12}$$

Now, for a given sequence $\varepsilon_n\to 0$, let $\bar{\lambda}=\varepsilon_n\hat{g}$. By (A.12) , $\bar{\lambda}=o_p\left(n^{-\zeta}\right)$, and so $\bar{\lambda}\in\Lambda_n^\zeta$ w.p.a.1. Then, as in (A.11) , we have

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

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. ■

Lemma A.4 *Suppose that Z_i is a sequence of iid random variables such that $\mathbb{E}\left|Z_i\right|^\alpha<\infty$. Then,* $\max_{1\leq i\leq n}|Z_i|=O_p\left(n^{1/\alpha}\right)$.

Proof of Lemma A.4: The result follows from

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

Lemma A.5 *Assume Assumptions 1 to 5. Let $g_n(x,\theta)=I\{x\in\mathcal{X}_n\}g(x,\theta)$ where*

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

where $\frac{1}{\alpha^2}<\xi<\frac{1}{2\alpha}$ and $\alpha>2$ as in Assumption 5. Define

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

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

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

Proof of Lemma A.5: By the mean value theorem,

$$
\begin{aligned}
& \sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{q_n\left(X_i,\theta,\nu\right)-\tilde{q}_n\left(X_i,\theta,\nu\right)\right\}\right| \\
= & \sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n^{-\zeta}u'(\theta,\nu)g\left(X_i,\theta\right)}{1+n^{-\zeta}u_*'\left(\theta,\nu\right)g\left(X_i,\theta\right)}\right)I\left\{X_i\notin\mathcal{X}_n\right\}\right| \qquad (A.13)\\
\leq & \max_{1\leq i\leq n}\sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{n^{-\zeta}u'\left(\theta,\nu\right)g\left(X_i,\theta\right)}{1+n^{-\zeta}u_*'\left(\theta,\nu\right)g\left(X_i,\theta\right)}\right|\frac{1}{n}\sum_{i=1}^{n}I\left\{\sup_{\theta\in\Theta}\left\|g\left(X_i,\theta\right)\right\|>n^{\xi}\right\}\\
\leq & \frac{1}{n^{\alpha\xi}}\left(\max_{1\leq i\leq n}\sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{n^{-\zeta}u'\left(\theta,\nu\right)g\left(X_i,\theta\right)}{1+n^{-\zeta}u_*'\left(\theta,\nu\right)g\left(X_i,\theta\right)}\right|\right)\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}\left\|g\left(X_i,\theta\right)\right\|^{\alpha}\right)
\end{aligned}
$$

where $u_*(\theta,\nu)$ is located between 0 and $u(\theta,\nu)$. The second term on the right-hand side of (A.13) can be bounded as follows. According to Lemma A.4

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

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

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

By Assumption 5 and the Markov inequality, the third term on the right-hand side of (A.13) is $O_p(1)$. Since $\frac{1}{\alpha^2}<\xi<\frac{1}{2\alpha}$, we are able to deduce that

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

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

$$G_n^{*(1)}\left(\beta\right)=\left[G_n^{*(1)}\left(\beta\right)_\theta',G_n^{*(1)}\left(\beta\right)_\nu',G_n^{*(1)}\left(\beta\right)_\lambda'\right]',\qquad (A.14)$$

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 $\beta_{n,0}$ the first derivatives simplify to

$$G_n^{*(1)}(\beta_{n,0})_\theta = 0, \quad G_n^{*(1)}(\beta_{n,0})_\nu = 0, \quad G_n^{*(1)}(\beta_{n,0})_\lambda = \frac{1}{n}\sum_{i=1}^n g(X_i,\theta_0) - M'\nu_{n,0} = n^{-1/2}Z_n,$$

which leads to the formula for $G_n^{*(1)}(\beta_{n,0})$ that appears in Equation (21) of the main text. We proceed by partitioning the matrix of second derivative as follows

$$G_n^{*(2)}(\beta) = \begin{pmatrix} G_n^{*(2)}(\beta)_{\theta\theta'} & G_n^{*(2)}(\beta)_{\theta\nu'} & G_n^{*(2)}(\beta)_{\theta\lambda'} \\ G_n^{*(2)}(\beta)_{\nu\theta'} & G_n^{*(2)}(\beta)_{\nu\nu'} & G_n^{*(2)}(\beta)_{\nu\lambda'} \\ G_n^{*(2)}(\beta)_{\lambda\theta'} & G_n^{*(2)}(\beta)_{\lambda\nu'} & G_n^{*(2)}(\beta)_{\lambda\lambda'} \end{pmatrix}, \tag{A.15}$$

where

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

At $\beta_{n,0}$ the second derivatives simplify to

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

which leads to the formula for $G_n^{*(2)}(\beta_{n,0})$ 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,

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

A.3.1 Main Results

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

$$\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n} \left\| G_n^{*(2)}(\beta) - G_n^{*(2)}(\beta_{n,0}) \right\| = o_p(1),$$

for every sequence $\gamma_n \longrightarrow 0$. $G_n^{*(2)}$ is defined in (A.15). To verify this sufficient condition we will subsequently show that

(i) $\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\leq\gamma_n}\left\|G_n^{*(2)}(\beta)_{\theta\theta'}-G_n^{*(2)}(\beta_{n,0})_{\theta\theta'}\right\|=o_p(1)$,

(ii) $\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\leq\gamma_n}\left\|G_n^{*(2)}(\beta)_{\lambda\theta'}-G_n^{*(2)}(\beta_{n,0})_{\lambda\theta'}\right\|=o_p(1)$,

(iii) $\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\leq\gamma_n}\left\|G_n^{*(2)}(\beta)_{\lambda\lambda'}-G_n^{*(2)}(\beta_{n,0})_{\lambda\lambda'}\right\|=o_p(1)$.

We begin by showing that

$$\sup_{\beta\in\mathcal{B}_n}\left|\frac{1}{1+\lambda' g(X_i,\theta)}\right|=O_p(1). \tag{A.16}$$

Since

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

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

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

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

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

which proves (A.16).

(i) Notice that

$$\begin{aligned}&\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\leq\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lambda_j g_j^{(2)}(X_i,\theta)}{1+\lambda' g(X_i,\theta)}\right)\right\|\\ \leq\ & \sup_{\lambda\in\Lambda_n^\zeta}|\lambda_j|\left(\sup_{\beta\in\mathcal{B}_n,1\leq i\leq n}\left|\frac{1}{1+\lambda' g(X_i,\theta)}\right|\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum\left\|g_j^{(2)}(X_i,\theta)\right\|\right)\\ =\ & O(n^{-\zeta})O_p(1)O_p(1)=o_p(1),\end{aligned}$$

where the last inequality holds by the definition of Λ_n^ζ, (A.16) and the ULLN under Assumption 6. Moreover,

$$\begin{aligned}&\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\leq\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g^{(1)}(X_i,\theta)'\lambda\lambda' g^{(1)}(X_i,\theta)}{(1+\lambda' g(X_i,\theta))^2}\right)\right\|\\ \leq\ & \sup_{\lambda\in\Lambda_n^\zeta}\|\lambda_k\|^2\left(\sup_{\beta\in\mathcal{B}_n,\ 1\leq i\leq n}\frac{1}{(1+\lambda' g(X_i,\theta))^2}\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\left\|g^{(1)}(X_i,\theta)\right\|\right)\\ =\ & O\left(n^{-2\zeta}\right)O_p(1)\,O_p(1)=o_p(1).\end{aligned}$$

The last inequality holds by the definition of Λ_n^ζ, (A.16) and the ULLN under Assumption 6.

(ii) Apply the triangle inequality to

$$\begin{aligned}
&\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^n\left(\frac{g^{(1)}(X_i,\theta)}{1+\lambda' g(X_i,\theta)}-g^{(1)}(X_i,\theta_{n,0})\right)\right\| \\
\le\ & \sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^n\left(\frac{g^{(1)}(X_i,\theta)}{1+\lambda' g(X_i,\theta)}-g^{(1)}(X_i,\theta)\right)\right\| \\
&+\sup_{\theta\in\Theta}\left\|\frac{1}{n}\sum_{i=1}^n\left(g^{(1)}(X_i,\theta)-\mathbb{E}\left[g^{(1)}(X_i,\theta)\right]\right)\right\| \\
&+\sup_{\theta\in\Theta:\|\theta-\theta_0\|\le\gamma_n}\left\|\mathbb{E}\left[g^{(1)}(X_i,\theta)\right]-E\left[g^{(1)}(X_i,\theta_0)\right]\right\| \\
&+\left\|\frac{1}{n}\sum_{i=1}^n\left(g^{(1)}(X_i,\theta_0)-\mathbb{E}\left[g^{(1)}(X_i,\theta_0)\right]\right)\right\| \\
=\ & I_d+o_p(1)+o_p(1)+o_p(1),
\end{aligned}$$

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

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

by Lemma A.1, (A.16), and the ULLN under Assumption 6. Moreover,

$$\begin{aligned}
&\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^n\frac{g(X_i,\theta)}{1+\lambda' g(X_i,\theta)}\frac{\lambda\prime g(X_i,\theta)}{1+\lambda' g(X_i,\theta)}\right\| \\
\le\ & \sup_{\lambda\in\Lambda_n^\zeta}\|\lambda\|\left(\sup_{\beta\in\mathcal{B}_n,1\le i\le n}\frac{1}{(1+\lambda' g(X_i,\theta))^2}\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum\|g(X_i,\theta)\|^2\right) \\
=\ & O\left(n^{-\zeta}\right)O_p(1)\,O_p(1)=o_p(1).
\end{aligned}$$

(iii) Similar as before, we have

$$
\begin{aligned}
&\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'\right)\right\| \\
\le\quad &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right)\right\| \\
&+\sup_{\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X_i,\theta\right)g\left(X_i,\theta\right)'-I\!E\left[g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right]\right)\right\| \\
&+\sup_{\Theta}\left\|I\!E\left[g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right]-I\!E\left[g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'\right]\right\| \\
&+\sup_{\Theta}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'-I\!E\left[g\left(X_i,\theta_0\right)g\left(X_i,\theta_0\right)'\right]\right)\right\| \\
=\quad &\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right)\right\|+o_p\left(1\right).
\end{aligned}
$$

Next,

$$
\begin{aligned}
&\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{g\left(X_i,\theta\right)g\left(X_i,\theta\right)'}{\left(1+\lambda' g\left(X_i,\theta\right)\right)^2}-g\left(X_i,\theta\right)g\left(X_i,\theta\right)'\right)\right\| \\
\le\quad &\sup_{\beta\in\mathcal{B}_n,1\le i\le n}\left|\lambda' g\left(X_i,\theta\right)\right|\left(\sup_{\beta\in\mathcal{B}_n,1\le i\le n}\frac{1}{\left|1+\lambda' g\left(X_i,\theta\right)\right|}\right) \\
&\times\left(\sup_{\beta\in\mathcal{B}_n,1\le i\le n}\frac{1}{\left|1+\lambda' g\left(X_i,\theta\right)\right|}+1\right)\left(\sup_{\theta\in\Theta}\frac{1}{n}\sum_{i=1}^{n}\left\|g\left(X_i,\theta\right)\right\|^2\right) \\
=\quad &o_p\left(1\right)O_p\left(1\right)O_p\left(1\right)O_p\left(1\right)=o_p\left(1\right). \quad\blacksquare
\end{aligned}
$$

Proof of Theorem 2 is omitted. ■

Proof of Theorem 3: (i) Follows from Lemma A.7.

(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

$$
0=\mathcal{G}_n^*(\hat\phi,0)\le\mathcal{G}_n^*(\hat\phi,\hat l(\hat\phi))\le\mathcal{G}_n^*(0,\hat l(0)). \tag{A.17}
$$

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

$$
\begin{aligned}
\mathcal{G}_n^*(\hat\phi,\hat l(\hat\phi)) &= \mathcal{G}_{nq}^*(\hat\phi,\hat l(\hat\phi))+(1+\|\hat\phi-\phi_0\|^2+\|\hat l(\hat\phi)\|^2)o_p(1) \qquad\qquad (A.18)\\
&= \frac{1}{2}(Z_n-R_n'(\hat\phi-\phi_0))'J_n^{-1}(Z_n-R_n'(\hat\phi-\phi_0)) \\
&\quad -\frac{1}{2}(\hat l(\hat\phi)-J_n^{-1}[Z_n-R_n'(\hat\phi-\phi_0)])'J_n(\hat l(\hat\phi)-J_n^{-1}[Z_n-R_n'(\hat\phi-\phi_0)]) \\
&\quad +(1+\|\hat\phi-\phi_0\|^2+\|\hat l(\hat\phi)\|^2)o_p(1) \\
&= \frac{1}{2}(Z_n-R_n'(\hat\phi-\phi_0))'J_n^{-1}(Z_n-R_n'(\hat\phi-\phi_0))+(1+\|\hat\phi-\phi_0\|^2+\|\hat l(\hat\phi)\|^2)o_p(1),
\end{aligned}
$$

where $\phi_0 = [0, u_0']'$. The last equality is a consequence of Lemma A.8. Similarly, we can deduce from Lemmas A.2, 2, and Theorem 2 that

$$\mathcal{G}_n^*(0, \hat{l}(0)) = -\frac{1}{2}\hat{l}\,(0)'\,J_n\hat{l}\,(0) + Z_n'\hat{l}\,(0) + (1 + \|\hat{l}(0)\|^2)o_p(1) = O_p\,(1)\,. \tag{A.19}$$

Hence, from (A.17), (A.18), and (A.19) we obtain the inequality

$$0 \leq \frac{1}{2}(Z_n + o_p(1) - R_n'(\hat{\phi} - \phi_0))'J_n^{-1}(Z_n + o_p(1) - R_n'(\hat{\phi} - \phi_0)) \leq O_p(1). \tag{A.20}$$

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.20) implies that $\hat{\phi} - \phi_0$ is stochastically bounded.

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

$$\begin{aligned} nG_n^*(\hat{\beta}_n) &= \mathcal{G}_{nq}^*(\sqrt{n}(\hat{\beta}_n - \beta_{n,0})) + (1 + \|\sqrt{n}(\hat{\beta}_n - \beta_{n,0})\|^2)o_p(1) \\ &= nG_{nq}^*(\hat{\beta}_n) + O_p(1)o_p(1). \end{aligned}$$

(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

$$\begin{aligned} nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq}) &= \mathcal{G}_n^*(\hat{\phi}, \hat{l}(\hat{\phi})) + o_p(1) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) \\ &\leq \mathcal{G}_n^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) - \mathcal{G}_{nq}^*(\tilde{\phi}_q, \hat{l}_q(\tilde{\phi}_q)) + o_p(1). \end{aligned} \tag{A.21}$$

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.21) can be rewritten as

$$\begin{aligned} \mathcal{G}_n^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) &= \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p(1)\left(1 + \|\hat{\phi}_q - \phi_0\|^2 + \|\hat{l}(\hat{\phi}_q)\|^2\right) \\ &= \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p(1). \end{aligned} \tag{A.22}$$

The second equality in (A.22) is a consequence of Lemmas A.2 and A.7. According to Lemma A.8

$$\hat{l}(\bar{\phi}) = (J_n + o_p(1))^{-1}[Z_n - (R_n' + o_p(1))(\bar{\phi} - \phi_0)]$$

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.7. 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.21) as

$$\mathcal{G}_{nq}^*(\tilde{\phi}_q, \hat{l}(\tilde{\phi}_q)) = \mathcal{G}_{nq}^*(\hat{\phi}_q, \hat{l}(\hat{\phi}_q)) + o_p\,(1) \tag{A.23}$$

Plugging (A.22) and (A.23) into (A.21) 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

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

Proof of Lemma A.6: 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. ■

Lemma A.7 *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.7:

Proof of (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

$$\bar{\mathcal{G}}^*_{nq}(\phi) = \frac{1}{2} (\phi - \phi_0 + A_{2n})' A_{1n} (\phi - \phi_0 + A_{2n}) + \frac{1}{2} A_{3n}.$$

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

$$\bar{\phi}_q = \text{argmin}_{\phi \in \mathbb{R}^{m+h_2}} \bar{\mathcal{G}}^*_{nq}(\phi) = \phi_0 - A_{2n} = O_p(1).$$

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,

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

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 $\tilde{l}_q(\tilde{\phi}_q) = O_p(1)$.

Proof of (ii): According to Lemma A.6 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 $\epsilon > 0$

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

where the $o(1)$ term in the last line holds by Lemma A.1(ii). The set $\sqrt{n}\Lambda^{\zeta}_n$ consists of the elements in Λ^{ζ}_n 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, the first m ordinates of Φ_n expand to $\mathbb{R}^m$. Ordinate $m + j$ expands to $\mathbb{R}$ if $\nu_{0,j} > 0$ and to $\mathbb{R}^+$ otherwise. Since $\tilde{b}_q = O_p(1)$, we deduce $P\{\tilde{b}_q \in B\backslash(\Phi_n \otimes \sqrt{n}\Lambda^{\zeta}_n)\} = o(1)$. Therefore $\hat{b}_q = \tilde{b}_q + o_p(1)$, as required. ■

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* $\bar{\nu} - \nu_{n,0} \xrightarrow{p} 0$. *Let* $\hat{l}(\bar{\phi}) = \sqrt{n}\hat{\lambda}(\bar{\theta}, \bar{\nu})$, *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*

$$0 = Z_n - (R'_n + o_p(1))(\bar{\phi} - \phi_0) - (J_n + o_p(1))\hat{l}(\bar{\phi}).$$

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 $\max_{\lambda\in\hat{\Lambda}(\bar{\theta})} G^*_n(\bar{\theta}, \bar{\nu}, \lambda)$:

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

We now apply the mean-value theorem and multiply by $\sqrt{n}$:

$$0 = \sqrt{n} G^{*(1)}_n(\beta_{n,0})_\lambda + G^{*(2)}_n(\beta_*)_{\lambda\theta'}\bar{s} - M'(\bar{u} - u_0) + G^{*(2)}_n(\beta_*)_{\lambda\lambda'}\hat{l},$$

where β_* lies on the line joining $\beta_{n,0}$ 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.14) and (A.15). 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. ■

Derivations for MSE: Define

$$\widetilde{\mathcal{P}} = \phi_0 + (RJ^{-1}R')^{-1}RJ^{-1}Z$$

and partition $\widetilde{\mathcal{P}} = [\widetilde{\mathcal{P}}_s', \widetilde{\mathcal{P}}_u']'$. We can write

$$\begin{aligned}
\mathcal{S} &= \widetilde{\mathcal{P}}_s I\{\widetilde{\mathcal{P}}_u \geq 0\} + (\widetilde{\mathcal{P}}_s - \Omega_{su}\widetilde{\mathcal{P}}_u) I\{\widetilde{\mathcal{P}}_u < 0\} \\
&= \Omega_{su}\widetilde{\mathcal{P}}_u I\{\widetilde{\mathcal{P}}_u \geq 0\} + \widetilde{\mathcal{P}}_{s.uu} \\
\mathcal{U} &= \widetilde{\mathcal{P}}_u I\{\widetilde{\mathcal{P}}_u \geq 0\},
\end{aligned}$$

where $\widetilde{\mathcal{P}}_{s.uu} = \widetilde{\mathcal{P}}_s - \Omega_{su}\widetilde{\mathcal{P}}_u$. Notice that

$$\widetilde{\mathcal{P}}_{s.uu} \sim \mathcal{N}(\Omega_{su}u_0, \Omega_{ss} - \Omega_{su}\Omega_{us})$$

and $\widetilde{\mathcal{P}}_{s.uu}$ is uncorrelated with $\widetilde{\mathcal{P}}_u$. Using the formulas for moments of a truncated normal distribution (e.g. Greene (2003) p. 763) the mean and variance of $\mathcal{S}$ reported in the text can be computed.

A.5 Inference

Proof of Corollary 1: omitted. ■

Proof of Corollary 2: omitted. ■

Proof of Theorem 7: The asymptotics of $\hat{\theta}_n^H$ and $\hat{\lambda}^H(\hat{\theta}_n^H, n^{-1/2}u_0)$ are well known (e.g., Newey and Smith (2004)) and follow from straightforward modifications of the proofs of Theorems 3, 4, and 5. We will denote the limit distribution of $[\hat{s}_n^{H'}, u^{H'}]'$ by $\mathcal{P}^H$ and begin by characterizing $\mathcal{P}$ and $\mathcal{P}^H$. The concentrated limit objective function is of the form

$$\begin{aligned}
\bar{\mathcal{G}}_q^*(\phi) &= \frac{1}{2}(Z - R'(\phi-\phi_0))'J^{-1}(Z - R'(\phi-\phi_0)) \\
&= \frac{1}{2}[(\phi-\phi_0) - (RJ^{-1}R')^{-1}RJ^{-1}Z]'RJ^{-1}R'[(\phi-\phi_0) - (RJ^{-1}R')^{-1}RJ^{-1}Z] \\
&\quad + g(J,R,Z),
\end{aligned}$$

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

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

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

$$\begin{aligned}\bar{\mathcal{G}}_q(s,u) &= \frac{1}{2}[(s-Z_s)-(QJ^{-1}Q')^{-1}(QJ^{-1}M')(u-u_0-Z_u)]' \\ &\quad \times QJ^{-1}Q'[(s-Z_s)-(QJ^{-1}Q')^{-1}(QJ^{-1}M')(u-u_0-Z_u)] \\ &\quad +\frac{1}{2}(u-u_0-Z_u)'M\Omega M'(u-u_0-Z_u)+g(J,R,Z).\end{aligned}$$

Under the assumption that $u^H = u_0$ we can deduce that

$$\begin{aligned}\mathcal{S}^H &= Z_s-(QJ^{-1}Q')^{-1}QJ^{-1}M'Z_u \qquad \text{(A.29)} \\ \mathcal{S} &= Z_s-(QJ^{-1}Q')^{-1}QJ^{-1}M'(Z_u-\widetilde{\mathcal{U}}) \\ \widetilde{\mathcal{U}} &= \underset{\tilde{u}\geq -u_0}{\operatorname{argmin}}\,(\tilde{u}-Z_u)'M\Omega M'(\tilde{u}-Z_u),\end{aligned}$$

where $\tilde{u} = u-u_0$ and $\widetilde{\mathcal{U}} = \mathcal{U}-u_0$. Then let $\mathcal{P}^H = [\mathcal{S}^{H'}, u_0']'$ and $\mathcal{P} = [\mathcal{S}', u_0' + \widetilde{\mathcal{U}}']'$. The limit distribution of the likelihood ratio statistic can be manipulated as follows

$$\begin{aligned}&2\big(\bar{\mathcal{G}}_q^*(\mathcal{P}^H)-\bar{\mathcal{G}}_q^*(\mathcal{P})\big) \\ &= Z_u'M\Omega MZ_u-(\widetilde{\mathcal{U}}-Z_u)'M\Omega M'(\widetilde{U}-Z_u) \\ &= -\widetilde{\mathcal{U}}'\Lambda^{-1}\widetilde{\mathcal{U}}+Z_u'\Lambda^{-1}\widetilde{\mathcal{U}}+\widetilde{\mathcal{U}}'\Lambda^{-1}Z_u,\end{aligned}$$

where $\Lambda = (M\Omega M')^{-1}$. We deduce from Theorems 4 and 5

$$\mathcal{LR}_n^u(u_0) \Longrightarrow 2\big(\bar{\mathcal{G}}_q^*(\mathcal{P}^0)-\bar{\mathcal{G}}_q^*(\mathcal{P})\big).$$

The statement of the theorem follows from defining . ■

Proof of Corollary 3: omitted. ■

References

Andrews, Donald W.K., Stephen Berry, and Panle Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Locations," *Manuscript*, Yale University, Department of Economics.

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.

Canova, Fabio, and G. De Nicoló (2002): "Monetary Disturbances Matter for Business Cycle Fluctuations in the G-7," *Journal of Monetary Economics* **49**, 1131-1159.

Cavanagh, Christopher L., Graham Elliott, and James H. Stock (1995): "Inference in Models with Nearly Integrated Regressors," *Econometric Theory*, **11**, 1131-1147.

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*, Northwestern 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.

Gourieroux, 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.

Greene, William H. (2003): *"Econometric Analysis,"* Fifth Edition, Pearson Education Inc., Upper Saddle River, New Jersey.

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

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

Luttmer, Erzo G.J. (1996): "Asset Pricing in Economies with Frictions," *Econometrica*, **64**, 1439-1467.

Luttmer, Erzo G.J. (1999): "What Level of Fixed Costs Can Reconcile Consumption and Stock Returns?" *Journal of Political Economy*, **107**, 969-997.

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

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.

Pakes, Ariel, Jack Porter, Kate Ho, and Joy Ishii (2005): "Moment Inequalities and Their Application," *Manuscript*, Harvard University, Department of Economics.

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

Pollard, David (2002): *"A User's Guide to Measure Theoretic Probability,"* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK.

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.

Stock, James H., Jonathan H. Wright, and Motohiro Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business Economics and Statistics*, **20**, 518-529.

Taylor, John B. (1999): *"Monetary Policy Rules"*, NBER Business Cycle Series, Vol 31, University of Chicago Press, Chicago.

Uhlig, Harald (2005): "What are the Effects of Monetary Policy on Output? Results from an Agnostic Identification Procedure," *Journal of Monetary Economics*, **52**, 381-419.

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

Woodford, Michael (2003): *"Interest and Prices,"* Princeton University Press, Princeton.

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

Table 1: Parameterizations of DGPs

	DGP 1	DGP 2	DGP 3
$\rho_{1,2}$	$[0.5, -0.1]'$	$[0.9, 0.3]'$	$[0.5, -0.1]'$
$\rho_{1,X}$	$[0.5, 0.5]'$	$[0.5, 0.5]'$	$[0.3, 0.3]'$
$\rho_{2,X}$	0.5	0.5	0.8
$\sqrt{v(\hat{\theta}_{(1)})}$	1.41	1.41	2.36
$\sqrt{v(\hat{u}_{(1)})}$	1.05	0.55	1.87

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

	$\hat{\theta}_{(0)}$		$\hat{\theta}_{(1)}$		$\hat{\theta}_{(12)}$	
u_0	Bias	MSE	Bias	MSE	Bias	MSE
			DGP 1			
0.00	-0.25	1.81	0.00	2.00	0.00	1.61
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	2.03
2.00	-0.01	1.98	0.00	2.00	1.30	3.30
3.00	0.00	2.00	0.00	2.00	1.95	5.42
5.00	0.00	2.00	0.00	2.00	3.26	12.21
10.00	0.00	2.00	0.00	2.00	6.52	44.08
			DGP 2			
0.00	0.23	1.84	0.00	2.00	0.00	1.67
0.50	0.02	1.97	0.00	2.00	-0.83	2.36
1.00	0.00	2.00	0.00	2.00	-1.66	4.44
2.00	0.00	2.00	0.00	2.00	-3.33	12.76
3.00	0.00	2.00	0.00	2.00	-5.00	26.64
5.00	0.00	2.00	0.00	2.00	-8.33	71.06
10.00	0.00	2.00	0.00	2.00	-16.66	279.34
			DGP 3			
0.00	-0.82	3.37	0.00	5.56	0.00	1.24
0.50	-0.56	3.52	0.00	5.56	0.57	1.57
1.00	-0.37	3.85	0.00	5.56	1.15	2.55
2.00	-0.13	4.63	0.00	5.56	2.29	6.47
3.00	-0.03	5.19	0.00	5.56	3.43	13.00
5.00	0.00	5.53	0.00	5.56	5.71	33.87
10.00	0.00	5.56	0.00	5.56	11.42	131.67

Notes: The table reports bias and mean squared error (MSE) based on the simulation of the limit distribution.

Table 3: Asymptotic Confidence Intervals for θ_0

	$\mathcal{CS}^{\theta}_{(0)}$		$\mathcal{CS}^{\theta}_{(1)}$	
u_0	$\sqrt{n}$ Length	Cov Prob	$\sqrt{n}$ Length	Cov Prob
		DGP 1		
0.00	4.42	0.90	4.64	0.90
0.50	4.51	0.90	4.64	0.90
1.00	4.58	0.90	4.64	0.90
2.00	4.65	0.90	4.64	0.90
3.00	4.66	0.90	4.64	0.90
5.00	4.66	0.90	4.64	0.90
10.00	4.66	0.90	4.64	0.90
		DGP 2		
0.00	4.45	0.90	4.64	0.90
0.50	4.62	0.90	4.64	0.90
1.00	4.65	0.90	4.64	0.90
2.00	4.65	0.90	4.64	0.90
3.00	4.65	0.90	4.64	0.90
5.00	4.65	0.90	4.64	0.90
10.00	4.65	0.90	4.64	0.90
		DGP 3		
0.00	5.81	0.90	7.77	0.90
0.50	6.15	0.93	7.77	0.90
1.00	6.48	0.93	7.77	0.90
2.00	7.05	0.91	7.77	0.90
3.00	7.44	0.90	7.77	0.90
5.00	7.75	0.90	7.77	0.90
10.00	7.79	0.90	7.77	0.90

Notes: The table reports the length, scaled by $\sqrt{n}$, and the coverage probabilities of the asymptotic confidence intervals.

Table 4: Asymptotic Confidence Intervals for u_0

	$\mathcal{CS}^u_{(0)}$		$\mathcal{CS}^u_{(1)}$	
u_0	$\sqrt{n}$ Length	Cov Prob	$\sqrt{n}$ Length	Cov Prob
		DGP 1		
0.00	1.64	0.90	1.91	0.90
0.50	2.03	0.90	2.21	0.90
1.00	2.37	0.90	2.51	0.90
2.00	2.86	0.90	2.97	0.90
3.00	3.08	0.90	3.23	0.90
5.00	3.16	0.90	3.45	0.90
10.00	3.16	0.90	3.46	0.90
		DGP 2		
0.00	0.61	0.90	1.03	0.90
0.50	0.95	0.90	1.31	0.90
1.00	1.11	0.90	1.54	0.90
2.00	1.15	0.90	1.76	0.90
3.00	1.15	0.90	1.80	0.90
5.00	1.15	0.90	1.81	0.90
10.00	1.15	0.90	1.81	0.90
		DGP 3		
0.00	3.11	0.90	3.43	0.90
0.50	3.49	0.90	3.71	0.90
1.00	3.87	0.90	4.01	0.90
2.00	4.56	0.90	4.59	0.90
3.00	5.11	0.90	5.08	0.90
5.00	5.74	0.90	5.68	0.90
10.00	5.99	0.90	6.14	0.90

Notes: The table reports the length, scaled by $\sqrt{n}$, and the coverage probabilities of the asymptotic confidence intervals.