Estimation with Overidentifying Inequality Moment Conditions

Hyungsik Roger Moon

University of Southern California

Frank Schorfheide*

University of Pennsylvania, CEPR and NBER

August 27, 2008

*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 Ron Gallant (co-editor), two anonymous referees, Sungbae An, Donald Andrews, Jinyong Hahn, Shakeeb Khan, Yuichi Kitamura, Masao Ogaki, Nick Souleles, and seminar participants at the 2003 Econometrics Society Winter Meetings, the 2006 European Econometrics Society Meetings, Columbia University, the Federal Reserve Bank of Atlanta, London School of Economics, Michigan State University, New York University, Penn State University, Princeton University, UC Irvine, UC Riverside, UT Austin, Texas A&M, and the Universities of Chicago, Montreal, Pennsylvania, and Virginia for helpful comments. Moon gratefully acknowledges financial support from the USC Faculty Development Award. Schorfheide gratefully acknowledges financial support from the Alfred P. Sloan Foundation. Gauss Programs that implement the Monte Carlo simulation are available at http://www.ssc.upenn.edu/~schorf.
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 uniformly valid tests and 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: C13, E52

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,

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

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. 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. According to our reading of the literature, from 2002 to 2005 the American Economic Review, the Journal of Political Economy, and the Quarterly Journal of Economics published more than 60 empirical studies that are based on instrumental variable regressions. In almost all of the papers the authors explicitly stated their beliefs about the sign of the correlation between endogenous regressor and error term, yet none of the authors exploited the resulting inequality moment condition in their estimation.\(^1\)

The lead example in our paper 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. 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 \( E[g_2(X_i, \theta)] \geq 0 \) can provide in a model in which \( \theta_0 \) is in principle identifiable based on the equality moment condition \( E[g_1(X_i, \theta)] = 0 \) alone. If it is the case that

\(^1\)The assumption that a subset of the instrumental variables is potentially correlated with the error term in the regression equation and the sign of this correlation is assumed to be known is closely related to Manski and Pepper’s (2000) notion of monotone instrumental variables.
some elements of the vector $E[g_2(X_i, \theta_0)]$ are near zero, then the second set of moment conditions provides additional information. The inequality condition constrains the limit objective function of the estimator of $\theta_0$ and hence reduces its variability.

A key difficulty in models with inequality moment conditions is to establish the uniform validity of asymptotic approximations.\(^2\) To fix ideas, consider the location model $Y_i \sim \text{iid}\mathcal{N}(\theta, 1)$ with the inequality constraint $\theta \geq 0$. The maximum-likelihood estimator for $\theta$ based on a sample of $n$ observations is given by $\hat{\theta}_n = \max \{0, \bar{\theta}_n\}$, where $\bar{\theta}_n$ denotes the sample mean. For any sample size $n$ the distribution of this estimator is

$$
\sqrt{n}(\hat{\theta}_n - \theta_0) \sim \max \{-\sqrt{n}\theta_0, Z\}, \quad \text{where} \quad Z \sim \mathcal{N}(0, 1).
$$

(2)

Thus, the slackness in the inequality moment condition, here equal to the location parameter $\theta_0$, scaled by $\sqrt{n}$ appears as a nuisance parameter in the sampling distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. Suppose we would like to find a fixed critical value $c_\alpha$ that produces a uniformly valid test of size $\alpha$ in large samples, then we have to show that

$$
\lim_{n \to \infty} \sup_{\theta_0 \geq 0} \sup P_{\theta_0}\{n(\hat{\theta}_n - \theta_0)^2 \geq c_\alpha\} = \alpha.
$$

(3)

An important objective of our paper is to establish that our proposed tests and confidence sets are uniformly valid asymptotically in the sense of (3).

According to Andrews and Guggenberger (2007a) it suffices to study

$$
\lim_{n \to \infty} P_{\theta_0,n}\{n(\hat{\theta}_n - \theta_{0,n})^2 \geq c_\alpha\}
$$

for parameter sequences $\sqrt{n}\theta_{0,n} \to u_0 \in [0, \infty]$ – with the understanding that $u_0 = \infty$ if $\sqrt{n}\theta_{0,n}$ is diverging – to establish the uniform validity of, say a test procedure. In turn, this argument suggests that it is essential to study the properties of $\hat{\theta}_n$ under local-to-zero data generating processes with particular emphasis on parameter sequences $\theta_{0,n} = u_0/\sqrt{n}$, which in the simple example leads to $\sqrt{n}(\hat{\theta}_n - \theta_0) \sim \max \{-u_0, Z\}$ and provides an exact characterization of the finite-sample distribution (2). In the context of our moment inequality model (1) an important step for establishing the uniform asymptotic validity of our inference procedures is to consider sequences of data generating processes for which $E[g_2(X_i, \theta_0)] = u_0/\sqrt{n}$, $u_0 \geq 0$.\(^3\)

The local-to-zero parameterization of the slackness in the second set of inequality moment conditions suggests the following comparison of our analysis to the weak instrument

\(^2\)Andrews and Guggenberger (2007a) provide additional examples of models in which uniformly valid approximations are difficult to obtain.

\(^3\)In the working paper version of this paper, Moon and Schorfheide (2006, CEPR Working Paper 5605) we used the drifting data generating process $E[g_2(X_i, \theta_0)] = u_0/\sqrt{n}$ as a starting point without claiming and verifying that such an analysis essentially establishes the uniform validity of our asymptotic approximations. In the current version of this paper, we use results from a recent paper by Andrews and Guggenberger (2007a) that enable us to show uniform validity with only a slight modification of the analysis laid out in our working paper.
literature, surveyed for instance in Stock, Wright, and Yogo (2002). Consider the following simple IV regression model

\[ y_i = \theta_0 x_i + \epsilon_{1,i}, \quad x_i = \gamma_{0,n} z_i + \vartheta_{0,n} \epsilon_{1,i} + \epsilon_{2,i}, \]

where \( \epsilon_{1,i} \), \( \epsilon_{2,i} \), and \( z_i \) are mutually uncorrelated. The weak instrument literature focuses on \( \gamma_{0,n} = c/\sqrt{n} \) and \( \vartheta_{n,0} = \vartheta_0 \), that is, the regressor \( x_i \) has a non-zero correlation with the error term \( \epsilon_{1,i} \) and the instrument \( z_i \) is orthogonal to \( \epsilon_{1,i} \) but only weakly correlated with the regressor \( x_i \). In applying our moment-inequality setup (1) to the IV regression problem, we essentially conduct large sample analysis under the assumption \( \gamma_{0,n} = \gamma_0 \), that is, regressor and instrument are correlated and asymptotically identify \( \theta_0 \) by \( E[g_2(X_i, \theta_0)] = 0 \). In addition, we let \( \vartheta_{0,n} = u_0/\sqrt{n} \geq 0 \), which means that the endogeneity is positive and potentially small and the regressor \( x_i \) could be used to construct an inequality moment condition that can provide additional information on \( \theta_0 \).

A variety of approaches exist to exploit the moment conditions (1) for the estimation of \( \theta_0 \). 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 \( \theta_0 \) and \( E[g_2(X_i, \theta_0)] \). 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

\[^4\text{In practice our idea of constructing inequality moment conditions for IV regressions is probably most useful in situations in which the valid instruments provide only weak identification of the parameter of interest. However, it is beyond the scope of the current paper to analyze models of the form } \gamma_{0,n} = c/\sqrt{n} \text{ and } \vartheta_{0,n} = u_0/\sqrt{n}.\]
likelihood ratio test statistics to obtain tests and confidence sets for \( \theta_0 \) and \( E[g_2(X, \theta_0)] \).

We construct fixed critical value as well as Bonferroni procedures and show that they are uniformly valid asymptotically.\(^5\)

The concentrated limit objective function of the EL estimator has the same first-order asymptotic approximation as a GMM estimator that uses an optimal weight matrix and handles the presence of inequality moment conditions through additional slackness parameters. Hence, our large sample results, in particular the efficiency gain through the inequality moment conditions, also apply to conventional GMM estimators. Since we could rewrite the inequality moment condition as \( E[g_2(X, \theta_0) - \vartheta_0] = 0 \), 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 uniformly valid tests and confidence intervals.

The special case of \( E[g_2(X, \theta_0)] = 0 \) translates into \( \vartheta_0 = 0 \), which means that \( \vartheta_0 \) 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 \( \vartheta_0 \) (using our notation), whereas we are particularly interested in inference about \( \theta_0 \), treating the slackness in the inequality moment condition, \( \vartheta_0 \), 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, for example, Andrews, Berry, and Jia (2004), Chernozhukov, Hong, and Tamer (2002), and Pakes, Porter, Ho, and Ishii (2005), Rosen (2005), and Romano and Shaikh (2005a,b, 2005a,b), Andrews and Guggenberger (2007b), and Andrews and Soares (2007), 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

\(^5\)In this paper we do not consider re-sampling methods such as bootstrapping or sub-sampling as in Andrews and Guggenberger (2007b), Andrews and Soares (2007), Romano and Shaikh (2005a,b).
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 5 studies uniformly valid likelihood ratio tests and confidence sets for $\theta_0$ and $E[g_2(X_i, \theta_0)]$. 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 6. Moreover, we make a comparison with the asymptotic properties of simple procedures that ignore the information in the inequality moment condition. Section 7 concludes and the most important steps of the proofs are presented in the Appendix of this paper. Further technical details are provided in an unpublished Appendix that is available on the internet.

We use the following notation throughout the paper: \( \overset{P}{\rightarrow} \) and \( \overset{\Rightarrow}{\rightarrow} \) 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 (1993, 1999), Clarida, Galí, and Gertler (2000), and Woodford (2003)). Such rules provide a fairly accurate description of actual central bank behavior in recent decades and in many environments a good approximation of optimal monetary policy. To illustrate our inference methods we consider the following policy rule

\[
\tilde{R}_i = \rho_R \tilde{R}_{i-1} + (1 - \rho_R) \psi_1 \tilde{\pi}_i + (1 - \rho_R) \psi_2 \tilde{x}_i + \epsilon_{R,i},
\]

where \( \tilde{R}_i \) is the nominal interest rate in period \( i \), controlled by the central bank through open-market operations, \( \tilde{\pi}_i \) is the inflation rate, and \( \tilde{x}_i \) is a measure of real activity, such as output deviations from trend or output growth. The shock \( \epsilon_{R,i} \) 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 \( E[\tilde{x}_i(-\epsilon_{R,i})] \geq 0 \). This
implication generates a moment inequality condition that can be exploited to sharpen the inference about the policy rule coefficients.⁶

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

\[ \tilde{y}_i = \mathbb{E}_i[\tilde{y}_{i+1}] - \frac{1}{\tau}(\tilde{R}_i - \mathbb{E}_i[\tilde{\pi}_{i+1}]) + (1 - \rho_g)\tilde{g}_i + \frac{\rho_z}{\tau} \tilde{z}_i \]  
(5)

\[ \tilde{\pi}_i = \beta\mathbb{E}_i[\tilde{\pi}_{i+1}] + \kappa(\tilde{y}_i - \tilde{g}_i) \]  
(6)

\[ \tilde{g}_i = \rho_g \tilde{g}_{i-1} + \epsilon_{g,i} \]  
(7)

\[ \tilde{z}_i = \rho_z \tilde{z}_{i-1} + \epsilon_{z,i}. \]  
(8)

Equation (5) represents the intertemporal Euler equation of a representative household. The parameter \( \tau \) can be interpreted as household’s intertemporal substitution elasticity. \( \tilde{y}_i \) denotes output (measured in percentage deviations from a trend) and \( \tilde{g}_i \) represents an exogenous demand shock. The exogenous process \( \tilde{z}_i \) captures the stochastic growth of the level of total factor productivity 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 forward-looking Phillips curve (6). The parameter \( \beta \) is the household’s discount factor.

The specification of the model can be completed with a distributional assumption for the innovations \( \epsilon_{R,i}, \epsilon_{g,i}, \) and \( \epsilon_{z,i}. \) Researchers often assume that the innovations are independently and identically distributed Gaussian random variables with mean zero and standard deviations \( \sigma_R, \sigma_g, \) and \( \sigma_z, \) respectively.

In principle, one could estimate the entire model using likelihood-based techniques (see An and Schorfheide (2007) 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 \( \mathbb{E}[\tilde{x}_i(-\epsilon_{R,i})] \geq 0 \) remains valid.

We will revisit the prototypical New Keynesian model in Section 6 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 \( \kappa. \) If \( \kappa \) is large, then

---

⁶In 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).
prices in the model economy are fairly flexible. Hence, monetary policy shocks have only small real effects, \( E[x_i(-\epsilon_{R,i})] \) 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

As explained in the Introduction, in order to establish that our asymptotic approximations are uniformly valid we develop a large sample theory for converging sequences of parameters. We assume that the random vectors \( X_i, \ldots, X_n \) have the structure of a triangular array. Given a sample of size \( n \),

\[
E_n[g_1(X_i, \theta_{n,0})] = 0 \quad \text{and} \quad E_n[g_2(X_i, \theta_{n,0})] = \nu_{n,0} \geq 0, \tag{9}
\]

where \( \theta_{n,0} \rightarrow \theta_0 \) and \( \nu_{n,0} \rightarrow \nu_0 \). Let \( \Theta \) be the domain of the parameter vector \( \theta \). The functions \( g_1 \) and \( g_2 \) are of dimension \( h_1 \times 1 \) and \( h_2 \times 1 \), respectively. Let \( h = h_1 + h_2 \), \( g(X_i, \theta) = [g_1(X_i, \theta)', g_2(X_i, \theta)']' \), and \( M = [0_{h_2 \times h_1} \ I_{h_2}] \). When the moment function \( g(X_i, \theta) \) is differentiable with respect to \( \theta \), we use \( g_j^{(1)}(X_i, \theta) \) and \( g_j^{(2)}(X_i, \theta) \) to denote the first and the second order partial derivatives of \( g_j(X_i, \theta) \), the \( j \)’th element of the vector \( g(X_i, \theta) \), with respect to \( \theta \). Moreover, we collect the first-order derivatives in the matrix \( g^{(1)}(X_i, \theta) = [g_1^{(1)}(X_i, \theta), \ldots, g_h^{(1)}(X_i, \theta)] \).

Define

\[
Z_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [g(X_i, \theta_{n,0}) - M'\nu_{n,0}] \tag{10}
\]

\[
Q_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g^{(1)}(X_i, \theta)
\]

\[
J_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} g(X_i, \theta)g(X_i, \theta)'.
\]

\( K \) 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 on a probability space \( (\Omega_n, \mathcal{F}_n, P_n) \);

(b) \( \Theta \) is an \( m \)-dimensional compact subset of \( \mathbb{R}^m \), where \( m \leq h_1 \), \( \theta_{n,0} \rightarrow \theta_0 \), \( \theta_{n,0} \in \Theta \) \( \forall \) \( n \), and \( \theta_0 \in \Theta \);

(c) \( g(x, \theta) \) is continuous at each \( \theta \in \Theta \) with probability one;

(d) \( E_n[g_1(X_i, \theta_{n,0})] = 0 \), and \( \inf_n \|E_n[g_1(X_i, \theta)]\| > 0 \) for \( \theta \neq \theta_{n,0} \);

(e) \( \nu_{n,0} \rightarrow \nu_0 \) and \( \sqrt{n}\nu_{n,0} \rightarrow u_0 \in [0, \infty[^{h_2} \).
The true parameter \( \theta \) for any \( g \) function \( V \) is Lipschitz. This assumption is satisfied in the interest-rate-rule example in Section 2. We assume in 2(e) that the second derivative of \( g(X) \) diverges. The asymptotic slackness parameter \( u \) measures the slackness of the inequality conditions. For the subsequent large sample analysis the behavior of \( \sqrt{n}u \) is crucial. We assume that \( \sqrt{n}u \rightarrow u_0 \) for every element \( j \) with the understanding that \( u_0 = \infty \) whenever \( \sqrt{n}u \) diverges. The asymptotic slackness parameter \( u_0 \) will enter the limit distribution of estimators and test statistics. Assumption 1(j) implies that the moment function \( g(X, \theta) \) is Lipschitz. This assumption is satisfied in the interest-rate-rule example in Section 2. We make the following assumptions with respect to the first and the second derivatives of the function \( g(X, \theta) \).

**Assumption 2** (a) The true parameter \( \theta_0 \) exists in an interior of \( \Theta \);  
(b) \( g(X, \theta) \) is twice continuously differentiable with respect to \( \theta \);  
(c) the minimum eigenvalue of \( \left( E_n[g_1^{(1)}(X, \theta)](E_n[g_1^{(1)}(X, \theta)])' \right) \) is bounded below by a constant \( K > 0 \);  
(d) \( E_n[\sup_{\theta \in \Theta} \|g^{(1)}(X, \theta)\|^2] \leq K < \infty \), \( E_n[\sup_{\theta \in \Theta} \|g_j^{(2)}(X, \theta)\|] \leq K < \infty \) for \( j = 1, \ldots, h \);  
(e) for any \( \theta \) and \( \theta^* \), \( \|g_j^{(2)}(X, \theta) - g_j^{(2)}(X, \theta^*)\| \leq L_j(X) l_j(||\theta - \theta^*||) \), for some measurable function \( L_j \) of \( X \) such that \( \sup_n E_n(L_j(X)) < \infty \), and \( l_j(y) \downarrow 0 \) as \( y \downarrow 0 \).

Under Assumption 2(c) the expected value of the first derivative matrix \( E_n[g_1^{(1)}(X, \theta)] \) has full rank uniformly in \( n \). We assume in 2(e) that the second derivative of \( g(X, \theta) \), \( g_j^{(2)}(X, \theta) \), is Lipschitz. In fact, it is straightforward to verify that under Assumption 2 \( g(X, \theta) \) and \( g^{(1)}(X, \theta) \) are Lipschitz functions, too. Let
\[
Q(\theta) = \lim_{n \to \infty} E_n[g^{(1)}(X, \theta)], \quad J(\theta) = \lim_{n \to \infty} E_n[g(X, \theta)g(X, \theta)'].
\]
Moreover, we use \( Q_n, Q, J_n, \) and \( J \) to denote \( Q_n(\theta_{n,0}), Q(\theta_0), J_n(\theta_{n,0}), \) and \( J(\theta_0), \) respectively.
Assumption 3  (a) For each $\theta$, $Q_n(\theta) \overset{p}{\rightarrow} Q(\theta)$ and $J_n(\theta) \overset{p}{\rightarrow} J(\theta)$. (b) For each $\theta$, 
\[ \frac{1}{n} \sum_{i=1}^{n} g_j(X_i, \theta) \overset{p}{\rightarrow} \mathbb{E}[g_j(X_i, \theta)]. \]
(c) $Z_n \Rightarrow Z$, where $Z \sim N(0, J - M' \nu_0 \nu_0' M)$.

In 3(a) and (b) we assume pointwise convergence of $Q_n(\theta)$ and $J_n(\theta)$ in probability. However, when the conditions in Assumptions 1 and 2, especially, the Lipschitz restrictions, are satisfied, it is well known that the pointwise convergence in probability becomes uniform convergence in probability (e.g., see Andrews (1992)). We will use the uniform convergence in probability to derive the key asymptotic results of the paper. Assumption 3 (c) is satisfied, if, for instance, \{g(X, \theta_n, 0) - M' \nu_n, 0\} is a Martingale Difference Sequence with respect to the natural filtration and satisfies proper moment conditions, for example as in Assumption 2. For notational convenience, we subsequently omit the $n$ subscript of the expectation operator unless its dependence on the sample size is critical.

### 3.1 Empirical Likelihood Formulation

Among the various methods that could be used to estimate $\theta_0$ 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

\[ L_{EL}(\theta, p) = \left\{ p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i g_1(X_i, \theta) = 0, \sum_{i=1}^{n} p_i g_2(X_i, \theta) \geq 0 \right\}, \quad (11) \]

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

\[ \{\hat{\theta}_{n, EL}, \hat{p}_{n, EL}\} = \arg\max_{\theta \in \Theta, p} L_{EL}(\theta, p). \quad (12) \]

If one concentrates out the vector of probabilities $p$ subject to the moment restrictions, one can express the empirical likelihood estimator as the saddlepoint of the function

\[ G_n(\theta, \lambda_1, \lambda_2) = \frac{1}{n} \sum_{i=1}^{n} \ln \left( 1 + \lambda_1' g_1(X_i, \theta) + \lambda_2' g_2(X_i, \theta) \right). \quad (13) \]

We stack the Kuhn-Tucker parameters to obtain the vector $\lambda = [\lambda'_1, \lambda'_2]'$. In finite samples we need to restrict the domain of $\lambda$ to ensure that the argument of the logarithm in (13) is strictly positive. Thus, we define

\[ \hat{\Lambda}_n(\theta) = \{ \lambda \in \mathbb{R}^h \mid \lambda' g(X_i, \theta) \geq -1 + \kappa, i = 1, \ldots, n \}. \]
for some $\kappa_* > 0$.

Moreover, we use $G_n(\theta, \lambda)$ to abbreviate $G_n(\theta, \lambda_1, \lambda_2)$. Then

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} \max_{\lambda \in \Lambda_n(\theta), \lambda_2 \leq 0} G_n(\theta, \lambda).$$

(14)

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 $\lambda_2$. Newey and Smith (2004) study a broader class of estimators, called Generalized Empirical Likelihood (GEL) estimators, that are obtained by generalizing the objective function $G_n(\theta, \lambda)$. 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, in which the first-step maximization is asymptotically unconstrained. We will subsequently study the limit distribution of the saddlepoint

$$\{\hat{\theta}_n, \hat{\nu}_n\} = \arg \min_{\theta \in \Theta, \nu \geq 0} \max_{\lambda \in \Lambda_n(\theta)} G^*_n(\theta, \nu, \lambda)$$

(15)

$$\hat{\lambda}(\theta, \nu) = \arg \max_{\lambda \in \Lambda_n(\theta)} G^*_n(\theta, \nu, \lambda),$$

where

$$G^*_n(\theta, \nu, \lambda) = G_n(\theta, \lambda) - \nu'M\lambda.$$  

(16)

Here $\nu \geq 0$ plays the role of a Kuhn-Tucker parameter for the constraint that $\lambda_2 \leq 0$ and $\hat{\nu}$ will asymptotically capture the slackness in the inequality moment condition $E_n[g_2(X_i, \theta_{n,0})]$. We show in Lemma A.1 (Appendix A.1) that the saddlepoints of $G_n(\theta, \lambda)$ and $G^*_n(\theta, \nu, \lambda)$ are equivalent provided that $\hat{\nu}$ is in the interior of $\mathcal{V}$.

3.2 GMM Formulation

As pointed out in the Introduction, one can introduce an additional $h_2 \times 1$ parameter vector $\vartheta = E[g_2(X_i, \theta)]$ that captures the slackness in the inequalities and express the second moment condition as $E[g_2(X_i, \theta_{n,0}) - \vartheta_{n,0}] = 0$. The restriction $E[g_2(X_i, \theta_{n,0})] \geq 0$ can now be expressed as a constraint on the parameter $\vartheta$: $\vartheta_{n,0} \geq 0$. A GMM estimator can be obtained by solving the following minimization problem jointly for $\theta$ and $\vartheta$:

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

(17)

In the actual implementation of the empirical likelihood estimator in Section 6 we do not specify a value for $\kappa_*$. Instead we use Owen’s (2001) $\ln^*$ function, defined as $\ln^*(x) = \ln(x)$ if $x \geq \epsilon$ and $\ln^*(x) = \ln(\epsilon) - 1.5 + 2x/\epsilon - x^2/(2\epsilon^2)$ otherwise. $\epsilon$ is a small number, e.g. $1E-10$. 


where \{W_n\} is a sequence of positive-definite \( h \times h \) weight matrices. Notice that we impose the sign-restriction for the slackness parameter \( \vartheta \) directly onto the extremum problem that defines the GMM estimator. The subsequent large sample analysis will focus on the saddlepoint problem (15) but we will return to the GMM formulation in our discussion in Section 5.

4 Large Sample Analysis

The large sample analysis proceeds in two steps. First, we establish the consistency of the saddlepoint estimator \( \hat{\theta}_n \). Second we construct a quadratic approximation, denoted by \( G^*_n(\theta, \nu, \lambda) \) of the objective function \( G^*_n(\theta, \nu, \lambda) \) in the neighborhood of \( \theta = \theta_0, \nu = \nu_0, \) and \( \lambda = 0 \). We show that the saddlepoint estimators defined on \( G^*_n(\theta, \nu, \lambda) \) and \( G^*_n(\theta, \nu, \lambda) \) 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 \( \theta_0 \) 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 A.1 (Appendix A.1) as derivative of \( G^*_n(\theta, \lambda_1, \lambda_2) \) with respect to \( \lambda_2 \), and \( \nu_{n,0} = \mathbb{E}[g_2(X_i, \theta_{n,0})] \) 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 \( \hat{\theta}_n - \theta_{n,0} \xrightarrow{p} 0 \) and \( \hat{\nu}_n - \nu_{n,0} \xrightarrow{p} 0 \). Moreover, \( \hat{\lambda}(\hat{\theta}_n, \hat{\nu}_n) \xrightarrow{p} 0 \).

4.2 Limit Distributions

We proceed with a second-order Taylor approximation of the objective function \( G^*_n \). Let \( \beta = [\theta', \nu', \lambda']', \beta_{n,0} = [\theta_{n,0}', \nu_{n,0}', 0_{1 \times h}]', \) and abbreviate \( G^*_n(\theta, \nu, \lambda) \) as \( G^*_n(\beta) \). The objective function is expanded around \( \beta_{n,0} \) as follows:

\[
G^*_n(\beta) = G^*_n(\beta_{n,0}) + \frac{1}{n} \mathcal{R}_n(\beta),
\]

where

\[
G^*_n(\beta) = G^*_n(\beta_{n,0}) \quad \text{and} \quad G^*_n(\beta_{n,0}) = G^*_n(\beta_{n,0}) + \frac{1}{2}(\beta - \beta_{n,0})' G^*_n(\beta_{n,0}) (\beta - \beta_{n,0}).
\]

We prove in the unpublished Appendix that the remainder term \( \mathcal{R}_n(\beta) \) is negligible in the following sense:
Lemma 1 Suppose Assumptions 1 to 2 are satisfied, then for all $\gamma_n \to 0$
\[
\sup_{\beta \in B_n: \|\beta - \beta_{n,0}\| \leq \gamma_n} \frac{|R_n(\beta)|}{(1 + \|\sqrt{n}(\beta - \beta_{n,0})\|^2)} = o_p(1),
\]
where $R_n(\beta)$ is the remainder term in (18).

The key step in the subsequent analysis is to appropriately re-center and re-scale the vector $\beta$. First, we transform the vectors $\theta$ and $\lambda$ as follows:
\[
s = \sqrt{n}(\theta - \theta_{n,0}), \quad l = \sqrt{n}\lambda.
\]
Second, we transform the slackness or Kuhn-Tucker parameter $\nu$. We distinguish two cases for each element $\nu_j$ of $\nu$, depending on whether $u_{0,j} < \infty$ or not. Roughly speaking, the first case corresponds to an inequality condition in which the slackness is small enough to affect the sampling distribution of our estimator or test statistic, whereas in the second case the slackness is large and the inequality condition provides no information. Using indicator function notation, we let
\[
u_j = \sqrt{n}(\nu_j - \nu_{j,n,0}I\{u_{j,0} = \infty\}), \quad u_{j,n,0} = \sqrt{n}\nu_{j,n,0}I\{u_{j,0} < \infty\}.
\]
Thus, if $u_{j,0} = \infty$ our definition of $u_j$ re-centers the parameter $\nu_j$, otherwise it only re-scales it. We can write
\[
u_j = \sqrt{n}(\nu - \nu_{n,0}).
\]
The domains for $s$, $u_j$, and $\lambda$ are defined as $^8$
\[
S_n = \sqrt{n}(\Theta - \theta_{n,0})
\]
\[
U_{n,j}(u_{j,0}) = \sqrt{n}(\Psi_j - \nu_{j,n,0}I\{u_{j,0} = \infty\})
\]
\[
L_n(s) = \sqrt{n}\Lambda_n(\theta_{n,0} + s/\sqrt{n}).
\]
It is important to note, that under our transformation the domain of $u_j$ differs depending on whether the slackness is small ($u_{j,0} < \infty$) or large ($u_{j,0} = \infty$). If $u_{j,0} < \infty$ the domain of $u_j$ expands to $\mathbb{R}^+$. If $u_{j,0} = \infty$ the domain expands to $\mathbb{R}$. For notational convenience we will stack the parameters $s$ and $u$ into the vector $\phi = [s', u']$ and define
\[
\Phi_n(u_0) = S_n \otimes U_{1,n}(u_{1,0}) \otimes \cdots \otimes U_{h_2,n}(u_{h_2,0}) \text{ and } \phi_{n,0} = [0_1 \times m, \ u_{h_2,0}'].
\]
We denote the transformed empirical likelihood estimator as $\hat{\phi} = [\hat{s}', \hat{u}']$ and $\hat{l} = (\hat{s}, \hat{u})$.

The quadratic approximation $G_{nq}^*(\beta)$ of the objective function can be expressed in terms of the transformed parameters:
\[
G_{nq}^*(\phi, l) = \frac{1}{2}(l - J_n^{-1}[Z_n - R_n'(\phi - \phi_{n,0})])'J_n(l - J_n^{-1}[Z_n - R_n'(\phi - \phi_{n,0})]) + \frac{1}{2}(Z_n - R_n'(\phi - \phi_{n,0}))'J_n^{-1}(Z_n - R_n'(\phi - \phi_{n,0})),
\]

$^8$For instance, we use the notation $S_n = \sqrt{n}(\theta - \theta_{n,0})$ to denote the set $\{s \in \mathbb{R}^m | (\theta_{n,0} + s/\sqrt{n}) \in \Theta\}$. 

where \( R_n = [-Q'_n, M'] \) and the matrices \( Q_n, J_n, \) and \( Z_n \) were defined in (10). To approximate the sampling distribution of the empirical likelihood estimator we use the saddlepoint of (20) with respect to the \( \phi \)-domain

\[
\Phi(u_0) = \left\{ \phi = [s', u']' \in \mathbb{R}^m \otimes \mathbb{R}^h_2 \mid u_j \geq 0 \text{ if } u_{j,0} < \infty \right\}, \tag{21}
\]

where \( \Phi(u_0) \) can be interpreted as the limit version of \( \Phi_n(u_0) \). Thus, we let

\[
\tilde{l}_q(\phi) = \arg\max_{l \in \mathbb{R}^h} G_{nq}^*(\phi, l), \quad \tilde{\phi}_q = \arg\min_{\phi \in \Phi(u_0)} G_{nq}^*(\phi, \tilde{l}_q(\phi)).
\]

From (20) it follows immediately that \( G_{nq}^*(\phi, l) \) is maximized with respect to \( l \in \mathbb{R}^h \) by

\[
\tilde{l}_q(\phi) = J_n^{-1}(Z_n - R_n^*(\phi - \phi_{n,0})). \tag{22}
\]

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

\[
G_{nq}^*(\phi) = G_{nq}^*(\phi, \tilde{l}_q(\phi)) = \frac{1}{2}(Z_n - R_n^*(\phi - \phi_{n,0}))'J_n^{-1}(Z_n - R_n^*(\phi - \phi_{n,0})). \tag{23}
\]

The limit distribution of \( \tilde{\phi}_q \) can be determined from \( G_{nq}^*(\phi) \). We then use (22) to obtain the distribution of \( \tilde{l}_q(\tilde{\phi}_q) \). According to Assumption 3 \( Z_n \Rightarrow Z, J_n \overset{p}{\rightarrow} J, \) and \( R_n \overset{p}{\rightarrow} R, \)

where \( R = [-Q', M]' \). Let \( u_0^\infty \) be composed of the elements \( u_{j,0}^\infty = u_{j,0}I\{u_{j,0} < \infty\} \) and define \( \phi_0 = [0_{1 \times m}, u_0^\infty]' \). The results are summarized in the following theorem.

**Theorem 2** Suppose Assumptions 1 – 3 are satisfied. (i) Then

\[
(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) \Rightarrow (\mathcal{P}, \mathcal{L}), \quad \text{and} \quad G_{nq}^*(\tilde{\phi}_q, \tilde{l}_q(\tilde{\phi}_q)) \Rightarrow G_q^*(\mathcal{P}, \mathcal{L}),
\]

where

\[
\begin{align*}
\mathcal{P} &= \arg\min_{\phi \in \Phi(u_0)} \frac{1}{2}(Z - R'(\phi - \phi_0))'J^{-1}(Z - R'(\phi - \phi_0)), \\
\mathcal{L} &= J^{-1}(Z - R'(\mathcal{P} - \phi_0)), \\
G_q^*(\mathcal{P}, \mathcal{L}) &= \frac{1}{2}(Z - R'(\mathcal{P} - \phi_0))'J^{-1}(Z - R'(\mathcal{P} - \phi_0)).
\end{align*}
\]

and the domain \( \Phi(u_0) \) has been defined in (21).

(ii) \( \tilde{\phi} - \phi = o_p(1) \) and \( \tilde{l}_q(\tilde{\phi}_q) - \tilde{l}_q(\tilde{\phi}_q) = o_p(1) \).

**Remark 1:** GMM and MELE are first-order asymptotically equivalent. The limit distribution derived in Theorem 2 also applies to the GMM estimator defined in (17). Let \( s = \sqrt{n}(\theta - \theta_0), \ u = u_{n,0} + \sqrt{n}(\vartheta - \nu_{n,0}), \) and \( \phi = [s', u']' \). Using our definitions of \( Z_n, R_n, \) and \( J_n \) and assuming that \( W_n - J_n^{-1} \overset{p}{\rightarrow} 0 \) it follows from the arguments in Andrews (1999) that the objective function of the GMM estimator has a quadratic approximation that is equal to the concentrated objective function \( G_{nq}^*(\phi) \) of the empirical likelihood estimator in
Equation (23). Therefore, the analysis in the remainder of the paper applies not only to empirical likelihood estimators but also to conventional GMM estimators.

**Remark 2:** The sampling distribution of \( \hat{\phi} \) does not converge uniformly to the fixed parameter asymptotic distribution. For the limit in Theorem 2 we consider asymptotics under a sequence of parameters \((\theta_{n,0}, \nu_{n,0})\), while a conventional asymptotic analysis assumes fixed parameters \( \theta_0, \nu_0 \). If the parameters are fixed at \((\theta_0, \nu_0)\), one can verify that the limit distribution of \( \hat{\phi}, \mathcal{P}^\dagger \), solves

\[
\mathcal{P}^\dagger = \arg\min_{\phi \in \Phi(v_0)} \frac{1}{2} (Z - R'(\phi - \phi_0))^J^{-1}(Z - R'(\phi - \phi_0)),
\]

where

\[
\Phi(v_0) = \mathbb{R}^m \otimes U(\nu_{1,0}) \otimes \cdots \otimes U(\nu_{n,0}),
\]

\[
U(\nu_{n,0}) = \begin{cases} \mathbb{R}^+ & \text{if } \nu_{h,0} = 0 \\ \mathbb{R} & \text{if } \nu_{h,0} > 0. \end{cases}
\]

The limit \( \mathcal{P}^\dagger \) depends on the true parameter \( v_0 \) discontinuously. As a consequence, the sampling distribution of \( \hat{\phi} \) does not converge uniformly to the limit \( \mathcal{P}^\dagger \) in the sense that for every \( c \)

\[
\lim_{n \to \infty} \sup_{v_0 \in V} \left| P_{\theta_0, v_0} \{ \hat{\phi} \leq c \} - P\{ \mathcal{P}^\dagger \leq c \} \right| \neq 0.
\]

To verify this statement, choose a \( u_0 \) and a converging sequence \( \nu_{n,0} \to v_0 \) such that \( \sqrt{n}\nu_{n,0} \to u_0 \) with \( 0 < u_0 < \infty \). Construct the bound

\[
\sup_{v_0 \in V} \left| P_{\theta_0, v_0} \{ \hat{\phi} \leq c \} - P\{ \mathcal{P}^\dagger \leq c \} \right|
\geq \left| P_{\theta_0, \nu_{n,0}} \{ \hat{\phi} \leq c \} - P\{ \mathcal{P} \leq c \} - (P\{ \mathcal{P} \leq c \} - P\{ \mathcal{P}^\dagger \leq c \}) \right|
\to \left| P\{ \mathcal{P} \leq c \} - P\{ \mathcal{P}^\dagger \leq c \} \right| > 0.
\]

Here \( \mathcal{P} \) denotes the limit distribution of \( \hat{\phi} \) under the previously chosen sequence \( \nu_{n,0} \) and \( \theta_{n,0} = \theta_0 \). According to Theorem 2 \( |P_{\theta_0, \nu_{n,0}} \{ \hat{\phi} \leq c \} - P\{ \mathcal{P} \leq c \}| \) converges to zero and the last inequality follows from the definitions of \( \mathcal{P} \) and \( \mathcal{P}^\dagger \).

**Remark 3:** Information contained in inequality moment conditions. Consider the case \( h_2 = 1 \). The concentrated asymptotic objective function becomes

\[
\tilde{G}_q^*([s', u']) = \frac{1}{2} (Z + Q's - M'(u - u_0I\{u_0 < \infty\}))'J^{-1}(Z + Q's - M'(u - u_0I\{u_0 < \infty\})) \quad (24)
\]

and has to be minimized subject to \( u \in \mathbb{R}^+ \) if \( u_0 < \infty \) and \( u \in \mathbb{R} \) otherwise. If we re-parameterize the minimization problem in terms of \( \tilde{u} = u - u_0I\{u_0 < \infty\} \) we obtain

\[
\tilde{G}_q^*([s', u_0' + \tilde{u}']) = \frac{1}{2} (Z + Q's - M'\tilde{u})'J^{-1}(Z + Q's - M'\tilde{u}) \quad (25)
\]
where \( \tilde{u} \geq -u_0 \) with the understanding that the domain of \( \tilde{u} \) is \( \mathbb{R} \) if \( u_0 = \infty \). Thus, the further \( \mathbb{E}[g_2(X_i, \theta_0)] \) is apart from zero (in the local metric) the less likely it is that the constraint on \( \tilde{u} \) is binding and that it affects the limit distribution of estimators and empirical likelihood ratio statistics.

Suppose we partition \( Z, R \), and \( J \) conforming with the partitioning \( g(x, \theta) = [g'_1(x, \theta), g'_2(x, \theta)]' \)

\[
Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}, \quad R' = \begin{bmatrix} -Q'_1 & 0 \\ -Q'_2 & I \end{bmatrix}, \quad J = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}.
\]

Using the formulas for marginal and conditional means and variances of a multivariate normal distribution it is straightforward to verify that the concentrated objective function can be rewritten as

\[
(Z - R' (\phi - \phi_0))' J^{-1} (Z - R' (\phi - \phi_0)) = (Z_1 + Q'_1 s)' J_{11}^{-1} (Z_1 + Q'_1 s) + [Z_2 + Q'_2 s - \bar{u} - J_{21} J_{11}^{-1} (Z_1 + Q'_1 s)]' 
\times (J_{22} - J_{21} J_{11}^{-1} J_{12})^{-1} [Z_2 + Q'_2 s - \bar{u} - J_{21} J_{11}^{-1} (Z_1 + Q'_1 s)].
\]

In the absence of a constraint on \( \tilde{u} \) the function (26) is minimized with respect to \( s \) at

\[
S_{(1)} = -(Q_1 J_{11}^{-1} Q'_1)^{-1} Q_1 J_{11}^{-1} Z_1 \equiv \mathcal{N}(0, (Q_1 J_{11}^{-1} Q'_1)^{-1}).
\]

Thus, the limit distribution reduces to the well-known case in which estimation and inference is based only on \( \mathbb{E}[g_1(X_i, \theta_0)] = 0 \). The analysis can be generalized to \( h_2 > 1 \).

**Remark 4:** A mixture representation of \( \hat{\theta}_n \). The empirical likelihood estimator creates a mixture of an estimator that ignores the second set of moment inequalities and an estimator that imposes both \( \mathbb{E}[g_1(X_i, \theta_{n,0})] = 0 \) and \( \mathbb{E}[g_2(X_i, \theta_{n,0})] = 0 \). We will refer to these two estimators as \( \hat{\theta}_{(1)} \) and \( \hat{\theta}_{(12)} \), respectively. Define

\[
\mathcal{U} = u_0 + Z_2 - Q'_2 (Q_1 J_{11}^{-1} Q'_1)^{-1} Q_1 J_{11}^{-1} Z_1 - [J_{11}^{-1} - J_{11}^{-1} Q'_1 (Q_1 J_{11}^{-1} Q_1)^{-1} Q_1 J_{11}^{-1}] Z_1.
\]

Suppose that \( h_2 = 1 \), then it can be verified based on (26) that the limit distribution of the empirical likelihood estimator takes the form

\[
\mathcal{S} = \begin{cases} 
S_{(1)} & \text{if } \mathcal{U} \geq 0 \\
S_{(12)} & \text{otherwise}
\end{cases},
\]

where \( S_{(1)} \) is defined in (27) and represents the limit distribution of \( \hat{\theta}_{(1)} \) and

\[
S_{(12)} = -(Q_1 J_{11}^{-1} Q'_1)^{-1} Q_1 J_{11}^{-1} (Z + M' u_0).
\]

\( S_{12} \) characterizes the limit distribution of \( \hat{\theta}_{(12)} \) with the understanding that this estimator is inconsistent if \( u_0 = \infty \). The closer \( u_0 \) the more likely it is that \( S = S_{(12)} \).
There are of course other ways of creating estimators that randomly select between $S_{(1)}$ and $S_{(12)}$. The most plausible alternative is a pre-test estimator that works as follows: (i) use $E[g_1(X_i, \theta_{n,0})] = 0$ to obtain $\hat{\theta}_{(1)}$; (ii) estimate the slackness in the second set of inequality moment conditions by

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

and treat $\hat{\nu}_{(1)}$ as a test statistic. The limit distribution of $\sqrt{n}\hat{\nu}_{(1)}$ is of the form

$$T = \max \left\{ 0, u_0 + Z_2 - Q_2'(Q_1 J_1^{-1} Q_1')^{-1} Q_1 J_1^{-1} Z_1 \right\}$$

with the understanding that $\sqrt{n}\hat{\nu}_{(1)}$ is diverging if $u_0 = \infty$. One rejects the null hypothesis if $T$ exceeds a critical value. In this case one selects $\hat{\theta}_{(1)}$ and ignores the second moment condition. If the test is unable to reject then $E[g_2(X_i, \theta_{n,0})] = 0$ is dogmatically imposed and $\hat{\theta}_{(12)}$ is selected. The critical value of the pre-test controls the frequency with which either of the two estimators is selected. The larger the critical value, the less likely it is that $\hat{\theta}_{(12)}$ is selected. It turns out that the distributions of our estimator and the pre-test estimator are difficult to compare analytically. Some Monte Carlo evidence reported in an earlier draft of this paper suggested that our moment-inequality estimator is not dominated by the pre-test estimator.

### 4.3 Mean-Squared-Error Comparisons

To assess the efficiency of the proposed inequality-moment-conditions estimator we derive an analytic formula for the asymptotic mean-squared-errors (MSE) of the estimators $\hat{\theta}_n$ and $\hat{\nu}_n$ for the special case of $h_2 = 1$. The case $h_2 = 1$ covers many applications, in particular instrumental variable estimation problems with one endogenous regressor for which the inequality moment condition is generated from the covariance of the endogenous regressor and the error term. More specifically, we approximate finite-sample MSEs

$$E_n[n(\hat{\theta}_n - \theta_{n,0})(\hat{\theta}_n - \theta_{n,0})']$$

and

$$E_n[n(\hat{\nu}_n - \nu_{n,0})^2]$$

for parameter sequences $\{\theta_{n,0}, \nu_{n,0}\}$ and compare them to the MSEs of $\hat{\theta}_{(1)}$, $\hat{\theta}_{(12)}$, and $\hat{\nu}_{(1)}$ discussed in Remark 4 of the previous subsection. We focus on $u_0 < \infty$. For $u_0 = \infty$ the estimators $\hat{\theta}_n$ and $\hat{\theta}_{(1)}$ are asymptotically equivalent and strictly dominate $\hat{\theta}_{(12)}$.

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

$$\bar{G}_Q(\phi) = \frac{1}{2}(\phi - \tilde{P})'Y^{-1}(\phi - \tilde{P}) + \frac{1}{2}Z'(J^{-1} - J^{-1} R'(R J^{-1} R')^{-1} R J^{-1})Z,$$
Suppose Assumptions 1–3 are satisfied, is always preferable, in an asymptotic MSE sense, to the estimator ˆθ(1). (iii) There exists a additional information. If IE MSE θ

show in Appendix A.6 that constraints ( ˆθ(12)) that ignores this additional information. If E[g2(Xi, θ0)] = 0 then it is preferable to impose two inequality constraints ( ˆθ(12)), namely E[g2(Xi, θ0)] ≤ 0 and E[g2(Xi, θ0)] ≥ 0, instead of just one inequality constraint ( ˆθn). However, the performance of ˆθ(12) deteriorates as the slackness in the inequality constraint increases and will be inferior to our proposed estimator ˆn for large values of θ0.

According to Theorem 3 the estimator that exploits the inequality moment condition is always preferable, in an asymptotic MSE sense, to the estimator ˆθ(1) that ignores this additional information. If E[g2(Xi, θ0)] = 0 then it is preferable to impose two inequality constraints ( ˆθ(12)), namely E[g2(Xi, θ0)] ≤ 0 and E[g2(Xi, θ0)] ≥ 0, instead of just one inequality constraint ( ˆθn). However, the performance of ˆθ(12) deteriorates as the slackness in the inequality constraint increases and will be inferior to our proposed estimator ˆn for large values of θ0.

Next we consider the estimation of the slackness in the inequality moment condition, u. In the special of h2 = 1 with a weakly informative moment restriction, we can deduce from Theorem 2 that the limit distribution of ˆu is a censored normal distribution

\[ \mathcal{U} = \mathcal{P}_u I(\mathcal{P}_u \geq 0), \]
where
\[ \tilde{P}_u \sim N\left(u_0, (M[J^{-1} - J^{-1}Q'(QJ^{-1}Q')^{-1}QJ^{-1}][M']^{-1}) \right) \]

The benchmark estimator \( \hat{\nu}(1) \) has also a truncated normal limit distribution:
\[ \sqrt{n}\hat{\nu}(1) \Rightarrow U(1) \sim N\left(u_0, J_{22} + Q_2'(Q_1J_{11}Q_1')^{-1}Q_2 - 2Q_2'(Q_1J_{11}Q_1)Q_1J_{11}J_2 \right) \]

The following theorem states that the slackness estimator that uses the inequality moment condition is more precise than the estimator that ignores it.

**Theorem 4** Suppose Assumptions 1 - 3 are satisfied, \( h_2 = 1 \), and \( u_0 < \infty \), then \( \text{MSE}(U) \leq \text{MSE}(U(1)) \).

## 5 Inference

Based on the results obtained in Section 4, we now analyze the limit distribution of likelihood ratio tests for the hypotheses that \( \nu_0 = \nu^H \) and \( \theta_0 = \theta^H \). We invert the likelihood ratio statistics to obtain confidence intervals for \( \nu_0 \) and \( \theta_0 \), respectively. Since the analysis in the previous section was conducted under a drifting data generating process with \( \theta_{n,0} \rightarrow \theta_0 \) and \( \nu_{n,0} \rightarrow \nu_0 \), we are able to show that the confidence intervals are uniformly valid. Our exposition focuses on joint sets for all the elements of the vectors \( \nu \) and \( \theta \), respectively. An extension to subset inference is conceptually straightforward but cumbersome in terms of notation. The main text contains heuristic derivations. formal proofs are provided in the Appendix.

### 5.1 A Confidence Set for \( \nu \)

The empirical likelihood statistic for the null hypothesis \( \nu_{n,0} = \nu^H_n \) is defined as follows. Let
\[ \hat{\theta}_n^H = \arg\min_{\theta} \max_{\lambda \in \Lambda_n(\theta)} G_n^*(\theta, \nu^H_n, \lambda) \]
and define the test statistic (omitting the \( n \) subscript from parameters and their estimators)
\[ LR_n^\nu(\nu^H) = 2n \left(G_n^*(\hat{\theta}_n^H, \nu^H, \hat{\lambda}(\hat{\theta}_n^H, \nu^H)) - G_n^*(\hat{\theta}_n, \hat{\nu}, \hat{\lambda}(\hat{\theta}_n, \hat{\nu})) \right) \]  (30)

The first term represents the constrained extremum of the empirical likelihood function, whereas the second corresponds to the unconstrained extremum. Assuming that \( (\theta_{n,0}, \nu_{n,0}) \rightarrow (\theta_0, \nu_0) \) and \( \sqrt{n}\nu_{n,0} \rightarrow u_0 \), the test statistic converges in distribution to
\[ LR_n^\nu(\nu_{n,0}) \Rightarrow LR^\nu(u_0) \]
under the null hypothesis. A characterization of the limit distribution is provided in Theorem 5 below. This limit distribution is a function of the slackness parameter \( u_0 \) which was defined as the (possibly infinite) limit of the sequence \( \sqrt{n}\nu_{n,0} \).

Based on the limit distribution, we construct a critical value function \( c^\nu_\alpha(u_0) \) with the property

\[
P\{ LR_n^\nu(u_0) \geq c^\nu_\alpha(u_0) \} = \alpha. \tag{31} \]

For values of \( \alpha \) less than some cut-off value \( \bar{\alpha} \) the critical value function is uniquely defined and continuous in \( u_0 \). A confidence interval for \( \nu \) can be obtained as follows

\[
CS_n^\nu(\alpha) = \left\{ \nu \left| LR_n^\nu(\nu) \leq c^\nu_\alpha(\sqrt{n}\nu) \right. \right\}. \tag{32} \]

Using the convergence results stated in Theorem 2 and the continuity property of the critical value function, it can be shown that

\[
\lim_{n \to \infty} P_{\theta_0,\nu_0}\{ LR_n^\nu(\nu_{n,0}) \geq c^\nu_\alpha(\sqrt{n}\nu_{n,0}) \} = \alpha
\]

for arbitrary parameter sequences such that \( (\theta_{n,0},\nu_{n,0}) \to (\theta_0,\nu_0) \), and \( \sqrt{n}\nu_{n,0} \to u_0 \). A slight modification of the arguments presented in Andrews and Guggenberger (2007b) lets us deduce the uniform validity of the proposed confidence interval:

\[
\lim \inf_{n \to \infty} \inf_{\theta_0,\nu_0} P_{\theta_0,\nu_0}\{ \nu_0 \in CS_n^\nu(\alpha) \} = 1 - \alpha.
\]

The inf is taken over \( \theta_0 \in \Theta \) and \( \nu_0 \in \mathbb{V} \). To simplify the notation we omit the domain. The results are summarized in the following theorem.

**Theorem 5** Suppose Assumptions 1 – 3 are satisfied. (i) Let \( (\theta_{n,0},\nu_{n,0}) \to (\theta_0,\nu_0) \) and \( \sqrt{n}\nu_{n,0} \to u_0 \). Then

\[
LR_n^\nu(\nu_{n,0}) \Rightarrow LR^\nu(u_0) \equiv Z_u^n \Lambda^{-1} Z_u - (\bar{U} - Z_u)' \Lambda^{-1} (\bar{U} - Z_u),
\]

where

\[
\bar{U} = \arg\min_{\bar{u} \geq -u_0} (\bar{u} - Z_u)' \Lambda^{-1} (\bar{u} - Z_u),
\]

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

(ii) For \( \alpha < \bar{\alpha} \) the critical value \( c^\nu_\alpha(u_0) \) defined in (31) exists, it is unique, and continuous in \( u_0 \).

(iii) Size: \( \lim \sup_{n \to \infty} \sup_{\theta_0,\nu_0} P_{\theta_0,\nu_0}\{ LR_n^\nu(\nu_0) \geq c^\nu_\alpha(\sqrt{n}\nu_0) \} = \alpha. \)

(iv) Coverage Probability: \( \lim \inf_{n \to \infty} \inf_{\theta_0,\nu_0} P_{\theta_0,\nu_0}\{ \nu_0 \in CS_n^\nu(\alpha) \} = 1 - \alpha. \)

We show in the Appendix that for \( u_0 = 0 \) the limit distribution of the likelihood ratio statistic has a point mass at zero. \( LR^\nu(0) \) takes the value 0 if all elements of the random variable \( Z_u \) are less than zero. The probability that this event occurs is \( 1 - \bar{\alpha} \). To guarantee that \( c^\nu_\alpha(0) \) is well defined, we have to require that \( \alpha \) cannot be too large. In practice, this qualification is not of particular concern, since researchers typically choose \( \alpha \) to be 0.1, 0.05, or 0.01.
5.2 Confidence Sets for $\theta$

The empirical likelihood ratio statistic for the null hypothesis $\theta_{n,0} = \theta^H_n$ is defined as follows. Let

$$
\hat{\nu}^H_n = \arg\min_{\nu \in \mathcal{Y}} \max_{\lambda \in \Lambda_n(\theta^H)} G^*_n(\theta^H_n, \nu, \lambda).
$$

We express the test statistic in terms of the function $G^*_n(\theta, \nu, \lambda)$ (omitting $n$ subscripts from the estimators):

$$
LR_{\theta}^n(\theta^H_n) = 2n \left( G^*_n(\theta^H_n, \hat{\nu}^H_n, \hat{\lambda}(\theta^H_n, \hat{\nu}^H_n)) - G^*_n(\hat{\theta}, \hat{\nu}, \hat{\lambda}(\hat{\theta}, \hat{\nu})) \right).
$$

(33)

As before, the first term represents the constrained extremum of the empirical likelihood function and the second term corresponds to the unconstrained extremum. Assuming that $(\theta_{n,0}, \nu_{n,0}) \rightarrow (\theta_0, \nu_0)$ and $\sqrt{n}\nu_{n,0} \rightarrow u_0$ one can use Theorem 2 to show that

$$
LR_{\theta}^n(\theta_{n,0}) \Rightarrow LR^\theta(u_0).
$$

The distribution of the limit random variable $LR^\theta(u_0)$ is characterized in Theorem 6 below. Unfortunately, the limit distribution depends through $u_0$ on the slackness associated with the inequality moment conditions, which complicates the derivation of uniformly valid tests and confidence sets.

As in Section 5.1, we use the limit distribution to construct a critical value function $c^\theta_\alpha(u_0)$ that satisfies

$$
P\{LR^\theta(u_0) \geq c^\theta_\alpha(u_0)\} = \alpha.
$$

(34)

This function is uniquely defined and continuous in $u_0$. However, since the null hypothesis underlying the empirical likelihood ratio test is essentially a composite hypothesis of the form $\theta_{n,0} = \theta^H$ and $\nu_{n,0} \in \mathcal{V}$, the critical value function $c^\theta_\alpha(u_0)$ is not directly usable.

We consider two approaches of dealing with the composite hypothesis testing problem: a fixed critical value approach and a Bonferroni approach. We first explore the fixed critical value approach. Define

$$
c^\theta_\alpha(fix) = \sup_{u_0 \in [0, \infty]^k} c^\theta_\alpha(u_0).
$$

Since

$$
I\{LR^\theta_n(\theta_{n,0}) \geq c^\theta_\alpha(u_0)\} \geq I\{LR^\theta_n(\theta_{n,0}) \geq c^\theta_\alpha(fix)\}
$$

we can use a dominated convergence argument to deduce that

$$
\lim_{n \rightarrow \infty} P_{\theta_{n,0}, \nu_{n,0}}\{LR^\theta_n(\theta_{n,0}) \geq c^\theta_\alpha(fix)\} \leq \alpha.
$$

If we define

$$
CS^\theta_n(fix)(\alpha) = \left\{ \theta \mid LR^\theta_n(\theta) \leq c^\theta_\alpha(fix) \right\}
$$

(35)
and apply an argument similar to the ones used in Andrews and Guggenberger (2007a), we can establish that

$$\liminf_{n \to \infty} \inf_{\theta_0,\nu_0} P_{\theta_0,\nu_0} \{ \theta_0 \in CS^\theta_{n,fix}(\alpha) \} = 1 - \alpha.$$ 

However, depending on the shape of the asymptotic critical value function $c^\theta_0(u_0)$ the use of a sup critical value is potentially conservative.

Even though $u_0$ cannot be consistently estimated, it is possible to obtain non-trivial bounds from a confidence interval for $\nu_{n,0}$. Thus, as an alternative to the fixed critical value confidence set, we are considering a Bonferroni approach. Let $CS^\nu_{n,0}(\alpha,\nu)$ be an asymptotically valid $1 - \alpha,\nu$ confidence interval for $\nu_{n,0}$ and define the sup critical value over a restricted set of $u_0$ values:

$$c^\theta_{\alpha,0}(\sqrt{n}CS^\nu_{n,0}(\alpha,\nu)) = \sup_{u_0 \in \sqrt{n}CS^\nu_{n,0}(\alpha,\nu)} c^\theta_{\alpha,0}(u_0).$$

The Bonferroni confidence set is given by

$$CS^\theta_{n,Bf}(\alpha,\nu) = \{ \theta \mid LR^\theta_{n}(\theta_0) \leq c^\theta_{\alpha,0}(\sqrt{n}CS^\nu_{n,0}(\alpha,\nu)) \}. \quad (36)$$

We now construct the following bound for the probability that $\theta_0$ is not included in the confidence set:

$$\limsup_{n \to \infty} \sup_{\theta_0,\nu_0} P_{\theta_0,\nu_0} \{ LR^\theta_{n}(\theta_0) \geq c^\theta_{\alpha,0}(\sqrt{n}CS^\nu_{n,0}(\alpha,\nu)) \}$$

$$\leq \limsup_{n \to \infty} \sup_{\theta_0,\nu_0} P_{\theta_0,\nu_0} \{ LR^\theta_{n}(\theta_0) \geq c^\theta_{\alpha,0}(\sqrt{n}CS^\nu_{n,0}(\alpha,\nu)) \} \{ \nu_0 \in CS^\nu_{n,0}(\alpha,\nu) \}$$

$$+ \limsup_{n \to \infty} \sup_{\theta_0,\nu_0} P_{\theta_0,\nu_0} \{ \nu_0 \notin CS^\nu_{n,0}(\alpha,\nu) \}$$

$$\leq \limsup_{n \to \infty} \sup_{\theta_0,\nu_0} P_{\theta_0,\nu_0} \{ LR^\theta_{n}(\theta_0) \geq c^\theta_{\alpha,0}(\sqrt{n}\nu_0) \}$$

$$+ \limsup_{n \to \infty} \sup_{\theta_0,\nu_0} P_{\theta_0,\nu_0} \{ \nu_0 \notin CS^\nu_{n,0}(\alpha,\nu) \}$$

$$\leq \alpha + \alpha.$$ 

These inequalities establish that the likelihood ratio test and the confidence interval obtained from inverting it are uniformly valid in large samples. The results are summarized in the following theorem.

**Theorem 6** Suppose Assumptions 1 – 3 are satisfied. (i) Let $(\theta_{n,0},\nu_{n,0}) \longrightarrow (\theta_0,\nu_0)$ and $\sqrt{n}\nu_{n,0} \longrightarrow u_0$. Then,

$$LR^\theta_{n}(\theta_0) \longrightarrow LR^\theta(u_0) = \left( \min_{\phi \in \Phi_H(u_0)} 2\tilde{G}^*_{\phi}(\phi) \right) - \left( \min_{\phi \in \Phi(u_0)} 2\tilde{G}^*_{\phi}(\phi) \right),$$

where

$$\Phi_H(u_0) = \left\{ s = 0, u \in \mathbb{R}^{h_2} \mid u_j \geq 0 \text{ if } u_{j,0} < \infty \right\}.$$
(ii) The critical value \( c^\theta_\alpha(u_0) \) defined in (34) exists, it is unique, and continuous in \( u_0 \).

(iii) Fixed critical value approach:

\[
\liminf_{n \to \infty} \inf_{\theta_0, \nu_0} P_{\theta_0, \nu_0} \{ \theta_0 \in CS_{n, fix}^\theta(\alpha) \} = 1 - \alpha.
\]

(iv) Bonferroni approach:

\[
\liminf_{n \to \infty} \inf_{\theta_0, \nu_0} P_{\theta_0, \nu_0} \{ \theta_0 \in CS_{n, BF}^\theta(\alpha_\theta, \alpha_\nu) \} \geq 1 - \alpha_\theta - \alpha_\nu.
\]

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 \( \theta_1 \) by \( \theta_1^H \). Let (omitting the \( n \) subscript from estimators)

\[
(\hat{\theta}_2^H, \hat{\nu}^H) = \arg\min_{\theta_2, \nu \in V} \max_{\lambda \in \Lambda(\hat{\theta}_1^H, \theta_2)} G_n^*(\theta_1^H, \theta_2, \nu, \lambda)
\]

and redefine the test statistic as

\[
\mathcal{L}R_n^\theta(\theta_1^H) = 2n \left( G_n^*(\theta_1^H, \hat{\theta}_2^H, \hat{\nu}^H, \hat{\lambda}(\theta_1^H, \hat{\theta}_2^H, \hat{\nu}^H)) - G_n^*(\hat{\theta}, \hat{\nu}, \hat{\lambda}(\hat{\theta}, \hat{\nu})) \right). \tag{38}
\]

The subsequent steps remain unchanged.

### 5.3 Implementation

The asymptotic critical value functions \( c^\theta_\alpha(u_0) \) and \( c^\nu_\alpha(u_0) \) 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:

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

Approximate asymptotic critical value functions \( \hat{c}^\nu_\alpha(u_0) \) and \( \hat{c}^\theta_\alpha(u_0) \) and can be obtained by simulating \( \mathcal{L}R_n^\theta(u_0) \) (Theorem 5) and \( \mathcal{L}R_n^\nu(u_0) \) (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 \( \nu_0 \) and \( \theta_0 \) can be constructed according to Equations (32), (35), and (36).

### 6 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 \( E[g_2(X_i, \theta_0)] \geq 0 \) is not of immediate interest, we focus on the estimation of \( \theta_0 \).
6.1 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 (4) is $\tilde{x}_i = \tilde{y}_i - \tilde{y}_{i-1} + \tilde{z}_i$, log total factor productivity has a stochastic trend and is given by $\tilde{A}_i = \tilde{A}_{i-1} + \tilde{z}_i$, and $\tilde{y}_i$ measures percentage deviations of output from the level of productivity. $\mathcal{M}_1$ consists of Equations (4) to (8).

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 $\tilde{x}_i = \tilde{y}_i$. Moreover, we regard log productivity as trend stationary process $\tilde{A}_i = \gamma_i + \tilde{A}_i^* + \tilde{z}_i$, and define $\tilde{y}_i$ as percentage deviations from the deterministic trend induced by the $\gamma_i$ term in the law of motion of $\tilde{A}_i$. Euler equation and Phillips curve are modified as follows:

$$\tilde{y}_i = \mathbb{E}_i[\tilde{y}_{i+1}] - \frac{1}{\tau}(\tilde{R}_i - \mathbb{E}_i[\pi_{i+1}]) + (1 - \rho_g)\tilde{g}_i$$

$$\pi_i = \beta\mathbb{E}_i[\pi_{i+1}] + \kappa(\tilde{y}_i - \tilde{g}_i - \tilde{z}_i).$$

Hence, $\mathcal{M}_2$ consists of Equations (4), (7), (8), (40), and (41).

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 $\tilde{R}_i$, $\pi_i$, and $\tilde{y}_i$ 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 (4). 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

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

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

\[ y_i = \text{INT}_i, \quad x_i = [\text{INT}_{i-1}, \text{INFL}_i, \text{OUTPUT}_i], \]
\[ z_{1,i} = [\text{INT}_{i-1}, \text{INFL}_{i-1}, \text{YGR}_{i-1}], \quad z_{2,i} = -\text{OUTPUT}_i \]

Let \( X_i = [y_i, x'_i, z'_1, i, z'_2, i]' \), 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 \( g_j(X_i, \theta) = z_{j,i}(y_i - x'_i(\theta)) \), \( j = 1, 2 \) and obtain the desired moment conditions (1).

Both model specifications \( M_1 \) and \( M_2 \) have been used in empirical work with DSGE models. In some applications, e.g. Lubik and Schorfheide (2007), aggregate output is modelled as unit root process and output growth is included as argument in the monetary policy rule (\( M_1 \)), whereas in other applications, e.g., Smets and Wouters (2003) and Del Negro and Schorfheide (2007), 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 (\( M_2 \)).

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 \( M_1 \) using a value of \( \rho_z = 0.5 \), whereas under \( M_2 \) \( \rho_z = 0.95 \). As a consequence, the correlation between instruments and regressors is larger for \( M_2 \) than it is for \( M_1 \). 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 \( \kappa \) controls the slackness in the inequality moment condition. If there is a low degree of price stickiness in the economy the value of \( \kappa \) will be large, monetary policy shocks have little effect on output and \( E[g_2(X_i, \theta_0)] \) will be close to zero. Vice versa, if \( \kappa \) is small the slackness in the inequality moment condition tends to be large. \( v_0 = E[g_2(X_i, \theta_0)] \), \( J \) 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. The limit theory was developed for sequences of parameters. To compare the distribution of finite sample estimation errors to their limiting distribution, we let \( u_0 = \sqrt{n}v_0 \), where \( \sqrt{n} \) is either 80 or 160.

\footnote{In principle we could include \( \text{INFL}_i \) also in the definition of \( z_{2,i} \), which would introduce a second nuisance parameter.}

\footnote{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.}
6.2 Point Estimation Results

We consider the three estimators \( \hat{\theta}, \hat{\theta}(1), \) and \( \hat{\theta}(12). \) Since the estimation problem is linear and the number of inequality moment conditions is \( h_2 = 1 \) the computational problem simplifies considerably. Based on \( E[g_1(X_i, \theta_0)] \) the parameters are exactly identified and the empirical likelihood estimator \( \hat{\theta}(1) \) corresponds to the linear IV estimator. If we also impose that \( E[g_2(X_i, \theta_0)] \) then the model is overidentified. For the small sample analysis we use a numerical optimization procedure to find the saddle point

\[
\hat{\theta}(12) = \arg\min_{\theta \in \Theta} \max_{\lambda_1, \lambda_2} G_n(\theta, \lambda_1, \lambda_2)
\]

without imposing a sign restriction on \( \lambda_2. \) If \( \lambda_2(\hat{\theta}(12)) \leq 0 \) we deduce that the estimator \( \hat{\theta} \) that treats the second moment condition as inequality equals \( \hat{\theta}(12). \) Alternatively, if \( \lambda_2(\hat{\theta}(12)) > 0 \) then \( \hat{\theta} = \hat{\theta}(1). \) As discussed in Section 4, \( \hat{\theta} \) is a mixture of \( \hat{\theta}(1) \) and \( \hat{\theta}(12). \)

Tables 2 to 4 summarize the performance of the three point estimators. 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 \( u_0 = \sqrt{n}E[g_2(X_i, \theta_0)] \) 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. To characterize the performance of the estimators we consider the following three robust statistics\( ^{13} \): 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 \( M_1 \) 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 \( E[g_2(X_i, \theta_0)] = 0 \) 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%.

\(^{12}\) 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 \( G_\theta(\theta, \lambda_1, \lambda_2) \) by the \( \ln^* \) function described in Owen (2001). Samples for which the numerical optimization fails in an obvious manner are disregarded.

\(^{13}\) 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.
In the $M_1$ example the best estimator is the one that incorrectly imposes $\mathbb{E}[g_2(X_i, \theta_0)] = 0$. 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 $M_2$ 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 $\mathbb{E}[g_2(X_i, \theta_0)]$.

At last, we consider a version of $M_2$ 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.

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.\footnote{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.}

### 6.3 Interval Estimation Results

We consider three interval estimators for $\theta_0$: (i) the fixed critical value interval $\mathcal{CS}_{n, \text{fix}}^\alpha(\alpha)$ with $\alpha = 0.10$; (ii) a Bonferroni interval $\mathcal{CS}_{n, \text{BF}}^\alpha(\alpha_\theta, \alpha_\nu)$ with $\alpha_\theta = \alpha_\nu = 0.05$; and (iii) $\mathcal{CS}_{n,(1)}^\alpha$, which is the Wald confidence interval constructed from $\hat{\theta}_{(1)}$ and based on estimates of the asymptotic standard errors. Our confidence intervals are constructed for individual parameters, instead of jointly for the entire parameter vector. Notice that $\mathcal{CS}_{n, \text{fix}}^\alpha(\alpha)$ can be interpreted as a Bonferroni interval with $\alpha_\nu = 0$.

Table 5 presents coverage probabilities and average lengths for the three confidence intervals. Data are generated from the version of $M_2$ 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 translates into a reduction of the length of the confidence intervals that exploit the inequality moment condition.
The computations are implemented as follows. To simulate the asymptotic behavior of $\mathcal{CS}^\theta$ we begin by evaluating $Q$ and $J$ as a function of $\theta_0$. Second, we calculate the critical value functions $c_\alpha^\nu(u)$ and $c_\alpha^\theta(u)$ for $u$ on a grid $\mathbb{U}$ based on 100,000 draws from the limit distributions $\mathcal{LR}^\nu(u)$ and $\mathcal{LR}^\theta(u)$ of the likelihood ratio statistics. The fixed critical value for the confidence interval of $\theta$ is defined as $c_\alpha^\nu(fix) = \sup_{u \in \mathbb{U}} c_\alpha^\nu(u)$. Third, we draw 1,000,000 $Z$’s to obtain realizations of the limit random variables $\mathcal{LR}^\nu(u_0)$ and $\mathcal{LR}^\theta(u_0)$, where $u_0 = \sqrt{n}E[g_2(X_i, \theta_0)]$. Notice that $\mathcal{LR}^\nu(u_0)$ and $\mathcal{LR}^\theta(u_0)$ defined in Theorems 5 and 6 characterize the limit distribution of the likelihood ratio statistics under the null hypothesis. In order to simulate the asymptotic behavior of the confidence sets we also need the limit distributions under local alternatives, which we denote by $\mathcal{LR}^\nu(u)$ and $\mathcal{LR}^\theta(s, u_0)$. The asymptotic analogues to (32), (35), and (36) are given by

\[
\mathcal{CS}^\nu(\alpha) = \left\{ u/\sqrt{n} \left| \mathcal{LR}^\nu(u, u_0) \leq c_\alpha^\nu(u) \right. \right\},
\]

\[
\mathcal{CS}_{fix}^\theta(\alpha) = \left\{ \theta_0 + s/\sqrt{n} \left| \mathcal{LR}^\theta(s, u_0) \leq c_\alpha^\theta(fix) \right. \right\},
\]

\[
\mathcal{CS}_{Bf}(\theta_0, \alpha_v) = \left\{ \theta_0 + s/\sqrt{n} \left| \mathcal{LR}^\theta(s, u_0) \leq c_\alpha^\theta(\sqrt{n}CS^\nu(\alpha_v)) \right. \right\}.
\]

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 $c_\alpha^\nu(u)$ and $c_\alpha^\theta(u)$ conditional on $\hat{J}$ and $\hat{Q}$ by simulating the limit distributions of the likelihood ratio statistics. Once we have obtained the critical value functions we use (32), (35), and (36) to determine the confidence sets.

The critical value function $c_\alpha^\nu(u_0)$ for the three parameters of the monetary policy rule is plotted in Figure 1. As $u_0$ increases, the moment $E[g_2(X_i, \theta)]$ becomes irrelevant and the critical value converges to the critical value of a $\chi^2$ 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 $\sqrt{160} \cdot 0.11 = 1.39$ in our simulation 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 target coverage $1 - \alpha_\theta$. 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 $\chi^2$ critical value. Since the critical value function is convergent as $u_0$ increases, it is preferable to set $\alpha_v = 0$ and use the fixed critical value confidence interval. Nevertheless, we also report results for the Bonferroni interval with $\alpha_v = 0.05$. Notice that in practice, the researcher can calculate the (asymptotic) critical value function before constructing the confidence interval and decide what value of $\alpha_v$ to choose.

The simulation of the limit distribution implies that the coverage probabilities of $\mathcal{CS}_{Bf}(\theta_0, \alpha_v)$ for the output coefficient is essentially equal to $1 - \alpha_\theta$. For the other two coefficients it is
slightly larger: 0.96 (0.91) if $\alpha_\nu = 0.05$ ($\alpha_\nu = 0$). As expected from the inspection of the critical value function, if $\alpha_\nu = 0.05$ the $CS^\theta$ intervals are longer than the $CS^\theta_{(1)}$ intervals because the Bonferroni approach leads to conservative intervals under this configuration of the data generating process. If we set $\alpha_\nu = 0$ and use the fixed critical value interval, then the average length decreases and the intervals that take advantage of the inequality moment condition are shorter than the Wald confidence intervals $CS^\theta_{(1)}$, 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.

7 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 uniformly valid fixed critical value and Bonferroni hypothesis tests and confidence sets for the parameters of interest, $\nu$ and $\theta$.

Finally, throughout the paper we focused on models in which the parameter $\theta$ is identifiable based on the equality moment condition $E[g_1(X_i, \theta_0)] = 0$. 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

The following lemma states that the functions \( G_n \) (Equation 13) and \( G^*_n \) (Equation 16) have the same saddlepoint.

Lemma A.1 (i) If \( \hat{\theta}, \lambda \) and \( \hat{\nu} \in \text{int}(V) \) is a saddlepoint of \( G^*_n(\theta, \nu, \lambda) \) and solve (15), then \( \hat{\theta} \) and \( \lambda \) is also a saddlepoint of \( G_n(\theta, \lambda) \) that solves (14). (ii) If \( \hat{\theta} \) and \( \hat{\lambda} \) is a saddlepoint of \( G_n(\theta, \lambda) \), solving (14), and \( \hat{\nu} = [\hat{\nu}_1, \ldots, \hat{\nu}_{h_2}]' \in \text{int}(V) \), where

\[
\hat{\nu}_j = \begin{cases} 
\frac{\partial G_n(\theta, \lambda_1, \lambda_2)}{\partial \lambda_{2,j}} \bigg|_{\hat{\theta}, \lambda_1, \lambda_2} & \text{if } \hat{\lambda}_{2,j} = 0 \\
0 & \text{if } \hat{\lambda}_{2,j} < 0, \quad j = 1, \ldots, h_2.
\end{cases}
\]

then \( \hat{\theta}, \hat{\lambda}, \) and \( \hat{\nu} \) is a saddlepoint of \( G^*_n(\theta, \nu, \lambda) \), solving (15).

Proof of Lemma A.1: (i) Notice that all elements of \( \hat{\lambda}_2 \) are non-positive. Suppose, to the contrary, that \( \lambda_{2,j} > 0 \). Since

\[
G^*_n(\hat{\theta}, \nu, \hat{\lambda}) = G_n(\hat{\theta}, \hat{\lambda}) - \sum_{i \neq j} \nu_i \hat{\lambda}_{2,i} - \nu_j \hat{\lambda}_{2,j}
\]

the minimum with respect to \( \nu_j \) is attained by choosing \( \nu_j \) as large as possible such that \( \hat{\nu} \) is on the boundary of \( V \), which contradicts the assumption.

Moreover, \( \hat{\nu}' \hat{\lambda}_2 = 0 \). We previously ruled out that \( \hat{\lambda}_{2,j} > 0 \). If \( \hat{\lambda}_{2,j} < 0 \), then \( \hat{\nu}_j = 0 \) minimizes \( G^*_n \) with respect to \( \nu_j \). Finally, if \( \hat{\lambda}_{2,j} = 0 \) then \( \hat{\nu}_j \hat{\lambda}_{2,j} = 0 \) because the domain of \( \nu \) is bounded.

We begin by checking the saddlepoint property with respect to \( \theta \). Since \( \hat{\nu}' \hat{\lambda}_2 = 0 \) we deduce:

\[
G_n(\hat{\theta}, \hat{\lambda}) = G^*_n(\hat{\theta}, \hat{\nu}, \hat{\lambda}) \leq G^*_n(\hat{\theta}, \hat{\nu}, \hat{\lambda}) = G_n(\hat{\theta}, \hat{\lambda})
\]

for all \( \theta \in \Theta \). We proceed by verifying the saddlepoint property with respect to \( \lambda \). Without loss of generality, we partition \( \lambda_2 = [\lambda_{2,1}', \lambda_{2,2}']' \) and \( \nu = [\nu', \nu_2] \) where \( \hat{\nu}_1 = 0 \) and \( \hat{\nu}_2 > 0 \). Also, \( \lambda_2 \) is restricted to be non-positive in the \( G_n \) problem. As before,

\[
G_n(\hat{\theta}, \hat{\lambda}) = G^*_n(\hat{\theta}, \hat{\nu}, \hat{\lambda})
\]
Define $\alpha > 0$ to Assumption 1(i), the constant $n$ such that

$$G_n(\hat{\theta}, \nu, \lambda_1, \lambda_2) \geq G_n(\hat{\theta}, \nu, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2) = G_n(\hat{\theta}, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2) \geq G_n(\hat{\theta}, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2).

The second line follows from the fact that $\hat{\nu}_1 = 0$ and the third line holds because $\hat{\nu}_2 \lambda_2, \lambda_2, \lambda_2 \leq 0$.

(ii) Suppose that $\hat{\theta}$ and $\lambda$ is a saddlepoint of $G_n$ and

$$\hat{\nu}_j = \begin{cases} \frac{\partial G_n(\theta, \lambda_1, \lambda_2)}{\partial \lambda_2} & \text{if } \hat{\lambda}_2, \lambda_2 > 0 \\ 0 & \text{if } \hat{\lambda}_2, \lambda_2 < 0, \quad j = 1, \ldots, h_2. \end{cases}

Notice that $\hat{\nu}_2 \lambda_2 = 0$. We begin by checking the saddlepoint property of $G_n^*$ with respect to $\theta$ and $\nu \geq 0$:

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

since $\hat{\lambda}_2 \leq 0$. We proceed by checking the saddlepoint property with respect to $\lambda$, adopting the same partitioning of $\nu$ and $\lambda_2$ as in the proof of Part (i):

$$G_n^*(\hat{\theta}, \nu, \lambda) = G_n(\hat{\theta}, \lambda) \geq G_n(\hat{\theta}, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2) \geq G_n(\hat{\theta}, \lambda_1, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2, \lambda_2) = G_n^*(\hat{\theta}, \nu, \lambda).

In the second-to-last inequality we use the fact that $G_n$ is globally concave in $\lambda$. The last inequality holds because $\hat{\nu}_1 = 0$. □

A.2 Consistency

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

$$\lim_{n \to \infty} P \left\{ \hat{\theta}_n \in B(\theta_n, 0, \delta), \hat{\nu}_n \in B(\nu_n, 0, \delta) \right\} = 1,$

where

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

Define

$$\Theta_0^c = \Theta \cap B(\theta_n, 0, \delta)^c \quad \text{and} \quad N_0^c = V \cap B(\nu_n, 0, \delta)^c.

To simplify the notation we omit the subscript $n$ from the sets $\Theta_0^c$ and $N_0^c$. Recall that according to Assumption 1(i), the constant $\alpha > 2$ is such that $E[\sup_{\theta \in \Theta} \| g(X, \theta) \|^\alpha] < K$. We show in the unpublished Appendix that the following two statements are true: (i) For a given $\varepsilon, \delta > 0$ and $\zeta$ such that $\frac{\delta}{2} < \zeta < \frac{\delta}{2}$, there exist positive constants $\eta$ and $k$ and $\bar{n}$ such that for $n \geq \bar{n}$

$$P \left\{ \hat{G}_n^*(\theta_n, 0, \nu_n, 0) \geq n^{-\zeta - \kappa} \eta \right\} < \frac{\varepsilon}{2} \tag{A.1}$$
where
\[ G_n^*(\theta, \nu) = \max_{\lambda \in \Lambda_n(\theta, \nu)} G_n^*(\theta, \nu, \lambda), \]
and (ii)
\[ P \left\{ \min_{\theta \in \Theta_0^0, \nu \in N_0^0} G_n^*(\theta, \nu) \leq n^{-\zeta} \eta \right\} \leq \frac{\varepsilon}{2}. \] (A.2)

Then, from (A.1) and (A.2) we deduce that there exists an \( \eta > \bar{n} \)

\[ P \left\{ \theta_n \in B(\theta, \delta) \right\} \geq \frac{\varepsilon}{2} \]

\[ \geq P \left\{ G_n^*(\theta_n, \nu_n) < n^{-\zeta} \eta, \min_{\theta \in \Theta_0^0, \nu \in N_0^0} G_n^*(\theta, \nu) > n^{-\zeta} \eta \right\} \geq 1 - \varepsilon. \]

### A.3 Quadratic Approximation of the Objective Function

The coefficient matrices for the quadratic approximation of the objective function can be derived as follows:
\[ G_n^* (\beta) = G_n^* (\beta, \theta_n) + G_n^{(1)} (\beta, \nu_n) (\beta - \beta_n, \theta_n) + \frac{1}{2} (\beta - \beta_n, \theta_n)' G_n^{(2)} (\beta, \theta_n) (\beta - \beta_n, \theta_n). \] (A.3)

A direct calculation shows that
\[ G_n^{(1)} (\beta) = \begin{bmatrix} G_n^{(1)} (\beta)_{\theta} & G_n^{(1)} (\beta)_{\nu} & G_n^{(1)} (\beta)_{\lambda} \end{bmatrix}, \] (A.4)

where
\[ G_n^{(1)} (\beta)_{\theta} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{g^{(1)} (X_i, \theta) \lambda}{1 + \lambda g (X_i, \theta)} \right), \]
\[ G_n^{(1)} (\beta)_{\nu} = -M \lambda, \]
\[ G_n^{(1)} (\beta)_{\lambda} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{g (X_i, \theta)}{1 + \lambda g (X_i, \theta)} \right) - M \nu. \]

At \( \beta_n, \theta_n \) the first derivatives simplify to
\[ G_n^{(1)} (\beta_n, \theta_n) = [0, 0, n^{-1/2} Z_n^0]. \] (A.5)

Now partition the matrix of second derivative as follows
\[ G_n^{(2)} (\beta) = \begin{bmatrix} G_n^{(2)} (\beta)_{\theta \theta} & G_n^{(2)} (\beta)_{\theta \nu} & G_n^{(2)} (\beta)_{\theta \lambda} \\ G_n^{(2)} (\beta)_{\nu \theta} & G_n^{(2)} (\beta)_{\nu \nu} & G_n^{(2)} (\beta)_{\nu \lambda} \\ G_n^{(2)} (\beta)_{\lambda \theta} & G_n^{(2)} (\beta)_{\lambda \nu} & G_n^{(2)} (\beta)_{\lambda \lambda} \end{bmatrix}, \]
(A.6)

where
\[ G_n^{(2)} (\beta)_{\theta \theta} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\sum_{j=1}^{n} \lambda_j g^{(2)} (X_i, \theta) \lambda_j g^{(1)} (X_i, \theta)'}{1 + \lambda g (X_i, \theta)} - \frac{g^{(1)} (X_i, \theta) \lambda g^{(1)} (X_i, \theta)'}{1 + \lambda g (X_i, \theta)^2} \right), \]
\[ G_n^{(2)} (\beta)_{\theta \nu} = 0, \quad G_n^{(2)} (\beta)_{\theta \lambda} = 0, \quad G_n^{(2)} (\beta)_{\nu \theta} = -M, \]
\[ G_n^{(2)} (\beta)_{\lambda \theta} = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{g^{(1)} (X_i, \theta)'}{1 + \lambda g (X_i, \theta)} - \frac{g (X_i, \theta) \lambda g^{(1)} (X_i, \theta)'}{1 + \lambda g (X_i, \theta)^2} \right), \]
\[ G_n^{(2)} (\beta)_{\lambda \nu} = \frac{1}{n} \sum_{i=1}^{n} \frac{g (X_i, \theta) g (X_i, \theta)'}{1 + \lambda g (X_i, \theta)^2}. \]
At $\beta_{n,0}$ the second derivatives simplify to

$$G^{*2}_n(\beta_{n,0}) = \begin{bmatrix}
0 & 0 & Q_n \\
0 & 0 & -M \\
Q'_n & -M' & J_n
\end{bmatrix}. \quad (A.7)$$

### A.4 $\sqrt{n}$ Consistency

In the unpublished Appendix we prove the following theorem:

**Theorem A.1** Suppose Assumptions 1–3 are satisfied. Then, (i) $\sqrt{n}(\tilde{\beta}_{nq} - \beta_{n,0}) = O_p(1)$, (ii) $\sqrt{n}(\beta_n - \beta_{n,0}) = O_p(1)$, (iii) $nG^*_n(\beta_n) = nG^*_n(\tilde{\beta}_{nq}) + o_p(1)$, (iv) $nG^*_n(\tilde{\beta}_{nq}) = nG^*_n(\beta_{nq}) + o_p(1)$, and (v) $nG^*_n(\beta_{nq}) = nG^*_n(\beta_{nq}) + o_p(1)$.

Theorem A.1 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 $G^*_n(\beta_{nq})$ evaluated at $\tilde{\beta}_{nq}$ vanishes. Thus, the large-sample behavior of likelihood ratios can be approximated by the behavior of $G^*_n(\beta_{nq})$.

### A.5 Limit Distribution

**Proof of Theorem 2:** (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 $\tilde{I}_q$ 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 A.1(iv):

$$G^*_n(\phi, \tilde{I}_q(\phi)) = G^*_n(\tilde{\phi}_q, \tilde{I}_q(\tilde{\phi}_q)) + o_p(1). \quad (A.8)$$

Since $\tilde{\phi} = O_p(1)$, it can be shown that

$$\tilde{I}_q(\phi) = \tilde{I}_q(\tilde{\phi}_q) + o_p(1) \quad (A.9)$$

and

$$G^*_n(\phi, \tilde{I}_q(\phi)) = G^*_n(\phi, \tilde{I}_q(\tilde{\phi}_q)) + o_p(1). \quad (A.10)$$

Recall that $G^*_n(\phi) = G^*_n(\phi, \tilde{I}_q(\phi))$. Combining (A.8), (A.9), and (A.10) then yields

$$G^*_n(\phi) = G^*_n(\tilde{\phi}_q) + o_p(1). \quad (A.11)$$

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.9) once more we conclude that

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

which completes the proof. ■
A.6 MSE Derivations

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

$$
(\phi - \bar{\phi})' \Upsilon^{-1}(\phi - \bar{\phi}) = [s - \bar{P}_s - \Upsilon_{su}(u - \bar{P}_u)]' (\Upsilon_{ss} - \Upsilon_{su}\Upsilon_{uu})^{-1} [s - \bar{P}_s - \Upsilon_{su}(u - \bar{P}_u)] + (u - \bar{P}_u)' (u - \bar{P}_u).
$$

Hence, we can write

$$
S = \bar{P}_s I[\bar{P}_u \geq 0] + (\bar{P}_s - \Upsilon_{su}\bar{P}_u) I[\bar{P}_u < 0] = \Upsilon_{su}\bar{P}_u I[\bar{P}_u \geq 0] + \bar{P}_{s,uu},
$$

where

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

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

$$
E[\bar{P}_s I[\bar{P}_u \geq 0]] = u_0 [1 - F_{\mathcal{N}}(-u_0)] + f_{\mathcal{N}}(-u_0)
$$

$$
V[\bar{P}_s I[\bar{P}_u \geq 0]] = \left(1 - F_{\mathcal{N}}(-u_0)\right) \left(1 - \frac{f_{\mathcal{N}}^2(-u_0)}{1 - F_{\mathcal{N}}(-u_0)^2} \right) - \frac{u_0 f_{\mathcal{N}}(-u_0)}{1 - F_{\mathcal{N}}(-u_0)}
$$

$$
+ \left(u_0 + \frac{f_{\mathcal{N}}(-u_0)}{1 - F_{\mathcal{N}}(-u_0)}\right)^2 F_{\mathcal{N}}(-u_0).
$$

We then use the facts that $\bar{P}_u$ and $\bar{P}_{s,uu}$ 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 $S$ reported in the text.

Proof of Theorem 3: Provided in main text. ■

Proof of Theorem 4: It can be verified by direct calculation that $(M^\prime \Omega M)^{-1} \leq \Xi(1)$. Hence, it suffices to prove the following: if $Y = Y^*1\{Y^* \geq 0\}$, where $Y^* \sim \mathcal{N}(\mu, \sigma^2)$, then the MSE of $Y$ is an increasing function of $\sigma^2$. 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 $\sigma^2$. Next, for $\mu > 0$

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

$$
= \frac{2\mu^4}{x^3(\sigma^2)^2} F_{\mathcal{N}}(x) > 0,
$$

since $x > 0$, which completes the proof. ■
A.7 Inference

Proof of Theorem 5: (i) The asymptotics of $\hat{\theta}_n^H$ and $\lambda^H(\hat{\theta}_n^H, \nu_n^H)$ are well known (e.g., Newey and Smith (2004)) and follow from straightforward modifications of the proofs of Theorems A.1 and 2. Let $\mathcal{P}^H$ be the limit distribution of $[\hat{s}_n^H, u_0]$.

We begin by characterizing $\mathcal{P}$ and $\mathcal{P}^H$. The concentrated limit objective function is of the form

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

$$= \frac{1}{2}[(\phi - \phi_0) - (RJ^{-1}R')^{-1}RJ^{-1}Z]'RJ^{-1}R'[(\phi - \phi_0) - (RJ^{-1}R')^{-1}RJ^{-1}Z] + g(J, R, Z),$$

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

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

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

$$Z_u = (M\Omega M')^{-1}M\Omega Z.$$ (A.12)

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

$$\tilde{G}_q(s, u) = \frac{1}{2}[(s - Z_s) - (QJ^{-1}Q')^{-1}(QJ^{-1}M')(u - u_0 - Z_u)]' \times QJ^{-1}Q'[\Lambda^{-1}(s - Z_s) - (QJ^{-1}Q')^{-1}(QJ^{-1}M')(u - u_0 - Z_u)] + \frac{1}{2}(u - u_0 - Z_u)'M\Omega M'(u - u_0 - Z_u) + g(J, R, Z).$$

Suppose $\sqrt{n}\nu_{n,0} \longrightarrow u_0$, then we can deduce that

$$S = Z_s - (QJ^{-1}Q')^{-1}QJ^{-1}M'(Z_u - \tilde{U})$$

$$\tilde{U} = \arg\min_{\tilde{u}\in Z_u} (\tilde{u} - Z_u)'\Lambda^{-1}(\tilde{u} - Z_u),$$

A.13

where $\tilde{u} = u - u_0, \tilde{U} = U - u_0$, and $\Lambda^{-1} = M\Omega M'$. If we set $u = u_0$, then we obtain

$$S^H = Z_s - (QJ^{-1}Q')^{-1}QJ^{-1}M'Z_u.$$ Now let $\mathcal{P} = [S', u_0' + \tilde{U}]'$ and $\mathcal{P}^H = [S^H', u_0']'$. The limit distribution of the likelihood ratio statistic under the null hypothesis $\nu_{n,0} = \nu_{0}^H$ then becomes

$$2(\tilde{G}_q^*(\mathcal{P}^H) - \tilde{G}_q^*(\mathcal{P})) = Z_s'\Lambda^{-1}Z_u - (\tilde{U} - Z_u)'\Lambda^{-1}(\tilde{U} - Z_u).$$ (A.14)

(ii) The proof requires four steps. (ii)-(a) Existence of a critical value: If $u_0 \neq 0$ then the limit distribution of $\mathcal{L}R^\nu(u_0)$ is continuous. If $u_0 = 0$ then the limit distribution is a mixture of a discrete and continuous distribution with a pointmass of size $1 - \alpha$ at the origin. Thus, the cdf of the likelihood ratio statistic for $\mathcal{L}R^\nu(u_0) > 0$ is continuous and for every admissible $\alpha$ we can find a $c$ such that $P\{\mathcal{L}R^\nu(u_0) \geq c\} = \alpha$. This claim can be verified as follows.
Notice that for every possible realization of the random variable $Z_u$ a subset of the constraints on $\tilde{u}$ in (A.13) will be binding. Suppose we introduce a vector $w$, which is a re-ordered version of $\tilde{u}$, partitioned into $w_1$ and $w_2$, where $w_1$ corresponds to those elements of $\tilde{u}$ for which the constraint is binding. Formally, let $M_1$ and $M_2$ be selection matrices such that

$$\tilde{u} = M_1 w_1 + M_2 w_2.$$ 

Thus, for every realization of $Z_u$ we can find matrices $M_1$ and $M_2$ such that the likelihood ratio statistic is of the form

$$\mathcal{L}R^*(u_0) = -w_{1,0}' M_1' \lambda_{11} M_1 w_{1,0} + 2Z_{1,w}' M_1' \lambda_{11} M_1 w_{1,0} + 2Z_{2,w}' M_2' \lambda_{21} M_1 w_{1,0}$$

$$- \min \frac{w_{1,0}' M_1' \lambda_{12} M_2 \tilde{w}_2}{\tilde{w}_2} + \tilde{w}_2 M_2' \lambda_{22} M_2 \tilde{w}_2^2 - 2Z_{1,w}' M_1' \lambda_{12} \tilde{w}_2 - 2Z_{2,w}' M_2' \lambda_{22} M_2 \tilde{w}_2$$

$$= -w_{1,0}' M_1' \lambda_{11} M_1 w_{1,0} + 2Z_{1,w}' M_1' \lambda_{11} M_1 w_{1,0}$$

$$+ Z_{2,w}' M_2' \lambda_{22} M_2 Z_{2,w} + 2Z_{2,w}' M_2' \lambda_{21} M_1 Z_{1,w}$$

$$+ (Z_{1,w} - w_{1,0})' M_1' \lambda_{12} M_2 (M_2' \lambda_{22} M_2)^{-1} M_2' \lambda_{21} M_1 (Z_{1,w} - w_{1,0}).$$

Here $Z_{1,w}$ and $Z_{2,w}$ are partitions of the re-ordered vector $Z_u$ and $\lambda_{ij}$ refers to the conforming partitions of $\Lambda^{-1}$. For the distribution of the likelihood ratio statistic to have a point mass, there must exist a set of $Z$’s with non-zero probability for which $\mathcal{L}R^*(u_0)$ as a function of $Z$ is flat. Differentiating the above expression with respect to $Z_{1,w}$ and $Z_{2,w}$ yields

$$\frac{\partial \mathcal{L}R^*(u_0)}{\partial Z_{1,w}} = 2w_{1,0}' M_1' \lambda_{12} M_1 + 2Z_{2,w}' M_2' \lambda_{22} M_1 + 2(Z_{1,w} - w_{1,0}) M_1' \lambda_{12} M_2 (M_2' \lambda_{22} M_2)^{-1} M_2' \lambda_{21} M_1$$

$$\frac{\partial \mathcal{L}R^*(u_0)}{\partial Z_{2,w}} = 2Z_{2,w}' M_2' \lambda_{22} M_2 + 2Z_{2,w}' M_2' \lambda_{21} M_1.$$ 

Notice that this expression is only constant as a function of the $Z$’s if $M_2$ is empty (all constraints are binding) and $w_{1,0} = 0$ (which requires $u_0 = 0$). Thus, the point mass is located at $\mathcal{L}R^*(u_0) = 0$. Its mass, $1 - \bar{\alpha}$, equals the probability that all constraints in the minimization (A.13) are binding.

(ii)-(b) **Uniqueness of the critical value:** To show that the cdf of the likelihood ratio statistic is strictly increasing we have to rule out that there exist values for which the density is zero. An inspection of (A.14) suggests that the likelihood ratio statistic is a continuous, piecewise quadratic function of $Z_u$. The function is unbounded from above and is able to attain the value zero. Hence, the density of the likelihood ratio statistic is non-zero on $\mathbb{R}^+$.

(ii)-(c) Given a value $u_0 \geq 0$, for every sequence $u_k \rightarrow u_0$ and every $\eta > 0$

$$P(\{\mathcal{L}R^*(u_0) - \mathcal{L}R^*(u_k)\} > \eta) \rightarrow 0.$$ 

To verify this claim, assume that the inequality moment conditions are ordered such that $u_0 = [u_{1,0}', u_{2,0}']'$, where $u_{2,0} = \infty$. The notation $u_k \rightarrow u_0$ means that $u_{1,k} \rightarrow u_{1,0} < \infty$ and that there exists an increasing sequence $\nu_{2,k} \uparrow \infty$ such that $u_{2,k} \geq \nu_{2,k}$. Notice that

$$\min_{\tilde{u} \geq u_0} (\tilde{u} - Z_u)' \Lambda^{-1} (\tilde{u} - Z_u) = \min_{\tilde{u}_1 \geq u_{1,0}} (\tilde{u}_1 - Z_{1,u})' \Lambda_{11}^{-1} (\tilde{u}_1 - Z_{1,u}).$$

It can be verified that

$$\begin{bmatrix} \min_{\tilde{u}_1 \geq u_{1,0}} (\tilde{u}_1 - Z_{1,u})' \Lambda_{11}^{-1} (\tilde{u}_1 - Z_{1,u}) \\ \min_{\tilde{u}_2 \geq u_{2,0}} (\tilde{u}_2 - Z_u)' \Lambda^{-1} (\tilde{u}_2 - Z_u) \end{bmatrix} - \begin{bmatrix} \min_{\tilde{u}_1 \geq u_{1,k}} (\tilde{u}_1 - Z_{1,u})' \Lambda_{11}^{-1} (\tilde{u}_1 - Z_{1,u}) \\ \min_{\tilde{u}_2 \geq u_{2,k}} (\tilde{u}_2 - Z_u)' \Lambda^{-1} (\tilde{u}_2 - Z_u) \end{bmatrix} \rightarrow 0.$$
and
\[
\min_{k_1 \geq w_{1,k}} (\tilde{u}_1 - Z_{1,u}) \Lambda_{11}^{-1}(\tilde{u}_1 - Z_{1,u}) - \min_{k_1 \geq w_{1,0}} (\tilde{u}_1 - Z_{1,u}) \Lambda_{11}^{-1}(\tilde{u}_1 - Z_{1,u}) \longrightarrow 0
\]
as \(k \longrightarrow \infty\), which leads to the desired continuity results.

(ii)-(d) Continuity of the critical value function: for every sequence \(u_k \longrightarrow u_0\) we obtain that \(c_{u_k}(u_k) \longrightarrow c_{u_0}(u_0)\). This claim can be proved by contradiction. The convergence in distribution of the likelihood ratio statistic and the definition of the critical value function imply that
\[
P\{\mathcal{L}R(u_k) \geq c_{u_k}(u_0)\} \longrightarrow P\{\mathcal{L}R(u_0) \geq c_{u_0}(u_0)\} = \alpha.
\]
Suppose, contrary to our claim, that \(c_{u_k}(u_k) \not\rightarrow c_{u_0}(u_0)\). Then we can choose as subsequence \(k_m\) and an \(\epsilon > 0\) such that \(|c_{u_k}(u_k) - c_{u_0}(u_0)| \geq \epsilon\). Consider the following two cases: (a) If \(c_{u_k}(u_{km}) \geq c_{u_k}(u_0) + \epsilon\), then
\[
\alpha = P\{\mathcal{L}R(u_{km}) \geq c_{u_k}(u_{km})\} \leq P\{\mathcal{L}R(u_{km}) \geq c_{u_0}(u_0) + \epsilon\} \longrightarrow P\{\mathcal{L}R(u_0) \geq c_{u_0}(u_0) + \epsilon\} < \alpha,
\]
which is a contradiction. (b) If \(c_{u_k}(u_{km}) \leq c_{u_0}(u_0) - \epsilon\) the same argument applies with reversed inequalities.

(iii) This part follows from arguments used by Andrews and Guggenberger (2007a) in the proof of their Theorem 2. According to (i), for arbitrary sequences \((\theta_{n,0}, \nu_{n,0}) \longrightarrow (\theta_0, \nu_0)\) and such that \(\sqrt{n} \nu_{n,0} \longrightarrow u_0\):
\[
\mathcal{L}R_{\nu_{n,0}}^{\nu}(\nu_{n,0}) \Longrightarrow \mathcal{L}R_{\nu}^{\nu_0}(u_0).
\]
Thus, the cdf of \(\mathcal{L}R_{\nu_{n,0}}^{\nu}(\nu_{n,0})\) converges to the cdf of \(\mathcal{L}R_{\nu}^{\nu_0}(u_0)\). In the remainder of the proof we construct a lower and an upper bound for the type-I error of the empirical likelihood ratio test for \(\nu\). (iii)-(a) For any \(u_0\) there exist converging sequences \(\{\theta_{n,0}, \nu_{n,0}\} \longrightarrow \{\theta_0, \nu_0\}\) such that \(\sqrt{n} \nu_{n,0} \longrightarrow u_0\). Hence,
\[
\limsup_{n \to \infty} \sup_{\theta_{n,0}, \nu_{n,0}} P_{\theta_{n,0}, \nu_{n,0}} \{\mathcal{L}R_{\nu_{n,0}}^{\nu}(\nu_{n,0}) \geq c_\nu(\sqrt{n} \nu_{n,0})\} \geq \limsup_{n \to \infty} \sup_{\theta_{n,0}, \nu_{n,0}} P_{\theta_{n,0}, \nu_{n,0}} \{\mathcal{L}R_{\nu_{n,0}}^{\nu}(\nu_{n,0}) \geq c_\nu(\sqrt{n} \nu_{n,0})\} \geq P\{\mathcal{L}R_{\nu_0}^{\nu_0}(u_0) \geq c_{\nu_0}(u_0)\} = \alpha.
\]
In anticipation of the confidence set analysis we also take the sup over \(\nu_{n,0}\), although this is not required for the analysis of the asymptotic size of the likelihood ratio test. Here we used the continuity property of the critical value function established in (ii) and the convergence in (A.15).

(iii)-(b) Let \(\{\theta_{n,0}^{*}, \nu_{n,0}^{*}\}\) be a sequence such that
\[
\limsup_{n \to \infty} \sup_{\theta_{n,0}, \nu_{n,0}} P_{\theta_{n,0}^{*}, \nu_{n,0}^{*}} \{\mathcal{L}R_{\nu_{n,0}^{*}}^{\nu}(\nu_{n,0}^{*}) \geq c_\nu(\sqrt{n} \nu_{n,0}^{*})\} = \limsup_{n \to \infty} \sup_{\theta_{n,0}, \nu_{n,0}} P_{\theta_{n,0}^{*}, \nu_{n,0}^{*}} \{\mathcal{L}R_{\nu_{n,0}^{*}}^{\nu}(\nu_{n,0}^{*}) \geq c_\nu(\sqrt{n} \nu_{n,0}^{*})\}.
\]
Such a sequence always exists. We will now turn the \(\limsup\) into a \(\lim\) by choosing an appropriate subsequence. Let \(\{k_n\}\) be a subsequence of \(\{n\}\) such that
\[
\lim_{n \to \infty} \sup_{\theta_{k_n,0}^{*}, \nu_{k_n,0}^{*}} P_{\theta_{k_n,0}^{*}, \nu_{k_n,0}^{*}} \{\mathcal{L}R_{\nu_{k_n,0}^{*}}^{\nu}(\nu_{k_n,0}^{*}) \geq c_\nu(\sqrt{k_n} \nu_{k_n,0}^{*})\} = \limsup_{n \to \infty} \sup_{\theta_{k_n,0}^{*}, \nu_{k_n,0}^{*}} P_{\theta_{k_n,0}^{*}, \nu_{k_n,0}^{*}} \{\mathcal{L}R_{\nu_{k_n,0}^{*}}^{\nu}(\nu_{k_n,0}^{*}) \geq c_\nu(\sqrt{k_n} \nu_{k_n,0}^{*})\}.
\]
This subsequence always exists. We can choose a further subsequence \(\nu_{n}^{*}\) such that \(\sqrt{n} \nu_{n}^{*} \longrightarrow u^*\). Thus,
\[
\lim_{n \to \infty} \sup_{\theta_{n,0}^{*}, \nu_{n}^{*}} P_{\theta_{n,0}^{*}, \nu_{n}^{*}} \{\mathcal{L}R_{\nu_{n}^{*}}^{\nu}(\nu_{n}^{*}) \geq c_\nu(\sqrt{n} \nu_{n}^{*})\} = \lim_{n \to \infty} P_{\theta_{n,0}^{*}, \nu_{n}^{*}} \{\mathcal{L}R_{\nu_{n}^{*}}^{\nu}(\nu_{n}^{*}) \geq c_\nu(\sqrt{n} \nu_{n}^{*})\} = \alpha.
\]
(iv) Follows from (iii).

**Proof of Theorem 6:** (i) Follows from straightforward modifications of the proofs of Theorems A.1 and 2.

(ii) The key step is to write

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

Hence,

\[ \mathcal{LR}^\theta(u_0) = \min_{a \geq -u_0} \left( \frac{1}{2} (Z - M'\tilde{u})^J(Z - M'\tilde{u}) \right) \]

\[ - \min_{s \geq -u_0} \left( \frac{1}{2} (Z + Q's - M'\tilde{u})^J(Z + Q's - M'\tilde{u}) \right). \]

The first (second) term corresponds to the constrained (unconstrained) extremum of the likelihood function. For every realization of \( Z \) there exist selection matrices \( M(c) \) and \( M(u) \) such that the inequality constraint \( \tilde{u} \geq -u_0 \) in the terms (c) and (uc) is binding for the elements \( M(c) \tilde{u} \) and \( M(u) \tilde{u} \) of \( \tilde{u} \). Following the steps in the proof of Theorem 5(ii), it can then be verified that \( \mathcal{LR}^\theta(u_0) \) as a function of \( Z \) has the properties that guarantee the existence, uniqueness, and continuity of the critical value function.

(iii) This part follows from arguments used by Andrews and Guggenberger (2007a) in the proof of their Theorem 2. The remainder of the proof has two steps. We will construct an upper and a lower bound for the type-I error.

(iii)-(a) A lower bound for the type-I error can be constructed as follows. Let \( u^*_0 \) be the value of the slackness parameter that generates the least-favorable distribution and defines the fixed critical value \( C^\alpha(fix) \). Pick a converging sequence \( \{\theta_{n,0}, \nu_{n,0}\} \) such that \( \sqrt{n} \nu_{n,0} \rightarrow u^*_0 \). Then,

\[ \lim_{n \rightarrow \infty} \sup_{\theta_{n,0}, \nu_{n,0}} P_{\theta_{n,0}, \nu_{n,0}} \{ \mathcal{LR}_n^\theta(\theta_0) \geq c_\alpha(fix) \} = \lim_{n \rightarrow \infty} P_{\theta_{n,0}, \nu_{n,0}} \{ \mathcal{LR}_n^\theta(\theta_0) \geq c_\alpha(fix) \} = P\{ \mathcal{LR}^\theta(u^*_0) \geq c_\alpha(u^*_0) \} = \alpha. \]

(iii)-(b) An upper bound is obtained as follows: Let \( \{\theta^*_n, \nu^*_n\} \) be a sequence such that

\[ \lim_{n \rightarrow \infty} \sup_{\theta^*_n, \nu^*_n} P_{\theta^*_n, \nu^*_n} \{ \mathcal{LR}_n^\theta(\theta^*_n, \nu^*_n) \geq c_\alpha(fix) \} = \lim_{n \rightarrow \infty} \sup_{\theta^*_n, \nu^*_n} P_{\theta^*_n, \nu^*_n} \{ \mathcal{LR}_n^\theta(\theta^*_n, \nu^*_n) \geq c_\alpha(fix) \}. \]

Such a sequence always exists. We will now turn the \( \limsup \) into a \( \lim \) by choosing an appropriate subsequence. Let \( \{k_n\} \) be a subsequence of \( \{n\} \) such that

\[ \lim_{n \rightarrow \infty} P_{\theta^*_n, \nu^*_n} \{ \mathcal{LR}_{k_n}^\theta(\theta^*_n, \nu^*_n) \geq c_\alpha(fix) \} = \lim_{n \rightarrow \infty} P_{\theta^*_n, \nu^*_n} \{ \mathcal{LR}_n^\theta(\theta^*_n, \nu^*_n) \geq c_\alpha(fix) \}. \]

This subsequence always exists. We can choose another subsequence such that \( \sqrt{n} \nu^*_n \rightarrow u_0 \).

Thus,

\[ \lim_{n \rightarrow \infty} P_{\theta^*_n, \nu^*_n} \{ \mathcal{LR}_{k_n}^\theta(\theta^*_n, \nu^*_n) \geq c_\alpha(fix) \} = P\left\{ \mathcal{LR}^\theta(u_0) \geq c_\alpha(fix) \right\} \leq P\left\{ \mathcal{LR}^\theta(u_0) \geq c_\alpha(u_0) \right\} = \alpha. \]
(iv) According to (37) in the main text it remains to show that

$$\limsup_{n \to \infty} \sup_{\theta_0, \nu_0} P_{\theta_0, \nu_0} \{ \mathcal{LR}_n^\theta(\theta_0) \geq c_{\alpha_\theta}(\sqrt{\nu_0}) \} \leq \alpha_\theta.$$ 

The proof is similar to the proof of Theorem 5 (iii). $\blacksquare$
References


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


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


Table 1: Parameterization of DGP

<table>
<thead>
<tr>
<th>Name</th>
<th>$\mathcal{M}_1$-flex</th>
<th>$\mathcal{M}_2$-sticky</th>
<th>$\mathcal{M}_2$-flex</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>0.60</td>
<td>0.01</td>
<td>0.60</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.995</td>
<td>0.995</td>
<td>0.995</td>
</tr>
<tr>
<td>$\tau$</td>
<td>2.00</td>
<td>2.00</td>
<td>2.00</td>
</tr>
<tr>
<td>$\psi_1$</td>
<td>1.50</td>
<td>1.50</td>
<td>1.50</td>
</tr>
<tr>
<td>$\psi_2$</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>$\rho_R$</td>
<td>0.70</td>
<td>0.70</td>
<td>0.70</td>
</tr>
<tr>
<td>$\rho_g$</td>
<td>0.95</td>
<td>0.85</td>
<td>0.85</td>
</tr>
<tr>
<td>$\rho_z$</td>
<td>0.50</td>
<td>0.95</td>
<td>0.95</td>
</tr>
<tr>
<td>$\sigma_R$</td>
<td>0.10</td>
<td>0.20</td>
<td>0.20</td>
</tr>
<tr>
<td>$\sigma_g$</td>
<td>0.30</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>$\sigma_z$</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>

$\mathbb{E}[g_2(X_i, \theta_0)]$ | 0.03 | 0.16 | 0.11 |

Notes: Asymptotic approximations are based on $u_0 = \sqrt{n} \mathbb{E}[g_2(X_i, \theta_0)]$. 
<table>
<thead>
<tr>
<th>Statistic</th>
<th>Parameter</th>
<th>Asymptotics</th>
<th>Small Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{s}$</td>
<td>$\hat{s}_{(1)}$</td>
</tr>
<tr>
<td>Median</td>
<td>$\rho_R$</td>
<td>-0.39</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>0.28</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>0.35</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>$\rho_R$</td>
<td>4.84</td>
<td>5.75</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>6.92</td>
<td>7.22</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>6.56</td>
<td>11.57</td>
</tr>
<tr>
<td>Median(SE)</td>
<td>$\rho_R$</td>
<td>1.01</td>
<td>1.39</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>2.05</td>
<td>2.19</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>0.47</td>
<td>5.62</td>
</tr>
</tbody>
</table>

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 $u_0 = \sqrt{n}E[g_2(X_i, \theta_0)]$. The entries in the columns labelled Small Sample are calculated based on $10,000$ samples of size $n$, simulated from the DSGE model.
### Table 3: $M_2$ – Prices Are Sticky

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Parameter</th>
<th>Asymptotics</th>
<th>Small Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{s}$</td>
<td>$\hat{s}_{(1)}$</td>
</tr>
<tr>
<td>Sample Size $n = 80$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>$\rho_R$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>$\rho_R$</td>
<td>2.10</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>3.26</td>
<td>3.26</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>5.30</td>
<td>5.30</td>
</tr>
<tr>
<td>Median(SE)</td>
<td>$\rho_R$</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>1.18</td>
<td>1.18</td>
</tr>
<tr>
<td>Sample Size $n = 160$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>$\rho_R$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>$\rho_R$</td>
<td>2.10</td>
<td>2.10</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>3.26</td>
<td>3.26</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>5.30</td>
<td>5.30</td>
</tr>
<tr>
<td>Median(SE)</td>
<td>$\rho_R$</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>$(1 - \rho_R)\psi_1$</td>
<td>0.45</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>1.18</td>
<td>1.18</td>
</tr>
</tbody>
</table>

**Notes:** See Table 2.
Table 4: \( \mathcal{M}_2 \) – Prices Are Nearly Flexible

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Parameter</th>
<th>Asymptotics</th>
<th>Small Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{s} )</td>
<td>( \hat{s}_{(1)} )</td>
<td>( \hat{s}_{(12)} )</td>
</tr>
<tr>
<td>Median</td>
<td>( \rho_R )</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(1 - ( \rho_R ))( \psi_1 )</td>
<td>-0.02</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>4(1 - ( \rho_R ))( \psi_2 )</td>
<td>-0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>( \rho_R )</td>
<td>0.94</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>(1 - ( \rho_R ))( \psi_1 )</td>
<td>1.18</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>4(1 - ( \rho_R ))( \psi_2 )</td>
<td>3.88</td>
<td>3.89</td>
</tr>
<tr>
<td>Median(SE)</td>
<td>( \rho_R )</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(1 - ( \rho_R ))( \psi_1 )</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>4(1 - ( \rho_R ))( \psi_2 )</td>
<td>0.63</td>
<td>0.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Parameter</th>
<th>Asymptotics</th>
<th>Small Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \hat{s} )</td>
<td>( \hat{s}_{(1)} )</td>
<td>( \hat{s}_{(12)} )</td>
</tr>
<tr>
<td>Median</td>
<td>( \rho_R )</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>(1 - ( \rho_R ))( \psi_1 )</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td></td>
<td>4(1 - ( \rho_R ))( \psi_2 )</td>
<td>-0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>Range</td>
<td>( \rho_R )</td>
<td>0.98</td>
<td>1.07</td>
</tr>
<tr>
<td></td>
<td>(1 - ( \rho_R ))( \psi_1 )</td>
<td>1.23</td>
<td>1.33</td>
</tr>
<tr>
<td></td>
<td>4(1 - ( \rho_R ))( \psi_2 )</td>
<td>3.89</td>
<td>3.89</td>
</tr>
<tr>
<td>Median(SE)</td>
<td>( \rho_R )</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(1 - ( \rho_R ))( \psi_1 )</td>
<td>0.06</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>4(1 - ( \rho_R ))( \psi_2 )</td>
<td>0.64</td>
<td>0.64</td>
</tr>
</tbody>
</table>

**Notes:** See Table 2.
Table 5: $\mathcal{M}_2$ – Prices Are Nearly Flexible

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\mathcal{CS}_{Bf}^\theta(0.05 + 0.05)$</th>
<th>$\mathcal{CS}_{f1x}^\theta(0.10)$</th>
<th>$\mathcal{CS}^\theta(0.10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lgth</td>
<td>Cov</td>
<td>Lgth</td>
<td>Cov</td>
</tr>
<tr>
<td>$\rho_R$</td>
<td>1.17 (0.96)</td>
<td>0.99 (0.91)</td>
<td>1.07 (0.90)</td>
</tr>
<tr>
<td>$(1 - \rho_R)\psi_1$</td>
<td>1.48 (0.96)</td>
<td>1.24 (0.91)</td>
<td>1.33 (0.90)</td>
</tr>
<tr>
<td>$4(1 - \rho_R)\psi_2$</td>
<td>4.66 (0.95)</td>
<td>3.90 (0.90)</td>
<td>3.89 (0.90)</td>
</tr>
</tbody>
</table>

Asymptotics, $n = 160$

| $\rho_R$  | 1.33 (0.96)                             | 1.12 (0.91)                     | 1.24 (0.90)             |
| $(1 - \rho_R)\psi_1$ | 1.69 (0.96)                             | 1.37 (0.90)                     | 1.50 (0.92)             |
| $4(1 - \rho_R)\psi_2$ | 5.79 (0.94)                             | 4.75 (0.90)                     | 4.83 (0.90)             |

Small Sample, $n = 160$

emphNotes: \textit{Lgth} refers to the average length of the confidence interval (scaled by $\sqrt{n}$) across repetitions. \textit{Cov} is the coverage probability. The target coverage probability of the intervals is 90%. \textit{Asymptotics} are based on 1,000,000 draws from the limit distribution; \textit{Small Sample} results are based 10,000 samples of size $n$, simulated from the DSGE model.
Figure 1: Critical Value Function $c^0_\alpha(u_0)$ for $\alpha = 0.10$