Boosting Your Instruments: Estimation with Overidentifying Inequality Moment Conditions

Hyungsik Roger Moon
University of Southern California

Frank Schorfheide*
University of Pennsylvania and CEPR

March 3, 2006

*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. An earlier version of this paper circulated under the title "Empirical Likelihood Estimation with Inequality Moment Constraints." We thank Sungbae An, 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, London School of Economics, Michigan State University, Penn State University, Princeton University, UC Irvine, UC Riverside, and the Universities of Chicago, Montreal, Pennsylvania, and Virginia for helpful comments.

Abstract

This paper derives limit distributions of empirical likelihood estimators for models in which inequality moment conditions provide overidentifying information. We show that the use of this information leads to a reduction of the asymptotic mean-squared estimation error and propose asymptotically valid confidence sets for the parameters of interest. While inequality moment conditions arise in many important economic models, we use a dynamic macroeconomic model as data generating process and illustrate our methods with instrumental variable estimators of monetary policy rules. The assumption that output does not fall in response to an expansionary monetary policy shock leads to an inequality moment condition that can substantially increase the precision with which the policy rule is estimated. The results obtained in this paper extend to conventional GMM estimators.

JEL CLASSIFICATION: C32

KEY WORDS: Empirical Likelihood Estimation, Generalized Method of Moments, Inequality Moment Conditions, Instrumental Variable Estimation, Monetary Policy Rules.

1 Introduction

This paper extends moment-based 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 in many important economic models. For instance, in an influential paper Zeldes (1989) studies whether the presence of borrowing constraints can explain households' violation of consumption Euler equations. Zeldes regards households with a low wealth-to-income ratio as potentially borrowing constrained. These households' current marginal utility of consumption might exceed the discounted expected future marginal utility, which leads to an inequality moment condition. Luttmer (1996, 1999) studies asset pricing in the presence of financial frictions, which turn conventional asset pricing relationships into inequality conditions. Pakes, Porter, Ho, and Ishii (2005) and Andrews, Berry, and Jia (2004) provide examples of inequality moment conditions derived from models of industrial organization. These models share the basic assumption that firms' actual choices yield higher ex-ante expected profits than alternative feasible choices.

Inequality moment conditions also arise in instrumental variable (IV) models in which a subset of the instrumental variables is potentially correlated with the error term in the regression equation, but the direction of this correlation is assumed to be known. Our lead example involves the estimation of an interest-rate feedback rule that describes the behavior of a central bank. A measure of output appears as endogenous regressor in the policy reaction function and renders the OLS estimator inconsistent. While in this time series setting lagged output and inflation can be used as instrumental variables, in practice these instruments are often poorly correlated with the endogenous regressors and lead to imprecise parameter estimates.[1] The methods developed in this paper allow us to augment the list of instruments by variables for which economic theory provides some guidance about the sign of their potential correlation with the error term. For instance, most New Keynesian dynamic stochastic general equilibrium (DSGE) models imply that output does not fall in response to an expansionary monetary policy shock (see Woodford (2003)). This implication leads to an inequality moment condition that can substantially increase the precision with which the reaction function is estimated.

Formally, our paper 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

[1] Since this problem of so-called weak instruments is widespread in both macro- and microeconomic applications, there exists an extensive econometrics literature (see, for instance, Stock, Wright, and Yogo (2002)) on how to conduct valid, albeit imprecise inference in such a setting.

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_0 and hence reduces its variability. The larger u_0, the less informative is the second set of moment conditions. 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 (EL) estimators have emerged as an attractive alternative to GMM, e.g., Owen (1988), Qin and Lawless (1994), Imbens (1997), Kitamura and Stutzer (1997), and Imbens, Spady, and Johnson (1998). 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. Although we do not extend higher-order optimality properties of EL procedures to the class of irregular models considered in this paper, we believe that these results provide a good reason for studying EL estimators. In fact, since moment conditions are imposed as parametric constraints on the empirical likelihood function, an extension to inequality conditions is quite natural.

Throughout the paper we focus on first-order asymptotic approximations and make three 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. We derive a quadratic approximation of the EL objective function and analyze the distribution of its saddlepoint. The inequality moment conditions translate into sign restrictions on the corresponding Kuhn-Tucker parameters. Second, 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. Third, 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 distribution of the EL estimator of θ_0 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.[2] 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

[2]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).

parameters. Hence, our large sample results, in particular the efficieny gain through the inequality moment conditions, also apply to conventional GMM estimators.

Since we can rewrite the inequality moment condition as $\mathbb{E}[g_2(X_i, \theta_0) - \vartheta_0] = 0$, where $\vartheta_0 \geq 0$, our work is related to the literature on estimation and inference in the presence of inequality parameter constraints, 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 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 $\mathbb{E}[g_2(X_i, \theta_0)] = 0$ 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. While Andrews (1999) considers estimators that are defined as extremum of an objective function, we extend some of his results to estimators that are defined as a saddlepoints. Moreover, Andrews (2001) focuses on inference for ϑ_0 (using our notation), whereas we are particularly interested in inference about θ_0, treating the slackness in the inequality moment condition, ϑ_0, as a nuisance parameter.

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-identified models has recently been studied by Andrews, Berry, and Jia (2004), Chernozhukov, Hong, and Tamer (2002), and Pakes, Porter, Ho, and Ishii (2005), Rosen (2005), and Shaikh (2005) and is not considered in our paper.

The plan of the paper is as follows. To illustrate how the methods proposed in this paper can be used to solve an important practical estimation problem in macroeconomics, we introduce our lead example of estimating a monetary policy rule in Section 2. Technical assumptions as well as the estimators' objective functions are stated in Section 3. Section 4 develops the asymptotic distribution theory for the EL estimator and its objective function in the presence of inequality moment conditions. In Section 5 some implications of the limit theory are discussed and the efficiency result is provided. Section 6 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 the policy rule example in Section 7. Moreover, we make a comparison with the asymptotic properties of simple procedures that ignore the information in the inequality moment condition. Section 8 concludes and the Appendix contains all proofs and technical Lemmas.

We use the following notation throughout the paper: "$\stackrel{p}{\longrightarrow}$" 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 An Example of Inequality Moment Conditions

In macroeconomics there is great interest in characterizing the behavior of central banks through interest rate feedback rules (see for instance, Taylor (1999) and Clarida, Galí, and Gertler (2000)). Such rules are built into vector autoregressions (VAR) as well as dynamic stochastic general equilibrium (DSGE) models. To illustrate our methods for inference with inequality moment conditions we consider the following policy rule

$$\widetilde{R}_t = \rho_R \widetilde{R}_{t-1} + (1-\rho_R)\psi_1 \widetilde{\pi}_t + (1-\rho_R)\psi_2 \widetilde{x}_t + \epsilon_{R,t}, \tag{2}$$

where $\widetilde{R}_t$ is the nominal interest rate, controlled by the central bank through open-market operations, $\widetilde{\pi}_t$ is the inflation rate, and $\widetilde{x}_t$ is a measure of real activity, such as output deviations from trend or output growth. The shock $\epsilon_{R,t}$ captures unanticipated (by the public) deviations from the systematic component of the policy rule. In equilibrium both inflation and output are likely to be a function of the monetary policy shock, which causes an endogeneity problem that can be addressed by IV estimation. Lagged values of inflation and output are natural candidates for instrumental variables. According to large class of monetary DSGE models, in particular New Keynesian models, output does not fall in response to a expansionary monetary policy shock, implying that $I\!E[\widetilde{x}_t(-\epsilon_{R,t})] \geq 0$. This implication generates a moment inequality condition that can be exploited to sharpen the inference about the policy rule coefficients.[3]

A prototypical New Keynesian DSGE model (see Woodford (2003)) can be described by the following additional equations:

$$\begin{aligned}
\widetilde{y}_t &= I\!E_t[\widetilde{y}_{t+1}] - \frac{1}{\tau}(\widetilde{R}_t - I\!E_t[\widetilde{\pi}_{t+1}]) + (1-\rho_g)\widetilde{g}_t + \frac{\rho_z}{\tau}\widetilde{z}_t && (3)\\
\widetilde{\pi}_t &= \beta I\!E_t[\widetilde{\pi}_{t+1}] + \kappa(\widetilde{y}_t - \widetilde{g}_t) && (4)\\
\widetilde{g}_t &= \rho_g \widetilde{g}_{t-1} + \epsilon_{g,t} && (5)\\
\widetilde{z}_t &= \rho_z \widetilde{z}_{t-1} + \epsilon_{z,t} && (6)
\end{aligned}$$

These equations can be derived as log-linearized equilibrium relationships from a fully-specified DSGE model. Equation (3) represents an intertemporal Euler equation obtained

[3] In the VAR literature such sign restrictions are often used to identify monetary policy shocks, e.g., Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005).

from the households' optimal choice of consumption and bond holdings. Since the underlying structural model has no investment, output $\widetilde{y}_t$ (measured in percentage deviations from stochastic trend A_t) is proportional to consumption up to an exogenous process $\widetilde{g}_t$ that can be interpreted as time-varying government spending or, more broadly, as preference change. The parameter τ can be interpreted as households' intertemporal substitution elasticity. The exogenous process $\widetilde{z}_t$ captures the stochastic growth of the level of total factor productivity, A_t, in the economy.

The production sector in the underlying economy is characterized by a continuum of monopolistically competitive firms, each of which faces a downward-sloping demand curve for its differentiated product. Prices are sticky due to quadratic adjustment costs for nominal prices or a Calvo-style rigidity that allows only a constant fraction of firms adjust their prices. The resulting dynamics are described by the expectational Phillips curve (4). The parameter β is the households' discount factor.

In principle, one could estimate the entire model using likelihood-based techniques (see An and Schorfheide (2005) for a survey) to obtain estimates of the policy rule coefficients. However, since the full-information estimator exploits cross-coefficients restrictions it is sensitive to model misspecification. For instance, the simple model abstracts from habit formation, investment, and wage rigidities, which have been found to be important to capture salient feature of U.S. and Euro Area data (see Smets and Wouters (2003) and Christiano, Eichenbaum, and Evans (2005)). Nevertheless, many of the richer specifications proposed in the literature share the basic property that unanticipated reductions of interest rates do not lower output. Hence, the moment condition $\mathbb{E}[\widetilde{x}_t(-\epsilon_{R,t})] \geq 0$ remains valid.

We will revisit the prototypical New Keynesian model in Section 7 when we conduct a small-scale simulation exercise to illustrate the proposed estimation and inference methods. The DSGE model will serve as a data generating mechanism. The slackness in the inequality moment condition is a function of the slope of the Phillips curve κ. If κ is large, then prices in the model economy are fairly flexible. Hence, monetary policy shocks have only small real effects, $\mathbb{E}[\widetilde{x}_t(-\epsilon_{R,t})]$ is near zero and there will be a substantial efficiency gain associated with the use of the inequality moment condition. Vice versa, if there is a lot of price stickiness in the economy, output responds strongly to monetary policy shocks and the inequality moment condition does not generate much additional information about the parameters of interest.

3 Moment-Based Estimation

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 $h = h_1 + h_2$, $g(X_i, \theta) = [g_1(X_i, \theta)', g_2(X_i, \theta)']'$, $M =$

$[0_{h_2\times h_1}\ \ I_{h_2}]$, $\mathbb{E}[g_2(X_i,\theta_0)] = \nu_{n,0}$, and $Z_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n [g(X_i,\theta_0) - M'\nu_{n,0}]$. K is used to denote a large constant. We begin by stating some fundamental assumptions. Assumption 1 is used for the consistency proof and Assumptions 2 and 3 to derive the limit distributions of estimators and empirical likelihood ratio statistics.

Assumption 1 *(a) X_i, $i = 1,\ldots,n$ are strictly stationary and ergodic on a probability space $(\Omega_n, \mathcal{F}_n, P_n)$; (b) Θ, the parameter space for θ, is an m-dimensional compact subset of $\mathbb{R}^m$; (c) $g(x,\theta)$ is continuous at each $\theta \in \Theta$ with probability one; (d) $\mathbb{E}[g_1(X_i,\theta_0)] = 0$, and $\mathbb{E}[g_1(X_i,\theta)] \neq 0$ for $\theta \neq \theta_0$; (e) $\nu_{n,0} = \nu_0 + n^{-1/2}u_0 \geq 0$; (f) $\mathbb{E}[g(X_i,\theta_0)g(X_i,\theta_0)'] \longrightarrow J$ is non-singular; (g) $Z_n = O_p(1)$; (h) $\mathbb{V} = \left\{\nu \in \mathbb{R}^{h_2} : \nu \geq 0 \text{ and } \|v\| \leq K\right\}$ and $\{\nu_{n,0}\}_n \subset \mathbb{V}$; (i) $\mathbb{E}\left[\sup_{\theta\in\Theta} \|g(X_i,\theta)\|^\alpha\right] \leq K < \infty$ for some $\alpha > 2$.*

When the moment function $g(X_i,\theta)$ is differentiable with respect to θ, 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 $g^{(1)}(X_i,\theta) = [g_1^{(1)}(X_i,\theta),\ldots,g_h^{(1)}(X_i,\theta)]$.

Assumption 2 *(a) The true parameter θ_0 exists in an interior of Θ; (b) $g(X_i,\theta)$ is twice continuously differentiable; (c) The matrix $\mathbb{E}[g_1^{(1)}(X_i,\theta_0)']$ has full column rank; (d) $\mathbb{E}\left[\sup_{\theta\in\Theta}\|g^{(1)}(X_i,\theta)\|^2\right] \leq K < \infty$, $\mathbb{E}\left[\sup_{\theta\in\Theta}\|g_j^{(2)}(X_i,\theta)\|\right] \leq K < \infty$ for $j = 1,\ldots,h$.*

Assumption 3 *$Z_n \Longrightarrow Z$, where $Z \sim \mathcal{N}(0, J - M'\nu_0\nu_0'M)$.*

In this paper, we assume that the sequence of observable random vectors X_i are strictly stationary and ergodic. Assumption 3 is satisfied, if, for instance, $\{g(X_i,\theta_0) - M'\nu_{n,0}\}$ is a Martingale Difference Sequence with respect to the natural filtration. We also assume that the parameter θ_0 is identifiable based on the equality moment condition $\mathbb{E}[g_1(X_i,\theta_0)] = 0$ (Assumption 1(d)). The expected value of $g_2(X_i,\theta_0)$ is denoted by $\nu_{n,0} \geq 0$. The parameter $\nu_{n,0}$ measures the slackness of the inequality conditions and 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 slackness parameter as $\nu_{n,0} = \nu_0 + n^{-1/2}u_0$. To accommodate the drift in our notation we indexed the probability space in Assumption 1(a) by the sample size n. We will show in Section 5 that moment conditions for which the corresponding element of ν_0 is strictly greater than zero do not affect the asymptotic 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.

3.1 Empirical Likelihood Formulation

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). In the case of *iid* observations the (constrained) empirical likelihood function is

$$L_{EL}(\theta, p) = \left\{ \prod_{i=1}^{n} p_i \;\middle|\; p_i > 0,\ \sum_{i=1}^{n} p_i = 1,\ \sum_{i=1}^{n} p_i g_1(X_i, \theta) = 0,\ \sum_{i=1}^{n} p_i g_2(X_i, \theta) \geq 0 \right\}, \tag{7}$$

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

$$\{\hat{\theta}_{n,EL}, \hat{p}_{n,EL}\} = \operatorname*{argmax}_{\theta \in \Theta,\ p} \ L_{EL}(\theta, p). \tag{8}$$

3.2 Saddlepoint Formulation

The empirical likelihood estimator can be expressed as the saddlepoint

$$\hat{\theta}_n = \operatorname*{argmin}_{\theta \in \Theta} \ \max_{\lambda_1,\ \lambda_2 \leq 0} \ G_n(\theta, \lambda_1, \lambda_2) \tag{9}$$

of the function

$$G_n(\theta, \lambda_1, \lambda_2) = \frac{1}{n} \sum_{i=1}^{n} \ln\left(1 + \lambda_1' g_1(X_i, \theta) + \lambda_2' g_2(X_i, \theta)\right). \tag{10}$$

While in the conventional moment-based estimation based on equality conditions the Kuhn-Tucker parameters are unconstrained, the inequality moment condition results in a non-positivity constraint for λ_2. Newey and Smith (2004) study a broader class of estimators, called Generalized Empirical Likelihood (GEL) estimators, that are obtained by generalizing the objective function $G_n(\theta, \lambda_1, \lambda_2)$. This class contains Kitamura and Stutzer's (1997) exponential tilting estimator as well as Hansen, Heaton, and Yaron's (1996) continuous updating GMM estimator. Our analysis has a straightforward extension to the GEL class, but we do not pursue the extension in this paper.

In order to facilitate the large-sample analysis we re-write our estimator as the solution of a modified saddle-point problem. Define $\lambda = [\lambda_1', \lambda_2']'$ and let

$$\hat{\Lambda}_n(\theta) = \{\lambda \in \mathbb{R}^h \mid \lambda' g(X_i, \theta) \geq -1 + \kappa, i = 1, \ldots, n\},$$

for some $\kappa > 0$. Moreover, we use $G_n(\theta, \lambda)$ to abbreviate $G_n(\theta, \lambda_1, \lambda_2)$. We will subsequently study the limit distribution of the saddlepoint

$$\begin{aligned} \{\hat{\theta}_n,\ \hat{\nu}_n\} &= \operatorname*{argmin}_{\theta \in \Theta,\ \nu \in \mathbb{V}} \ \max_{\lambda \in \hat{\Lambda}_n(\theta)} \ G_n^*(\theta, \nu, \lambda) \\ \hat{\lambda}(\theta, \nu) &= \operatorname*{argmax}_{\lambda \in \hat{\Lambda}_n(\theta)} \ G_n^*(\theta, \nu, \lambda), \end{aligned} \tag{11}$$

where

$$G_n^*(\theta,\nu,\lambda) = G_n(\theta,\lambda) - \nu' M\lambda. \tag{12}$$

Here ν plays the role of a Kuhn-Tucker parameter for the constraint that $\lambda_2 \le 0$. We show in Lemma 1 (Appendix A.1) that the saddlepoints of $G_n(\theta,\lambda)$ and $G_n^*(\theta,\nu,\lambda)$ are equivalent. Moreover, $\hat{\nu}$ will asymptotically capture the slackness in the inequality moment condition $\mathbb{E}[g_2(X_i,\theta_0)]$.

3.3 GMM Formulation

As pointed out in the Introduction, one can introduce an additional $h_2 \times 1$ parameter vector $\vartheta = \mathbb{E}[g_2(X_i,\theta)]$ that captures the slackness in the inequalities and express the second moment condition as $\mathbb{E}[g_2(X_i,\theta_0) - \vartheta_{n,0}] = 0$, where $\vartheta_{n,0} \ge 0$. A GMM estimator can be obtained by solving

$$\min_{\theta\in\Theta,\vartheta\ge 0} \quad \frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta) - M'\vartheta\right)' W_n \left(\frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta) - M'\vartheta\right), \tag{13}$$

where $\{W_n\}$ is a sequence of positive-definite $h \times h$ weight matrices. The subsequent large sample analysis will focus on the saddlepoint problem (11) but we will return to the GMM formulation in our discussion in Section 5.

4 Large Sample Analysis

The large sample analysis proceeds in three steps. First, we establish the consistency of the saddlepoint estimator $\hat{\theta}_n$. 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. Finally, we show that the estimators obtained from G_n^* and and its quadratic approximation G_{nq}^* are distributionally equivalent in large samples and characterize their limit distributions.

4.1 Consistency

It is well known that the MELE with equality moment conditions is consistent. Since Assumption 1(d) 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 (Appendix A.1) as derivative of $G_n(\theta,\lambda_1,\lambda_2)$ with respect to λ_2, and $\nu_{n,0} = \mathbb{E}[g_2(X_i,\theta_0)]$ 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 Assumption 1 is satisfied. Then* $\hat{\theta}_n \xrightarrow{p} \theta_0$ *and* $\hat{\nu}_n - \nu_{n,0} \xrightarrow{p} 0$. *Moreover,* $\hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{p} 0$.

4.2 Quadratic Approximation of Objective Function

We proceed with a second-order Taylor approximation of the objective function G_n^*. Let $\beta = [\theta', \nu', \lambda']'$, $\beta_{n,0} = [\theta_0', \nu_{n,0}', 0_{1\times h}]'$, and abbreviate $G_n^*(\theta, \nu, \lambda)$ as $G_n^*(\beta)$. The domain of β is given by

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

where $\Lambda_n^\zeta = \left\{ \lambda \in \mathbb{R}^h : \|\lambda\| \le n^{-\zeta} \right\}$. 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 (Appendix A.3) that this domain restriction asymptotically does not affect $\hat{\lambda}$. According to Assumption 2, the moment function $g(X, \theta)$ is twice continuously differentiable and we can write the objective function as

$$G_n^*(\beta) = G_{nq}^*(\beta) + \frac{1}{n}\mathcal{R}_n(\beta), \tag{14}$$

where $\frac{1}{n}\mathcal{R}_n(\beta)$ is the remainder term of the Taylor approximation. We show in Lemma 2 (Appendix A.4) that the remainder term $\mathcal{R}_n(\beta)$ is uniformly ignorable in a shrinking neighborhood of the true parameter $\beta_{n,0}$.

It is convenient to re-parameterize the problem as follows. Let $b = [s', u', l']' = \sqrt{n}(\beta - \beta_0)$, where $\beta_0 = [\theta_0', \nu_0', 0_{1\times h}]'$. The domain of b, denoted by B_n, is defined such that

$$s \in S_n = \sqrt{n}(\Theta - \theta_0),\ u \in U_n = \sqrt{n}\left(\mathbb{V} - \nu_0\right),\ l \in L_n(s) = \{l \mid l/\sqrt{n} \in \hat{\Lambda}_n(\theta_0 + s/\sqrt{n})\}.$$

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. For notational convenience we will stack the parameters s and u into the vector $\phi = [s', u']'$ with domain $\Phi_n = S_n \otimes U_n$ and $\phi_0 = [0_{1\times m}, u_0']'$. The Taylor series expansion of $G_n^*(\beta)$ leads to the quadratic approximation

$$\begin{aligned} \mathcal{G}_{nq}^*(\phi, l) \quad = \quad & -\frac{1}{2}(l - J_n^{-1}[Z_n - R_n'(\phi - \phi_0)])' J_n (l - J_n^{-1}[Z_n - R_n'(\phi - \phi_0)]) \\ & +\frac{1}{2}(Z_n - R_n'(\phi - \phi_0))' J_n^{-1}(Z_n - R_n'(\phi - \phi_0)), \end{aligned} \tag{15}$$

where $R_n = [-Q_n', M']'$ and

$$Q_n = \frac{1}{n}\sum_{i=1}^n g^{(1)}(X_i, \theta_0), \quad J_n = \frac{1}{n}\sum_{i=1}^n g(X_i, \theta_0) g(X_i, \theta_0)'.$$

We now consider two estimators: $\hat{\beta}_n$ is the actual empirical likelihood estimator. The second estimator is $\tilde{\beta}_{nq} = \beta_0 + \tilde{b}_q/\sqrt{n}$, where $\tilde{b}_q = [\tilde{\phi}_q', \tilde{l}_q(\tilde{\phi}_q)']'$ is obtained by solving a saddlepoint problem based on the objective $\mathcal{G}_{nq}^*(\phi, l)$ without restricting b to lie in B_n:

$$\tilde{l}_q(\phi) \quad = \quad \mathrm{argmax}_{l \in \mathbb{R}^h}\, \mathcal{G}_{nq}^*(\phi, l), \quad \tilde{\phi}_q = \mathrm{argmin}_{\phi \in \Phi(\nu_0)}\, \mathcal{G}_{nq}^*(\phi, \tilde{l}_q(\phi)),$$

where

$$\Phi(\nu_0) = \left\{ \phi = [s', u'] \in \mathbb{R}^m \otimes \mathbb{R}^{h_2} \mid u_j \geq 0 \text{ if } \nu_{0,j} = 0 \right\}. \tag{16}$$

According to Assumption 3 $Z_n \Longrightarrow Z$. Moreover, we can deduce from Assumptions 1 and 2 and the Ergodic Theorem that $J_n \xrightarrow{p} J$ and $R_n \xrightarrow{p} R$, where $R = [-Q', M']'$ and $Q = \mathbb{E}[g^{(1)}(X_i, \theta_0)]$. We obtain the following theorem:

Theorem 2 *Suppose Assumptions 1 – 3 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)* $nG_{nq}^*(\hat{\beta}_n) = nG_{nq}^*(\tilde{\beta}_{nq}) + o_p(1)$, *and (v)* $nG_n^*(\hat{\beta}_n) = nG_{nq}^*(\tilde{\beta}_{nq}) + o_p(1)$.

Theorem 2 establishes that $\hat{\beta}_n$ and $\tilde{\beta}_{nq}$ are $\sqrt{n}$-consistent. Moreover, the theorem states that the discrepancy between $G_n^*(\beta)$ evaluated at $\hat{\beta}_n$ and $G_{nq}^*(\beta)$ evaluated at $\tilde{\beta}_{nq}$ vanishes. Thus, the large-sample behavior of likelihood ratios can be approximated by the behavior of $G_{nq}^*(\tilde{\beta}_{nq})$.

4.3 Limit Distribution

We begin by studying the limit distribution of $\tilde{b}_q$. From (15) it follows immediately that $\mathcal{G}_{nq}^*(\phi, l)$ is maximized with respect to $l \in \mathbb{R}^h$ by

$$\tilde{l}_q(\phi) = J_n^{-1}\left(Z_n - R_n'(\phi - \phi_0)\right). \tag{17}$$

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

$$\bar{\mathcal{G}}_{nq}^*(\phi) = \mathcal{G}_{nq}^*(\phi, \tilde{l}_q(\phi)) = \frac{1}{2}(Z_n - R_n'(\phi - \phi_0))' J_n^{-1}(Z_n - R_n'(\phi - \phi_0)). \tag{18}$$

The limit distribution of $\tilde{\phi}_q$ can be determined from $\bar{\mathcal{G}}_{nq}^*(\phi)$. We then use (17) to obtain the distribution of $\tilde{l}_q(\tilde{\phi}_q)$. Finally, it can be shown that $\hat{b}$ and $\tilde{b}_q$ are asymptotically equivalent. The results are summarized in the following theorem.

Theorem 3 *Suppose Assumptions 1 – 3 are satisfied. (i) Then*

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

where

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

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

5 Implications of the Limit Distribution Results

We will now explore the limit distribution of $\hat{b}$ in more detail. First, we discuss the relationship between the conventional GMM method and the EL approach pursued in the previous section. Second, 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. Third, if $\nu_0 = 0$ and $I\!E[g_2(X_i, \theta_0)] = n^{-1/2}u_0$, 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. Fourth, for the case $h_2 = 1$ we derive the asymptotic means and the variances of $\hat{s}$ and $\hat{u}$ with a weakly informative inequality restriction, and compare them to the means and variances of some of benchmark estimators.

5.1 GMM with Inequality Moment Conditions

The limit distribution derived in Theorem 3 also applies to the GMM estimator defined in (13). 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 and assuming that $W_n - J_n^{-1} \stackrel{p}{\longrightarrow} 0$ 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 (18). Therefore, the analysis in the remainder of the paper applies not only to empirical likelihood estimators but also to conventional GMM estimators.

5.2 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

$$\begin{aligned}
&(Z - R'(\phi - \phi_0))' J^{-1} (Z - R'(\phi - \phi_0)) \\
&= (Z_1 + Q_1's)' J_{11}^{-1} (Z_1 + Q_1's) \qquad (19) \\
&\quad + [Z_2 + Q_2's - (u - u_0) - J_{21}J_{11}^{-1}(Z_1 + Q_1's)]' \\
&\quad \times (J_{22} - J_{21}J_{11}^{-1}J_{12})^{-1} [Z_2 + Q_2's - (u - u_0) - J_{21}J_{11}^{-1}(Z_1 + Q_1's)].
\end{aligned}$$

If $\nu_0 > 0$ then the limit distribution of $\hat{u}$ is obtained by minimizing (19) 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 (19) 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,

$$2\mathcal{G}_q^*(\mathcal{P},\mathcal{L}) = Z_1'[J_{11}^{-1} - J_{11}^{-1}Q_1'(Q_1J_{11}^{-1}Q_1')^{-1}Q_1J_{11}^{-1}]Z_1, \tag{20}$$

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

5.3 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

$$\bar{\mathcal{G}}_q^*([s',u']') = \frac{1}{2}(Z + Q's - M'(u-u_0))'J^{-1}(Z + Q's - M'(u-u_0)) \tag{21}$$

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

$$\bar{\mathcal{G}}_q^*([s',u_0' + \tilde{u}']') = \frac{1}{2}(Z + Q's - M'\tilde{u})'J^{-1}(Z + Q's - M'\tilde{u}) \tag{22}$$

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.

5.4 Mean-Squared-Error Comparisons

The main goal of this section is to derive an analytic formula for the asymptotic mean-squared-errors (MSE) of the estimator $\hat{\theta}_n$ and $\hat{\nu}_n$ for the special case of $h_2 = 1$, when the inequality moment conditions are weakly informative. We will compare $\hat{\theta}_n$ to the following two alternative estimators: $\hat{\theta}_{(1)}$ is based only on $\mathbb{E}[g_1(X_i,\theta)] = 0$, and $\hat{\theta}_{(12)}$ is obtained by imposing $\mathbb{E}[g_1(X_i,\theta)] = 0$ and $\mathbb{E}[g_2(X_i,\theta)] = 0$. A natural benchmark for the evaluation of $\hat{\nu}_n$ can be obtained as follows: use the equality moment condition $\mathbb{E}[g_1(X_i,\theta_0)] = 0$ to calculate $\hat{\theta}_{(1)}$, then take a sample average of $g_2(X_i,\hat{\theta}_{(1)})$ to obtain the estimator $\hat{\nu}_{(1)}$.

Define $\widetilde{\mathcal{P}} = \phi_0 + (RJ^{-1}R')^{-1}RJ^{-1}Z$. The concentrated limit objective function for ϕ can be written as:

$$\bar{\mathcal{G}}_q^*(\phi) = \frac{1}{2}(\phi - \tilde{\mathcal{P}})'\Upsilon^{-1}(\phi - \tilde{\mathcal{P}}) + \frac{1}{2}Z'(J^{-1} - J^{-1}R'(RJ^{-1}R')^{-1}RJ^{-1})Z,$$

where

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

The partitions of Υ conform with the partition $\phi = [s', u']'$. Without loss of generality we are re-normalizing the inequality moment condition such that $\Upsilon_{uu} = 1$. Let $f_{\mathcal{N}}(\cdot)$ denote the probability density function and $F_{\mathcal{N}}(\cdot)$ the cumulative density function of a $\mathcal{N}(0,1)$. We show in Appendix A.6 that

$$\begin{aligned} \mathbb{E}[\mathcal{S}] &= \Upsilon_{su}[f_{\mathcal{N}}(u_0) - u_0(1 - F_{\mathcal{N}}(u_0))] \\ V[\mathcal{S}] &= \Upsilon_{ss} + \Upsilon_{su}\Upsilon_{us}F_{\mathcal{N}}(u_0)\Bigg(1 - \frac{f_{\mathcal{N}}^2(u_0)}{F_{\mathcal{N}}^2(u_0)} - \frac{u_0 f_{\mathcal{N}}(u_0)}{F_{\mathcal{N}}(u_0)} \\ &\quad + \left(u_0 + \frac{f_{\mathcal{N}}(u_0)}{F_{\mathcal{N}}(u_0)}\right)^2 [1 - F_{\mathcal{N}}(u_0)]\Bigg) - \Upsilon_{su}\Upsilon_{us} \end{aligned}$$

and the mean-squared-error is given by

$$MSE(\mathcal{S}) = \Upsilon_{ss} + \Upsilon_{su}\Upsilon_{us}[(u_0^2 - 1)(1 - F_{\mathcal{N}}(u_0)) - u_0 f_{\mathcal{N}}(u_0)].$$

The limit distributions of the estimator $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(12)}$ can be expressed as

$$\mathcal{S}_{(1)} \sim \mathcal{N}\Big(0, \Upsilon_{ss}\Big) \quad \text{and} \quad \mathcal{S}_{(12)} \sim \mathcal{N}\Big((QJ^{-1}Q')^{-1}QJ^{-1}M'u_0, \Upsilon_{ss} - \Upsilon_{su}\Upsilon_{us}\Big). \tag{23}$$

Since (see Pollard (2002, page 317))

$$(u_0^2 - 1)(1 - F_{\mathcal{N}}(u_0)) - u_0 f_{\mathcal{N}}(u_0) \begin{cases} = -\frac{1}{2} & \text{if } u_0 = 0 \\ < -\frac{1}{u_0} f_{\mathcal{N}}(u_0) & \text{if } u_0 > 0 \end{cases}$$

we obtain the following efficiency result:

Theorem 4 *Suppose Assumptions 1 – 3 are satisfied and $h_2 = 1$. Then (i) $MSE(\mathcal{S}) \le MSE(\mathcal{S}_{(1)})$ for all values of u_0. (ii) If $u_0 = 0$ then $MSE(\mathcal{S}_{(12)}) \le MSE(\mathcal{S})$. (iii) There exists a $\bar{u}_0 > 0$ such that $MSE(\mathcal{S}) \le MSE(\mathcal{S}_{(12)})$ for $u_0 \ge \bar{u}_0$.*

According to Theorem 4 the estimator that exploits the inequality moment condition is always preferable, in an asymptotic MSE sense, to the estimator $\hat{\theta}_{(1)}$ that ignores this additional information. If $\mathbb{E}[g_2(X_i, \theta_0)] = 0$ then it is preferable to impose it. However, the performance of $\hat{\theta}_{(12)}$ deteriorates as the slackness in the inequality constraint increases and will be inferior to our proposed estimator $\hat{\theta}_n$ for large values of u_0.

Next we consider the estimation of the slackness in the inequality moment condition, u. In the special of $h_2 = 1$ with a weakly informative moment restriction, we can deduce from Theorem 3 that the limit distribution of $\hat{u}$ is a censored normal distribution

$$\mathcal{U} = \widetilde{\mathcal{P}}_u I\{\widetilde{\mathcal{P}}_u \geq 0\}$$

where

$$\widetilde{\mathcal{P}}_u \sim \mathcal{N}\left(u_0, (M\Omega M')^{-1}\right), \quad \Omega = J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}.$$

The benchmark estimator $\hat{\nu}_{(1)}$ described above can be formally expressed as

$$\hat{\nu}_{(1)} = \frac{1}{n}\sum_{i=1}^{n} g_2\left(X_i, \hat{\theta}_{(1)}\right) I\left\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_2\left(X_i, \hat{\theta}_{(1)}\right) \geq 0\right\}.$$

This estimator has been used, for instance, by Zeldes (1989) to test whether low wealth-to-income households are in fact borrowing constrained. It can be verified that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g_2\left(X_i, \hat{\theta}_{(1)}\right) \Longrightarrow \mathcal{U}_{(1)} \sim \mathcal{N}\Big(u_0, \Xi_{(1)}\Big),$$

where

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

Hence, the limit distribution of the benchmark estimator $\hat{\nu}_{(1)}$ is also a truncated normal distribution. The following theorem states that the slackness estimator that uses the inequality moment condition is more precise than the estimator that ignores it.

Theorem 5 *Suppose Assumptions 1 – 3 are satisfied and $h_2 = 1$, then $MSE(\mathcal{U}) \leq MSE(\mathcal{U}_{(1)})$.*

6 Inference

Based on the results obtained in Section 4, we will proceed by deriving asymptotically valid confidence sets for θ and ν.

6.1 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

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

The test statistic is given by

$$\mathcal{LR}_n^\theta(\theta^H) = 2n\bigg(G_n^*(\theta^H, \hat{\nu}_n^H, \hat{\lambda}(\theta^H, \hat{\nu}_n^H)) - G_n^*(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n)) \bigg). \tag{24}$$

Denote the concentrated limit objective function by

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

Define the set

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

The limit distribution under H_0 can be easily obtained as a corollary from Theorem 3.

Corollary 1 *Suppose Assumptions 1 – 3 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*

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

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:

$$\mathcal{CS}_n^\theta(u_0, \alpha) = \left\{ \theta_0 \in \Theta \mid \mathcal{LR}_n^\theta(\theta_0) \leq c_\alpha^\theta(u_0) \right\}. \tag{26}$$

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

Corollary 2 *Suppose Assumptions 1 – 3 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 $\mathcal{CS}_n^u(\alpha_u)$ be a confidence set for u_0 with coverage probability $1 - \alpha_u$. Define,

$$\mathcal{CS}_n^\theta(\alpha) = \bigcup_{u \in \mathcal{CS}_n^u(\alpha_u)} \mathcal{CS}_n^\theta(u, \alpha_s). \tag{27}$$

Then,

$$\begin{aligned} P_{u_0}\Big\{\theta_0 \notin \mathcal{CS}_n^\theta(\alpha)\Big\} &\leq P_{u_0}\bigg(\Big\{\theta_0 \notin \mathcal{CS}_n^\theta(\alpha)\Big\} \cap \Big\{u_0 \in \mathcal{CS}_n^u(\alpha_u)\Big\}\bigg) + P_{u_0}\Big\{u_0 \notin \mathcal{CS}_n^u(\alpha_u)\Big\} \\ &\leq P_{u_0}\Big\{\theta_0 \notin \mathcal{CS}_n^\theta(u_0, \alpha_s)\Big\} + P_{u_0}\Big\{u_0 \notin \mathcal{CS}_n^u(\alpha_u)\Big\} \longrightarrow \alpha_s + \alpha_u. \end{aligned}$$

The Bonferroni confidence interval raises two questions. First, how should one construct the confidence set $\mathcal{CS}_n^u(\alpha_u)$, and second, how large should its coverage probability be. The next subsection discusses confidence intervals for u_0. The choice of α_u will be discussed in Section 7.

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

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

and redefine the test statistic as

$$\mathcal{LR}_n^\theta(\theta_1^H) = 2n\Big(G_n^*(\theta_1^H, \hat{\theta}_{2,n}^H, \hat{\nu}_n^H, \hat{\lambda}(\theta_1^H, \hat{\theta}_{2,n}^H, \hat{\nu}_n^H)) - G_n^*(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n))\Big). \tag{28}$$

The subsequent steps remain unchanged.

6.2 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

$$\mathcal{LR}_n^u(u^H) = 2n\Bigg(G_n^*(\hat{\theta}_n^H, n^{-1/2}u^H, \hat{\lambda}(\hat{\theta}_n^H, n^{-1/2}u^H)) - G_n^*(\hat{\theta}_n, \hat{\nu}_n, \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n))\Bigg). \tag{29}$$

We summarize its limit distribution in the following theorem.

Theorem 6 *Suppose Assumptions 1 – 3 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

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

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

$$P_{u_0}\Bigg\{\mathcal{LR}^u(u_0) \leq c_\alpha^u(u_0)\Bigg\} = 1 - \alpha.$$

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:

$$\mathcal{CS}_n^u(\alpha) = \Bigg\{u \geq 0 \mid \mathcal{LR}_n^u(u) \leq c_\alpha^u(u)\Bigg\}. \tag{30}$$

We can deduce from Theorem 6 that the confidence set has the desired coverage probability.

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

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

6.3 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:

$$\hat{J}_n = \frac{1}{n}\sum_{i=1}^n g(X_i, \hat{\theta}_n)g(X_i, \hat{\theta}_n)', \quad \hat{Q}_n = \frac{1}{n}\sum_{i=1}^n g^{(1)}(X_i, \hat{\theta}_n), \quad \hat{R}_n' = [-\hat{Q}_n', M']. \tag{31}$$

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 6) 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 (26) and (30).

7 Policy-Rule Estimation Revisited

In the remainder of this paper we provide a numerical example to illustrate the large sample results that we derived previously. We also conduct a small-scale Monte Carlo experiment to assess the finite-sample performance of our proposed estimation and inference procedure. Since in the context of monetary policy rules the slackness in the inequality moment condition $I\!E[g_2(X_i, \theta_0)] \geq 0$ is not of immediate interest, we focus on the estimation of θ_0.

7.1 Data Generating Process

We consider two versions of the prototypical New Keynesian DSGE model discussed in Section 2 as data generating processes. We refer to the first version, $\mathcal{M}_1$, as output growth rule specification. For $\mathcal{M}_1$ the measure of output used in the monetary policy rule (2) is $\widetilde{x}_t = \widetilde{y}_t - \widetilde{y}_{t-1} + \widetilde{z}_t$, log total factor productivity has a stochastic trend and is given by $\widetilde{A}_t = \widetilde{A}_{t-1} + \widetilde{z}_t$, and $\widetilde{y}_t$ measures percentage deviations of output from the level of productivity. $\mathcal{M}_1$ consists of Equations (2) to (6).

The second version of the model, $\mathcal{M}_2$, will be called output gap rule version. The measure of output used in the policy rule is $\widetilde{x}_t = \widetilde{y}_t$. Moreover, we regard log productivity as trend stationary process $\widetilde{A}_t = \gamma t + \widetilde{A}_* + \widetilde{z}_t$, and define $\widetilde{y}_t$ as percentage deviations from a deterministic trend. Euler equation and Phillips curve are modified as follows:

$$\widetilde{y}_t = I\!E_t[\widetilde{y}_{t+1}] - \frac{1}{\tau}(\widetilde{R}_t - I\!E_t[\widetilde{\pi}_{t+1}]) + (1-\rho_g)\widetilde{g}_t \quad (32)$$

$$\widetilde{\pi}_t = \beta I\!E_t[\widetilde{\pi}_{t+1}] + \kappa(\widetilde{y}_t - \widetilde{g}_t - \widetilde{z}_t) \quad (33)$$

Hence, $\mathcal{M}_2$ consists of Equations (2), (5), (6), (32), and (33).

Models $\mathcal{M}_1$ and $\mathcal{M}_2$ are written as linear rational expectations systems that can be solved with standard techniques, e.g. Sims (2002), to derive a law of motion for interest rates, inflation, and output. We assume that a time period corresponds to one quarter. The models can be completed by defining a set of measurement equations that relate $\widetilde{R}_t$, $\widetilde{\pi}_t$, and $\widetilde{y}_t$ to a set of observables. For our analysis, we assume that we have observations on annualized quarter-to-quarter inflation rates (INFL), and annualized nominal interest rates (INT) in percentages. For specification $\mathcal{M}_1$ we observe quarter-to-quarter per capita GDP growth rates (YGR) and for specification $\mathcal{M}_2$ we have observations of percentage deviations of GDP from a deterministic trend (YGAP).

We will simulate samples of sizes $n = 80$ and $n = 160$ from models $\mathcal{M}_1$ and $\mathcal{M}_2$ and estimate the coefficients of the monetary policy rule (2). The sample sizes are consistent with the number of observations used in actual applications. Many industrial countries experienced disinflation episodes and monetary policy shifts in the 1980s. Hence $n = 80$ can be thought of as a post-disinflation sample, whereas an $n = 160$ sample would contain observations from the 1960s to the present. The interest-rate feedback rule can be expressed in terms of observables as (switching from t to i subscript)

$$INT_i = \rho_R INT_{i-1} + (1-\rho_R)\psi_1 INFL_i + 4(1-\rho_R)\psi_2 OUTPUT_i + 4\epsilon_{R,i}, \qquad (34)$$

where $OUTPUT$ is either YGR ($\mathcal{M}_1$) or $YGAP$ ($\mathcal{M}_2$). A common approach in practice is to use lagged inflation and a measure of lagged output as instrumental variables. We define[4]

$$\begin{aligned} y_i &= INT_i, \quad x_i = [INT_{i-1}, INFL_i, OUTPUT_i]', \\ z_{1,i} &= [INT_{i-1}, INFL_{i-1}, YGR_{i-1}]', \quad z_{2,i} = -OUTPUT_i \end{aligned}$$

Let $X_i = [y_i, x_i', z_i']'$, where $z_i = [z_{1,i}', z_{2,i}']'$, and $\theta = [\rho_R, (1-\rho_R)\psi_1, 4(1-\rho_R)\psi_2]'$. Moreover, we define $g_j(X_i,\theta) = z_{j,t}(y_i - x_i'\theta)$, $j = 1, 2$ and obtain the desired moment conditions (1).

Both model specifications $\mathcal{M}_1$ and $\mathcal{M}_2$ have been used in empirical work with DSGE models. In some applications, e.g. Lubik and Schorfheide (2006), aggregate output is modelled as unit root process and output growth is included as argument in the monetary policy rule ($\mathcal{M}_1$), whereas in other applications, e.g., Smets and Wouters (2003) and Del Negro and Schorfheide (2005), output is detrended prior to estimation by either linear trend extraction or HP-filtering and the monetary policy rule is written as a function of detrended output ($\mathcal{M}_2$).[5]

From an econometric perspective it is interesting to consider the two versions of the DSGE model for the following reason. The serial correlation of output growth rates, for instance, in U.S. data is fairly small, whereas output deviations from trends tend to be highly autocorrelated. To capture these properties of the actual data we parameterize $\mathcal{M}_1$ using a value of $\rho_z = 0.5$, whereas under $\mathcal{M}_2$ $\rho_z = 0.95$. As a consequence, the correlation between instruments and regressors is larger for $\mathcal{M}_2$ than it is for $\mathcal{M}_1$. This suggests that MSE reductions due to the use of the inequality moment condition are potentially large for the output growth rule version of the DSGE model.

Numerical values for the remaining structural parameters are provided in Table 1 and are in general in line with estimates obtained from U.S. or Euro Area data. The parameter κ controls the slackness in the inequality moment condition. If there is a low degree of price stickiness in the economy the value of κ will be large, monetary policy shocks have little effect on output and $I\!E[g_2(X_i,\theta_0)]$ will be close to zero. Vice versa, if κ is small the slackness in the inequality moment condition tends to be large. $I\!E[g_2(X_i,\theta_0)]$, J and Q, which are needed to simulate the limit distributions, can be calculated as a function of the structural parameters from the solution of the log-linearized DSGE models.

[4] In principle we could include $INFL_i$ also in the definition of $z_{2,i}$, which would introduce a second nuisance parameter.

[5] The theoretical literature on New Keynesian models defines output gap as the deviation of actual output from the level of output that would prevail in the absence of nominal rigidities. However, in the empirical literature on the estimation of monetary policy rules it is common to define the output gap as deviations of output from a smooth trend. For instance, the U.S. potential output series constructed by the Congressional Budget Office closely resembles the trend that is extracted by HP-filtering U.S. GDP.

7.2 Interpretation of Local-to-Zero Framework

In Sections 3 to 6 we assumed that the slackness in the inequality moment constraint, $\mathbb{E}[g_2(X_i, \theta_0)]$, is asymptotically small in the sense that $\nu_{n,0} = n^{-1/2}u_0$. The local-to-zero representation is a technical device that allows us to capture in the asymptotic calculation the notion that in finite samples the slackness in the inequality moment condition may be sufficiently small to provide overidentifying information about the parameters of interest. In the context of the policy function estimation the local-to-zero setup should not be interpreted as the belief that the price stickiness in the economy will vanish as the sample size tends to infinity. In fact, all structural parameters, including the slope of the Phillips curve κ, are fixed in the DSGE model. For the subsequent analysis we compute $\mathbb{E}[g_2(X_i, \theta_0)]$ as a function of the parameter values reported in Table 1. We use the limit distribution obtained under $u_0 = \sqrt{n}\mathbb{E}[g_2(X_i, \theta_0)]$ to approximate the finite sample properties of estimators and confidence intervals for sample sizes $n = 80$ and $n = 160$.

7.3 Alternative Estimators and Confidence Sets

To assess the performance of the proposed point estimator $\hat{\theta}$ we consider $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(12)}$ defined in Section 5.4 as alternatives. Inference with respect to θ_0 is based on the following two 90% confidence sets: (i) $\mathcal{CS}^{\theta}(\alpha)$ is obtained from $\mathbb{E}[g_1(X_i, \theta)] = 0$ and $\mathbb{E}[g_2(X_i, \theta)] \geq 0$ as described in Section 6. In computing the Bonferroni interval we use $\alpha_u = 0.05$ and $\alpha_u = 0$. (ii) $\mathcal{CS}^{\theta}_{(1)}$ uses only $\mathbb{E}[g_1(X_i, \theta)] = 0$ and is the Wald confidence interval based on estimates of the asymptotic standard errors. Our confidence intervals are constructed for individual parameters, instead of jointly for the entire parameter vector.

7.4 Point Estimation Results

Tables 2 to 4 summarize the performance of the three point estimators $\hat{\theta}$, $\hat{\theta}_{(1)}$, and $\hat{\theta}_{(12)}$. All results are reported in terms of the transformed parameter vector $s = \sqrt{n}(\theta - \theta_0)$. The entries in the columns labelled *Asymptotics* are calculated based on 1,000,000 draws from the limit distribution, where $u_0 = \sqrt{n}\mathbb{E}[g_2(X_i, \theta_0)]$ as discussed above. The entries under the heading *Small Sample* are obtained by applying the estimation procedures to 10,000 samples of size n, simulated from the DSGE model.

Since the estimation problem is linear and the number of inequality moment conditions is $h_2 = 1$ the computational problem simplifies considerably. Based on $\mathbb{E}[g_1(X_i, \theta_0)]$ the parameters are exactly identified and the empirical likelihood estimator $\hat{\theta}_{(1)}$ corresponds to the linear IV estimator. If we also impose that $\mathbb{E}[g_2(X_i, \theta_0)]$ then the model is overidentified. For the small sample analysis we use a numerical optimization procedure to find the saddle point

$$\hat{\theta}_{(12)} = \text{argmin}_{\theta \in \Theta} \max_{\lambda_1, \lambda_2} G_n(\theta, \lambda_1, \lambda_2)$$

without imposing a sign restriction on λ_2.[6] If $\hat{\lambda_2}(\hat{\theta}_{(12)}) \leq 0$ we deduce that the estimator $\hat{\theta}$ that treats the second moment condition as inequality equals $\hat{\theta}_{(12)}$. Alternatively, if $\hat{\lambda}_2(\hat{\theta}_{(12)}) > 0$ then $\hat{\theta} = \hat{\theta}_{(1)}$. This insight can also be applied to the simulation of the limit distribution. To characterize the performance of the estimators we consider the following three robust statistics[7]: the median of $\hat{s}$, the distance between the 5th and the 95th percentile, and the median of the squared estimation error $\hat{s}^2$.

Table 2 is based on a parameterization of $\mathcal{M}_1$ in which prices are nearly flexible and the slackness in the inequality moment condition is small. Since the policy rule is specified in terms of output growth, which is only weakly correlated with lagged output growth, inflation, and interest rates, the estimator $\hat{\theta}_{(1)}$ performs poorly, in particular with respect to the output growth coefficient. While imposing incorrectly that $I\!E[g_2(X_i, \theta_0)] = 0$ introduces a bias in the estimation of the output growth coefficient, the variability of the estimator drops considerably. According to the limit distribution, the median of the squared error drops from 5.62 to 0.11 for $n = 80$. Using the second moment condition as inequality also leads to a considerable improvements in performance. Across the board, $\hat{s}$ dominates $\hat{s}_{(1)}$ both asymptotically and in finite samples. The median squared error of the output growth coefficient is reduced by approximately 90%.

In the $\mathcal{M}_1$ example the best estimator is the one that incorrectly imposes $I\!E[g_2(X_i, \theta_0)] = 0$. However, imposing invalid moment conditions can also generate very misleading estimates as we will illustrate in our second simulation. Table 3 is based on a parameterization of $\mathcal{M}_2$ in which prices are sticky, implying that the slackness in the inequality moment condition is large. It turns out that $\hat{\theta}_{(12)}$ is severely biased and performs very poorly. Our inequality moment estimator, on the other hand, proves to be robust. However, since the inequality condition is not binding, we are unable to extract overidentifying information and $\hat{\theta}$ is essentially equal to the estimator $\hat{\theta}_{(1)}$ which ignores $I\!E[g_2(X_i, \theta_0)]$.

At last, we consider a version of $\mathcal{M}_2$ in which prices are nearly flexible, which reduces the slackness in the inequality moment condition compared to the second experiment. Results are summarized in Table 4. Under this parameterization $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(12)}$ perform about equally well. The former estimator is slightly more variable, but the latter has a larger bias. For all three parameters, our inequality-based estimator performs no worse than $\hat{\theta}_{(1)}$ and $\hat{\theta}_{(12)}$. In fact, in some instances $\hat{\theta}$ beats its two competitors, albeit with a small margin.

[6] The optimization is carried out with a version of the BFGS quasi-Newton algorithm, written originally by Chris Sims for the ML estimation of a DSGE model. The algorithm uses a fairly simple line search and randomly perturbs the search direction if it reaches a cliff. We replace the ln-function in the definition of $G_n(\theta, \lambda_1, \lambda_2)$ by the $\ln_*$ function described in Owen (2001). Samples for which the numerical optimization fails in an obvious manner are disregarded.

[7] Since the small sample distribution of all 3 estimators exhibits fat tails (see Mariano (1982) for the non-existence of finite sample moments in the classical simultaneous equations model), we report robust statistics.

To summarize, both according to the limit distribution and the small sample simulation results, the inequality-based estimator performs no worse than $\hat{\theta}_{(1)}$. In situations in which there is additional information contained in the inequality moment condition, our estimator is able to exploit that information. At the same time the estimator is robust to large values of $\mathbb{E}[g_2(X_i, \theta_0)]$ and, unlike $\hat{\theta}_{(12)}$ its performance does not break down as u_0 increases. Despite the small sample sizes considered, the asymptotic results proved to be a fairly reliable indicator of small sample performance.[8]

7.5 Interval Estimation Results

Table 5 presents coverage probabilities and average lengths for the confidence intervals $\mathcal{CS}^\theta(\alpha)$ and $\mathcal{CS}^\theta_{(1)}(\alpha)$, where $\alpha = 0.1$. Data are generated from the version of $\mathcal{M}_2$ in which prices are nearly flexible. Recall from Table 4 that for this data generating process the median squared error of $\hat{s}$ approximately equals that of $\hat{s}_{(1)}$. However, the variability of $\hat{s}$ is smaller than the variability of $\hat{s}_{(1)}$ which essentially will translate into a reduction of the length of the confidence interval that exploits the inequality moment condition.

The computations are implemented as follows. To simulate the asymptotic behavior of $\mathcal{CS}^\theta(\alpha)$ we begin by evaluating Q and J as a function of θ_0. Second, we calculate the critical value functions $c^\theta_{\alpha_s}(u_0)$ and $c^u_{\alpha_u}(u_0)$ for u_0 on a grid $\mathbb{U}$ based on 100,000 draws from the limit distribution of the likelihood ratio statistics. Third, we draw 1,000,000 Z's and compute confidence intervals. Specifically, for each Z we evaluate $\mathcal{LR}^u(u)$ for $u \in \mathbb{U}$ and use the previously calculated critical value function to obtain $\mathcal{CS}^u(\alpha_u)$. To obtain the Bonferroni interval $\mathcal{CS}^\theta(\alpha)$ we determine the supremum of the critical values $c^\theta_\alpha = \sup_{u \in \mathcal{CS}^u(\alpha_u)} \; c^\theta_{\alpha_s}(u)$ and find the boundaries of the confidence set $\mathcal{CS}^\theta(\alpha)$ by numerically solving $\mathcal{LR}^\theta(\theta) = c^\theta_\alpha$. Our small sample analysis is based on 1,000 samples of 160 observations. For each sample we begin by computing the point estimator $\hat{\theta}$ as well as the estimates $\hat{J}$ and $\hat{Q}$ described in Section 6.3. We then calculate the critical value functions $c^\theta_{\alpha_s}(u_0)$ and $c^u_{\alpha_u}(u_0)$ conditional on $\hat{J}$ and $\hat{Q}$ and proceed with the computation of the confidence intervals as in the simulation of the limit distribution.

The critical value function $c^\theta_{\alpha_s}(u_0)$ for the three parameter of the monetary policy rule is plotted in Figure 1. As u_0 increases, the moment $\mathbb{E}[g_2(X_i, \theta)]$ becomes irrelevant and the critical value converges to the critical value of a χ^2 distribution with one degree of freedom. In our example the critical value function has an inverted hump shape. Since the true value of u_0 is $\sqrt{160} \cdot 0.11 = 1.39$ we expect the Bonferroni interval for the interest rate and the inflation coefficients in the policy rule (Parameters 1 and 2) to be slightly conservative, whereas the intervals for the output coefficient should have essentially the

[8] Indeed, if the sample size is increased to $n = 500$ or $n = 1000$ the finite sample behavior is well approximated by the limit distribution, not just for the robust statistics reported in the tables, but also for means, standard deviations, and MSEs.

target coverage $1 - \alpha_s$. It turns out that in our application the confidence intervals for the slackness parameter are fairly large and tend to cover zero as well as values of u_0 for which the critical value essentially equals the χ^2 critical value. Since the critical value function is convergent as u_0 increases it is preferable to set $\alpha_u = 0$. Nevertheless, we also report results for $\alpha_u = 0.05$.

As expected, the simulation of the limit distribution implies that the coverage probabilities of $\mathcal{CS}^\theta$ for the output coefficient is essentially equal to $1 - \alpha_s$. For the other two coefficients it is slightly larger: 0.96 and 0.91, respectively. If $\alpha_u = 0.05$ the $\mathcal{CS}^\theta$ intervals are longer than the $\mathcal{CS}^\theta_{(1)}$ intervals because the Bonferroni approach leads to very conservative intervals. If we set $\alpha_u = 0$ then the average length of $\mathcal{CS}^\theta$ decreases and the intervals that take advantage of the inequality moment condition are shorter than the Wald confidence intervals $\mathcal{CS}^\theta_{(1)}$, both asymptotically and in finite samples. On average, the asymptotic confidence intervals are slightly shorter than the finite sample intervals, but the coverage probabilities are very similar.

8 Conclusion

This paper developed a limit distribution theory for moment-based estimators when some of the moment conditions take the form of inequalities. If the slackness in the inequality moment condition is small our estimator is able to translate the additional information provided by the inequality to a mean-squared-error reduction. If on the other hand, the slackness in the inequality moment conditions is large, our estimator performs no worse than an estimator that ignores the moment inequalities. 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.

Finally, throughout the paper we focused on models in which the parameter θ is identifiable based on the equality moment condition $\mathbb{E}[g_1(X_i, \theta_0)] = 0$. We think that this is an important class of models and provided a substantive illustration with the estimation of monetary policy rules. Our procedures are especially attractive for instrumental variable estimation problems in which there are only a few valid instruments available that suffer from a small correlation with the endogenous regressors. Our procedures can also be used to sharpen inference in the estimation of intertemporal optimality conditions in the presence of financial frictions.

A Appendix: 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 Kitamura, Tripathi, and Ahn (2004) and 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 Equivalence of Saddlepoint Problems

One could also rewrite the second moment condition as

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

and restrict the auxiliary parameter $\vartheta_{n,0}$ to be nonnegative. The estimators $\hat{\theta}$ and $\hat{\vartheta}$ can be defined as the saddlepoint

$$\min_{\theta\in\Theta,\ \vartheta\geq 0}\ \max_{\lambda_1\in\hat{\Lambda}_{n,1}(\theta),\ \lambda_2\in\hat{\Lambda}_{n,2}(\theta)} \tilde{G}_n(\theta,\vartheta,\lambda_1,\lambda_2), \tag{A.1}$$

where

$$\tilde{G}_n(\theta,\vartheta,\lambda_1,\lambda_2) = \frac{1}{n}\sum_{i=1}^{n}\ln(1+\lambda_1' g_1(X_i,\theta)+\lambda_2'[g_2(X_i,\theta)-\vartheta]). \tag{A.2}$$

The partition of $\hat{\Lambda}_n(\theta)$ into $\hat{\Lambda}_{n,1}(\theta)$ and $\hat{\Lambda}_{n,2}(\theta)$ conforms with the partition of $\lambda=[\lambda_1',\lambda_2']'$. As in (11) 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 (Equation 10), G_n^* (Equation (12), and $\tilde{G}_n$ have the same saddlepoints.

Lemma 1 *$\hat{\theta}$, $\hat{\lambda}_1$, $\hat{\lambda}_2$ are a saddlepoint of $G_n(\theta,\lambda_1,\lambda_2)$ and solve (9),*

(i) if and only if $\hat{\theta}$, $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\nu}$ are a saddlepoint of $G_n^*(\theta,\nu,[\lambda_1',\lambda_2']')$ and solve (11);

(ii) if and only if $\hat{\theta}$, $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\vartheta}$ are a saddlepoint of $\tilde{G}_n(\theta,\vartheta,\lambda_1,\lambda_2)$ and solve (A.1).

The elements of the $h_2\times 1$ vectors $\hat{\nu}$ and $\hat{\vartheta}$ are defined as

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

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 $\hat{\Lambda}_{n,2}(\theta)$ and satisfies the first-order condition

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

If $\hat{\lambda}_{2,j} \neq 0$ 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 Preliminaries

Throughout the appendix we are frequently using the following results. Notice that Assumptions 1(a), (b), (c), and (i) imply that

$$\max_{1\leq i\leq n} \sup_{\theta\in\Theta} \|g(X_i,\theta)\| = O_p\left(n^{1/\alpha}\right), \tag{A.3}$$

$$\frac{1}{n}\sum_{i=1}^{n} \sup_{\theta\in\Theta} \|g(X_i,\theta)\|^{\alpha} = O_p(1), \tag{A.4}$$

$$\sup_{\theta\in\Theta} \left\| \frac{1}{n}\sum_{i=1}^{n} \left\{ g(X_i,\theta)\, g(X_i,\theta)' - I\!E\left[g(X_i,\theta)\, g(X_i,\theta)' \right] \right\} \right\| = o_p(1). \tag{A.5}$$

According to Lemma 2.4 of McFadden and Newey (1994), under Assumptions 1 and 2,

$$I\!E\left[g^{(1)}(X_i,\theta)\right] \text{ and } I\!E\left[g_j^{(2)}(X_i,\theta)\right] \text{ are uniformly continuous,} \tag{A.6}$$

and

$$\sup_{\theta\in\Theta} \left\| \frac{1}{n}\sum_{i=1}^{n} \left\{ g^{(1)}(X_i,\theta) - I\!E\left[g^{(1)}(X_i,\theta)\right] \right\} \right\| = o_p(1) \tag{A.7}$$

$$\sup_{\theta\in\Theta} \left\| \frac{1}{n}\sum_{i=1}^{n} \left\{ g_j^{(2)}(X_i,\theta) - I\!E\left[g_j^{(2)}(X_i,\theta)\right] \right\} \right\| = o_p(1) \text{ for all } j = 1, ..., h.$$

A.3 Consistency

A.3.1 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

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

Define

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

To simplify the notation we omit the subscript n from the set N_0^c. Recall that according to Assumption 1(i), the constant $\alpha > 2$ is such that $I\!E[\sup_{\theta\in\Theta}\|g(X_i,\theta)\|^{\alpha}] < K$. 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 $n \geq \bar{n}$

$$P\left\{\bar{G}_n^*(\theta_0,\nu_{n,0}) \geq n^{-\zeta-\kappa}\eta\right\} < \frac{\varepsilon}{2}, \tag{A.8}$$

where

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

and (ii)

$$P\left\{ \min_{\theta \in \Theta_0^c,\, \nu \in N_0^c} \bar{G}_n^*\left(\theta, \nu\right) \leq n^{-\zeta}\eta \right\} < \frac{\varepsilon}{2}. \tag{A.9}$$

Then, from (A.8) and (A.9) we deduce that there exists an $\eta > 0$ such that for $n \geq \bar{n}$:

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

Proof of (i): By Lemma A.2 $\bar{G}_n^*(\theta_0, \nu_{n,0}) \leq O_p(1/n)$. 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

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

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

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

We then replace the terms

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

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

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

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

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

and

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

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.4 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

$$n^{\zeta}q_n(x,\theta,\nu) = u(\theta,\nu)'(g_n(x,\theta)-M'\nu) - \frac{1}{2}\frac{n^{-\zeta}u'(\theta,\nu)g_n(x,\theta)g_n(x,\theta)'u(\theta,\nu)}{\left(1+n^{-\zeta}u_*'(\theta,\nu)g_n(x,\theta)\right)^2}, \tag{A.10}$$

where $u_*(\theta,\nu)$ lies between zero and $u(\theta,\nu)$. The second-order term of the Taylor approximation (A.10) can be bounded as follows. For given x, θ, and ν

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

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

$$\sup_{\theta\in\Theta,\ \nu} n^{-\zeta}\frac{u(\theta,\nu)'g_n(x,\theta)g_n(x,\theta)'u(\theta,\nu)}{\left(1+n^{-\zeta}u_*(\theta,\nu)'g_n(x,\theta)\right)^2} \le \sup_{\theta\in\Theta,\ \nu} n^{-\zeta}\frac{\|g_n(x,\theta)\|^2\|u(\theta,\nu)\|^2}{(1-n^{-\zeta/2})^2} \le n^{-\zeta+2\xi} = o(1). \tag{A.11}$$

Now consider the expected value of $n^{\zeta}q_n(x,\theta,\nu)$. From (A.10), (A.11), and by the dominated convergence theorem, we have

$$\begin{aligned}
n^{\zeta}I\!E\left[q_n(X_i,\theta,\nu)\right] &= u'(\theta,\nu)(I\!E[g_n(X_i,\theta)]-M'\nu)+o(1) \qquad (A.12)\\
&= \begin{cases} o(1) & \text{if } \theta=\theta_0,\ \nu=\nu_{n,0} \\ \|I\!E[g(X_i,\theta)]-M'\nu\|+o(1)>0 & \text{otherwise} \end{cases}.
\end{aligned}$$

The $o(1)$ terms absorb the second-order term of the Taylor approximation and the discrepancy between $I\!E[g_n(X,\theta)]$ and $I\!E[g(X,\theta)]$, which vanishes as $\mathcal{X}_n$ expands. From (A.12) and the monotone convergence theorem we can deduce that

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

Since Θ and $\mathbb{V}$ are compact by assumption, the sets $\Theta\cap\mathbb{B}(\theta_0,\delta)^c$ and $\mathbb{V}\cap\mathbb{B}(\nu_{n,0},\delta)^c$ are compact. We can cover both $\Theta\cap\mathbb{B}(\theta_0,\delta)^c$ and $\mathbb{V}\cap\mathbb{B}(\nu_{n,0},\delta)^c$ 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

$$n^{\zeta}I\!E\left[\inf_{\theta\in\Theta_j,\ \nu\in N_j} q_n(X_i,\theta,\nu)\right] \ge 2\eta_j,\ n\ge n_j \tag{A.13}$$

for some positive numbers $\eta_j = \eta_j(\delta)$, $j = 1, \ldots, J$. By the WLLN[9] and (A.13), for a given $\varepsilon > 0$, we can find $\bar{n}_j' s$ such that $n \geq \bar{n}_j$ implies that

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

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

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

as required part (ii)-(b).

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

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

as required for (ii).

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

A.3.2 Technical Lemmas

Lemma A.1 *Suppose that Assumption 1 is 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 Assumption 1 is 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,*

[9]Notice that

$$I\!\!E\left[\left(n^{\zeta}\inf_{\theta\in\Theta_j,\, \nu\in N_j} q_n(X_i,\theta,\nu)\right)^2\right] \leq I\!\!E\left[\sup_{\theta\in\Theta} 2\left\|g(X_i,\theta)\right\|^2\right] + 2K + n^{-2\zeta+4\xi} < \infty. \tag{A.14}$$

(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: The proof is similar to that of Lemma A2 in Newey and Smith (2004).

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 (A.5) and Assumption 1(f). The remainder of the proof follows the proof of Lemma A2 in Newey and Smith (2004). ■

Lemma A.3 *Suppose Assumption 1 is 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: The proof is similar to that of Lemma A.3 in Newey and Smith (2004).

Let $\hat{g}_i = g\left(X_i, \hat{\theta}\right) - M'\hat{\nu}$ and $\hat{g} = \frac{1}{n}\sum_{i=1}^n \left[g\left(X_i, \hat{\theta}\right) - M'\hat{\nu}\right]$. Define $\hat{u}\left(\hat{\theta}, \hat{\nu}\right) = n^{-\zeta}\frac{\hat{g}}{\|\hat{g}\|}$. (Recall the definition of $u\left(\theta, \nu\right)$ in the proof of consistency.) Approximation $G_n^*\left(\theta, \nu, \lambda\right)$ with respect to λ around $\lambda = 0$ at $(\theta, \nu, \lambda) = \left(\hat{\theta}, \hat{\nu}, \hat{u}\left(\hat{\theta}, \hat{\nu}\right)\right)$. 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 $\hat{u}(\hat{\theta}, \hat{\nu})$.

Notice that $\max_{1\leq i\leq n}\left|\hat{u}'(\hat{\theta},\hat{\nu})\hat{g}_i\right| \to_p 0$ 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, under Assumption 1 $\left\|\frac{1}{n}\sum_{i=1}^n \hat{g}_i\hat{g}_i'\right\| \leq 2\left(\frac{1}{n}\sum_{i=1}^n \sup_{\theta\in\Theta}\|g(X_i,\theta)\|^2 + K\right) = O_p(1)$. Then, w.p.a.1, for some constant C,

$$\begin{aligned}
&\hat{g}'\hat{u}\left(\hat{\theta},\hat{\nu}\right) - \frac{1}{2}\hat{u}'\left(\hat{\theta},\hat{\nu}\right)\left(\frac{1}{n}\sum_{i=1}^n \frac{\hat{g}_i\hat{g}_i'}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right) \\
&= n^{-\zeta}\|\hat{g}\| - \frac{1}{2}\hat{u}'\left(\hat{\theta},\hat{\nu}\right)\left(\frac{1}{n}\sum_{i=1}^n \frac{\hat{g}_i\hat{g}_i'}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right) \\
&\geq n^{-\zeta}\|\hat{g}\| - \frac{1}{2}\max_{1\leq i\leq n}\left(\frac{1}{\left(1+\ddot{\lambda}'\hat{g}_i\right)^2}\right)\hat{u}'\left(\hat{\theta},\hat{\nu}\right)\left(\frac{1}{n}\sum_{i=1}^n \hat{g}_i\hat{g}_i'\right)\hat{u}\left(\hat{\theta},\hat{\nu}\right) \\
&\geq n^{-\zeta}\|\hat{g}\| - Cn^{-2\zeta}.
\end{aligned} \tag{A.15}$$

Then,

$$n^{-\zeta}\|\hat{g}\| - Cn^{-2\zeta} \leq G_n^*\left(\hat{\theta},\hat{\nu},\hat{u}\left(\hat{\theta},\hat{\nu}\right)\right) \leq G_n^*\left(\hat{\theta},\hat{\nu},\hat{\lambda}\right) \leq \sup_{\lambda\in\hat{\Lambda}_n(\theta_0)} G_n^*\left(\theta_0,\nu_{n,0},\lambda\right) \leq O_p\left(\frac{1}{n}\right), \tag{A.16}$$

where the first inequality is from (A.15), 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

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

by Assumption 1(g). Also, by $\zeta < \frac{1}{2}$, $\zeta - 1 < -\frac{1}{2} < -\zeta$. Solving (A.16) for $\|\hat{g}\|$ gives

$$\|\hat{g}\| \leq O_p\left(n^{-\zeta}\right). \tag{A.17}$$

For a given sequence $\varepsilon_n \longrightarrow 0$, let $\bar{\lambda} = \varepsilon_n\hat{g}$. According to (A.17) $\bar{\lambda} = o_p\left(n^{-\zeta}\right)$. Hence, $\bar{\lambda} \in \Lambda_n^\zeta$ w.p.a.1. Then, as in (A.16), we have

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

For large enough n the term $1-C\varepsilon_n$ is bounded away from zero and 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 Assumption 1 is satisfied. Let* $g_n(x,\theta) = I\{x\in\mathcal{X}_n\}g(x,\theta)$ *where*

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

where $\frac{1}{\alpha^2} < \xi < \frac{1}{2\alpha}$ *and* $\alpha > 2$ *as in Assumption 1(i). Define*

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

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

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

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

$$\begin{aligned}
&\sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{1}{n}\sum_{i=1}^{n}\left\{q_n\left(X_i,\theta,\nu\right)-\tilde{q}_n\left(X_i,\theta,\nu\right)\right\}\right| \\
&= \sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\frac{n^{-\zeta}u'(\theta,\nu)g\left(X_i,\theta\right)}{1+n^{-\zeta}u_*'\left(\theta,\nu\right)g\left(X_i,\theta\right)}\right)I\left\{X_i\notin\mathcal{X}_n\right\}\right| \qquad \text{(A.18)}\\
&\leq \max_{1\leq i\leq n}\sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{n^{-\zeta}u'\left(\theta,\nu\right)g\left(X_i,\theta\right)}{1+n^{-\zeta}u_*'\left(\theta,\nu\right)g\left(X_i,\theta\right)}\right|\frac{1}{n}\sum_{i=1}^{n}I\left\{\sup_{\theta\in\Theta}\left\|g\left(X_i,\theta\right)\right\|>n^{\xi}\right\}\\
&\leq \frac{1}{n^{\alpha\xi}}\left(\max_{1\leq i\leq n}\sup_{\theta\in\Theta,\ \nu\geq 0}\left|\frac{n^{-\zeta}u'\left(\theta,\nu\right)g\left(X_i,\theta\right)}{1+n^{-\zeta}u_*'\left(\theta,\nu\right)g\left(X_i,\theta\right)}\right|\right)\left(\frac{1}{n}\sum_{i=1}^{n}\sup_{\theta\in\Theta}\left\|g\left(X_i,\theta\right)\right\|^{\alpha}\right)
\end{aligned}$$

where $u_*(\theta,\nu)$ is located between 0 and $u(\theta,\nu)$. The second term on the right-hand side of (A.18) can be bounded as follows. According to (A.3)

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

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

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

By Assumption 1(i) and the Markov inequality, the third term on the right-hand side of (A.18) is $O_p(1)$. Since $\frac{1}{\alpha^2}<\xi<\frac{1}{2\alpha}$, we are able to deduce that

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

as required. ■

A.4 Quadratic Approximation of the Objective Function

We begin by deriving the coefficient matrices for the quadratic approximation of the objective function

$$G_{nq}^{*}\left(\beta\right) = G_{n}^{*}\left(\beta_{n,0}\right)+G_{n}^{*(1)}\left(\beta_{n,0}\right)'\left(\beta-\beta_{n,0}\right)+\frac{1}{2}\left(\beta-\beta_{n,0}\right)'G_{n}^{*(2)}\left(\beta_{n,0}\right)\left(\beta-\beta_{n,0}\right). \qquad \text{(A.19)}$$

A direct calculation shows that

$$G_{n}^{*(1)}\left(\beta\right) = \left[G_{n}^{*(1)}\left(\beta\right)_{\theta}',G_{n}^{*(1)}\left(\beta\right)_{\nu}',G_{n}^{*(1)}\left(\beta\right)_{\lambda}'\right]', \qquad \text{(A.20)}$$

where

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

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

$$G_n^{*(1)}(\beta_{n,0}) = [0, 0, n^{-1/2} Z_n']. \tag{A.21}$$

We proceed by partitioning the matrix of second derivative as follows

$$G_n^{*(2)}(\beta) = \begin{pmatrix} G_n^{*(2)}(\beta)_{\theta\theta'} & G_n^{*(2)}(\beta)_{\theta\nu'} & G_n^{*(2)}(\beta)_{\theta\lambda'} \\ G_n^{*(2)}(\beta)_{\nu\theta'} & G_n^{*(2)}(\beta)_{\nu\nu'} & G_n^{*(2)}(\beta)_{\nu\lambda'} \\ G_n^{*(2)}(\beta)_{\lambda\theta'} & G_n^{*(2)}(\beta)_{\lambda\nu'} & G_n^{*(2)}(\beta)_{\lambda\lambda'} \end{pmatrix}, \tag{A.22}$$

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

$$G_n^{*(2)}(\beta_{n,0}) = \begin{bmatrix} 0 & 0 & Q_n \\ 0 & 0 & -M \\ Q_n' & -M' & -J_n \end{bmatrix}. \tag{A.23}$$

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

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

A.4.1 Main Results

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

$$\sup_{\beta\in\mathcal{B}_n:\|\beta-\beta_{n,0}\|\le\gamma_n} \frac{|\mathcal{R}_n(\beta)|}{(1+\|\sqrt{n}(\beta-\beta_{n,0})\|^2)} = o_p(1), \tag{A.24}$$

where $\mathcal{R}_n(\beta)$ is the remainder term in (14).

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

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

for every sequence $\gamma_n \longrightarrow 0$. $G_n^{*(2)}$ is defined in (A.22). To verify this sufficient condition we will subsequently show that

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

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

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

We begin by showing that

$$\sup_{\beta\in\mathcal{B}_n}\left|\frac{1}{1+\lambda'g(X_i,\theta)}\right|=O_p(1). \tag{A.25}$$

For any given $0<\delta<\frac{1}{2}$, set $K=\frac{1}{1-\delta}$. Then, since $\sup_{1\leq i\leq n,\ \beta\in\mathcal{B}_n}\left|\lambda'g(X_i,\theta)\right|\leq\delta$ implies $\sup_{1\leq i\leq n,\ \beta\in\mathcal{B}_n}\left|\frac{1}{1+\lambda'g(X_i,\theta)}\right|\leq K$,

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

which proves (A.25). The convergence result for the upper bound can be deduced from Lemma A.1.

(i) Notice that

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

where the last inequality holds by the definition of Λ_n^ζ, (A.25) and (A.7). Moreover,

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

The last inequality holds by the definition of Λ_n^ζ, (A.25) and (A.7).

(ii) Apply the triangle inequality to

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

where the last equality holds by (A.7) and (A.6). Next,

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

by Lemma A.1, (A.25), and (A.7). Moreover,

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

(iii) Similar as before, we have

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

Next,

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

Proof of Theorem 2: (i) Follows from Lemma A.6.

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

$$0=\mathcal{G}_n^*(\hat{\phi},0)\le\mathcal{G}_n^*(\hat{\phi},\hat{l}(\hat{\phi}))\le\mathcal{G}_n^*(0,\hat{l}(0)). \tag{A.26}$$

Then using the quadratic approximation (14), the bound for the remainder term given in Lemma 2 and the definition of $\hat{l}$ and $\hat{\phi}$ we obtain

$$\begin{aligned}\mathcal{G}_n^*(\hat{\phi},\hat{l}(\hat{\phi})) &= \mathcal{G}_{nq}^*(\hat{\phi},\hat{l}(\hat{\phi})) + (1+\|\hat{\phi}-\phi_0\|^2+\|\hat{l}(\hat{\phi})\|^2)o_p(1) \\ &= \frac{1}{2}(Z_n - R_n'(\hat{\phi}-\phi_0))'J_n^{-1}(Z_n - R_n'(\hat{\phi}-\phi_0)) \\ &\quad -\frac{1}{2}(\hat{l}(\hat{\phi}) - J_n^{-1}[Z_n - R_n'(\hat{\phi}-\phi_0)])'J_n(\hat{l}(\hat{\phi}) - J_n^{-1}[Z_n - R_n'(\hat{\phi}-\phi_0)]) \\ &\quad +(1+\|\hat{\phi}-\phi_0\|^2+\|\hat{l}(\hat{\phi})\|^2)o_p(1) \\ &= \frac{1}{2}(Z_n - R_n'(\hat{\phi}-\phi_0))'J_n^{-1}(Z_n - R_n'(\hat{\phi}-\phi_0)) + (1+\|\hat{\phi}-\phi_0\|^2+\|\hat{l}(\hat{\phi})\|^2)o_p(1),\end{aligned} \tag{A.27}$$

where $\phi_0 = [0, u_0']'$. The last equality is a consequence of Lemma A.7. Similarly, we can deduce from Lemmas A.2, 2, and Assumptions 2 and 3 that

$$\mathcal{G}_n^*(0,\hat{l}(0)) = -\frac{1}{2}\hat{l}\,(0)'\,J_n\hat{l}\,(0) + Z_n'\hat{l}\,(0) + (1+\|\hat{l}(0)\|^2)o_p(1) = O_p\,(1)\,. \tag{A.28}$$

Hence, from (A.26), (A.27), and (A.28) we obtain the inequality

$$0 \le \frac{1}{2}(Z_n + o_p(1) - R_n'(\hat{\phi}-\phi_0))'J_n^{-1}(Z_n + o_p(1) - R_n'(\hat{\phi}-\phi_0)) \le O_p(1). \tag{A.29}$$

Notice that $Z_n + o_p(1) = O_p\,(1)$. According to Assumption 1, R_n is full rank and J_n is positive definite w.p.a. 1. Therefore, (A.29) 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) we can rewrite the differential as

$$\begin{aligned} nG_{nq}^*(\hat{\beta}_n) - nG_{nq}^*(\tilde{\beta}_{nq}) &= \mathcal{G}_n^*(\hat{\phi},\hat{l}(\hat{\phi})) + o_p(1) - \mathcal{G}_{nq}^*(\tilde{\phi}_q,\tilde{l}_q(\tilde{\phi}_q)) \\ &\le \mathcal{G}_n^*(\hat{\phi}_q,\hat{l}(\hat{\phi}_q)) - \mathcal{G}_{nq}^*(\tilde{\phi}_q,\hat{l}(\tilde{\phi}_q)) + o_p(1).\end{aligned} \tag{A.30}$$

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.30) can be rewritten as

$$\begin{aligned} \mathcal{G}_n^*(\hat{\phi}_q,\hat{l}(\hat{\phi}_q)) &= \mathcal{G}_{nq}^*(\hat{\phi}_q,\hat{l}(\hat{\phi}_q)) + o_p(1)\left(1+\|\hat{\phi}_q-\phi_0\|^2+\|\hat{l}(\hat{\phi}_q)\|^2\right) \\ &= \mathcal{G}_{nq}^*(\hat{\phi}_q,\hat{l}(\hat{\phi}_q)) + o_p(1).\end{aligned} \tag{A.31}$$

The second equality in (A.31) is a consequence of Lemmas A.2 and A.6. According to Lemma A.7

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

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

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

by Lemma A.6. Since $\mathcal{G}_{nq}^*(\phi,l)$ is continuous in its arguments we can now express the second term on the right-hand side of (A.30) as

$$\mathcal{G}_{nq}^*(\tilde{\phi}_q,\hat{l}(\tilde{\phi}_q)) = \mathcal{G}_{nq}^*(\hat{\phi}_q,\hat{l}(\hat{\phi}_q)) + o_p\,(1) \tag{A.32}$$

Plugging (A.31) and (A.32) into (A.30) 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.4.2 Technical Lemmas

Lemma A.5 *Suppose Assumptions 1 to 3 are satisfied. Then, $\tilde{b}_q$ exists uniquely w.p.a. 1.*

Proof of Lemma A.5: The subsequent statements are true w.p.a. 1. Notice that $\bar{\mathcal{G}}^*_{nq}(\phi)$, defined in (18), is strictly convex function of ϕ because $R'_n = [-Q'_n, M']$ is a full rank matrix under Assumption 2(c) and J_n^{-1} is positive definite. 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 (17) we deduce that $\tilde{l}_q$ exists uniquely. ■

Lemma A.6 *Suppose Assumptions 1 to 3 are satisfied. Then*

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

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

Proof of Lemma A.6:

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 (18) 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 Assumptions 2 and 3. Moreover based on Assumption 1, 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 (17) we can deduce that $\tilde{l}_q(\tilde{\phi}_q) = O_p(1)$.

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

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

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

Lemma A.7 *Suppose that Assumptions 1 to 3 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* $\hat{l}\left(\bar{\phi}\right) = \sqrt{n}\hat{\lambda}\left(\bar{\theta}, \bar{\nu}\right)$*, 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.7: In view of Lemmas A.1(ii) and A.2, we deduce that $\hat{\lambda}(\bar{\theta}, \bar{v})$ is in the interior of $\hat{\Lambda}(\bar{\theta})$ w.p.a. 1. Hence, $\hat{\lambda}$ satisfies the first-order conditions associated with $\max_{\lambda\in\hat{\Lambda}(\bar{\theta})} G_n^*\left(\bar{\theta}, \bar{\nu}, \lambda\right)$:

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

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

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

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.20) and (A.22). Using the same arguments as in the proof of Lemma 2 and the definitions of J_n, Q_n, R_n, and Z_n we obtain the desired result. ■

A.5 Limit Distribution

Proof of Theorem 3: (i) 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.

(ii) According to Theorem 2(iv):

$$\mathcal{G}^*_{nq}(\hat{\phi}, \hat{l}(\hat{\phi})) = \mathcal{G}^*_{nq}(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) + o_p(1). \tag{A.33}$$

Since $\hat{\phi} = O_p(1)$ we can deduce from Lemma A.7 that

$$\hat{l}(\hat{\phi}) = \tilde{l}_q(\hat{\phi}) + o_p(1). \tag{A.34}$$

and

$$\mathcal{G}^*_{nq}(\hat{\phi}, \hat{l}(\hat{\phi})) = \mathcal{G}^*_{nq}(\hat{\phi}, \tilde{l}_q(\hat{\phi})) + o_p(1). \tag{A.35}$$

Recall that $\bar{\mathcal{G}}^*_{nq}(\phi) = \mathcal{G}^*_{nq}(\phi, \tilde{l}_q(\phi))$. Combining (A.33), (A.34), and (A.35) then yields

$$\bar{\mathcal{G}}^*_{nq}(\hat{\phi}) = \bar{\mathcal{G}}^*_{nq}(\tilde{\phi}_q) + o_p(1). \tag{A.36}$$

The required result $\hat{\phi} = \tilde{\phi}_q + o_p(1)$ follows from an argument similar to the one used in the proof of Theorem 3(i) in Andrews (1999). Using (A.34) once more we conclude that

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

which completes the proof. ■

A.6 MSE Derivations

Partition $\widetilde{\mathcal{P}} = [\widetilde{\mathcal{P}}'_s, \widetilde{\mathcal{P}}'_u]'$ and use the formula for the factorization of a joint normal pdf into a conditional and a marginal pdf to verify that

$$\begin{aligned} &(\phi - \tilde{\mathcal{P}})'\Upsilon^{-1}(\phi - \tilde{\mathcal{P}}) \\ &= [s - \tilde{\mathcal{P}}_s - \Upsilon_{su}(u - \tilde{\mathcal{P}}_u)]'(\Upsilon_{ss} - \Upsilon_{su}\Upsilon_{us})^{-1}[s - \tilde{\mathcal{P}}_s - \Upsilon_{su}(u - \tilde{\mathcal{P}}_u)] \\ &\quad +(u - \tilde{\mathcal{P}}_u)'(u - \tilde{\mathcal{P}}_u) \end{aligned}$$

Hence, we can write

$$\mathcal{S} = \widetilde{\mathcal{P}}_s I\{\widetilde{\mathcal{P}}_u \geq 0\} + (\widetilde{\mathcal{P}}_s - \Upsilon_{su}\widetilde{\mathcal{P}}_u)I\{\widetilde{\mathcal{P}}_u < 0\} = \Upsilon_{su}\widetilde{\mathcal{P}}_u I\{\widetilde{\mathcal{P}}_u \geq 0\} + \widetilde{\mathcal{P}}_{s.uu},$$

where

$$\widetilde{\mathcal{P}}_u \sim \mathcal{N}(u_0, 1) \quad \text{and} \quad \widetilde{\mathcal{P}}_{s.uu} = \widetilde{\mathcal{P}}_s - \Upsilon_{su}\widetilde{\mathcal{P}}_u \sim \mathcal{N}(-\Upsilon_{su}u_0, \Upsilon_{ss} - \Upsilon_{su}\Upsilon_{us}).$$

From the formulas for moments of a censored normal distribution (e.g. Greene (2003, p. 763) we obtain

$$\begin{aligned} \mathbb{E}[\widetilde{\mathcal{P}}_u I\{\widetilde{\mathcal{P}}_u \geq 0\}] &= u_0[1 - F_{\mathcal{N}}(-u_0)] + f_{\mathcal{N}}(-u_0) \\ V[\widetilde{\mathcal{P}}_u I\{\widetilde{\mathcal{P}}_u \geq 0\}] &= [1 - F_{\mathcal{N}}(-u_0)]\Bigg(1 - \frac{f^2_{\mathcal{N}}(-u_0)}{[1 - F_{\mathcal{N}}(-u_0)]^2} - \frac{u_0 f_{\mathcal{N}}(-u_0)}{1 - F_{\mathcal{N}}(-u_0)} \\ &\quad + \left(u_0 + \frac{f_{\mathcal{N}}(-u_0)}{1 - F_{\mathcal{N}}(-u_0)}\right)^2 F_{\mathcal{N}}(-u_0)\Bigg). \end{aligned}$$

We then use the facts that $\widetilde{\mathcal{P}}_u$ and $\widetilde{\mathcal{P}}_{s.uu}$ are independent, $f_{\mathcal{N}}(-u_0) = f_{\mathcal{N}}(u_0)$, and $1 - F_{\mathcal{N}}(-u_0) = F_{\mathcal{N}}(u_0)$ to compute the mean and variance of $\mathcal{S}$ reported in the text.

Proof of Theorem 4: Provided in main text. ■

Proof of Theorem 5: It can be verified by direct calculation that $(M\Omega M')^{-1} \leq \Xi_{(1)}$. Hence, it suffices to prove the following: if $Y = Y^* 1\{Y^* \geq 0\}$, where $Y^* \sim N(\mu, \sigma^2)$, then the MSE of Y is an increasing function of σ^2. Using the formulas for moments of a censored normal distribution once more we have

$$MSE\,[Y] = \sigma^2 F_{\mathcal{N}}(x)\left[x^2 + x\frac{f_{\mathcal{N}}(x)}{F_{\mathcal{N}}(x)} + 1\right],$$

where $x = \mu/\sigma$. First, for $\mu = 0$,

$$MSE\,[Y] = \sigma^2 F_{\mathcal{N}}\,(0)$$

is an increasing function of σ^2. Next, for $\mu > 0$

$$\begin{aligned}\frac{\partial MSE[Y]}{\partial \sigma^2} &= \mu^2 \left(f_{\mathcal{N}}(x) - \frac{1}{x^2} f_{\mathcal{N}}(x) + \frac{1}{x} f'_{\mathcal{N}}(x) + \frac{1}{x^2} f_{\mathcal{N}}(x) - \frac{2}{x^3} F_{\mathcal{N}}(x) \right) \left(-\frac{\mu^2}{(\sigma^2)^2} \right) \\ &= \frac{2\mu^4}{x^3(\sigma^2)^2} F_{\mathcal{N}}(x) > 0,\end{aligned}$$

since $x > 0$, which completes the proof. ■

A.7 Inference

Proof of Corollary 1: omitted. ■

Proof of Corollary 2: omitted. ■

Proof of Theorem 6: 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 2 and 3. 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

$$Z_u = (M\Omega M')^{-1} M\Omega Z. \tag{A.37}$$

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

$$\begin{aligned}\mathcal{S}^H &= Z_s - (QJ^{-1}Q')^{-1}QJ^{-1}M'Z_u \qquad \text{(A.38)} \\ \mathcal{S} &= Z_s - (QJ^{-1}Q')^{-1}QJ^{-1}M'(Z_u - \widetilde{\mathcal{U}}) \\ \widetilde{\mathcal{U}} &= \underset{\tilde{u} \geq -u_0}{\operatorname{argmin}}\, (\tilde{u} - Z_u)'\Lambda^{-1}(\tilde{u} - Z_u),\end{aligned}$$

where $\tilde{u} = u - u_0$, $\widetilde{\mathcal{U}} = \mathcal{U} - u_0$, and $\Lambda^{-1} = M\Omega M'$. 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 then becomes

$$2\big(\bar{\mathcal{G}}_q^*(\mathcal{P}^H) - \bar{\mathcal{G}}_q^*(\mathcal{P})\big) = Z_u'\Lambda^{-1}Z_u - (\widetilde{\mathcal{U}} - Z_u)'\Lambda^{-1}(\widetilde{U} - Z_u).$$

We deduce from Theorem 3

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

Proof of Corollary 3: omitted. ■

References

An, Sungbae, and Frank Schorfheide (2005): "Bayesian Analysis of DSGE Models," *Manuscript*, University of Pennsylvania.

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.

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.

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

Canova, Fabio, and Gianni 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.

Christiano, Lawrence , Martin Eichenbaum, and Charles Evans (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, **113**, 1-45.

Clarida, Richard, Jordi Galí, and Mark Gertler (2000): "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, **115**(1), 147-180.

Del Negro, Marco, and Frank Schorfheide (2005): "Monetary Policy Analysis with Potentially Misspecified Models," *Manuscript*, University of Pennsylvania.

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.

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.

Faust, Jon (1998): "The Robustness of Identified VAR Conclusions about Money". *Carnegie Rochester Conference Series*, **49**, 207-244.

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

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

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

Hansen, Lars P., John Heaton, and Amir Yaron (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics,* **14**, 262-280.

Imbens, Guido (1997): "One-Step Estimators for Over-Identified Generalized Method of Moments Models," *Review of Economics Studies*, **64**, 359-383.

Imbens, Guido, Richard Spady, and Phillip Johnson (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, **66**, 333-357.

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

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

Kitamura, Yuichi, Gautam Tripathi, and Hyungtaik Ahn (2004): "Empirical Likelihood-Based Inference in Conditional Moment Restriction Models," *Econometrica*, **72**, 1667-1714.

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

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

Lubik, Thomas A., and Frank Schorfheide (2006): "Do Central Banks Respond to Exchange Rate Fluctuation – A Structural Investigation," *Journal of Monetary Economics*, forthcoming.

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.

Mariano, Roberto S. (1982): "Analytical Small-Sample Distribution Theory in Econometrics: The Simultaneous Equations Case," *International Economic Review*, **23**, 503-534.

McFadden, Daniel L., and Whitney K. Newey (1994): "Large Sample Estimation and Hypothesis Testing," in Robert F. Engle and Daniel L. McFadden (eds.) *Handbook of Econometrics*, **4**, Elsevier, Amsterdam.

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

Rosen, Adam M. (2005): "Confidence Sets for Partially Identified Parameters that Satisfies a Finite Number of Moment Inequalities," *Manuscript*, Northwestern University.

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.

Shaikh, Azeem M. (2005): "Inference for Partially Identified Econometric Models," *Manuscript*, Stanford University.

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

Sims, Christopher A. (2002): "Solving Linear Rational Expectations Models," *Computational Economics*, **20**, 1-20.

Smets, Frank, and Raf Wouters (2003): "An Estimated Stochastic Dynamic General Equilibrium Model for the Euro Area," *Journal of the European Economic Association*, **1**, 1123-1175.

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: PARAMETERIZATION OF DGP

Name	$\mathcal{M}_1$-flex	$\mathcal{M}_2$-sticky	$\mathcal{M}_2$-flex
κ	0.60	0.01	0.60
β	.995	.995	.995
τ	2.00	2.00	2.00
ψ_1	1.50	1.50	1.50
ψ_2	0.50	0.50	0.50
ρ_R	0.70	0.70	0.70
ρ_G	0.95	0.85	0.85
ρ_Z	0.50	0.95	0.95
σ_R	0.10	0.20	0.20
σ_G	0.30	1.00	1.00
σ_Z	1.00	1.00	1.00
$\mathbb{E}[g_2(X_i, \theta_0)]$	0.03	0.16	0.11

Notes: The slackness of the inequality moment constraint, ν_0, is calculated as a function of the DSGE model parameters.

Table 2: $\mathcal{M}_1$ – Prices Are Nearly Flexible

Statistic	Parameter	Asymptotics $\hat{s}$	Asymptotics $\hat{s}_{(1)}$	Asymptotics $\hat{s}_{(12)}$	Small Sample $\hat{s}$	Small Sample $\hat{s}_{(1)}$	Small Sample $\hat{s}_{(12)}$
			Sample Size $n = 80$				
Median	ρ_R	-0.39	0.00	0.08	-0.23	-0.08	0.27
	$(1-\rho_R)\psi_1$	0.28	0.00	-0.05	0.26	-0.13	0.35
	$4(1-\rho_R)\psi_2$	0.35	0.00	-0.23	0.43	0.12	-0.16
Range	ρ_R	4.84	5.75	4.24	12.38	21.63	8.83
	$(1-\rho_R)\psi_1$	6.92	7.22	6.75	20.51	31.02	14.63
	$4(1-\rho_R)\psi_2$	6.56	11.57	1.41	12.39	30.52	2.16
Median(SE)	ρ_R	1.01	1.39	0.76	1.96	2.95	1.13
	$(1-\rho_R)\psi_1$	2.05	2.19	1.92	4.45	6.46	2.95
	$4(1-\rho_R)\psi_2$	0.47	5.62	0.11	0.54	5.90	0.15
			Sample Size $n = 160$				
Median	ρ_R	-0.37	0.00	0.11	-0.26	0.04	0.27
	$(1-\rho_R)\psi_1$	0.27	0.00	-0.07	0.34	0.13	0.24
	$4(1-\rho_R)\psi_2$	0.29	0.00	-0.32	0.50	0.24	-0.26
Range	ρ_R	4.86	5.75	4.24	10.59	15.47	6.71
	$(1-\rho_R)\psi_1$	6.93	7.22	6.75	15.13	19.65	10.96
	$4(1-\rho_R)\psi_2$	6.65	11.57	1.41	13.43	28.10	1.80
Median(SE)	ρ_R	1.02	1.39	0.76	1.85	2.75	1.08
	$(1-\rho_R)\psi_1$	2.05	2.19	1.92	4.04	5.17	2.80
	$4(1-\rho_R)\psi_2$	0.55	5.62	0.14	0.68	6.56	0.17

Notes: Range refers to the distance between the 5th and the 95th percentile. *Median(SE)* is the median of the squared estimation error $\hat{s}^2$. The entries in the columns labelled *Asymptotics* are calculated based on 1,000,000 draws from the limit distribution where $u_0 = \sqrt{n}\mathbb{E}[g_2(X_i, \theta_0)]$. The entries in the columns labelled *Small Sample* are calculated based 10,000 samples of size n, simulated from the DSGE model.

Table 3: $\mathcal{M}_2$ – Prices Are Sticky

Statistic	Parameter	Asymptotics $\hat{s}$	Asymptotics $\hat{s}_{(1)}$	Asymptotics $\hat{s}_{(12)}$	Small Sample $\hat{s}$	Small Sample $\hat{s}_{(1)}$	Small Sample $\hat{s}_{(12)}$
			Sample Size $n = 80$				
Median	ρ_R	0.00	0.00	0.38	-0.26	-0.26	0.04
	$(1-\rho_R)\psi_1$	0.00	0.00	-1.07	0.35	0.35	-0.71
	$4(1-\rho_R)\psi_2$	0.00	0.00	-2.80	0.47	0.47	-2.84
Range	ρ_R	2.10	2.10	2.05	2.48	2.48	2.66
	$(1-\rho_R)\psi_1$	3.26	3.26	2.99	4.15	4.15	4.37
	$4(1-\rho_R)\psi_2$	5.30	5.30	4.08	6.90	6.90	5.09
Median(SE)	ρ_R	0.19	0.19	0.25	0.25	0.25	0.27
	$(1-\rho_R)\psi_1$	0.45	0.45	1.20	0.63	0.63	1.01
	$4(1-\rho_R)\psi_2$	1.18	1.18	7.87	1.62	1.63	8.08
			Sample Size $n = 160$				
Median	ρ_R	0.00	0.00	0.54	-0.24	-0.24	0.19
	$(1-\rho_R)\psi_1$	0.00	0.00	-1.52	0.34	0.34	-1.14
	$4(1-\rho_R)\psi_2$	0.00	0.00	-3.97	0.42	0.42	-3.96
Range	ρ_R	2.10	2.10	2.05	2.35	2.35	2.56
	$(1-\rho_R)\psi_1$	3.26	3.26	2.99	3.79	3.79	4.04
	$4(1-\rho_R)\psi_2$	5.30	5.30	4.08	6.17	6.17	4.80
Median(SE)	ρ_R	0.19	0.19	0.35	0.23	0.23	0.29
	$(1-\rho_R)\psi_1$	0.45	0.45	2.31	0.59	0.59	1.57
	$4(1-\rho_R)\psi_2$	1.18	1.18	15.73	1.47	1.47	15.64

Notes: See Table 2.

Table 4: $\mathcal{M}_2$ – Prices Are Nearly Flexible

Statistic	Parameter	Asymptotics $\hat{s}$	Asymptotics $\hat{s}_{(1)}$	Asymptotics $\hat{s}_{(12)}$	Small Sample $\hat{s}$	Small Sample $\hat{s}_{(1)}$	Small Sample $\hat{s}_{(12)}$
			Sample Size $n = 80$				
Median	ρ_R	0.02	0.00	-0.18	-0.01	-0.05	-0.20
	$(1-\rho_R)\psi_1$	-0.02	0.00	0.21	0.01	0.05	0.22
	$4(1-\rho_R)\psi_2$	-0.02	0.00	0.09	0.02	0.03	0.10
Range	ρ_R	0.94	1.07	0.64	1.13	1.42	0.87
	$(1-\rho_R)\psi_1$	1.18	1.33	0.87	1.37	1.73	1.18
	$4(1-\rho_R)\psi_2$	3.88	3.89	3.87	5.25	5.52	5.58
Median(SE)	ρ_R	0.04	0.05	0.04	0.05	0.06	0.05
	$(1-\rho_R)\psi_1$	0.06	0.07	0.06	0.07	0.09	0.07
	$4(1-\rho_R)\psi_2$	0.63	0.64	0.63	0.95	1.03	1.06
			Sample Size $n = 160$				
Median	ρ_R	0.01	0.00	-0.26	-0.01	-0.02	-0.26
	$(1-\rho_R)\psi_1$	-0.01	0.00	0.30	0.03	0.05	0.33
	$4(1-\rho_R)\psi_2$	-0.01	0.00	0.13	0.04	0.06	0.17
Range	ρ_R	0.98	1.07	0.64	1.07	1.22	0.81
	$(1-\rho_R)\psi_1$	1.23	1.33	0.87	1.33	1.51	1.08
	$4(1-\rho_R)\psi_2$	3.89	3.89	3.87	4.70	4.78	5.05
Median(SE)	ρ_R	0.04	0.05	0.07	0.05	0.06	0.07
	$(1-\rho_R)\psi_1$	0.06	0.07	0.10	0.07	0.08	0.11
	$4(1-\rho_R)\psi_2$	0.64	0.64	0.64	0.85	0.87	0.95

Notes: See Table 2.

Table 5: $\mathcal{M}_2$ – Prices Are Nearly Flexible

(α_u, α_s)	Parameter	Asymptotics				Small Sample			
		$\mathcal{CS}^\theta$		$\mathcal{CS}^\theta_{(1)}$		$\mathcal{CS}^\theta$		$\mathcal{CS}^\theta_{(1)}$	
		Lgth	Cov	Lgth	Cov	Lgth	Cov	Lgth	Cov
Sample Size $n = 160$									
(0.05, 0.05)	ρ_R	1.17	(0.96)	1.07	(0.90)	1.33	(0.96)	1.24	(0.90)
	$(1-\rho_R)\psi_1$	1.48	(0.96)	1.33	(0.90)	1.69	(0.96)	1.49	(0.92)
	$4(1-\rho_R)\psi_2$	4.66	(0.95)	3.89	(0.90)	5.79	(0.94)	4.83	(0.90)
(0.00, 0.10)	ρ_R	0.99	(0.91)	1.07	(0.90)	1.12	(0.91)	1.24	(0.90)
	$(1-\rho_R)\psi_1$	1.24	(0.91)	1.33	(0.90)	1.37	(0.90)	1.50	(0.92)
	$4(1-\rho_R)\psi_2$	3.90	(0.90)	3.89	(0.90)	4.75	(0.90)	4.83	(0.90)

Notes: Lgth refers to the average length of the confidence interval (scaled by $\sqrt{n}$) across repetitions. *Cov* is the coverage probability. The target coverage probability of the intervals is 90%. *Asymptotics* are based on 1,000,000 draws from the limit distribution; *Small Sample* results are based 10,000 samples of size n, simulated from the DSGE model.

Figure 1: Critical Value Function $c^\theta_\alpha(u_0)$ for $\alpha = 0.10$