

The Empirical Saddlepoint Approximation for GMM Estimators

Fallaw Sowell*

May 2007 (first draft June 2006)

Abstract

The empirical saddlepoint distribution provides an approximation to the sampling distributions for the GMM parameter estimates and the statistics that test the overidentifying restrictions. The empirical saddlepoint distribution permits asymmetry, non-normal tails, and multiple modes. If identification assumptions are satisfied, the empirical saddlepoint distribution converges to the familiar asymptotic normal distribution. In small sample Monte Carlo simulations, the empirical saddlepoint performs as well as, and often better than, the bootstrap.

The formulas necessary to transform the GMM moment conditions to the estimation equations needed for the saddlepoint approximation are provided. Unlike the absolute errors associated with the asymptotic normal distributions and the bootstrap, the empirical saddlepoint has a relative error. The relative error leads to a more accurate approximation, particularly in the tails.

KEYWORDS: Generalized method of moments estimator, test of overidentifying restrictions, sampling distribution, empirical saddlepoint approximation, asymptotic distribution.

*Tepper School of Business, Carnegie Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213. Phone: (412)-268-3769. fs0v@andrew.cmu.edu

Helpful comments and suggestions were provided by Don Andrews, Roni Israelov, Steven Lugauer and seminar participants at Carnegie Mellon University, University of Pittsburgh, Concordia, Yale, University of Pennsylvania and the 2006 Summer Meetings of the Econometric Society in Minneapolis. This research was facilitated through an allocation of advanced computing resources by the Pittsburgh Supercomputing Center through the support of the National Science Foundation.

Contents

1	INTRODUCTION	3
2	MODEL AND FIRST ORDER ASYMPTOTIC ASSUMPTIONS	6
3	MOMENT CONDITIONS TO ESTIMATION EQUATIONS	8
3.1	NOTES	10
4	THE SADDLEPOINT APPROXIMATION	13
4.1	DENSITY RESTRICTIONS	14
4.2	ADDITIONAL ASSUMPTIONS	15
4.3	ASYMPTOTIC NORMAL AND EMPIRICAL SADDLEPOINT APPROX- IMATIONS	23
5	BEHAVIOR IN FINITE SAMPLES	26
6	ON THE SPECTRAL DECOMPOSITION	33
7	FINAL ISSUES	34
7.1	IMPLEMENTATION AND NUMERICAL ISSUES	34
7.2	EXTENSIONS	35
8	APPENDIX	36
8.1	THEOREM 1: ASYMPTOTIC DISTRIBUTION	36
8.2	THEOREM 2: INVARIANCE OF ASYMPTOTIC DISTRIBUTION	39
8.3	THEOREM 3: THE SADDLEPOINT APPROXIMATION	41
8.4	THEOREM 4: EMPIRICAL SADDLEPOINT APPROXIMATION	44
8.5	THEOREM 5: EMPIRICAL SADDLEPOINT DENSITY STRUCTURE	47
8.6	THEOREM 6: INVARIANCE OF SADDLEPOINT APPROXIMATION	50
8.7	THEOREM 7: DERIVATIVE OF THE DECOMPOSITION	55
8.8	A DIFFERENTIAL GEOMETRY PERSPECTIVE	57
8.9	STEPS TO A SADDLEPOINT APPROXIMATION	58
9	REFERENCES	60

1 INTRODUCTION

The empirical saddlepoint density provides an approximation to the sampling distribution for parameters estimated with the generalized method of moments and the statistics that test overidentifying restrictions. Traditional GMM (Hansen (1982)) relies on a central limit theorem. The saddlepoint approach uses a generalization of the central limit theorem and provides a more accurate approximation.

There are three key differences between the results using the commonly first order normal approximation and the saddlepoint approximation. First, the parameters and the statistics that test the overidentifying restrictions are no longer forced to be independent. The asymptotic normal approximation uses a linear approximation that can be orthogonally decomposed into the identifying space and the overidentifying space, see Sowell (1996). The asymptotic normal approximation inherits the orthogonality from the tangent space; hence, $\hat{\theta}$ and the J statistic are independent. In contrast, the saddlepoint uses a different approximation at each parameter value. Each approximation generates an orthogonally decomposed tangent space. However, the union of these approximations is not required to give an orthogonal decomposition over the entire space. This lack of independence is a common feature in approximations with higher orders of accuracy. In Newey and Smith (2004) and Rilstone, Srivastava and Ullah (1996), the higher order bias for the parameter estimates involve the parameters for the statistics that test the overidentifying restrictions and vice versa. The lack of independence means the marginal densities are no longer equal to the conditional densities. As simulations in section 5 demonstrate, it is informative to report both the marginal density and the conditional densities (i.e. the sampling distribution for the parameters conditional on the overidentifying restrictions being satisfied).

The second difference is the sampling density does not have to be unimodal. If the GMM objective function has a unique local minimum, then the saddlepoint approximation will be unimodal. However, if the GMM objective function has multiple local minima, then the saddlepoint approximation can have multiple modes and confidence regions can be disjoint. The first order normal approximation using the location and convexity of the global minimum will provide misleading inference concerning the parameter values associated with the other local minima. However, the saddlepoint approximation can be interpreted as the density for the location of local minima of the GMM objective function and will more accurately summarize the available information. These issues are not new to Econometrics. The s -sets of Stock and Wright

(2000) can also lead to disjoint confidence intervals.

The third difference is that the saddlepoint approximation has a relative error instead of the absolute error that occurs with the first order normal approximation. When the underlying distribution is unknown and the empirical distribution is used, the empirical saddlepoint approximation results in a relative error of order¹ $N^{-1/2}$. The relative error results in an improvement over the absolute error in the tails of the distribution. This is important in the calculation of p -values. If $f(\alpha)$ denotes the density being approximated, the asymptotic approximation can be written

$$\hat{f}_A(\alpha) = f(\alpha) + \mathcal{O}(N^{-1/2}).$$

The empirical saddlepoint approximation can be written

$$\hat{f}_S(\alpha) = f(\alpha) \{1 + \mathcal{O}(N^{-1/2})\}.$$

When there is a relative error, the error gets small with the density in the tails of the distribution. When the error is absolute, the same error holds in the tails as at the modes of the density. Hall and Horowitz (1996) shows that the bootstrap approximation results in an absolute error of order N^{-1} . Hence, theoretically, neither the empirical saddlepoint approximation nor the bootstrap approximation dominate.

The first contribution of this paper is a new procedure to use Economic moment conditions (m moment conditions to estimate k parameters) to create a set of m estimation equations in m unknowns. The m equations permit the calculation of the empirical saddlepoint distribution. The second contribution is the extension of the saddlepoint approximation to include statistics to test the validity of the overidentifying restrictions. Previous work only focused on the parameter estimates and testing hypotheses concerning parameters. The third contribution is a saddlepoint approximation that allows for models and samples where the GMM objective function has multiple local minima. Multiple local minima are common for GMM objective functions, see Dominguez and Lobato (2004). Previous saddlepoint distribution research restricted attention to a unique local minimum. The final contribution is the interpretation of the saddlepoint density as the joint density for the parameter estimates and

¹When the underlying distribution is known, the relative error from the saddlepoint approximation is typically of order N^{-1} . In some special cases, the relative error can be of order $N^{-3/2}$. These dramatic improvements cannot be realized with empirical applications in Economics when the underlying distribution is unknown.

the statistics that test the validity of the overidentifying restrictions. In finite samples these will not be independent. The saddlepoint distribution gives a well defined density that accounts for this dependence.

The saddlepoint approximation's history in the Statistics literature starts with Escher (1932) and Daniels (1954). General introductions are available in Reid (1988), Goutis and Casella (1999) and Huzurbazar (1999) and the books Field and Ronchetti (1990), Jensen (1995) and Kolassa (1997). The saddlepoint distribution was first used to approximate the sampling distribution of the sample mean when the underlying distribution was known, e.g. Phillips (1978). It was then generalized to maximum likelihood estimators. Field (1982) gives the saddlepoint approximation to the density for parameter estimates defined as the solution to a system of equations where the underlying distribution is known. Ronchetti and Welsch (1994) presents the empirical saddlepoint distribution where the known distribution is replaced with the empirical distribution. In Economic applications the underlying distribution is typically unknown so the empirical saddlepoint distribution is appropriate. If there exists a unique local minimum, this empirical saddlepoint approximation is an alternative approximation to the GMM sampling density. Field (1982), Ronchetti and Welsch (1994) and Almudevar, Field and Robinson (2000) are the natural bridges to the results in this paper.

The closest paper to the current research is Ronchetti and Trojani (2003), denoted RT. RT shows how to go from the GMM moment conditions to estimation equations used in a saddlepoint approximation. RT also extends the parameter testing results from Robinson, Ronchetti and Young (2003) to situations where the moment conditions are overidentified.

This paper takes a different approach to the saddlepoint approximation that results in a system with only m equations. While the procedure presented in RT uses $(\frac{m}{2} + k)(m + 1)$ equations. In this paper, all the parameters are tilted instead of only a subset. The resulting saddlepoint density approximates the joint density of the parameter estimates and statistics that test the validity of the overidentifying restrictions.

The next section reviews basic GMM notation and the assumptions necessary to obtain first order asymptotic results. Section 3 is the derivation of the estimation equations from the GMM moment conditions. Section 4 presents the saddlepoint approximation including extensions needed to apply the saddlepoint approximation to the estimation equations built on moment conditions. Section 4 ends with an expla-

nation of how the first order normal approximation is related to the empirical saddlepoint approximation. Section 5 presents Monte Carlo simulations, which demonstrate the small sample performance of the empirical saddlepoint approximation. Situations where the empirical saddlepoint approximation is superior to currently available alternatives are noted. Section 6 shows that the saddlepoint approximation is well defined using the estimation equations presented in section 3. Section 7 concludes with additional interpretations, implementation issues, and directions for future work.

The random vector x will be defined on the probability space (Ω, \mathcal{B}, F) . Let a vector's norm be denoted $\|x\| \equiv \sqrt{x'x}$. Let a matrix's norm be denoted $\|M\| \equiv \sup \|Mx\|/\|x\|$. A δ ball centered at x will be denoted, $B_\delta(x) \equiv \{y : \|x - y\| < \delta\}$ for $\delta > 0$. Let $m(\cdot)$ denote the Lebesgue measure.

2 MODEL AND FIRST ORDER ASYMPTOTIC ASSUMPTIONS

The saddlepoint approximation requires stronger assumptions than those necessary for the first order normal approximation. The basic assumptions for first order asymptotic normality are recorded for the basic model.

Consider an m -dimensional set of moment conditions

$$g_i(\theta) \equiv g(x_i, \theta)$$

where θ is a k -dimensional set of parameters with $k \leq m$. According to the Economic theory, the moment conditions have expectation zero at the population parameter value, i.e.

$$E[g(x_i, \theta_0)] = 0.$$

The error process driving the system is assumed to be iid.

Assumption 1 . *The p -dimensional series x_i is iid from a distribution $F(x)$.*

The functions that create the moment conditions satisfy regularity conditions.

Assumption 2. *$g(x, \theta)$ is continuously partially differentiable in θ in a neighborhood of θ_0 . The functions $g(x, \theta)$ and $\frac{\partial g(x, \theta)}{\partial \theta'}$ are measurable functions of x for each $\theta \in \Theta$, and $E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial g(x, \theta)}{\partial \theta'} \right\| \right] < \infty$. $E[g(x, \theta_0)'g(x, \theta_0)] < \infty$ and $E[\sup_{\theta \in \Theta} \|g(x, \theta)\|] < \infty$. Each element of $g(x, \theta)$ is uniformly square integrable.*

Assumption 3. *The parameter space Θ is a compact subset of R^k . The population parameter value θ_0 is in the interior of Θ .*

To obtain an estimate from a finite sample, the weighted inner product of the sample analogue of the moment condition

$$G_N(\theta) = \frac{1}{N} \sum_{i=1}^N g(x_i, \theta)$$

is minimized conditional on, W_N , a given symmetric positive definite weighting matrix

$$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}} G_N(\theta)' W_N G_N(\theta).$$

The foundation for the first order asymptotic results is a central limit theorem distributional assumption.

Assumption 4. *The moment conditions evaluated at the population parameter values satisfy the central limit theorem*

$$\sqrt{N} G_N(\theta_0) \sim N(0, \Sigma_g).$$

Attention will be restricted to GMM estimates with minimum asymptotic variance.

Assumption 5. *The weighting matrix is selected so*

$$W_N \rightarrow \Sigma_g^{-1}.$$

Assumption 5 is usually satisfied by performing a first round estimation using an identity matrix as the weighting matrix.

Assumption 6. *The matrix $M(\theta) \equiv E \left[\frac{\partial g(x_i, \theta)}{\partial \theta'} \right]$ has full column rank for all θ in a neighborhood of θ_0 . The matrix $\Sigma(\theta) \equiv E [g(x_i, \theta)g(x_i, \theta)']$ is positive definite for all θ in a neighborhood of θ_0 .*

The parameter can be identified from the moment conditions.

Assumption 7. *Only θ_0 satisfies $E [g(x_i, \theta)] = 0$.*

These assumptions are sufficient to ensure that the GMM estimator is a root- N consistent estimator of θ_0 . The GMM estimates are asymptotically distributed as

$$\sqrt{N} (\hat{\theta} - \theta_0) \sim N \left(0, (M(\theta_0)' \Sigma_g^{-1} M(\theta_0))^{-1} \right).$$

The Economic theory implies all m moment conditions should equal zero. The first order conditions

$$M_N(\hat{\theta})' W_N G_N(\hat{\theta}) = 0$$

set k of the moments to exactly zero, where $M_N(\theta) = \frac{\partial G_N(\theta)}{\partial \theta'}$. The remaining $(m - k)$ moments can be used to test the Economic theory. The statistic

$$J = N G_N(\hat{\theta})' W_N G_N(\hat{\theta})$$

tests these overidentifying restrictions. The J -statistic is asymptotically distributed χ_{m-k}^2 when the null hypothesis of the Economic theory being true is correct.

3 MOMENT CONDITIONS TO ESTIMATION EQUATIONS

The saddlepoint approximation uses a just identified system of estimation equations. This section shows how a just identified system of equations can be created from the moment conditions implied by the Economic theory. The estimation equations are created by augmenting the parameters with a set of $(m - k)$ parameters denoted λ . The λ parameters are a local coordinate system², spanning the space of overidentifying restrictions. The λ parameters are selected so that under the null hypothesis of the Economic theory being true, the population parameter values are $\lambda_0 = 0$. Thus, the overidentifying restrictions can be tested by the hypothesis $H_0 : \lambda = 0$.

For each value of θ the sample moment conditions $G_N(\theta)$ is an m -dimensional vector. As θ takes different values, the moment conditions create a k -dimensional manifold. For a fixed value of θ the space spanned by the derivative of the k -dimensional manifold will be called the identifying space. The orthogonal complement of the identifying space is called the overidentifying space. This decomposition is a generalization of the decomposition used in Sowell (1996) where the tangent space at $\hat{\theta}$ was decomposed into a k -dimensional identifying space and an $(m - k)$ -dimensional space of overidentifying restrictions. The generalization is defining the decomposition at each value³ of θ , not only at $\hat{\theta}$.

For each value of θ , let $\overline{M}_N(\theta)$ denote the derivative of $G_N(\theta)$ with respect to θ

²An introduction to the use of local coordinate systems can be found in Boothby (2003). Application of these tools in Statistics and Econometrics can be found in Amari (1985) and Marriott and Salmon (2000).

³When attention is restricted to the empirical saddlepoint approximation, then the decomposition only needs to exist for parameters in neighborhoods of the local minima.

scaled (standardized) by the Cholesky decomposition of the weighting matrix,

$$\overline{M}_N(\theta) = W_N^{1/2} \frac{\partial G_N(\theta)}{\partial \theta'}.$$

Using this notation, the GMM first order conditions can be written

$$\overline{M}_N(\hat{\theta})' W_N^{1/2} G_N(\hat{\theta}) = 0.$$

The columns of $\overline{M}_N(\hat{\theta})$ define the k linear combinations used to identify and estimate θ .

The orthogonal complement of the space spanned by the columns of $\overline{M}_N(\hat{\theta})$ will be the $(m - k)$ -dimensional space used to test the validity of the overidentifying restrictions and will be spanned by λ . The augmenting parameters are determined by performing the decomposition for every value of θ . Denote the projection matrix for the space spanned by $\overline{M}_N(\theta)$ as

$$P_{\overline{M}(\theta),N} = \overline{M}_N(\theta) \left(\overline{M}_N(\theta)' \overline{M}_N(\theta) \right)^{-1} \overline{M}_N(\theta)'$$

$P_{\overline{M}(\theta),N}$ is a real symmetric positive semidefinite matrix, which is also idempotent. Denote a spectral decomposition⁴

$$P_{\overline{M}(\theta),N} = C_N(\theta) \Lambda C_N(\theta)' = \begin{bmatrix} C_{1,N}(\theta) & C_{2,N}(\theta) \end{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & 0_{(m-k)} \end{bmatrix} \begin{bmatrix} C_{1,N}(\theta)' \\ C_{2,N}(\theta)' \end{bmatrix}$$

where $C_N(\theta)' C_N(\theta) = I_m$. The column span of $C_{1,N}(\theta)$ is the same as the column span of $\overline{M}_N(\theta)$, and the columns of $C_{2,N}(\theta)$ span the orthogonal complement at θ . Hence, for each value of θ , the m -dimensional space containing $G_N(\theta)$ can be locally parameterized by

$$\Psi_N(\theta, \lambda) = \begin{bmatrix} C_{1,N}(\theta)' W_N^{1/2} G_N(\theta) \\ \lambda - C_{2,N}(\theta)' W_N^{1/2} G_N(\theta) \end{bmatrix}.$$

The first set of equations are the k -dimensions of $W_N^{1/2} G_N(\theta)$ that locally vary with θ . The parameters θ are local coordinates for these k -dimensions. The second set of equations gives the $(m - k)$ -dimensions of $W_N^{1/2} G_N(\theta)$ that are locally orthogonal to

⁴The spectral decomposition is not unique, raising a potential concern. However, the invariance of inference with respect to alternative spectral decompositions is documented in **Theorem 2** and **Theorem 6**.

θ . The parameters λ are local coordinates for these $(m - k)$ -dimensions. For each value of θ , the parameters λ span the space that is the orthogonal complement of the space spanned by θ .

This parameterization of the m -dimensional space can be used to obtain parameter estimates by solving

$$\Psi_N(\theta, \lambda) = 0.$$

This set of estimation equations will be used in the saddlepoint approximation. This function can also be written $\Psi_N(\theta, \lambda) = \frac{1}{N} \sum_{i=1}^N \psi(x_i, \theta, \lambda)$ where

$$\psi(x_i, \theta, \lambda) = \begin{bmatrix} C_{1,N}(\theta)' W_N^{1/2} g_i(\theta) \\ \lambda - C_{2,N}(\theta)' W_N^{1/2} g_i(\theta) \end{bmatrix}.$$

A generic value of $\psi(x_i, \theta, \lambda)$ will be denoted $\psi(x, \theta, \lambda)$. These estimation equations give a just identified system of m equations in m unknowns. The moment conditions $\psi(x, \theta, \lambda)$ summarize the first order conditions for GMM estimation and the overidentifying restrictions statistics.

The column span of $C_{1,N}(\theta)$ and $\bar{M}_N(\theta)$ are the same. So, the system of equations

$$C_{1,N}(\theta)' W_N^{1/2} G_N(\theta) = 0$$

is equivalent to

$$\bar{M}_N(\theta)' W_N^{1/2} G_N(\theta) = 0,$$

and both imply the same parameter estimates, $\hat{\theta}$. This system of equations is solved independently of λ . The system of equations $\lambda - C_{2,N}(\theta)' W_N^{1/2} G_N(\theta) = 0$ imply the estimate

$$\hat{\lambda} = C_{2,N}(\hat{\theta})' W_N^{1/2} G_N(\hat{\theta}).$$

An alternative differential geometric explanation of the estimation equations is presented in the Appendix section 8.8.

3.1 NOTES

1. The inner product of $\Psi_N(\theta, \lambda)$ will be called the extended-GMM objective function. The estimates are the parameters that minimize the inner product, which

can be rewritten

$$\begin{aligned}\mathcal{Q}_N(\theta, \lambda) &= \Psi_N(\theta, \lambda)' \Psi_N(\theta, \lambda) \\ &= G_N(\theta)' W_N G_N(\theta) - 2\lambda' C_{2,N}(\theta)' W_N^{1/2} G_N(\theta) + \lambda' \lambda.\end{aligned}$$

Because $\Psi_N(\theta, \lambda)$ is a just identified system of equations there is no need for an additional weighting matrix. When $\lambda = 0$ the extended-GMM objective function reduces to the traditional GMM objective function. The extended-GMM objective function is created from the traditional GMM objective function by appending a quadratic function in the tangent spaces of overidentifying restrictions.

2. The parameter estimates using the extended-GMM objective function agree with traditional results.

Theorem 1. *If Assumptions 1-7 are satisfied, the asymptotic distribution for the extended-GMM estimators is*

$$\sqrt{N} \begin{pmatrix} \hat{\theta} - \theta_0 \\ \hat{\lambda} \end{pmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} (M(\theta_0)' \Sigma_g^{-1} M(\theta_0))^{-1} & 0 \\ 0 & I_{(m-k)} \end{bmatrix} \right).$$

Proofs are in the appendix.

3. The estimates $\hat{\theta}$ and $\hat{\lambda}$ are independent up to first order asymptotics.
4. The overidentifying restrictions can be tested with

$$\begin{aligned}N\hat{\lambda}'\hat{\lambda} &= NG_N(\hat{\theta})' W_N^{1/2}' C_{2,N}(\hat{\theta}) C_{2,N}(\hat{\theta})' W_N^{1/2} G_N(\hat{\theta}) \\ &= NG_N(\hat{\theta})' W_N^{1/2}' P_{M(\hat{\theta})}^\perp W_N^{1/2} G_N(\hat{\theta}) \\ &= NG_N(\hat{\theta})' W_N^{1/2}' W_N^{1/2} G_N(\hat{\theta}) \\ &= NG_N(\hat{\theta})' W_N G_N(\hat{\theta}) \\ &= J.\end{aligned}$$

5. The asymptotic inference is invariant to the selection of the spectral decomposition.

Theorem 2. *For the extended-GMM estimators the asymptotic distribution for*

the parameters, θ , and the parameters that test the overidentifying restriction, λ , are invariant to the spectral decomposition that spans the tangent space.

6. To reduce notation let $\alpha \equiv \left[\theta' \quad \lambda' \right]'$. A generic element $\psi(x_i, \alpha)$ will be denoted $\psi(x, \alpha)$. The estimation equations used in the saddlepoint approximation can be denoted

$$\Psi_N(\alpha) = 0.$$

The extended-GMM objective function will be denoted

$$Q_N(\alpha) = \Psi_N(\alpha)' \Psi_N(\alpha).$$

7. The point of departure for the saddlepoint approximation is a system of m estimation equations in m parameters. These will be the first order conditions from minimization of the extended-GMM objective function.

The first order conditions imply

$$\frac{\partial Q_N(\hat{\alpha})}{\partial \alpha} = 2 \frac{\partial \Psi_N(\hat{\alpha})'}{\partial \alpha} \Psi_N(\hat{\alpha}) = 0.$$

Assuming $M(\theta)$ and W_N are full rank implies the first order conditions are equivalent to

$$\Psi_N(\hat{\alpha}) = 0.$$

8. A minimum or root of the extended-GMM objective function can be associated with either a local maximum or a local minimum of the original GMM objective function. Attention must be focused on only the minima of the extended-GMM objective function associated with the local minima of the original GMM objective function. This will be done with an indicator function

$$I_{\text{pos}}(Q_N(\theta)) = \begin{cases} 1, & \text{if } Q_N(\theta) \text{ is positive definite} \\ 0, & \text{otherwise.} \end{cases}$$

This indicator function uses the original GMM objective function not the extended-GMM objective function.

4 THE SADDLEPOINT APPROXIMATION

The saddlepoint density replaces the central limit theorem in the traditional GMM distribution theory. The central limit theorem uses information about the location and convexity of the GMM objective function at the global minimum. The saddlepoint approximation uses information about the convexity of the objective function at each point in the parameter space. The central limit theorem is built on a two-term Taylor series expansion, i.e. a linear approximation, of the characteristic function about the mean. A higher order Taylor series expansion about the mean can be used to obtain additional precision. This results in an Edgeworth expansion. Because the expansion is at the distribution's mean, the Edgeworth expansion gives a significantly better approximation at the mean of the distribution, $\mathcal{O}(N^{-1})$ versus $\mathcal{O}(N^{-1/2})$. Unfortunately, the quality of the approximation can deteriorate significantly for values away from the mean. The saddlepoint approximation exploits these characteristics of the Edgeworth expansion to obtain an improved approximation. Instead of a single Taylor series expansion, the saddlepoint uses multiple expansions to obtain improved accuracy, one expansion at every value in the parameter space.

The significantly improved approximation of the Edgeworth expansion only occurs at the mean of the distribution. To obtain this improvement at an arbitrary value in the parameter space, a conjugate distribution is used. For the parameter value α the conjugate density is

$$h_{N,\beta}(x) = \frac{\exp\left\{\frac{\beta'}{N}\psi(x, \alpha)\right\} dF(x)}{\int \exp\left\{\frac{\beta'}{N}\psi(w, \alpha)\right\} dF(w)}.$$

The object of interest is the distribution of $\Psi(\alpha)$ not an individual element $\psi(\alpha)$, hence the parameter β is normalized by N .

At the parameter value of interest, α , the original distribution is transformed to a conjugate distribution. The conjugate distribution is well defined for arbitrary values of β . This is a degree of freedom, i.e., β can be selected optimally for a given α . A specific conjugate distribution is selected so its mean is transformed back to the original distribution at the value of interest. Thus, if β is selected to satisfy the saddlepoint equation

$$\int \psi(x, \alpha) \exp\left\{\frac{\beta'}{N}\psi(x, \alpha)\right\} dF(x) = 0.$$

Denote the solution to the saddlepoint equation as $\beta(\alpha)$. An Edgeworth expansion is calculated for the conjugate distribution defined by $\beta(\alpha)$. This Edgeworth expansion is then transformed back to give the saddlepoint approximation to the original distribution at the parameter value of interest, α .

The basic theorems from the Statistics literature are in Almudevar, Field and Robinson (2000), Field and Rochetti (1994), Field (1982) and Rochetti and Welsch (1994). To date the saddlepoint distribution theory in the Statistics literature is not well suited for empirical Economics. There are two problems. The first is the assumption that the objective function has a single extreme value. The second is the assumption that the saddlepoint equation has a solution for every value in the parameter space. A common feature of GMM objective functions is the existence of more than one local minimum. In addition, the nonlinearity of the moment conditions can make it impossible to solve the saddlepoint equation for an arbitrary value in the parameter space. The basic theorems from the Statistics literature need slight generalizations to allow for multiple local minima and the non-existence of a solution to the saddlepoint equation. The generalizations are contained in **Theorems 3** and **4**. The next two subsections elaborate.

4.1 DENSITY RESTRICTIONS

The empirical saddlepoint density applied to GMM moments requires two restrictions: the local behavior of the objective function must be associated with a local minimum and the parameters must be consistent with the observed data.

Identification for GMM implies there will only be one minimum, asymptotically. However, in finite samples there may not be enough data to accurately distinguish this asymptotic structure, i.e., often there are multiple local minima. Traditional GMM ignores multiple minima and restricts attention to the global minimum. The traditional asymptotic normal approximation to the GMM estimator is an approximation built on the local structure (location and convexity) of the global minimum. The saddlepoint approximation takes a different approach. The saddlepoint density approximates the sampling density for the location of solutions to the estimation equations. These include both local maxima and local minima. As with traditional GMM estimation, as the sample size increases one of the local minima will become the unique minimum. For the empirical saddlepoint density, attention is focused on the local minima by setting the saddlepoint density to zero if the first derivative of

the estimation equation is not positive definite. The term

$$I_{\text{pos}}(Q_N(\theta))$$

will be used to denote the indicator function for the original GMM objective function being positive definite at the θ value in α . A similar restriction was used in Skovgaard (1990)⁵.

The second restriction for the saddlepoint applied to the GMM estimation equations concerns the fact that the empirical saddlepoint equation may not have a solution. In this case the approximate density is set equal to zero. The lack of a solution to the empirical saddlepoint equation means the observed data is inconsistent with the selected parameter, α . The indicator

$$I(\beta(\alpha))$$

equals one when the saddlepoint equation for α has a solution.

This type of restriction has occurred recently in the Statistics and Econometrics literature. The inability to find a solution to the empirical saddlepoint equation is equivalent to the restrictions on the sample mean in the selection of the parameters in the exponential tilting/maximum entropy estimation of Kitamura and Stutzer (1997). For the simple case of estimating the sample mean, the parameters must be restricted to the support of the observed sample. It is impossible to select nonnegative weights (probabilities) to have the weighted sum of the sample equal a value outside its observed range.

4.2 ADDITIONAL ASSUMPTIONS

To justify the empirical saddlepoint approximation, four results are needed.

1. The distribution for the local minima of the extended-GMM objective function associated with the local minima of the original GMM objective function.
2. The tilting relationship between the density for the zeros (local minima) of the extended-GMM objective function associated with the local minima of the GMM objective function when the observations are drawn from the population distribution

⁵In Skovgaard (1990) the restriction was applied to the equivalent of the extended-GMM objective function not the equivalent of the original GMM objective function, i.e., the Skovgaard (1990) density would include both the local maxima and the local minima of the original GMM objective function.

and the density when the observations are drawn from the conjugate distribution.

3. The saddlepoint approximation, assuming the distribution for the observed series is known.

4. The empirical saddlepoint approximation where the distribution for the observed series is replaced with the empirical distribution.

The first three are achieved by minor changes to the results in Almudevar, Field and Robinson (2000), denoted AFR. The changes concern the need to restrict attention to solutions of the estimation equations associated with local minima of the original GMM objective function. In AFR the density is for all solutions, even those associated with local maxima. Another change is to set the approximation to zero if the empirical saddlepoint equation does not have a solution.

The fourth result is achieved by minor changes to results in Ronchetti and Welsh (1994). The first change is to set the approximation to zero if the empirical saddlepoint equation does not have a solution. This is achieved by including an indicator function for the existence of a solution to the empirical saddlepoint equation. The other change is to allow multiple solutions to the estimation equations. This is achieved by applying the result in Ronchetti and Welsh (1994) to each solution of the estimation equation associated with a local minima of the original GMM objective function.

Unlike traditional Econometrics, the saddlepoint approximation is well defined for some models lacking identification. Traditionally, it is assumed that the limiting objective function uniquely identifies the population parameter, e.g. **Assumption 7**. The saddlepoint approximation does not require such a strong assumption. There may be multiple solutions to the moment conditions associated with local minima of the GMM objective function. Define the set $\mathcal{T} = \{\theta \in \Theta : E[g(x_i, \theta)] = 0\}$. If \mathcal{T} is a singleton, then this reduces to the traditional identification assumption. An individual element in \mathcal{T} will typically be denoted θ_0^* .

The empirical saddlepoint approximation requires stronger assumptions than the assumptions needed for the first order normal approximation. These additional restrictions concern the existence and integrability of higher order derivatives of the moments. The saddlepoint density requires that the asymptotic behavior in the neighborhood of the global minimum of the GMM objective function also holds for the neighborhood of each element in \mathcal{T} . Assumptions are also needed to ensure that the Edgeworth expansion of the conjugate density is well defined. Finally, the empirical saddlepoint approximation requires the existence of higher order moments and smoothness to ensure the sample averages converge uniformly to their limits.

The following assumptions are stated in terms of the moment conditions. The assumptions in AFR and RW are stated in terms of a just identified system of estimation equations. The “proofs” in the appendix show that the following assumptions are strong enough to ensure that the assumptions in AFR and RW are satisfied for the estimation equations derived above in section 3.

Assumption 1’ . *The p -dimensional series x_i is iid from a distribution $F(x)$.*

The moment conditions satisfy regularity conditions.

Assumption 2’. *The function $g(x, \theta)$ is uniformly continuous in x and has three derivatives with respect to θ which are uniformly continuous in x .*

The functions $g(x, \theta)$ and $\frac{\partial g(x, \theta)}{\partial \theta'}$ are measurable functions of x for each $\theta \in \Theta$, and $E \left[\sup_{\theta \in \Theta} \left\| \frac{\partial g(x, \theta)}{\partial \theta} \right\| \right] < \infty$. $E [g(x, \theta_0)' g(x, \theta_0)] < \infty$ and $E [\sup_{\theta \in \Theta} \|g(x, \theta)\|] < \infty$. Each element of $g(x, \theta)$ is uniformly square integrable.

Assumption 2’ implies that N times the GMM objective function will uniformly converge to the nonstochastic function $E [g(x, \theta)]' E [g(x, \theta)]$.

Assumption 3’. *The parameter space Θ is a compact subset of R^k .*

$\theta_0 \in \mathcal{T}$.

Each element of \mathcal{T} is in the interior of Θ .

There exists a $\delta > 0$ such that for any two unique elements in \mathcal{T} , $\theta_{0,j}^$ and $\theta_{0,i}^*$, $|\theta_{0,j}^* - \theta_{0,i}^*| > \delta$.*

By **Assumption 3’**, the population parameter is in the set of minima of the extended-GMM objective function associated with the local minima of the original GMM objective function. This assumption also ensures that the limiting objective function can be expanded in a Taylor series about each of its local minima. Finally, this assumption requires that any identification failure results in disjoint solutions.

Assumptions 1’-3’ are sufficient to ensure that the GMM estimator is a root- N consistent estimator of an element in \mathcal{T} .

The traditional distributional assumption must hold for each local minima.

Assumption 4’. *The moment conditions evaluated at $\theta_0^* \in \mathcal{T}$ satisfy the central limit theorem*

$$\sqrt{N}G_N(\theta_0^*) \sim N(0, \Sigma_g(\theta_0^*))$$

where $\Sigma_g(\theta_0^)$ may be different for each value of $\theta_0^* \in \mathcal{T}$.*

Assumption 5'. The weighting matrix is selected so that it is always positive definite and

$$W_N \rightarrow \Sigma_g(\theta_0^*)^{-1}$$

where $\theta_0^* \in \mathcal{T}$.

This assumption can be satisfied by performing a first round estimation using a positive definite matrix as the weighting matrix. If \mathcal{T} is a singleton this ensures the first order distribution will be efficient.

The objective function must satisfy restrictions in a neighborhood of each solution to the extended-GMM objective function associated with a local minima of the original GMM objective function.

Assumption 6'. For each $\theta_0^* \in \mathcal{T}$,

1. The matrix $M(\theta) \equiv E \frac{\partial g(x, \theta)}{\partial \theta'}$ is continuously differentiable and has full column rank for all θ in a neighborhood of θ_0^* .
2. The matrix $\Sigma(\theta) \equiv E [g(x, \theta)g(x, \theta)']$ is continuous and positive definite for all θ in a neighborhood of θ_0^* .

3. The function

$$\int \frac{\partial g(x, \theta)}{\partial \theta'} \exp \{ \beta' g(x, \theta) \} dF(x)$$

exists for β in a set containing the origin.

4. For $1 \leq i, j, s_1, s_2, s_3 \leq m$ the integrals

$$\begin{aligned} & \int \left\{ \frac{\partial g_i(x, \alpha)}{\partial \alpha_{s_3}} \right\}^2 dF(x), & \int \left\{ \frac{\partial g_i(x, \alpha)}{\partial \alpha_{s_3}} g_j(x, \alpha) \right\}^2 dF(x), \\ & \int \left\{ \frac{\partial^2 g_i(x, \alpha)}{\partial \alpha_{s_2} \partial \alpha_{s_3}} \right\}^2 dF(x), & \int \left\{ \frac{\partial^2 g_i(x, \alpha)}{\partial \alpha_{s_2} \partial \alpha_{s_3}} g_j(x, \alpha) \right\} dF(x), \\ & \int \left\{ \frac{\partial^3 g_i(x, \alpha)}{\partial \alpha_{s_1} \partial \alpha_{s_2} \partial \alpha_{s_3}} \right\} dF(x) \end{aligned}$$

are finite.

Assumption 6'.1 and **Assumption 6'.2** are straightforward. **Assumption 6'.3** ensures the saddlepoint equation will have a solution in the neighborhood of

θ_0^* . **Assumption 6'.4** with **Assumption 2'** ensures the uniform convergence of the sample averages to their limits.

The next assumption restricts attention to parameter values in a τ ball of a fixed parameter value. Calculate the maximum distance between the derivative at α_0 and the derivative of any other parameter value in $B_\tau(\alpha_0)$. If the derivative is zero for any parameter value in $B_\tau(\alpha_0)$, set the distance to infinity.

Define the random variable⁶

$$z(\theta_0^*, \tau) = \begin{cases} \sup_{\theta \in B_\tau(\theta_0^*)} \left\| \frac{\partial^2 g(x, \theta_0^*)' W_N g(x, \theta_0^*)}{\partial \theta' \partial \theta} - \frac{\partial^2 g(x, \theta)' W_N g(x, \theta)}{\partial \theta' \partial \theta} \right\|, & \frac{\partial^2 g(x, \theta)' W_N g(x, \theta)}{\partial \theta' \partial \theta} \text{ is} \\ & \text{positive definite,} \\ & \forall \theta \in B_\tau(\theta_0^*) \\ \infty, & \text{otherwise} \end{cases}$$

Now define the events in Ω such that this maximum deviation between the first derivative is below some value such as $\gamma > 0$.

$$H(\theta_0^*, \gamma, \tau) = \{z(\theta_0^*, \tau) < \gamma\} \subset \Omega$$

This restricts attention to events where the objective function has an invertible derivative and is fairly smooth.

Now define

$$R^*(x, \theta) = \begin{cases} \left(\frac{\partial^2 g(x, \theta_0^*)' W_N g(x, \theta_0^*)}{\partial \theta' \partial \theta} \right)^{-1} \frac{\partial g(x, \theta)'}{\partial \theta} W_N g(x, \alpha), & \frac{\partial^2 g(x, \theta)' W_N g(x, \theta)}{\partial \theta' \partial \theta} \text{ is positive} \\ & \text{definite, } \forall \theta \in B_\tau(\theta_0^*) \\ \infty, & \text{otherwise} \end{cases}$$

and the density

$$f_{R^*(x, \theta)}(z; H(\theta, \gamma, \tau)) = \frac{\Pr \{ \{R^*(x, \theta) \in B_\tau(z)\} \cap H(\theta, \gamma, \tau) \}}{m(B_\tau(z))}.$$

Assumption 7'. For any compact set $A \subset \mathcal{A}$ and for any $0 < \gamma < 1$, there exists $\tau > 0$ and $\delta > 0$ such that $f_{R^*(x, \theta)}(z; H(\theta, \gamma, \tau))$ exists and is continuous and bounded by some fixed constant K for any $\alpha \in A$ and $z \in B_\delta(0)$.

⁶Almudevar, Field and Robinson (2000) uses $\|B^{-1}A - I\|$ as the distance between the two matrices, A and B . This paper will use the alternative distance measure $\|A - B\|$, which is defined for none square matrices. This definition is more commonly used in the Econometrics literature, e.g., see Assumption 3.4 in Hansen (1982).

This is a high level assumption requiring the moment conditions have a bounded density in the neighborhood of zero. This is required to establish the existence of the density for the location of the local minima of the original GMM objective function.

The next assumption permits the moment conditions (and hence estimation equations) to possess a mixture of continuous and discrete random variables. For each θ let $D_\theta \subset \mathbb{R}^m$ be a set with Lebesgue measure zero such that by the Lebesgue decomposition theorem,

$$\Pr \left\{ \frac{\partial g(x, \theta)'}{\partial \theta} W_N g(x, \theta) \in A \right\} = \Pr \left\{ \frac{\partial g(x, \theta)'}{\partial \theta} W_N g(x, \theta) \in A \cap D_\theta \right\} + \int_A f_\theta dm$$

where f_θ may be an improper density. Let $I_{i\theta} = 0$ if $g(x_i, \theta) \in D_\theta$ and 1, otherwise. Assume the moment conditions can be partitioned into contributions from the continuous components and the discrete components.

Assumption 8'. *There are iid random vectors $U_{i\theta} = \begin{pmatrix} W_{i\theta} & V_{1i\theta} & V_{2i\theta} \end{pmatrix}$, where $W_{i\theta}$ are jointly continuous random vectors of dimension m and $V_{1i\theta}$ and $V_{2i\theta}$ are random vectors of dimension m and m^* , respectively, such that*

$$N\bar{T}_\theta = \sum_{i=1}^N \frac{\partial g(x, \theta)'}{\partial \theta} W_N g(x_i, \theta) = \sum_{i=1}^N I_{i\theta} W_{i\theta} + \sum_{i=1}^N (1 - I_{i\theta}) V_{1i\theta}$$

$$\text{vec}(N\bar{S}_\theta) = \text{vec} \left(\sum_{i=1}^N \frac{\partial^2 g(x_i, \theta)' W_N g(x_i, \theta)}{\partial \theta' \partial \theta} \right) = A_\theta \sum_{i=1}^N I_{i\theta} U_{i\theta},$$

where A_θ is of dimension m^2 by $2m + m^*$.

Let $U'_{j\theta} = \begin{pmatrix} W'_{j\theta} & V'_{j\theta} \end{pmatrix}$ have the distribution of $U_{j\theta}$ conditional on $I_{j\theta} = 1$ and $V''_{1j\theta}$ have the distribution of $V_{1j\theta}$ conditional on $I_{j\theta} = 0$.

Define

$$\tilde{U}_\theta = \begin{pmatrix} \tilde{W}'_{j\theta} & \tilde{V}'_{j\theta} \end{pmatrix} = \frac{1}{N} \sum_{i=1}^K U'_{j\theta},$$

where K is a draw from a binomial distribution with parameters N and $\rho = \Pr \{I_{i\theta} = 1\}$ conditional on being positive. Define

$$\tilde{T}_\theta = \frac{1}{N} \sum_{j=1}^K W'_{j\theta} + \frac{1}{N} \sum_{j=K+1}^N V''_{1j\theta}$$

when $0 < K < N$. Finally, define

$$\text{vec} \left(N \tilde{S}_\theta \right) = A_\theta \frac{1}{N} \sum_{j=1}^K U'_{j\theta}.$$

Assumption 9'. $\det \left(\tilde{S}_\theta \right) \neq 0$ and that the transformation \tilde{T}_θ to $\tilde{S}_\theta^{-1} \tilde{T}_\theta$ given \tilde{V}_θ is one-to-one with probability one.

Assumption 10'. $E \left[\exp \left\{ \beta^T U_\theta \right\} \right] < \infty$ for $\|\beta\| < \gamma$ for some $\gamma > 0$ and for all θ .

Let

$$\Lambda(F; \theta', \theta) = \begin{cases} \left\| E \left[\frac{\partial^2 g(x, \theta_0^*)' W_N g(x, \theta_0^*)}{\partial \theta' \partial \theta} \right] - E \left[\frac{\partial^2 g(x, \theta)' W_N g(x, \theta)}{\partial \theta' \partial \theta} \right] \right\|, & E \left[\frac{\partial^2 g(x, \theta)' W_N g(x, \theta)}{\partial \theta' \partial \theta} \right] \text{ is} \\ & \text{positive definite,} \\ \infty, & \text{otherwise} \end{cases}$$

and

$$\Lambda^*(F; \theta, \tau) = \sup_{\theta' \in B_\tau(\theta)} \Lambda(F; \theta', \theta).$$

Assumption 11'. Given $0 < \gamma < 1$, there is a τ such that $\sup_{\theta \in \tilde{B}_\tau(\theta_0)} \Lambda^*(F_0; \theta, \tau) < \gamma$.

Assumption 12'. For fixed $\theta \in B_\tau(\theta_0)$, $\Lambda^*(\cdot; \cdot, \tau)$ is continuous at (F_0, θ) in the product topology.

Theorem 3. (Almudevar, Field and Robinson (2000)) *Under Assumptions 1'-12'* for $\theta_0^* \in \mathcal{T}$, there is, with probability $1 - e^{-cN}$ for some $c > 0$, a uniquely defined M -estimate $\hat{\alpha}$ on $B_\tau(\alpha_0^*)$ which has a density, restricted to $B_\tau(\theta_0^*)$,

$$\begin{aligned} f_N(\alpha) &= K_N \times I_{\text{pos}}(Q_N(\theta)) \times I(\beta(\alpha)) \\ &\quad \times \left(\frac{N}{2\pi} \right)^{\frac{m}{2}} \left| E[\psi'(x, \alpha)] \right| \left| E[\psi(x, \alpha)\psi(x, \alpha)'] \right|^{-1/2} \\ &\quad \times \exp \{ N \kappa_N(\beta(\alpha), \alpha) \} (1 + \mathcal{O}(N^{-1})) \end{aligned}$$

where $\beta(\alpha)$ is the solution of the saddlepoint equation

$$\int \psi(x, \alpha) \exp \left\{ \frac{\beta'}{N} \psi(x, \alpha) \right\} dF(x) = 0,$$

$$\kappa_N(\beta(\alpha), \alpha) = \int \exp \left\{ \frac{\beta(\alpha)'}{N} \psi(x, \alpha) \right\} dF(x),$$

the expectations are with respect to the conjugate density

$$h_N(x) = \frac{\exp \left\{ \frac{\beta(\alpha)'}{N} \psi(x, \alpha) \right\} dF(x)}{\int \exp \left\{ \frac{\beta(\alpha)'}{N} \psi(w, \alpha) \right\} dF(w)},$$

the term $I_{\text{pos}}(Q_N(\theta))$ is an indicator function that sets the approximation to zero if the original GMM objective function is not positive definite at the θ value in α , the term $I(\beta(\alpha))$ is an indicator function that sets the approximation to zero if the saddlepoint equation does not have a solution, and K_N is a constant ensuring the approximation integrates to one.

This theorem shows how the saddlepoint approximation is calculated. The saddlepoint approximation is nonnegative and gives a faster rate of convergence than the asymptotic normal approximation.

The calculation of the saddlepoint density requires knowledge of the distribution $F(x)$. In most Economic applications this is unknown. Replacing the distribution with the observed empirical distribution results in the empirical saddlepoint approximation. This replaces the expectations with respect to the distribution $F(x)$ with sample averages, i.e., the expectation with respect to the empirical distribution. The empirical saddlepoint density is defined to be

$$\begin{aligned} \hat{f}_N(\alpha) &= K_N \times I_{\text{pos}}(Q_N(\theta)) \times I(\hat{\beta}_N(\alpha)) \\ &\times \left(\frac{N}{2\pi} \right)^{\frac{m}{2}} \left| \sum_{i=1}^N \frac{\partial \psi(x_i, \alpha)'}{\partial \alpha} p_i(\alpha) \right| \left| \sum_{i=1}^N \psi(x_i, \alpha) \psi(x_i, \alpha)' p_i(\alpha) \right|^{-1/2} \\ &\times \exp \left\{ N \ln \left(\frac{1}{N} \sum_{i=1}^N \exp \left\{ \frac{\hat{\beta}_N(\alpha)'}{N} \psi(x_i, \alpha) \right\} \right) \right\}, \end{aligned}$$

where $\hat{\beta}_N(\alpha)$ is the solution of

$$\sum_{i=1}^N \psi(x_i, \alpha) \exp \left\{ \frac{\beta'}{N} \psi(x_i, \alpha) \right\} = 0$$

and

$$p_i(\alpha) = \frac{\exp \left\{ \frac{\hat{\beta}_N(\alpha)'}{N} \psi(x_i, \alpha) \right\}}{\sum_{j=1}^N \exp \left\{ \frac{\hat{\beta}_N(\alpha)'}{N} \psi(x_j, \alpha) \right\}}.$$

The empirical saddlepoint approximation is scaled by its (numerical) integral, K_N , to give a density that integrates to one.

Using the empirical distribution instead of a known distribution gives a nonparametric procedure. It results in a reduction in accuracy as noted in the next theorem. The appropriate rate of convergence is achieved by restricting attention to parameters in a shrinking neighborhood of the local minima of the estimation equations, $|\alpha - \alpha_0| < \Delta/\sqrt{N}$ for some $\Delta < \infty$. This can also be thought of as obtaining the density for $u = \sqrt{N}(\alpha - \alpha_0)$.

Theorem 4. (Ronchetti and Welsh (1994)) *If **Assumptions 1'-12'** are satisfied for $\theta_0^* \in \mathcal{T}$,*

$$\frac{f_N \left(\alpha_0^* + u/\sqrt{N} \right)}{\hat{f}_N \left(\hat{\alpha}_N^* + u/\sqrt{N} \right)} = 1 + \mathcal{O}_p \left(N^{-1/2} \right)$$

where the convergence is uniform for u in any compact set.

4.3 ASYMPTOTIC NORMAL AND EMPIRICAL SADDLEPOINT APPROXIMATIONS

This section compares the empirical saddlepoint approximation with the asymptotic normal approximation. The two densities have similar structures. Their differences concern the “means” and “covariances.”

From **Theorem 2**, the asymptotic normal approximation can be written

$$\hat{f}_A(\alpha) = (2\pi)^{-\frac{m}{2}} \left| \frac{\hat{\Sigma}}{N} \right|^{-1/2} \exp \left\{ -\frac{1}{2} \begin{bmatrix} \theta - \hat{\theta} \\ \lambda - \hat{\lambda} \end{bmatrix}' \left(\frac{\hat{\Sigma}}{N} \right)^{-1} \begin{bmatrix} \theta - \hat{\theta} \\ \lambda - \hat{\lambda} \end{bmatrix} \right\}$$

where $\hat{\theta}$ is the GMM estimator, i.e. the global minima of the objective function, $\hat{\lambda} = C_2(\hat{\theta})W^{-1/2}G_N(\hat{\theta})$ and $\hat{\Sigma}$ is the covariance matrix given **Theorem 2**.

The covariance matrix is the observed second derivative (convexity) of the extended-GMM objective function at the global minima and can be estimated with $\hat{\Sigma} =$

$(A'B^{-1}A)^{-1}$ where

$$A = \frac{1}{N} \sum_{i=1}^N \frac{\partial \psi(x_i, \hat{\alpha})}{\partial \alpha'} \quad \text{and} \quad B = \frac{1}{N} \sum_{i=1}^N \psi(x_i, \hat{\alpha}) \psi(x_i, \hat{\alpha})'.$$

The asymptotic normal approximation is built on the local behavior (location and convexity) of the objective function at its extreme value. The asymptotic normal approximation delivers a familiar, but restricted, structure for the approximation to the sampling distribution: unimodal, symmetric with the thin tails associated with the normal distribution. The asymptotic normal approximation is constructed from a linear approximation to the first order conditions. Linear first order conditions occur when the objective function is quadratic. So the asymptotic normal approximation is built on a quadratic approximation to the objective function. This asymptotic normal approximation is a poor approximation to the sampling distribution if the objective function is not well approximated by a quadratic over relevant parameter values. The local behavior (location and convexity) of the objective function at its extreme value will not contain enough information to accurately approximate the objective function in a large enough region.

Instead of focusing only on the global minimum, the saddlepoint approximation summarizes the information in the sample using the global shape of the objective function. The similar structure to the asymptotic normal approximation is shown in the following theorem.

Theorem 5. *The empirical saddlepoint approximation can be written*

$$\begin{aligned} \hat{f}_S(\alpha) &= K_N \times I_{\text{pos}}(Q_N(\alpha)) \times I(\beta(\alpha)) \\ &\times (2\pi)^{-\frac{m}{2}} \left| \frac{\hat{\Sigma}_C(\alpha)}{N} \right|^{-1/2} \exp \left\{ -\frac{1}{2} \begin{bmatrix} \theta - \theta^*(\alpha) \\ \lambda - \lambda^*(\alpha) \end{bmatrix}' \left(\frac{\hat{\Sigma}_e(\alpha)}{N} \right)^{-1} \begin{bmatrix} \theta - \theta^*(\alpha) \\ \lambda - \lambda^*(\alpha) \end{bmatrix} \right\} \\ &\times \left\{ 1 + \mathcal{O}\left(N^{-\frac{1}{2}}\right) \right\}. \end{aligned}$$

The empirical saddlepoint approximation uses the convexity of the objective function at each point in the parameter space: $\hat{\Sigma}_C(\alpha)$ and $\hat{\Sigma}_e(\alpha)$. The matrix $\hat{\Sigma}_e(\alpha)$ is the convexity of the objective function at the parameter values estimated using the

empirical distribution, i.e., $\widehat{\Sigma}_e(\alpha) = (A_e(\alpha)'B_e(\alpha)^{-1}A_e(\alpha))^{-1}$ where

$$A_e(\alpha) = \sum_{i=1}^N \frac{1}{N} \frac{\partial \psi(x_i, \alpha)}{\partial \alpha'} \quad \text{and} \quad B_e(\alpha) = \sum_{i=1}^N \frac{1}{N} \psi(x_i, \alpha) \psi(x_i, \alpha)'.$$

The matrix $\widehat{\Sigma}_C(\alpha)$ is the convexity of the objective function at the parameter values estimated using the conjugate distribution, i.e., $\widehat{\Sigma}_C(\alpha) = (A_C(\alpha)'B_C(\alpha)^{-1}A_C(\alpha))^{-1}$ where

$$A_C(\alpha) = \sum_{i=1}^N p_i(\alpha) \frac{\partial \psi(x_i, \alpha)}{\partial \alpha'} \quad \text{and} \quad B_C(\alpha) = \sum_{i=1}^N p_i(\alpha) \psi(x_i, \alpha) \psi(x_i, \alpha)'$$

and

$$p_i(\alpha) = \frac{\exp \{ \beta(\alpha)' \psi(x_i, \alpha) \}}{\sum_{j=1}^N \exp \{ \beta(\alpha)' \psi(x_j, \alpha) \}}.$$

In addition, the empirical saddlepoint approximation uses each of the local minima as the “mean” for the parameters connected by a path where the objective function remains positive definite. This “mean” is denoted $\alpha^*(\alpha) = \begin{bmatrix} \theta^*(\alpha) \\ \lambda^*(\alpha) \end{bmatrix}$.

The saddlepoint approximation is a natural generalization of the asymptotic normal approximation. If the objective function is quadratic then the saddlepoint approximation will be equal to the asymptotic normal approximation. However, if the objective function is not quadratic, the saddlepoint approximation will incorporate the global structure of the objective function.

If the objective function is not quadratic, but \mathcal{T} is a singleton, consistency implies that the mass of the sampling distribution converges to a shrinking neighborhood of the population parameter value. In this neighborhood the objective function converges to a quadratic. Hence, the saddlepoint approximation will converge to the familiar normal approximation.

The empirical saddlepoint approximation is the sampling distribution for the location of the local minima of the GMM objective function. This may be asymmetric and does not force the tails behavior associated with the normal approximation. A given sample may not be informative enough to distinguish the general location of the population parameter values. If the GMM objective function does not have a unique local minimum then the empirical saddlepoint approximation can have multiple modes.

5 BEHAVIOR IN FINITE SAMPLES

Monte Carlo simulations demonstrate the performance of the empirical saddlepoint distribution for the model presented in Hall and Horowitz (1996). Four different confidence intervals/tests derived from the saddlepoint approximation are compared with confidence intervals/tests created from the bootstrap, two-step GMM and the s-sets. For these simulations, the empirical saddlepoint is superior to the alternatives. Although, the bootstrap occasionally has performance comparable to the empirical saddlepoint approximation.

The model presented in Hall and Horowitz (1996) has been used in Imbens, Spady Johnson (1998), Kitamura (2001), Schennach (2007) and elsewhere. The one parameter model is estimated with the two moment conditions

$$g(\theta) = \begin{bmatrix} \exp\{\mu - \theta(X + Z) + 3Z\} - 1 \\ Z(\exp\{\mu - \theta(X + Z) + 3Z\} - 1) \end{bmatrix}$$

where $\theta_0 = 3$, X and Z are iid scalars drawn from $N(0, s^2)$ and μ is a known constant set equal to $\theta_0^2 s^2 / 2$. With one overidentifying restriction, λ is a scalar for these simulations. The simulations consider sample sizes of $N = 50$ and 100 and s set equal to .2, .3, .4, .5 and .6. The s parameter controls the noise in the system of equations. For a fixed sample size and population parameter value, larger values of s make inference more difficult. This model can possess multiple local minima, a characteristic common in moment estimation. Regularly, two local minima occur because the first moment condition does not uniquely identify θ_0 . The first moment condition has two roots: $\theta = 0$ and $\theta = \theta_0$. The second moment condition has a unique root at $\theta = \theta_0$. Random variability frequently results in a GMM objective function with two local minima.

In finite samples, the sampling distribution of the parameters of interest and the statistics that judge the validity of the overidentifying restrictions are not independent. Hence, both the marginal and conditional distributions are reported. Under the null, the marginal and conditional distributions converge to the same asymptotic distribution. However, in finite samples it is informative to report both. Asymptotically the statistics that test the validity of the overidentifying restrictions converge to zero. So the distribution of θ conditional on $\lambda = 0$ is reported. The simulations demonstrate, this may generate appealing confidence intervals.

The empirical saddlepoint density is evaluated on an evenly spaced grid of $341 \times$

101 points on $(\theta, \lambda) \in [-7, 10] \times [-1.5, 1]$. The saddlepoint equation was declared impossible to solve if its final objective function value was above .0001. The results are robust to larger lengths for the grid and a finer grid. In addition, the results are robust to the cutoff value for the saddlepoint equation taking values between .01 and .000001.

The simulation results focus on the empirical size of tests for θ_0 and the average length of the associated confidence intervals. The confidence intervals/tests considered are:

1. SP conditional short: The shortest confidence region from the empirical saddlepoint density conditional on $\lambda = 0$. This confidence region may be disjoint.
2. SP conditional sym: The symmetric confidence interval from the empirical saddlepoint density conditional on $\lambda = 0$. The confidence interval will not be disjoint.
3. SP marginal short: The shortest confidence region from the marginal empirical saddlepoint density. This confidence region may be disjoint.
4. SP marginal sym: The symmetric confidence interval from the marginal empirical saddlepoint density. The confidence interval will not be disjoint.
5. Bootstrap: The symmetrical bootstrap confidence interval. This is the procedure presented in Hall and Horowitz (1996) with 100 bootstrap samples for each simulated series.
6. GMM Wald: Wald statistic from optimal two-step GMM.
7. S-set: The s-set for the θ parameter. Stock and Wright (2000) shows how to construct confidence intervals using the objective function from the continuously updated GMM estimator. These confidence intervals are robust to weak instruments. The inversion of the objective function results in the possibility of disjoint confidence regions. Because the confidence regions may be infinite their length is not reported.

The test was performed by comparing the scaled optimal value of the continuously updated GMM objective function with the critical value for a chi-square distribution with two degrees of freedom.

The results are reported in Table 1. The empirical size is always above the nominal size. Empirical sizes and relative confidence intervals are plotted⁷ in Figure 1 for sample size $N = 50$ and in Figure 2 for sample size $N = 100$. The average lengths are plotted relative to the average length of the bootstrap confidence intervals.

Table 1 gives the empirical size and average length of the confidence intervals. Consider the results for sample size $N = 50$, when $s = .4$ and the tests were performed with nominal size 1%. The GMM Wald test confidence interval is the shortest with an average length of .64. However, these short confidence intervals did not contain the population parameter value 13.8% of the time, instead of the 1% size that the test should have achieved. The bootstrap confidence interval has an average length of 4.16 and an empirical size of 6.5%. The saddlepoint marginal symmetric confidence interval has an average length of 2.99 and has the best empirical size of 3.7%. The saddlepoint conditional confidence interval has an average length of 1.96 and an empirical size of 6.3%.

The GMM Wald test always results in the shortest confidence intervals. However, the empirical size shows these confidence intervals are unrealistically short. The local information used for the normal approximation in the Wald test does not accurately summarize the information in the objective function, i.e. the objective function is not quadratic. Because the saddlepoint and the bootstrap use global information, their sampling distributions contain information about all local minima.

For $s = .5$ and $s = .6$, the saddlepoint marginal with symmetric confidence interval is the only confidence interval with an empirical size close to the nominal size. As the noise in the system increases (i.e. as s increases) the empirical size of the other confidence intervals deteriorate. However, the saddlepoint marginal symmetric confidence intervals are robust. This is demonstrated on the top panels of Figure 1 and Figure 2. The bootstrap's poor performance is driven by differences in the covariance estimates for the bootstrap samples relative to the covariance estimate for the original sample.

For $s = .2$, $.3$ or $.4$, the empirical size of all the saddlepoint confidence intervals and the bootstrap confidence intervals are similar; no estimation procedure dominates. However, the saddlepoint conditional density with the shortest coverage region achieves dramatically shorter confidence intervals, particularly for $\alpha = .01$. Its length is always smaller, sometimes by over 50 percent. This is demonstrated on the lower

⁷The s -set is not plotted because its empirical size is comparable to the GMM Wald test's. The saddlepoint conditional symmetric is not plotted because it is almost always dominated by the saddlepoint conditional shortest. The saddlepoint marginal shortest is not plotted because it is almost always dominated by the saddlepoint marginal symmetric.

panels of Figure 1 and Figure 2.

The dramatically shorter confidence interval calculated from the conditional distribution is explained by the existence of two local minima in the GMM objective function: one near 3 and the other near 0. The marginal distribution contains two modes: one near 0 and one near 3. This bimodality results in relatively large confidence intervals. A local minima near 0 usually has an overidentifying restriction statistic value that is away from zero. The local minima near 3 usually has a statistic that tests the overidentifying restriction test statistic value closer to zero. The saddlepoint density for θ conditional on $\lambda = 0$ is typically unimodal with most of its mass near 3.

s		$N = 50$						$N = 100$					
		$\alpha = .10$.05		.01		.10		.05		.01	
		size	length										
.2	SP conditional short	.144	1.29	.078	1.57	.027	2.21	.126	.91	.072	1.10	.019	1.49
	SP conditional sym	.155	1.32	.091	1.61	.033	2.28	.137	.91	.081	1.10	.023	1.49
	SP marginal short	.156	1.46	.094	1.80	.032	2.58	.137	.98	.075	1.18	.021	1.61
	SP marginal sym	.161	1.52	.094	1.88	.031	2.70	.141	.98	.079	1.19	.022	1.62
	Bootstrap	.137	1.66	.073	2.04	.024	3.08	.137	1.04	.080	1.25	.015	1.77
	GMM Wald	.196	.64	.117	.77	.049	1.01	.155	.30	.105	.36	.037	.47
	S-Set	.167		.104		.036		.135		.077		.023	
.3	SP conditional short	.160	1.21	.108	1.48	.047	2.06	.118	.87	.069	1.04	.025	1.41
	SP conditional sym	.177	1.24	.121	1.52	.055	2.12	.131	.86	.076	1.04	.031	1.41
	SP marginal short	.161	1.50	.105	1.86	.043	2.64	.114	.98	.068	1.19	.026	1.62
	SP marginal sym	.153	1.63	.093	1.99	.035	2.81	.121	1.00	.070	1.21	.026	1.66
	Bootstrap	.146	1.64	.099	2.06	.042	3.21	.126	1.05	.071	1.28	.023	1.87
	GMM Wald	.229	.50	.167	.59	.097	.78	.170	.25	.102	.30	.045	.39
	S-Set	.199		.139		.064		.145		.097		.047	
.4	SP conditional short	.185	1.17	.126	1.42	.063	1.96	.161	.84	.109	1.01	.043	1.36
	SP conditional sym	.204	1.20	.141	1.46	.071	2.01	.180	.84	.116	1.01	.051	1.37
	SP marginal short	.180	1.53	.128	1.90	.062	2.69	.152	1.00	.102	1.22	.042	1.68
	SP marginal sym	.148	1.79	.095	2.17	.037	2.99	.151	1.14	.095	1.37	.038	1.86
	Bootstrap	.154	1.89	.112	2.52	.065	4.16	.151	1.11	.098	1.44	.050	2.45
	GMM Wald	.254	.41	.197	.49	.138	.64	.226	.22	.158	.26	.084	.34
	S-Set	.255		.201		.131		.225		.170		.099	
.5	SP conditional short	.245	1.11	.186	1.35	.114	1.84	.195	.82	.133	.99	.074	1.32
	SP conditional sym	.266	1.14	.200	1.38	.125	1.88	.209	.82	.149	.99	.079	1.33
	SP marginal short	.219	1.59	.165	1.97	.098	2.78	.172	1.07	.115	1.31	.060	1.81
	SP marginal sym	.144	2.11	.099	2.50	.045	3.31	.151	1.44	.092	1.69	.040	2.20
	Bootstrap	.203	2.17	.150	3.03	.104	5.37	.153	1.43	.113	1.98	.069	3.66
	GMM Wald	.353	.33	.288	.39	.199	.51	.254	.18	.199	.22	.122	.28
	S-Set	.364		.297		.210		.269		.221		.131	
.6	SP conditional short	.341	1.06	.283	1.29	.200	1.75	.281	.80	.212	.97	.130	1.29
	SP conditional sym	.368	1.10	.305	1.33	.215	1.78	.296	.81	.228	.97	.140	1.30
	SP marginal short	.284	1.67	.230	2.06	.149	2.88	.224	1.17	.170	1.43	.104	1.96
	SP marginal sym	.155	2.47	.109	2.88	.053	3.66	.151	1.84	.099	2.12	.051	2.64
	Bootstrap	.258	2.71	.214	3.95	.166	7.05	.207	1.82	.157	2.68	.115	5.06
	GMM Wald	.455	.27	.380	.32	.308	.43	.355	.16	.305	.19	.214	.25
	S-Set	.450		.391		.283		.372		.312		.2430	

Table 1: Empirical size and average length of different confidence intervals from 1000 simulations with sample size N . The s parameter controls the noise in the system. The population parameter value is held fixed. Higher values of s make inference more difficult. In this table α denotes the nominal size of the tests.

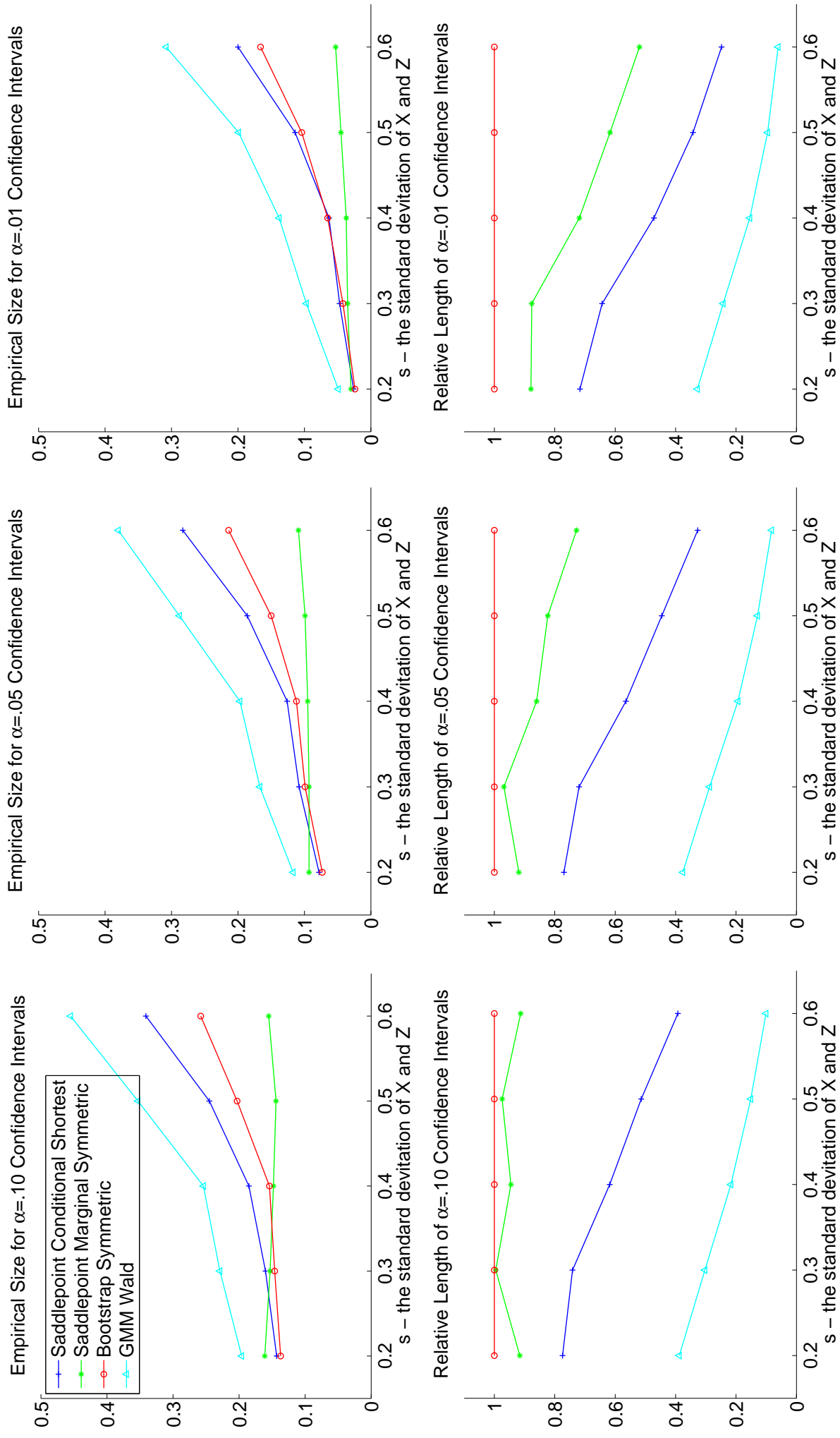


Figure 1: The empirical size and the average length for different confidence intervals relative to the average length of the bootstrap confidence interval. These graphs were created from 1000 simulated samples on $N=50$ observations.

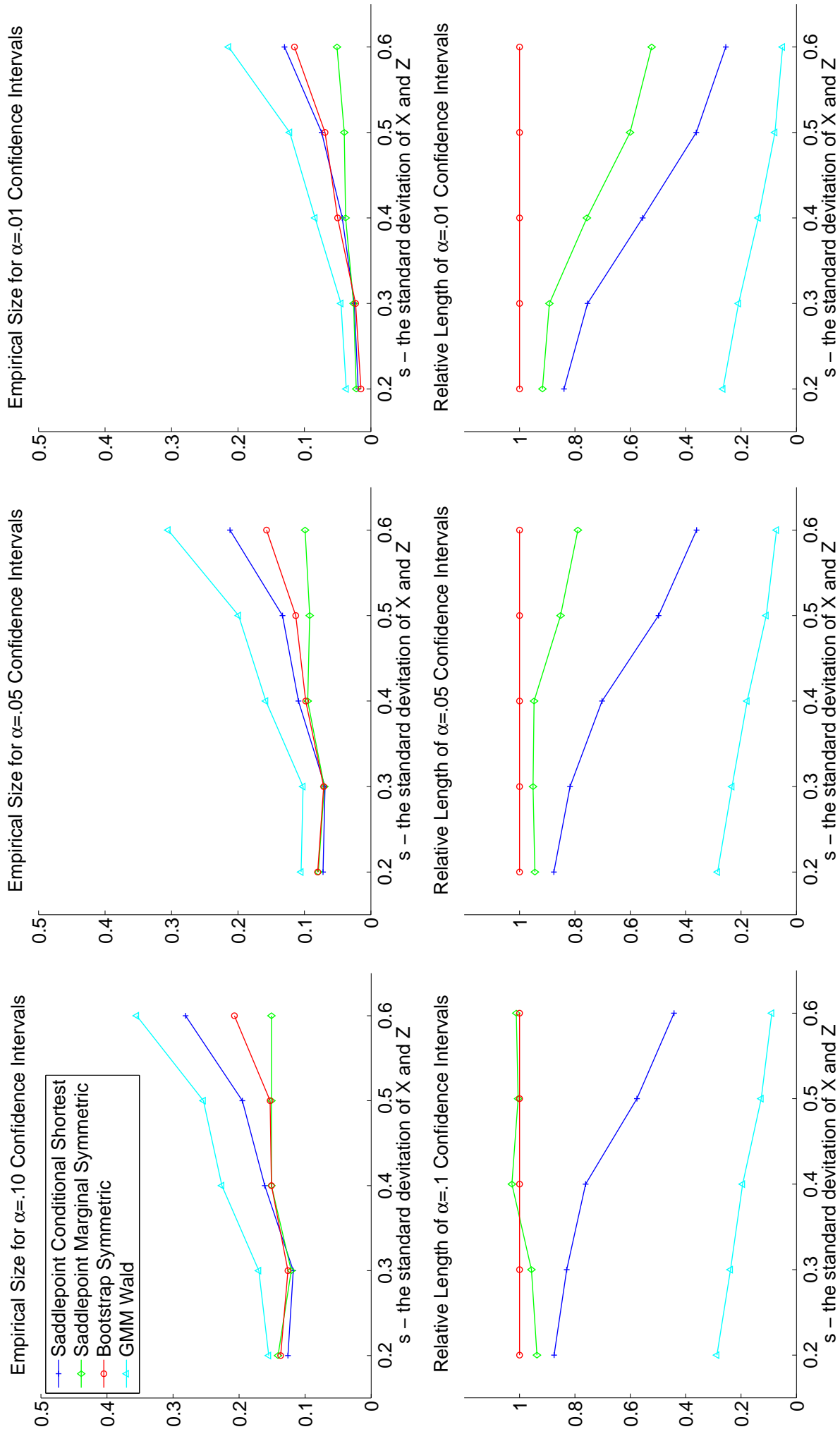


Figure 2: The empirical size and the average length for different confidence intervals relative to the average length of the bootstrap confidence interval. These graphs were created from 1000 simulated samples on $N=100$ observations.

6 ON THE SPECTRAL DECOMPOSITION

Using the spectral decomposition to determine the estimation equations raises two questions. Are the approximate densities well defined? What is its derivative with respect to θ ? The spectral decomposition is not unique. Fortunately, the approximations are invariant to the spectral decomposition selected.

Theorem 6. *The saddlepoint approximation is invariant to the spectral decomposition of the tangent spaces.*

The saddlepoint density possesses a large amount of symmetry similar to the normal density. The value of an individual λ is not well defined. However, inference is well defined in three cases (i) conditional on $\lambda = 0$, (ii) marginal, i.e. with λ integrated, and (iii) inference built on the inner product $\lambda'\lambda$, such as the J statistic.

The estimation equations are built on the matrices $C_{i,N}(\theta)$ for $i = 1, 2$ from the spectral decomposition of the projection matrix $P_{\overline{M}(\theta)}$. The proof of the asymptotic normality of the extended-GMM estimators and the empirical saddlepoint density require the differentiability of the estimation equations. This in turn requires differentiation of the spectral decomposition, $C_{i,N}(\theta)$. This is demonstrated in the next theorem. The result and method of proof are of interest in their own right.

Let $D[\bullet]$ denote taking the derivative wrt θ of the $m \times m$ matrix in the square bracket, post multiplied by an $m \times 1$ vector, say δ , e.g. $D[B(\theta)]\delta \equiv \frac{\partial(B(\theta)\delta)}{\partial\theta'}$ is an $m \times m$ matrix.

Theorem 7. *Let $\overline{M}(\theta)$ be an $m \times k$ matrix with full column rank for every value of θ . Let $C(\theta) = \begin{bmatrix} C_1(\theta) & C_2(\theta) \end{bmatrix}$ be the orthonormal basis of R^m defined by the spectral decompositions*

$$P_{\overline{M}(\theta)} = \overline{M}(\theta) \left(\overline{M}(\theta)' \overline{M}(\theta) \right)^{-1} \overline{M}(\theta)' = C_1(\theta)C_1(\theta)'$$

and

$$P_{\overline{M}(\theta)}^\perp = I_m - P_{\overline{M}(\theta)} = C_2(\theta)C_2(\theta)'$$

The derivatives of $C_i(\theta)'$, $i = 1, 2$, times the constant vector δ , with respect to θ are

$$D [C_1(\theta)'] \delta = C_1(\theta)' \overline{M}(\theta) \left(\overline{M}(\theta)' \overline{M}(\theta) \right)^{-1} D [\overline{M}(\theta)'] C_2(\theta)C_2(\theta)'\delta$$

and

$$D [C_2(\theta)'] \delta = -C_2(\theta)' D [\overline{M}(\theta)] \left(\overline{M}(\theta)' \overline{M}(\theta) \right)^{-1} \overline{M}(\theta)'\delta.$$

7 FINAL ISSUES

7.1 IMPLEMENTATION AND NUMERICAL ISSUES

1. Calculations require a fixed orthonormal basis. To be specific, $\begin{bmatrix} C_1(\theta) & C_2(\theta) \end{bmatrix}$ are determined by the Gram-Schmidt orthogonalization of

$$ME(\theta) = \begin{bmatrix} m_1(\theta) & m_2(\theta) & \dots & m_k(\theta) & e_1 & e_2 & \dots & e_{m-k} \end{bmatrix},$$

where $m_i(\theta)$ is the i^{th} column of $\overline{M}(\theta)$ and e_i is the i^{th} standard basis element,

$$C_1(\theta) = \text{diag} \left(\frac{1}{\sqrt{m_j(\theta)' P_{j-1}^\perp(\theta) m_j(\theta)}} \right) [P_{j-1}^\perp(\theta) m_j(\theta)]$$

where $P_j^\perp(\theta)$ denotes the projection onto the orthogonal complement of the space spanned by the first j columns of $ME(\theta)$, for $j = 1, 2, \dots, m$. Let $P_0^\perp = I_k$.

$C_2(\theta)$ is

$$C_2(\theta) = \text{diag} \left(\frac{1}{\sqrt{e'_{j-k} P_{j-1}^\perp(\theta) e_{j-k}}} \right) [P_{j-1}^\perp(\theta) e_{j-k}]$$

for $j = k + 1, k + 2, \dots, m$.

2. The calculation of the saddlepoint density requires repeatedly solving systems of m nonlinear equations in m unknowns. Each value in the grid over the parameter space \mathcal{A} requires the solution to a system of equations. Fortunately, the continuity of the density implies that the solutions from nearby parameter values will be excellent starting values for the nonlinear search.
3. The system of m equations in m unknowns has a structure that is similar to the structure of an exponential regression. Regardless of the structure of the original GMM moment conditions, the empirical saddlepoint equations have a simple structure that allows the analytic calculation of first and second derivatives. This results in significant time savings in the calculation of the empirical saddlepoint density.
4. For highly parameterized problems the number of calculations can become large. The saddlepoint for each value in the parameter space can be solved indepen-

dently. This structure is ideal for distributed computing with multiple processors.

7.2 EXTENSIONS

The local parameterization of the overidentifying space and the associated extended-GMM objective function have applications beyond the saddlepoint approximation. The S -sets of Stock and Wright (2000) can be applied to the estimation equations to provide confidence intervals, which include parameters that test the validity of overidentifying restrictions. Because the estimation equations are a just identified system the associated S -sets will never be empty.

When a unique local minima exists, the formulas of Lugannani and Rice (1980) can more efficiently evaluate the tail probabilities.

Recent results in saddlepoint estimation (see Robinson, Ronchetti and Young (2003) and Ronchetti and Trojani (2003)) relate to using the saddlepoint parameters $\hat{\beta}_N(\alpha)$ to test hypotheses about the parameters θ . It should be possible to extend this to include new tests for the validity of the overidentifying restrictions, i.e. use $\hat{\beta}_N(\alpha)$ to test the hypotheses $H_0 : \lambda_0 = 0$.

The saddlepoint approximations need to be extended to models with dependent data. Recent work suggests that parameter estimates using moment conditions generated from minimum divergence criteria have better small sample properties. The FOC's from these estimation problems can be used to create estimation equations suitable for saddlepoint approximations. This will replace the CLT's with the saddlepoint approximations to provide more accurate inference in small samples.

8 APPENDIX

8.1 THEOREM 1: ASYMPTOTIC DISTRIBUTION

Proof:

To reduce notation the matrices $C_{i,N}(\theta)$ dependence on sample size will be suppressed. The first order conditions for the extended GMM estimator are

$$\Psi_N(\hat{\theta}, \hat{\lambda}) = \begin{bmatrix} C_1(\hat{\theta})' W_N^{1/2} G_N(\hat{\theta}) \\ \hat{\lambda} - C_2(\hat{\theta})' W_N^{1/2} G_N(\hat{\theta}) \end{bmatrix} = 0.$$

Expand the estimation equations about the population parameter values and evaluate at the parameter estimates

$$\Psi_N(\hat{\theta}, \hat{\lambda}) = \Psi_N(\theta_0, \lambda_0) + \begin{bmatrix} \frac{\partial \Psi_N(\theta_0, \lambda_0)}{\partial \theta} & \frac{\partial \Psi_N(\theta_0, \lambda_0)}{\partial \lambda} \end{bmatrix} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\lambda} \end{bmatrix} + \mathcal{O}_p(N^{-1}).$$

The first order conditions imply the LHS will be zero. For sufficiently large N it is possible to solve for the parameters

$$\begin{aligned} \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\lambda} \end{bmatrix} &= - \begin{bmatrix} \frac{\partial \Psi_N(\theta_0, \lambda_0)}{\partial \theta} & \frac{\partial \Psi_N(\theta_0, \lambda_0)}{\partial \lambda} \end{bmatrix}^{-1} \Psi_N(\theta_0, \lambda_0) + \mathcal{O}_p(N^{-1}) \\ &= - \begin{bmatrix} \text{plim}_{N \rightarrow \infty} \frac{\partial \Psi_N(\theta_0, \lambda_0)}{\partial \theta} & \text{plim}_{N \rightarrow \infty} \frac{\partial \Psi_N(\theta_0, \lambda_0)}{\partial \lambda} \end{bmatrix}^{-1} \Psi_N(\theta_0, \lambda_0) + \mathcal{O}_p(N^{-1}). \end{aligned}$$

The θ parameters enter the first order conditions in both $G(\theta)$ and $C_i(\theta)$ for $i = 1, 2$. The product rule implies

$$\frac{\partial C_i(\theta)' W^{1/2} G_N(\theta)}{\partial \theta'} = C_i(\theta)' W^{1/2} M_N(\theta) + D[C_i(\theta)'] W^{1/2} G_N(\theta).$$

The form of $D[C_i(\theta)']$ is explicitly given in **Theorem 7**. Sufficient assumptions are made to ensure each term is finite in a neighborhood of θ_0 hence the law of large numbers implies $\text{plim}_{N \rightarrow \infty} D[C_i(\theta)'] W_N^{1/2} G_N(\theta) = 0$. Use this result and substitute terms

$$\begin{aligned}
& \begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\lambda} \end{bmatrix} \\
= & - \begin{bmatrix} C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) & 0 \\ -C_2(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) & I \end{bmatrix}^{-1} \begin{bmatrix} C_1(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \\ C_2(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \end{bmatrix} + \mathcal{O}_p(N^{-1}) \\
= & - \begin{bmatrix} \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} & 0 \\ -C_2(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} & I \end{bmatrix} \begin{bmatrix} C_1(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \\ C_2(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \end{bmatrix} + \mathcal{O}_p(N^{-1}) \\
= & \begin{bmatrix} - \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} C_1(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \\ C_2(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} C_1(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) + C_2(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \end{bmatrix} \\
& + \mathcal{O}_p(N^{-1}) \tag{1}
\end{aligned}$$

The first k elements of (1) can be written

$$\begin{aligned}
& - \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} C_1(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \\
= & - \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} C_1(\theta_0)' P_{\overline{M}(\theta_0)} \Sigma_g^{-1/2} G_N(\theta_0) \\
= & - \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} C_1(\theta_0)' \overline{M}(\theta_0) \left(\overline{M}(\theta_0)' \overline{M}(\theta_0) \right)^{-1} \overline{M}(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \\
= & - \left(C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \right)^{-1} C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0) \left(\overline{M}(\theta_0)' \overline{M}(\theta_0) \right)^{-1} \overline{M}(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \\
= & - \left(\overline{M}(\theta_0)' \overline{M}(\theta_0) \right)^{-1} \overline{M}(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0)
\end{aligned}$$

The last $(m - k)$ elements of (1) can be written

$$\begin{aligned}
& C_2(\theta_0)' \left((P_{\overline{M}^\perp} + P_{\overline{M}}) \Sigma_g^{-1/2} M(\theta_0) (C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0))^{-1} C_1(\theta_0)' - I \right) \Sigma_g^{-1/2} G_N(\theta_0) \\
& \quad + \mathcal{O}_p(N^{-1}) \\
= & C_2(\theta_0)' \left(P_{\overline{M}} \Sigma_g^{-1/2} M(\theta_0) (C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0))^{-1} C_1(\theta_0)' - I \right) \Sigma_g^{-1/2} G_N(\theta_0) + \mathcal{O}_p(N^{-1}) \\
= & C_2(\theta_0)' \left(C_1(\theta) C_1(\theta)' \Sigma_g^{-1/2} M(\theta_0) (C_1(\theta_0)' \Sigma_g^{-1/2} M(\theta_0))^{-1} C_1(\theta_0)' - I \right) \Sigma_g^{-1/2} G_N(\theta_0) \\
& \quad + \mathcal{O}_p(N^{-1}) \\
= & C_2(\theta_0)' (C_1(\theta) C_1(\theta)' - I) \Sigma_g^{-1/2} G_N(\theta_0) + \mathcal{O}_p(N^{-1}) \\
= & C_2(\theta_0)' (P_{\overline{M}} - I) \Sigma_g^{-1/2} G_N(\theta_0) + \mathcal{O}_p(N^{-1}) \\
= & -C_2(\theta_0)' P_{\overline{M}^\perp} \Sigma_g^{-1/2} G_N(\theta_0) + \mathcal{O}_p(N^{-1}) \\
= & -C_2(\theta_0)' C_2(\theta_0) C_2(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) + \mathcal{O}_p(N^{-1}) \\
= & -C_2(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) + \mathcal{O}_p(N^{-1})
\end{aligned}$$

Combining these gives

$$\begin{bmatrix} \hat{\theta} - \theta_0 \\ \hat{\lambda} \end{bmatrix} = \begin{bmatrix} -(\overline{M}(\theta_0)' \overline{M}(\theta_0))^{-1} \overline{M}(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \\ -C_2(\theta_0)' \Sigma_g^{-1/2} G_N(\theta_0) \end{bmatrix} + \mathcal{O}_p(N^{-1})$$

Multiply both sides by \sqrt{N} and use the CLT to obtain the covariance.

■

8.2 THEOREM 2: INVARIANCE OF ASYMPTOTIC DISTRIBUTION

The derivation of the estimation equations require the calculation of a spectral decomposition for the matrix projection matrix

$$P_{\overline{M}(\theta),N} = C\Lambda C' = \begin{bmatrix} C_{1,N}(\theta) & C_{2,N}(\theta) \end{bmatrix} \begin{bmatrix} I_k & 0 \\ 0 & 0_{(m-k)} \end{bmatrix} \begin{bmatrix} C_{1,N}(\theta)' \\ C_{2,N}(\theta)' \end{bmatrix}.$$

The spectral decomposition is not unique. Hence it is necessary to show that the final results are invariant to the decomposition selected.

Proof:

Theorem 1 shows that the overidentifying restriction tests are asymptotically independent from the parameters of interests. The proof can be established by showing the result separately for the parameters of interest and the statistics that test the validity of the overidentifying restrictions.

First, consider the parameters of interest. The asymptotic distribution of θ does not depend on the decomposition selected and is independent of the statistics that test the validity of the overidentifying restrictions. Hence the decomposition selected for the estimation equations does not affect the asymptotic distribution for θ .

Finally, consider the statistics that test the validity of the overidentifying restrictions. A key result is that two different decompositions are related by an $m \times m$ unitary matrix. Recall for a unitary matrix, say U , that $U'U = I_m$ and $|U| = 1$. Consider two different decompositions $C = \begin{bmatrix} C_1(\theta) & C_2(\theta) \end{bmatrix}$ and $D = \begin{bmatrix} D_1(\theta) & D_2(\theta) \end{bmatrix}$. For $P_{\overline{M}} = C_1(\theta)C_1(\theta)'$ and $P_{\overline{M}}^\perp = C_2(\theta)C_2(\theta)'$, the estimation equations are

$$\begin{bmatrix} C_1(\theta)'W^{1/2}G(\theta) \\ \lambda - C_2(\theta)'W^{1/2}G(\theta) \end{bmatrix}$$

and the parameters for the overidentifying space are $\lambda = C_2(\theta)'W^{1/2}G(\theta)$. For the diagonalization $P_{\overline{M}} = D_1(\theta)D_1(\theta)'$ and $P_{\overline{M}}^\perp = D_2(\theta)D_2(\theta)'$, the estimation equations are

$$\begin{bmatrix} D_1(\theta)'W^{1/2}G(\theta) \\ L - D_2(\theta)'W^{1/2}G(\theta) \end{bmatrix}$$

and the parameters for the overidentifying space are $L = D_2(\theta)'W^{1/2}G(\theta)$.

The parameters λ and L are linearly related

$$\begin{aligned}
\lambda &= C_2(\theta)'W^{1/2}G(\theta) \\
&= C_2(\theta)'P_M^\perp W^{1/2}G(\theta) \\
&= C_2(\theta)'D_2(\theta)D_2(\theta)'W^{1/2}G(\theta) \\
&= C_2(\theta)'D_2(\theta)L.
\end{aligned}$$

The matrix $C_2(\theta)'D_2(\theta)$ is unitary

$$\begin{aligned}
(C_2(\theta)'D_2(\theta))'C_2(\theta)'D_2(\theta) &= D_2(\theta)'C_2(\theta)C_2(\theta)'D_2(\theta) \\
&= D_2(\theta)'P_M^\perp D_2(\theta) \\
&= D_2(\theta)'D_2(\theta) \\
&= I_{m-k}.
\end{aligned}$$

To transform from the density for λ to the density for L , the Jacobian of the transformation will be

$$\begin{aligned}
d\lambda &= |C_2(\theta)'D_2(\theta)| dL \\
&= dL
\end{aligned}$$

because $C_2(\theta)'D_2(\theta)$ is unitary, its determinant is one.

The asymptotic distribution for λ is standard normal hence

$$\begin{aligned}
(2\pi)^{-\frac{(m-k)}{2}} \exp\left\{\frac{-\lambda'\lambda}{2}\right\} &= (2\pi)^{-\frac{(m-k)}{2}} \exp\left\{\frac{-\lambda'I_{m-k}\lambda}{2}\right\} \\
&= (2\pi)^{-\frac{(m-k)}{2}} \exp\left\{\frac{-\lambda'(C_2(\theta)'D_2(\theta))'C_2(\theta)'D_2(\theta)\lambda}{2}\right\} \\
&= (2\pi)^{-\frac{(m-k)}{2}} \exp\left\{\frac{-L'L}{2}\right\}
\end{aligned}$$

The last form is the asymptotic distribution for L . Hence, the same probability statements are made regardless of the decomposition. Testing the overidentifying restrictions with either $N\lambda'\lambda$ or $NL'L$ will give the same test statistic value and the same p -value from the asymptotic distribution.

■

8.3 THEOREM 3: THE SADDLEPOINT APPROXIMATION

The proof is given in Almudevar, Field and Robinson (2000), denoted ARF. The 8 assumptions made in ARF are satisfied by **Assumptions 1'-12'** for the estimation equations derived in section 3.

There are two minor changes between the current model and the model considered in ARF. The changes concern the need to restrict attention to only the solutions to the estimation equations associated with the local zeros of the original GMM objective function and addressing parameter values where the saddlepoint equation does not have a solution. In ARF the density is for all solutions to the estimation equations and hence it includes local maxima of the original GMM objective function. Using the notation in ARF, this additional restriction is satisfied by setting $\Psi^*(X, \theta)$ and $z(\theta_0, \tau)$ to infinity if $\Psi'(X, \theta)$ is not positive definite. With this change, the same steps of the proof are satisfied. The other change in the model is setting the approximate density to zero if the saddlepoint equation does not have a solution.

The model considered in this paper is built on a set of (overidentified) moment conditions. The model in AFR is a just identified system of estimation equations. The assumptions in the current paper concerning moment conditions imply that the assumptions in AFR are satisfied for the set of estimation equations derived in section 3.

A1 Assumption A1 in ARF is satisfied by **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1**. Assumption A1 in ARF is stated in terms of the estimation equation and **Assumption 2'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite. The diffeomorphism defined by

$$M(\theta^\circ)\theta \quad \text{and} \quad C_1(\theta^\circ)t$$

ensures the differentiability of the estimation equations.

A2 Assumption A2 in ARF is satisfied by **Assumption 7'**, **Assumption 5'** and **Assumption 6'.1**. Assumption A2 in ARF is stated in terms of the estimation equation and **Assumption 7'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients

in the linear functions are differentiable; hence, the density for the estimation equations can be obtained by a Jacobian transformation.

- A3 Assumption A3 in ARF is satisfied by **Assumption 5'**, **Assumption 6'.1** and **Assumption 6'.2**. Assumption A3 in ARF is stated in terms of the estimation equation and **Assumption 6'.2** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite.
- A4 Assumption A4 in ARF is satisfied by **Assumption 8'**. Assumption A4 in ARF is stated in terms of the estimation equation and **Assumption 8'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite and differentiable. Hence the restriction on the moment conditions imply the restriction on the estimation equations.
- A5 Assumption A5 in ARF is satisfied by **Assumption 9'**. Assumption A5 in ARF is stated in terms of the estimation equation and **Assumption 9'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite and invertible. Hence the restriction on the moment conditions imply the restriction on the estimation equations.
- A6 Assumption A6 in ARF is stated in terms of the estimation equation and **Assumption 10'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite and invertible.

The existence of β for the moment conditions will imply the existence of a $\tilde{\beta}$ for

the estimation equations

$$\begin{aligned}
\beta' g_i(\theta) &= \beta' W^{1/2} W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} I_m W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} \left(P_{\overline{M}(\theta)} + P_{\overline{M}(\theta)}^\perp \right) W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} P_{\overline{M}(\theta)} W^{-1/2} g_i(\theta) + \beta' W^{1/2} P_{\overline{M}(\theta)}^\perp W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} C_1(\theta) C_1(\theta)' W^{-1/2} g_i(\theta) + \beta' W^{1/2} C_2(\theta) C_2(\theta)' W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} C_1(\theta) C_1(\theta)' W^{-1/2} g_i(\theta) \\
&\quad - \beta' W^{1/2} C_2(\theta) (\lambda - C_2(\theta)' W^{-1/2} g_i(\theta)) + \beta' W^{1/2} C_2(\theta) \lambda \\
&= \beta' W^{1/2} \begin{bmatrix} C_1(\theta) & -C_2(\theta) \end{bmatrix} \psi_i(\theta) + \beta' W^{1/2} C_2(\theta) \lambda.
\end{aligned}$$

The existence of β' for the moment conditions implies $\tilde{\beta}' \equiv \beta' W^{1/2} \begin{bmatrix} C_1(\theta) & -C_2(\theta) \end{bmatrix}$ holds for the estimation equations. The term $\beta' W^{1/2} C_2(\theta) \lambda$ does not depend on x .

A7 Assumption A7 in ARF is stated in terms of the estimation equation and **Assumption 11'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite and differentiable. Hence the restriction on the moment conditions imply the restriction on the estimation equations.

A8 Assumption A8 in ARF is stated in terms of the estimation equation and **Assumption 12'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite. Hence the restriction on the moment conditions imply the restriction on the estimation equations.

■

8.4 THEOREM 4: EMPIRICAL SADDLEPOINT APPROXIMATION

The steps of the proof are given in Ronchetti and Welsh (1994), denote RW. The 5 assumptions made in RW are satisfied by **Assumptions 1'-12'** for the estimation equations presented in section 3.

There are two minor changes between the current model and the model considered in RW.

The first change is to set the saddlepoint approximation to zero if the empirical saddlepoint equation does not have a solution. This is achieved by including an indicator function for the existence of a solution to the empirical saddlepoint equation.

The second change is to allow for multiple solutions to the estimation equations. This is achieved by applying the result in RW to each solution of the estimation equation that is associated with a minimum of the original GMM objective function.

Just as with the previous proof, the model considered in this paper is built on a set of (overidentified) moment conditions. The model in RW is a just identified system of estimation equations. The assumptions in the current paper concerning moment conditions imply that the assumptions in RW are satisfied for the set of estimation equations derived in section 3.

- (a) In RW it is assumed that there is only one solution to the estimation equations. In this paper multiple solutions are allowed. The needed generalization of Assumption (a) in RW is that each local minima of the extended-GMM objective function associated with a local minima of the original GMM objective function converge at rate $N^{-1/2}$ to an element in \mathcal{T} . This “root- N consistency” is implied by **Assumption 1'**, **Assumption 2'** and **Assumption 3'**.

- (b) Assumption (b) in RW is satisfied by **Assumption 6'.3** and **Assumption 10'**. Assumption (b) in RW is stated in terms of the estimation equation as is **Assumption 10'**.

Assumption (b) in RW is stated in terms of the estimation equation and **Assumption 6'.3** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite.

The existence of β for the moment conditions will imply the existence of a $\tilde{\beta}$ for the estimation equations

$$\begin{aligned}
\beta' g_i(\theta) &= \beta' W^{1/2} W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} I_m W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} \left(P_{\overline{M}(\theta)} + P_{\overline{M}(\theta)}^\perp \right) W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} P_{\overline{M}(\theta)} W^{-1/2} g_i(\theta) + \beta' W^{1/2} P_{\overline{M}(\theta)}^\perp W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} C_1(\theta) C_1(\theta)' W^{-1/2} g_i(\theta) + \beta' W^{1/2} C_2(\theta) C_2(\theta)' W^{-1/2} g_i(\theta) \\
&= \beta' W^{1/2} C_1(\theta) C_1(\theta)' W^{-1/2} g_i(\theta) \\
&\quad - \beta' W^{1/2} C_2(\theta) (\lambda - C_2(\theta)' W^{-1/2} g_i(\theta)) + \beta' W^{1/2} C_2(\theta) \lambda \\
&= \beta' W^{1/2} \begin{bmatrix} C_1(\theta) & -C_2(\theta) \end{bmatrix} \psi_i(\theta) + \beta' W^{1/2} C_2(\theta) \lambda.
\end{aligned}$$

The existence of β' for the moment conditions implies $\tilde{\beta}' \equiv \beta' W^{1/2} \begin{bmatrix} C_1(\theta) & -C_2(\theta) \end{bmatrix}$ holds for the estimation equations. The term $\beta' W^{1/2} C_2(\theta) \lambda$ does not depend on x .

- (c) Assumption (c) in RW is satisfied by **Assumption 6'.2** and **Assumption 5'**. Assumption (c) in RW is stated in terms of the estimation equation and **Assumption 6'.2** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite.
- (d) Assumption (d) in RW is satisfied by **Assumption 2'** and **Assumption 5'**. Assumption (d) in RW is stated in terms of the estimation equation and **Assumption 2'** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment conditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite. The structure of the derivative of the coefficients wrt to the parameters are given in **Theorem 7**.
- (e) Assumption (e) in RW is satisfied by **Assumption 6'.4**, **Assumption 5'** and **Assumption 2'**. Note that Assumption (e) in RW is stated in terms of the estimation equation and **Assumption 6'.4** is stated in terms of the moment conditions. The estimation equations are linear functions of the moment con-

ditions. **Assumption 2'**, **Assumption 5'** and **Assumption 6'.1** ensure that the coefficients in the linear functions are finite. The structure of the derivative of the coefficients wrt to the parameters are given in **Theorem 7**.

■

8.5 THEOREM 5: EMPIRICAL SADDLEPOINT DENSITY STRUCTURE

Proof:

The empirical saddlepoint density can be written

$$\begin{aligned}
& K_N \times I_{\text{pos}}(Q_N(\theta)) \times I(\beta(\alpha)) \\
& \times \left(\frac{N}{2\pi} \right)^{\frac{m}{2}} \left| E_C \left[\frac{\partial \psi(\alpha)'}{\partial \alpha} \right] (E_C [\psi(\beta)\psi(\beta)'])^{-1} E_C \left[\frac{\partial \psi(\alpha)}{\partial \alpha'} \right] \right|^{-\frac{1}{2}} \exp \{N\kappa(\beta(\alpha))\} \\
= & K_N \times I_{\text{pos}}(Q_N(\theta)) \times I(\beta(\alpha)) \\
& \times \left(\frac{1}{2\pi} \right)^{\frac{m}{2}} \left| \frac{E_C \left[\frac{\partial \psi(\alpha)'}{\partial \alpha} \right] (E_C [\psi(\beta)\psi(\beta)'])^{-1} E_C \left[\frac{\partial \psi(\alpha)}{\partial \alpha'} \right]}{N} \right|^{-\frac{1}{2}} \exp \{N\kappa(\beta(\alpha))\}
\end{aligned}$$

where the expectations, E_C , are wrt to the conjugate density. However, the cumulate generating function is calculated with the distribution of the empirical distribution, i.e. E_e .

The proof will be completed by showing that

$$\begin{aligned}
& \exp \{N\kappa(\beta(\alpha))\} \\
= & \exp \left\{ -\frac{1}{2} (\alpha - \alpha_*)' \left(\frac{E_e \left[\frac{\partial \psi(\alpha)'}{\partial \alpha} \right] (E_e [\psi(\beta)\psi(\beta)'])^{-1} E_e \left[\frac{\partial \psi(\alpha)}{\partial \alpha'} \right]}{N} \right)^{-1} (\alpha - \alpha_*) \right\} \\
& \times \left\{ 1 + \mathcal{O} \left(N^{-\frac{1}{2}} \right) \right\}
\end{aligned}$$

where α^* denotes the local minima associated with α . This will be achieved by showing that the cumulant generating function can be written

$$\kappa(\beta(\alpha)) = -\frac{1}{2} (\alpha - \alpha_*)' \left(E_e \left[\frac{\partial \psi(\alpha)'}{\partial \alpha} \right] (E_e [\psi(\beta)\psi(\beta)'])^{-1} E_e \left[\frac{\partial \psi(\alpha)}{\partial \alpha'} \right] \right)^{-1} (\alpha - \alpha_*) + \mathcal{O} \left(N^{-\frac{3}{2}} \right).$$

The cumulate generating function for $\psi(\alpha)$ is

$$\begin{aligned}
\kappa(\beta) &= \ln (E_e [\exp \{\beta' \psi(\alpha)\}]) \\
&= \ln (M(\beta))
\end{aligned}$$

where $M(\beta)$ is the moment generating function of $\psi(\alpha)$ under the distribution of the observed data. Expand this in a three term Taylor series about $\beta = \beta(\alpha_*) = 0$. Note

that

$$\begin{aligned}
\kappa(\beta)|_{\beta=0} &= \ln(M(0)) \\
&= 0 \\
\kappa(\beta)'|_{\beta=0} &= \frac{M'(0)}{M(0)} \\
&= E_e[\psi(\alpha)] \\
\kappa(\beta)''|_{\beta=0} &= \frac{M''(0)M(0) - M'(0)M'(0)^T}{M(0)^2} \\
&= E_e[\psi(\alpha)\psi(\alpha)'] - E_e[\psi(\alpha)]E_e[\psi(\alpha)]'.
\end{aligned}$$

The expansion can be written

$$\kappa(\beta) = \beta' E_e[\psi(\alpha)] + \frac{1}{2} \beta' \left[E_e[\psi(\alpha)\psi(\alpha)'] - E_e[\psi(\alpha)]E_e[\psi(\alpha)]' \right] \beta + \mathcal{O}(|\beta|^3).$$

Linearize the saddlepoint equation about $\beta = 0$. Recall the expectation is wrt the distribution of the observed data

$$E_e[\psi(\alpha) \exp\{\beta' \psi(\alpha)\}] = E_e[\psi(\alpha)] + E_e[\psi(\alpha)\psi(\alpha)'] \beta + \mathcal{O}(|\beta|^2).$$

Evaluate this at $\beta(\alpha)$, solve for

$$\beta(\alpha) = - (E_e[\psi(\alpha)\psi(\alpha)'])^{-1} E_e[\psi(\alpha)] + \mathcal{O}(|\beta(\alpha)|^2)$$

and substitute into the expansion

$$\begin{aligned}
\kappa(\beta(\alpha)) &= -E_e[\psi(\alpha)]' (E_e[\psi(\alpha)\psi(\alpha)'])^{-1} E_e[\psi(\alpha)] \\
&\quad + \frac{1}{2} E_e[\psi(\alpha)]' (E_e[\psi(\alpha)\psi(\alpha)'])^{-1} E_e[\psi(\alpha)] \\
&\quad - \frac{1}{2} \left(E_e[\psi(\alpha)]' (E_e[\psi(\alpha)\psi(\alpha)'])^{-1} E_e[\psi(\alpha)] \right)^2 \\
&\quad + \mathcal{O}(|\beta|^3) + \mathcal{O}(|\beta(\alpha)|^2 |E_e[\psi(\alpha)]|).
\end{aligned}$$

Expand the estimation equation about α and evaluate the expansion at the closest local minimum that is connected by a path that is always positive definite⁸, denote

⁸If such a path does not exist, set the approximate density to zero.

this local minimum by α_*

$$\psi(\alpha_*) = \psi(\alpha) + \frac{\partial\psi(\alpha)}{\partial\alpha'} (\alpha_* - \alpha) + \mathcal{O}(|\alpha - \alpha_*|^2).$$

Because α_* is a local minimum, the estimation equations will be satisfied. This implies

$$\psi(\alpha) = -\frac{\partial\psi(\alpha)}{\partial\alpha'} (\alpha_* - \alpha) + \mathcal{O}(|\alpha - \alpha_*|^2).$$

Take the expectation wrt the distribution of the empirical distribution

$$E_e[\psi(\alpha)] = -E_e\left[\frac{\partial\psi(\alpha)}{\partial\alpha'}\right] (\alpha_* - \alpha) + \mathcal{O}(|\alpha - \alpha_*|^2)$$

and substitute into the expansion

$$\begin{aligned} \kappa(\beta(\alpha)) &= -\frac{1}{2} (\alpha_* - \alpha)' E_e \left[\frac{\partial\psi(\alpha)'}{\partial\alpha} \right] (E_e[\psi(\alpha)\psi(\alpha)'])^{-1} E_e \left[\frac{\partial\psi(\alpha)}{\partial\alpha'} \right] (\alpha_* - \alpha) \\ &\quad + \mathcal{O}(|\beta|^3) + \mathcal{O}(|\beta(\alpha)|^2 |\alpha - \alpha_*|) + \mathcal{O}(|\alpha - \alpha_*|^3). \end{aligned}$$

Because attention is restricted to a \sqrt{N} neighborhood of α_* and because $\beta(\alpha)$ is continuous with $\beta(\alpha_*) = 0$, all the error terms can be collected into a single $\mathcal{O}\left(N^{-\frac{3}{2}}\right)$ term. As was to be shown.

■

8.6 THEOREM 6: INVARIANCE OF SADDLEPOINT APPROXIMATION

Use the same notation introduced in **Theorem 2**. The solution to the saddlepoint equation for the decomposition $D(\theta)$ is written in terms of the solution to the saddlepoint equation for the decomposition $C(\theta)$. This is used to show that the different terms that compose the saddlepoint density give the same value for both decompositions.

Proof:

The saddlepoint equation for the decomposition $C(\theta) = \begin{bmatrix} C_1(\theta) & C_2(\theta) \end{bmatrix}$ and λ is

$$\begin{aligned}
0 &= \int \begin{bmatrix} C_1(\theta)'W^{1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \exp \left\{ \frac{\beta'(\alpha)}{N} \begin{bmatrix} C_1(\theta)'W^{1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \right\} dF(x) \\
&= \int \begin{bmatrix} (D_1(\theta)'C_1(\theta))' D_1(\theta)'C_1(\theta) & 0 \\ 0 & (D_2(\theta)'C_2(\theta))' D_2(\theta)'C_2(\theta) \end{bmatrix} \begin{bmatrix} C_1(\theta)'W^{1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \\
&\quad \times \exp \left\{ \frac{\beta'(\alpha)}{N} \begin{bmatrix} (D_1(\theta)'C_1(\theta))' D_1(\theta)'C_1(\theta) & 0 \\ 0 & (D_2(\theta)'C_2(\theta))' D_2(\theta)'C_2(\theta) \end{bmatrix} \right. \\
&\quad \quad \left. \times \begin{bmatrix} C_1(\theta)'W^{1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \right\} dF(x) \\
&= \int \begin{bmatrix} C_1(\theta)'D_1(\theta) & 0 \\ 0 & C_2(\theta)'D_2(\theta) \end{bmatrix} \begin{bmatrix} D_1(\theta)'C_1(\theta)C_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'C_2(\theta)C_2(\theta)'W^{1/2}g_i(\theta) - D_2(\theta)'C_2(\theta)\lambda \end{bmatrix} \\
&\quad \times \exp \left\{ \frac{\beta(\alpha)'}{N} \begin{bmatrix} C_1(\theta)'D_1(\theta) & 0 \\ 0 & C_2(\theta)'D_2(\theta) \end{bmatrix} \right. \\
&\quad \quad \left. \times \begin{bmatrix} D_1(\theta)'C_1(\theta)C_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'C_2(\theta)C_2(\theta)'W^{1/2}g_i(\theta) - D_2(\theta)'C_2(\theta)\lambda \end{bmatrix} \right\} dF(x) \\
&= \begin{bmatrix} C_1(\theta)'D_1(\theta) & 0 \\ 0 & C_2(\theta)'D_2(\theta) \end{bmatrix} \int \begin{bmatrix} D_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'W^{1/2}g_i(\theta) - L \end{bmatrix} \\
&\quad \times \exp \left\{ \frac{\tilde{\beta}(\alpha)'}{N} \begin{bmatrix} D_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'W^{1/2}g_i(\theta) - L \end{bmatrix} \right\} dF(x)
\end{aligned}$$

where

$$\tilde{\beta}(\alpha)' \equiv \beta(\alpha)' \begin{bmatrix} C_1(\theta)'D_1(\theta) & 0 \\ 0 & C_2(\theta)'D_2(\theta) \end{bmatrix}.$$

The leading matrix is of full rank and hence the final equation implies

$$0 = \int \left[\begin{array}{c} D_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'W^{1/2}g_i(\theta) - L \end{array} \right] \exp \left\{ \frac{\tilde{\beta}(\alpha)'}{N} \left[\begin{array}{c} D_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'W^{1/2}g_i(\theta) - L \end{array} \right] \right\} dF(x).$$

So $\tilde{\beta}(\theta)$ will be the solution to the saddlepoint equation for the decomposition $D(\theta)$ with parameter L . Use a superscript to denote the different spectral decompositions. Then, it is possible to show the invariance of the conjugate density wrt the selection of the spectral decomposition

$$h_{N,\beta}^{C(\theta)}(x) = \frac{\exp \left\{ \frac{\beta(\alpha)'}{N} \psi^{C(\theta)}(x, \alpha) \right\} dF(x)}{\int \exp \left\{ \frac{\beta(\alpha)'}{N} \psi^{C(\theta)}(w, \alpha) \right\} dF(w)} = \frac{\exp \left\{ \frac{\tilde{\beta}(\alpha)'}{N} \psi^{D(\theta)}(x, \alpha) \right\} dF(x)}{\int \exp \left\{ \frac{\tilde{\beta}(\alpha)'}{N} \psi^{D(\theta)}(w, \alpha) \right\} dF(w)} = h_{N,\tilde{\beta}}^{D(\theta)}(x).$$

The saddlepoint approximation for $C(\theta)$ and λ with $\alpha = \left[\begin{array}{cc} \theta' & \lambda' \end{array} \right]'$

$$f_N(\alpha) = \left(\frac{1}{2\pi} \right)^{\frac{m}{2}} |E [\psi^{C(\theta)}(\alpha) \psi^{C(\theta)}(\alpha)']|^{-\frac{1}{2}} \left| E \left[\frac{\partial \psi^{C(\theta)}(\alpha)}{\partial \alpha'} \right] \right| \exp \{ N \kappa(\beta(\alpha), \alpha) \}$$

is composed of three terms, where $\beta(\alpha)$ is the solution to the saddlepoint equation and the expectation is wrt the conjugate density $h_{N,\beta}^{C(\theta)}(x)$.

The approximation depends on three terms that contain $C_1(\theta)$, $C_2(\theta)$ and/or λ . The proof will be completed once it is shown that each of the three terms is equal to the value if $D_1(\theta)$, $D_2(\theta)$ and L had been used.

1. Consider the term $\kappa(\beta(\alpha), \alpha)$.

$$\begin{aligned}
& \kappa(\beta(\alpha), \alpha) \\
&= \int \exp \left\{ \frac{\beta'(\alpha)}{N} \begin{bmatrix} C_1(\theta)'W^{1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \right\} dF(x) \\
&= \int \exp \left\{ \frac{\beta'(\alpha)}{N} \begin{bmatrix} (D_1(\theta)'C_1(\theta))' D_1(\theta)'C_1(\theta) & 0 \\ 0 & (D_2(\theta)'C_2(\theta))' D_2(\theta)'C_2(\theta) \end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix} C_1(\theta)'W^{1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \right\} dF(x) \\
&= \int \exp \left\{ \frac{\beta(\alpha)'}{N} \begin{bmatrix} C_1(\theta)'D_1(\theta) & 0 \\ 0 & C_2(\theta)'D_2(\theta) \end{bmatrix} \right. \\
&\quad \left. \times \begin{bmatrix} D_1(\theta)'C_1(\theta)C_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'C_2(\theta)C_2(\theta)'W^{1/2}g_i(\theta) - D_2(\theta)'C_2(\theta)\lambda \end{bmatrix} \right\} dF(x) \\
&= \exp \left\{ \frac{\tilde{\beta}(\alpha)'}{N} \begin{bmatrix} D_1(\theta)'W^{1/2}g_i(\theta) \\ D_2(\theta)'W^{1/2}g_i(\theta) - L \end{bmatrix} \right\} dF(x) \\
&= \kappa(\tilde{\beta}(\alpha), \alpha).
\end{aligned}$$

Hence this term is invariant to the decomposition selected.

2. Consider the term $\left| E \left[\frac{\partial \psi^{C(\theta)}(\alpha)}{\partial \alpha'} \right] \right|$. To reduce notation let U denote the unitary matrix

$$U = \begin{bmatrix} C_1(\theta)'D_1(\theta) & 0 \\ 0 & C_2(\theta)'D_2(\theta) \end{bmatrix}$$

and hence $|U| = 1$. Use **Theorem 7** to write

$$\begin{aligned}
& \left| E_\beta \left[\frac{\partial \psi^{C(\theta)}(\alpha)}{\partial \alpha'} \right] \right| \\
&= \left| \int \begin{bmatrix} C_1(\theta)' \overline{M}_i(\theta) + C_1(\theta)' \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D[\overline{M}(\theta)'] P_{\overline{M}(\theta)}^\perp W^{1/2} g_i(\theta) & 0 \\ C_2(\theta)' \overline{M}_i(\theta) - C_2(\theta)' D[\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' W^{1/2} g_i(\theta) & -I_{m-k} \end{bmatrix} \right. \\
&\quad \left. \times h_{N,\beta}^{C(\theta)}(x) dx \right| \\
&= \left| \int \begin{bmatrix} C_1(\theta)' \overline{M}_i(\theta) + C_1(\theta)' \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D[\overline{M}(\theta)'] P_{\overline{M}(\theta)}^\perp W^{1/2} g_i(\theta) & 0 \\ C_2(\theta)' \overline{M}_i(\theta) - C_2(\theta)' D[\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' W^{1/2} g_i(\theta) & -I_{m-k} \end{bmatrix} \right. \\
&\quad \left. \times h_{N,\tilde{\beta}}^{D(\theta)}(x) dx \right| \\
&= |U| \left| \int \begin{bmatrix} D_1(\theta)' \overline{M}_i(\theta) + D_1(\theta)' \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D[\overline{M}(\theta)'] P_{\overline{M}(\theta)}^\perp W^{1/2} g_i(\theta) & 0 \\ D_2(\theta)' \overline{M}_i(\theta) - D_2(\theta)' D[\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' W^{1/2} g_i(\theta) & -D_2(\theta)' C_2(\theta) \end{bmatrix} \right. \\
&\quad \left. \times h_{N,\tilde{\beta}}^{D(\theta)}(x) dx \right| \\
&= \left| \int \begin{bmatrix} D_1(\theta)' \overline{M}_i(\theta) + D_1(\theta)' \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D[\overline{M}(\theta)'] P_{\overline{M}(\theta)}^\perp W^{1/2} g_i(\theta) & 0 \\ D_2(\theta)' \overline{M}_i(\theta) - D_2(\theta)' D[\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' W^{1/2} g_i(\theta) & -I_{m-k} \end{bmatrix} \right. \\
&\quad \left. \times h_{N,\tilde{\beta}}^{D(\theta)}(x) dx \right| \\
&= \left| E_{\tilde{\beta}} \left[\frac{\partial \psi^{D(\theta)}(\alpha)}{\partial \alpha'} \right] \right|.
\end{aligned}$$

The first and last equalities come from the definitions of $E_\beta \left[\frac{\partial \psi^{C(\theta)}(\alpha)}{\partial \alpha'} \right]$, $E_{\tilde{\beta}} \left[\frac{\partial \psi^{D(\theta)}(\alpha)}{\partial \alpha'} \right]$ and **Theorem 7**. The second equality is justified with the same steps used above to derive $\tilde{\beta}(\alpha)$. The third equality is from the definition of U . The fourth equality occurs because U and $D_2(\theta)' C_2(\theta)$ are unitary and hence $|U| = |D_2(\theta)' C_2(\theta)| = 1$. Note that the upper right matrix being zero implies that the determinant will be the product of the the determinant of the matrices on the main diagonal.

Hence this term is invariant to the decomposition selected.

3. Finally consider the term $|E_\beta [\psi^{C(\theta)}(\alpha)\psi^{C(\theta)}(\alpha)']|$

$$\begin{aligned}
& E_\beta [\psi^{C(\theta)}(\alpha)\psi^{C(\theta)}(\alpha)'] \\
&= \int \begin{bmatrix} C_1(\theta)'W^{-1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \begin{bmatrix} C_1(\theta)'W^{-1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix}' h_{N,\beta}^{C(\theta)}(x) dx \\
&= \int \begin{bmatrix} C_1(\theta)'W^{-1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix} \begin{bmatrix} C_1(\theta)'W^{-1/2}g_i(\theta) \\ C_2(\theta)'W^{1/2}g_i(\theta) - \lambda \end{bmatrix}' h_{N,\tilde{\beta}}^{D(\theta)}(x) dx \\
&= U \int \begin{bmatrix} D_1(\theta)'W^{-1/2}g_i(\theta) \\ D_2(\theta)'W^{1/2}g_i(\theta) - L \end{bmatrix} \begin{bmatrix} D_1(\theta)'W^{-1/2}g_i(\theta) \\ D_2(\theta)'W^{1/2}g_i(\theta) - L \end{bmatrix}' h_{N,\tilde{\beta}}^{D(\theta)}(x) dx U' \\
&= U E_{\tilde{\beta}} [\psi^{D(\theta)}(\alpha)\psi^{D(\theta)}(\alpha)'] U'
\end{aligned}$$

The first and last equalities are from the definitions of $E_\beta [\psi^{C(\theta)}(\alpha)\psi^{C(\theta)}(\alpha)']$ and $E_{\tilde{\beta}} [\psi^{D(\theta)}(\alpha)\psi^{D(\theta)}(\alpha)']$. The second equality is from the same steps that were used to define $\tilde{\beta}(\alpha)$ and the equality $\kappa(\beta(\alpha), \alpha) = \kappa(\tilde{\beta}(\alpha), \alpha)$. The third equality is from the definition of U . Since, U is a unitary matrix, $|U| = 1$. So $|E_\beta [\psi^{C(\theta)}(\alpha)\psi^{C(\theta)}(\alpha)']| = |E_{\tilde{\beta}} [\psi^{D(\theta)}(\alpha)\psi^{D(\theta)}(\alpha)']|$. Hence the value is invariant to the decomposition selected.

The saddlepoint density is invariant to the decomposition selected.

■

8.7 THEOREM 7: DERIVATIVE OF THE DECOMPOSITION

The sufficient conditions for the existence of the saddlepoint density require the derivative of the estimation equations with respect to the parameters. This requires the derivative with respect to θ of the spectral decomposition, $C_i(\theta)$ for $i = 1, 2$. Recall $D[\bullet]$ denotes taking the derivative wrt θ of the $m \times m$ matrix in the square bracket that is post multiplied by an $m \times 1$ vector, say δ , where δ does not depend on θ .

Proof:

For $C_2(\theta)'$ the proof is demonstrated by evaluating $D \left[P_{\overline{M}(\theta)}^\perp \right] \delta$ two different ways. These are then set equal to determine $D [C_2(\theta)] \delta$.

Consider

$$\begin{aligned} D[P_{\overline{M}(\theta)}^\perp] \delta &= D[C_2(\theta)C_2(\theta)'] \delta \\ &= C_2(\theta)D[C_2(\theta)'] \delta + D[C_2(\theta)]C_2(\theta)' \delta. \end{aligned} \quad (2)$$

Now consider the same derivative written as

$$\begin{aligned} D[P_{\overline{M}(\theta)}^\perp] \delta &= D[I - \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)'] \delta \\ &= -D [\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' \delta \\ &\quad - \overline{M}(\theta) D \left[(\overline{M}(\theta)' \overline{M}(\theta))^{-1} \right] \overline{M}(\theta)' \delta \\ &\quad - \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D [\overline{M}(\theta)'] \delta \\ &= -D [\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' \delta \\ &\quad + \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' D [\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' \delta \\ &\quad + \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D [\overline{M}(\theta)'] \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' \delta \\ &\quad - \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D [\overline{M}(\theta)'] \delta \\ &= -P_{\overline{M}(\theta)}^\perp D [\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' \delta \\ &\quad - \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D [\overline{M}(\theta)'] P_{\overline{M}(\theta)}^\perp \delta \\ &= -C_2(\theta)C_2(\theta)' D [\overline{M}(\theta)] (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)' \delta \\ &\quad - \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} D [\overline{M}(\theta)'] C_2(\theta)C_2(\theta)' \delta. \end{aligned} \quad (3)$$

Comparing equations (2) and (3) gives the result

$$\begin{aligned}
D[C_2(\theta)']\delta &= -C_2(\theta)'D[\overline{M}(\theta)](\overline{M}(\theta)'\overline{M}(\theta))^{-1}\overline{M}(\theta)'\delta \\
&= -C_2(\theta)'D[\overline{M}(\theta)](\overline{M}(\theta)'\overline{M}(\theta))^{-1}\overline{M}(\theta)'P_{\overline{M}}\delta \\
&= -C_2(\theta)'D[\overline{M}(\theta)](\overline{M}(\theta)'\overline{M}(\theta))^{-1}\overline{M}(\theta)'C_1(\theta)C_1(\theta)'\delta.
\end{aligned}$$

Now consider $C_1(\theta)'$. The derivative of the projection matrix gives

$$\begin{aligned}
D[P_{\overline{M}(\theta)}]\delta &= D[C_1(\theta)C_1(\theta)']\delta \\
&= C_1(\theta)D[C_1(\theta)']\delta + D[C_1(\theta)]C_1(\theta)'\delta.
\end{aligned} \tag{4}$$

For all θ

$$P_{\overline{M}(\theta)} + P_{\overline{M}(\theta)}^\perp = I, \quad P_{\overline{M}(\theta)}\delta + P_{\overline{M}(\theta)}^\perp\delta = \delta$$

and hence

$$D[P_{\overline{M}(\theta)}]\delta + D[P_{\overline{M}(\theta)}^\perp]\delta = 0.$$

This implies

$$D[P_{\overline{M}(\theta)}]\delta = -D[P_{\overline{M}(\theta)}^\perp]\delta.$$

Substitute from equations (3)

$$\begin{aligned}
D[P_{\overline{M}(\theta)}]\delta &= C_2(\theta)C_2(\theta)'D[\overline{M}(\theta)](\overline{M}(\theta)'\overline{M}(\theta))^{-1}\overline{M}(\theta)'\delta \\
&\quad + \overline{M}(\theta)(\overline{M}(\theta)'\overline{M}(\theta))^{-1}D[\overline{M}(\theta)']C_2(\theta)C_2(\theta)'\delta \\
&= C_2(\theta)C_2(\theta)'D[\overline{M}(\theta)](\overline{M}(\theta)'\overline{M}(\theta))^{-1}\overline{M}(\theta)'P_{\overline{M}}\delta \\
&\quad + P_{\overline{M}}\overline{M}(\theta)(\overline{M}(\theta)'\overline{M}(\theta))^{-1}D[\overline{M}(\theta)']C_2(\theta)C_2(\theta)'\delta \\
&= C_2(\theta)C_2(\theta)'D[\overline{M}(\theta)](\overline{M}(\theta)'\overline{M}(\theta))^{-1}\overline{M}(\theta)'C_1(\theta)C_1(\theta)'\delta \\
&\quad + C_1(\theta)C_1(\theta)'\overline{M}(\theta)(\overline{M}(\theta)'\overline{M}(\theta))^{-1}D[\overline{M}(\theta)']C_2(\theta)C_2(\theta)'\delta.
\end{aligned} \tag{5}$$

Comparing equations (4) and (5) gives the result

$$\begin{aligned}
D[C_1(\theta)']\delta &= C_1(\theta)'\overline{M}(\theta)(\overline{M}(\theta)'\overline{M}(\theta))^{-1}D[\overline{M}(\theta)']C_2(\theta)C_2(\theta)'\delta \\
&= C_1(\theta)'\overline{M}(\theta)(\overline{M}(\theta)'\overline{M}(\theta))^{-1}D[\overline{M}(\theta)']P_{\overline{M}(\theta)}^\perp\delta.
\end{aligned}$$

■

8.8 A DIFFERENTIAL GEOMETRY PERSPECTIVE

$W_N^{1/2}G_N(\theta)$ is a k -dimensional manifold in \mathbb{R}^m . The m -dimensional tangent space for each value in the parameter space can be decomposed into the k -dimensional space spanned by $\overline{M}(\theta)$ and its $(m - k)$ -dimensional orthogonal complement. The orthogonal complement spans an orthonormal basis, denoted $C_2(\theta)$, with coefficient $\lambda \in \mathbb{R}^{(m-k)}$.

Let the columns of $C_1(\theta)$ be an orthonormal basis spanning the k -dimensional tangent space spanned by the columns of $\overline{M}(\theta)$. The m columns of $\begin{bmatrix} C_1(\theta) & C_2(\theta) \end{bmatrix}$ are an orthonormal basis for each value of θ . Typically, there are a continuum of orthonormal bases, each equally valid.

For each parameter value, θ° , a linear approximation to the limiting function $E \left[\Sigma_g^{-1/2} g(x_i, \theta) \right]$ is defined by

$$E \left[\Sigma_g^{-1/2} g(x_i, \theta) \right] \approx E \left[\Sigma_g^{-1/2} g(x_i, \theta^\circ) \right] + \overline{M}(\theta^\circ) (\theta - \theta^\circ).$$

A reparameterization of the tangent space can be defined by an orthonormal basis of the space spanned by the columns of $\overline{M}(\theta^\circ)$. Let $C_1(\theta^\circ)$ define such an orthonormal basis. Then, the new parameterization can be written

$$t = C_1(\theta^\circ)' \overline{M}(\theta^\circ) \theta.$$

The inverse of the reparameterization is given by

$$\theta = (\overline{M}(\theta^\circ)' \overline{M}(\theta^\circ))^{-1} \overline{M}(\theta^\circ)' C_1(\theta^\circ) t.$$

This gives a local diffeomorphism between θ and t . $E \left[\Sigma_g^{-1/2} g(x_i, \theta) \right]$ is an element of \mathbb{R}^m but θ and t are elements of \mathbb{R}^k the remaining dimensions $(m - k)$ -dimensions are spanned by the orthonormal basis $C_2(\theta^\circ)$.

At each value of θ the estimation equations have a natural interpretation. The parameter θ is selected to set the normalized moment conditions projected onto $C_1(\theta)$ to zero. The parameters λ are the normalized moment conditions projected onto $C_2(\theta)$.

8.9 STEPS TO A SADDLEPOINT APPROXIMATION

The steps to calculate the saddlepoint approximation for the θ parameters and the λ statistics are listed below. The terms' dependence on N will be suppressed in this section.

1. Calculate an optimal weighting matrix using the original GMM objective function.
 - (a) Calculate the first step GMM estimate using an identity as the weighting matrix.
 - (b) Calculate an estimate of the moments' covariance matrix at the first step GMM estimate, \widehat{W} .
 - (c) Calculate a Cholesky decomposition of the estimate, $\widehat{W}^{1/2}$.
2. Over \mathcal{A} , the parameter space for $\alpha = [\theta' \lambda']'$, select a grid on which the approximation will be calculated.
3. Loop through the grid over the θ parameters.
 - (a) Determine the spectral decomposition $C_i(\theta)$ for $i = 1, 2$.
 - i. Calculate $M(\theta)$.
 - ii. Calculate $\overline{M}(\theta) = \widehat{W}^{1/2} M(\theta)$.
 - iii. Calculate $P_{\overline{M}(\theta)} = \overline{M}(\theta) (\overline{M}(\theta)' \overline{M}(\theta))^{-1} \overline{M}(\theta)'$.
 - iv. Calculate $C_1(\theta)$.
 - v. Calculate $P_{\overline{M}(\theta)}^\perp = I - P_{\overline{M}(\theta)}$.
 - vi. Calculate $C_2(\theta)$.
 - (b) Loop through the grid over the λ parameters.
 - i. Calculate the estimation equation

$$\Psi_N(\alpha) = \begin{bmatrix} C_1(\theta)' \widehat{W}^{1/2} G_N(\theta) \\ \lambda - C_2(\theta)' \widehat{W}^{1/2} G_N(\theta) \end{bmatrix}.$$
 - ii. Solve the saddlepoint equation to determine $\beta(\alpha)$.
 - iii. Evaluate $A_e(\alpha)$, $B_e(\alpha)$, $A_C(\alpha)$ and $B_C(\alpha)$.

- iv. Evaluate the saddlepoint approximation for this value in the parameter space.
 - v. End the loop over λ .
- (c) End the loop over θ .
4. Scale the estimated density so that it integrates to one.

9 REFERENCES

1. Almudevar, Anthony, Chris Field and John Robinson (2000) "The Density of Multivariate M-Estimates" *The Annals of Statistics* vol. 28, no. 1, pp. 275-297.
2. Amari, Shu-ichi (1985) *Differential-Geometrical Methods in Statistics* Berlin Heidelberg Springer-Verlag, Lecture Notes in Statistics, 28.
3. Anatolyev, Stanislav (2005) "GMM, GEL, Serial Correlation, and Asymptotic Bias," *Econometrica*, vol. 73(3), pp. 983-1002.
4. Bhattacharya, R. N. and J. K. Ghosh (1978) "On the Validity of the Formal Edgeworth Expansion" *Annals of Statistics* vol 6, pp. 434-451.
5. Boothby, William M. (2003) *An Introduction to Differentiable Manifolds and Riemannian Geometry* revised second edition, Academic Press.
6. Daniels, H.E. (1954) "Saddlepoint approximations in statistics" *Annals of Mathematical Statistics* vol. 25, pp. 631-650.
7. Dominguez, M. A., and I. N. Lobato (2004) "Consistent Estimation of Models Defined by Conditional Moment Restrictions," *Econometrica*, vol. 72, pp. 1601-1615.
8. Esscher, F. (1932) "The probability function in the collective theory of risk" *Skand. Aktuarietidskr. Tidskr.*, vol. 15, pp. 175-195.
9. Feuerverger, Andrey (1989) "On the Empirical Saddlepoint Approximation" *Biometrika* vol. 76, pp. 457-464.
10. Field, C. A.(1982) "Small Sample Asymptotics for Multivariate M-Estimates" *Annals of Statistics*, vol. 10, pp. 672-689.
11. Field, C. A. and E. Ronchetti (1990) *Small Sample Asymptotics* Hayward, CA. IMS Monograph Series, vol 13.
12. Goutis, Constantin and George Casella (1999) "Explaining the Saddlepoint Approximation" *The American Statistician* vol. 53, no. 3. pp. 216-224.
13. Hall, Peter and Joel L. Horowitz (1996) "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators" *Econometrica* vol 64, no. 4, pp. 891-916.
14. Hansen, L.P. (1982) "Large Sample Properties of Generalized Method of Moments Estimators" *Econometrica* vol. 50.
15. Hansen, L.-P., John Heaton, and A. Yaron (1996) "Finite Sample Properties of Some Alternative GMM Estimators", *Journal of Business and Economic Statistics*, vol. 14 no.3.

16. Huzurbazar, S. (1999) "Practical Saddlepoint Approximations" *The American Statistician*, vol. 53, no. 3, pp. 225-232.
17. Jensen, J. L. (1995) *Saddlepoint Approximations*, Oxford University Press, Oxford.
18. Kitamura, Yuichi and Michael Stutzer (1997) "An Information-Theoretic Alternative to Generalized Method of Moments Estimation" *Econometrica* vol. 65, no. 4, pp. 861-874.
19. Kolassa, J.E. (1997) *Series Approximation Methods in Statistics* 2nd edition, New York, Springer-Verlag, Lecture Notes in Statistics, 88.
20. Lugannani, R. and S. Rice (1980) "Saddle Point Approximations for the Distribution of the Sum of Independent Random Variables" *Advances in Applied Probability* vol. 12, pp. 475-490.
21. Marriott, Paul and Mark Salmon (2000) *Applications of Differential Geometry to Econometrics* Cambridge, Cambridge University Press.
22. Newey, Whitney and Richard J. Smith (2004) "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators" *Econometrica* vol. 72, no. 1, pp. 219-255.
23. Phillips, P. C. B. (1978) "Edgeworth and Saddlepoint Approximations in the First-Order Noncircular Autoregression," *Biometrika*, 65, pp. 91-98.
24. Reid, N. (1988) "Saddlepoint Methods and Statistical Inference" *Statistical Science* vol 3, no. 2, pp. 213-227.
25. Robinson, J. E. Ronchetti and G. A. Young (2003) "Saddlepoint Approximations and Tests based on Multivariate M-estimates" *Annals of Statistics* vol. 31, no. 4, pp. 1154-1169.
26. Rochetti, Elvezio and Fabio Trojani (2003) "Saddlepoint Approximations and Test Statistics for Accurate Inference in Overidentified Moment Conditions Models" National Centre of Competence in Research Financial Valuation and Risk Management, working paper 27.
27. Rochetti, Elvezio and A. H. Welsch (1994) "Empirical Saddlepoint Approximations for Multivariate M-Estimators" *Journal of the Royal Statistical Society, Series B*, vol. 56, pp. 313-326.
28. Rilstone, Paul, V. K. Srivastava and Aman Ullah (1996) "The Second-Order Bias and Mean Squared Error of Nonlinear Estimators" *Journal of Econometrics*, vol 75, pp. 369-395.

29. Skovgaard, I. M. (1990) "On the Density of minimum contrast estimators", *Annals of Statistics*, vol. 18, pp. 779-789.
30. Sowell, F. B. (1996) "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework" *Econometrica* vol. 64, no. 5. pp. 1085-1107.
31. Stock, James and Jonathan Wright (2000) "GMM with Weak Identification," *Econometrica*, vol. 68(5), pp. 1055-1096.