Carnegie Mellon University
Research Showcase @ CMU

Tepper School of Business

7-2009

The Empirical Saddlepoint Likelihood Estimator Applied to Two-Step GMM

Fallaw Sowell
Carnegie Mellon University, fs0v@andrew.cmu.edu

Follow this and additional works at: http://repository.cmu.edu/tepper

Part of the Economic Policy Commons, and the Industrial Organization Commons

This Working Paper is brought to you for free and open access by Research Showcase @ CMU. It has been accepted for inclusion in Tepper School of Business by an authorized administrator of Research Showcase @ CMU. For more information, please contact research-showcase@andrew.cmu.edu.

The Empirical Saddlepoint Likelihood Estimator Applied to Two-Step GMM

Fallaw Sowell*
Tepper School of Business
Carnegie Mellon University
July 2009
(first draft February 2009)

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

*5000 Forbes Ave., Pittsburgh, PA 15213. Phone: (412)-268-3769. fs0v@andrew.cmu.edu

Helpful comments and suggestions were provided by Maria Chaderina, Dennis Epple, Benjamin Holcblat, Jason Imbrogno, Richard Lowery, Artem Neklyudov, Luis Quintero and seminar participants at Carnegie Mellon University.

Abstract

The empirical saddlepoint likelihood (ESPL) estimator is introduced. The ESPL provides improvement over one-step GMM estimators by including additional terms to automatically reduce higher order bias. The first order sampling properties are shown to be equivalent to efficient two-step GMM. New tests are introduced for hypothesis on the model's parameters. The higher order bias is calculated and situations of practical interest are noted where this bias will be smaller than for currently available estimators.

As an application, the ESPL is used to investigate an overidentified moment model. It is shown how the model's parameters can be estimated with both the ESPL and a conditional ESPL (CESPL), conditional on the overidentifying restrictions being satisfied. This application leads to several new tests for overidentifying restrictions.

Simulations demonstrate that ESPL and CESPL have smaller bias than currently available one-step GMM estimators. The simulations also show new tests for overidentifying restrictions that have performance comparable to or better than currently available tests. The computations needed to calculate the ESPL estimator are comparable to those needed for a one-step GMM estimator.

1 Introduction

The empirical saddlepoint density for a just-identified system of estimation equations can be considered an objective function. The parameter value where this objective function takes its highest value will define the empirical saddlepoint likelihood (ESPL) estimator. The asymptotic distribution of this estimator is shown to be equivalent to the sampling distribution of efficient two-step GMM (Hansen (1982)) and the one-step GMM estimators (Newey and Smith (2004)). The higher order bias of the ESPL estimator is different from the higher order bias of the GEL estimators (Newey and Smith (2004)). In some situations of practical interest, the higher order bias of the ESPL estimator is smaller than the higher order bias of the GEL estimators or the ETEL (Schennach (2007)) estimator. For these situations, a variation of the ESPL estimator is presented that will be higher order unbiased.

The intuition for how the ESPL estimator achieves its improvement can be understood by the estimation problem built on the moment condition $E\left[g(x_i,\theta_0)\right]=0$ and an iid sample $x_1,\ldots,x_n$. The sample moment condition $G_n(\theta)=\frac{1}{n}\sum_{i=1}^n g(x_i,\theta)$ evaluated at the population parameter value will satisfy a central limit theorem $\sqrt{n}G_n(\theta_0)\sim^a N(0,\Sigma(\theta_0))$. One way to estimate the parameters of interest would be to maximize the log-likelihood function implied by the asymptotic behavior of the moment conditions

$$\theta_{ml}\equiv\underset{\theta}{\operatorname{argmax}}\qquad -\frac{1}{2}\ln\left(|\Sigma(\theta)|\right)-\frac{n}{2}G_n(\theta)'\Sigma(\theta)^{-1}G_n(\theta).$$

An alternative estimate that ignores the determinant leads to the GMM estimator

$$\theta_{gmm}\equiv\underset{\theta}{\operatorname{argmin}}\qquad G_n(\theta)'\hat{\Sigma}(\theta_0)^{-1}G_n(\theta).$$

As long as $E\left[g(x_i,\theta)\right]$ identifies θ_0 and $\hat{\Sigma}(\theta_0)$ is a consistent estimate of $\Sigma(\theta_0)$, both estimators have the same efficient asymptotic distribution. If attention is restricted to the asymptotic distribution nothing is gained by considering the determinant. However, differences in the ML and GMM estimators can be observed by considering other properties such as higher order bias. The determinant in the log-likelihood function converges at a faster rate and hence does not contribute to the asymptotic distribution. However, it converges slow enough to contribute to the higher order bias. This is why the ML estimator typically has smaller higher order bias relative to the GMM estimator.

The ESPL estimator achieves a higher order bias similar to the ML estimator because terms in its objective function efficiently and nonparametrically estimate the determinant in the asymptotic normal approximation. The ESPL objective function basically adds terms to the exponential tilting (ET) objective function (Imbens (1997) and Kitamura and Stutzer (1997)) to reduce the higher order bias.

Because the objective function is a density, the ESPL estimator can be thought of as a maximum likelihood estimator. This structure leads to the usual test statistics:

likelihood ratio (LR), Wald and score. The LR and score test statistics require a conditional ESPL (CESPL) estimator that is of separate interest. Because the empirical saddlepoint density is defined using a tilting parameter, an additional test statistic is created using the tilting parameter.

The estimation and testing of an overidentified system of moment conditions is presented as an application of the ESPL estimator. The traditional two-step GMM estimator ignores information contained in the overidentifying restrictions when selecting the parameter estimates. Only after the GMM parameter estimates are determined are the overidentifying restrictions used to test the theory. Application of the ESPL estimator requires a different approach. The space of overidentifying restrictions is parameterized. This results in a just-identified system of moment conditions and an extended GMM objective function. Minimizing the extended GMM objective function *simultaneously* selects the parameters of interest from the original GMM objective function and the parameters that test the overidentifying restrictions. The empirical saddlepoint density approximates the distribution of the parameters that solve the extended moment conditions and are associated with a local minimum of the original GMM objective function. The parameters that maximize this empirical saddlepoint density define the ESPL estimator.

Simulation study shows that the ESPL has smaller bias than currently used estimators, e.g. EL, ET and ETEL. In addition, the simulations show that new tests for the overidentifying restrictions have comparable or better agreement with their asymptotic distributions relative to the tests considered in Imbens, Spady and Johnson (1998).

The ESPL estimator can be thought of as a natural extension for three different literatures: saddlepoint approximations, nonparametric maximum likelihood and information theoretic estimators.

Saddlepoint Density and Approximations: The ESPL estimator is a natural extension of the saddlepoint density literature. The saddlepoint approximation[1] was originally designed to give an improved approximation to the sampling distribution for maximum likelihood estimators when the distribution of the data is known. Recent work considers parameters estimated by just-identified estimation equations (Almudevar, Field and Robinson (2000)). These results require strong assumptions to ensure the existence of the saddlepoint density. Unfortunately, the needed assumptions are too strong to be appropriate for many applications in empirical economics.

Instead of focusing on sufficient conditions to ensure the existence of the saddlepoint density, this paper takes the literature in a new direction. The saddlepoint density is well defined for data drawn from a multinomial distribution. This permits defining the empirical saddlepoint density for the empirical distribution of a finite sample. The ESPL estimator is defined as the parameter values that maximize the empirical saddlepoint density. The saddlepoint density motivates the objective func-

[1]General introductions to saddlepoint methods are available in Reid (1988), Field and Ronchetti (1990), Jensen (1995), Kolassa (1997), Goutis and Casella (1999), Huzurbazar (1999), and Butler (2007).

tion used to define this estimator. However, the ESPL estimator is defined under *weaker conditions than are required for the existence of the saddlepoint density.*

Nonparametric Maximum Likelihood: The ESPL estimator is a natural extension of the nonparametric maximum likelihood literature. The EL[2] estimator is defined by maximizing the empirical density conditional on a set of moment conditions. The existence of the EL estimator does not require that the data are drawn from a multinomial distribution. The ESPL estimator takes a similar approach. The saddlepoint density can be viewed as an approximation to the distribution of parameters that solve the moment conditions, eg. see Skovgaard (1990), Jensen and Woods (1998) and Almudevar, Field and Robinson (2000). The EL creates a function over the parameters in the moment conditions and the EL estimator is the parameter value that maximizes this function. Note that even asymptotically the EL is not a density over the parameters. Alternatively, the saddlepoint density transforms the empirical distribution over the observed data and the moment conditions into a *density* over the parameters in the moment conditions. The ESPL estimator is the parameter that maximizes this density over the parameters. Hence, the ESPL estimator is a maximum likelihood estimator. The transformation to a density over the parameter values before the optimization is performed results in more accurate information about the parameter values that explain the data and the moment conditions.

Information Theoretic Estimators: The ESPL estimator is a natural extension of the information theoretic estimators literature. Instead of focusing on the maximum likelihood interpretation of the EL estimator, this literature focuses on minimizing alternative information criteria for the observed data conditional on a set of moment conditions (Imbens (1997), Newey and Smith (2004) and Schennach (2007)). A large class of information theoretic criteria functions give the identical first order asymptotic distribution. Hence, this literature has focused on higher order bias as a measure to distinguish between the estimators. The suggestions are to either use the estimator with the fewest terms in its higher order bias or to explicitly estimate these terms and calculate bias corrected estimators. The empirical saddlepoint density can be thought of as adding additional terms to the objective function and hence performing an automatic nonparametric correction to reduce the higher order bias. An advantage of the ESPL estimator is that it does not require the explicit calculation of terms in the higher order bias.

In Section 2 the ESPL estimator is defined and characterized with different representations. Section 2 also presents the first order properties, different tests for hypothesis concerning the model's parameters and the estimators higher order bias. Section 3 is an application of the ESPL estimator to a model of overidentified moment conditions. A just-identified system of estimation equations is presented and the ESPL is used to estimate the model's parameters. New tests of the overidentifying

[2]General introductions to empirical likelihood are available in Owen (1990), Imbens (1997) and Imbens (2002).

restrictions are introduced. Section 4 reports simulations that demonstrate that the ESPL and the CESPL estimator have smaller bias than currently available estimators and that new tests for the validity of the overidentifying restrictions have sampling properties comparable to, or better than, currently available tests. The final section summarizes the results and highlights directions for future research.

In this paper, all sums will be from 1 to n. Convergence to an asymptotic distribution will be denoted $\sim^a$. For a full column rank matrix Z let the projection matrix onto the space spanned by its columns be denoted P_Z and its orthogonal complement $P_Z^{\perp}$. For Σ a symmetric positive matrix let $\Sigma^{1/2}$ denote the upper triangular Cholesky decomposition, $\Sigma^{1/2\prime}\Sigma^{1/2} = \Sigma$. The generalized inverse of a matrix Ξ will be denoted $(\Xi)^{-g}$. Proofs are presented in the appendix.

2 The ESPL

This section presents the ESPL estimator and its asymptotic properties. The ESPL estimator will be defined as the parameter value that maximizes the empirical saddlepoint density. The empirical saddlepoint density is the saddlepoint density where the empirical distribution is used instead of the true distribution of the observed data. After the saddlepoint density is introduced, the form of the empirical saddlepoint density is presented. This form gives the intuition for the selection of the objective function used to define the ESPL estimator. Finally, the first order asymptotic results and the higher order bias are presented. (Section 2.1 motivates the objective function that defines the ESPL estimator. This motivation can be skipped by moving directly to section 2.2.)

2.1 Empirical Saddlepoint Overview/Introduction

The saddlepoint approximation was originally proposed to give an improved approximation to the sampling distribution for maximum likelihood estimators when the distribution of the data is known (Daniels (1954)). It was then extended to account for parameters estimated by general estimation equations where again the distribution of the data is known. Finally, the most recent work is concerned with parameters estimated by just-identified estimation equations where the data are generated under weak enough assumptions on the data to ensure that the saddlepoint density is well defined. This is the work that forms the foundation for the results presented in this paper. The basic theorems from the statistics literature are in Almudevar, Field and Robinson (2000), Field and Ronchetti (1990), Field (1982) and Ronchetti and Welsch (1994). To date, the saddlepoint distribution theory in the statistics literature is not well suited for empirical economics. Hence, 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 needed generalizations are presented in Sowell (2007).

The point of departure is the system of m estimation equations that form a just-identified system

$$\Psi_n(\alpha) \equiv n^{-1} \sum_i \psi(z_i, \alpha) = 0 \tag{1}$$

that is used to estimate the m parameters α where the observed data are $z_i \sim iid\ F(z)$.

The saddlepoint density replaces the asymptotic distribution implied by the central limit theorem in the traditional first order distribution theory. The normal approximation uses information about the shape of the estimation equations only at the selected solution. The saddlepoint approximation on the other hand uses information about the shape of the estimation equations at each point in the parameter space. The central limit theorem is built on a linear approximation of the characteristic function about the mean. A higher order approximation can be used to calculate an

Edgeworth expansion. Because the expansion is at the distribution's mean, the first order Edgeworth expansion gives a significantly better approximation at the mean of the distribution, $\mathcal{O}\left(n^{-1}\right)$ versus $\mathcal{O}\left(n^{-1/2}\right)$. Unfortunately, the quality of the first order Edgeworth expansion approximation can deteriorate significantly for values away from the mean. The saddlepoint approximation exploits this characteristic of the Edgeworth expansion. Instead of a single linear expansion, the saddlepoint uses multiple linear expansions to obtain improved accuracy, one expansion for every value in the parameter space.

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

$$dH_{n,\tau,\alpha}(z) = \frac{\exp\left\{\frac{\tau'}{n}\psi(z,\alpha)\right\} dF(z)}{\int \exp\left\{\frac{\tau'}{n}\psi(\zeta,\alpha)\right\} dF(\zeta)}.$$

Since the object of interest is the distribution of $\Psi_n(\alpha)$ and not an individual element $\psi(z_i,\alpha)$, the parameter τ is normalized by n.

At the parameter value of interest, α, the conjugate distribution is well defined for arbitrary values of τ. This is a degree of freedom, i.e. τ can be selected optimally for each value of α. A specific conjugate distribution is selected so its mean is transformed back to the original distribution at the value of interest. This will occur if τ is selected to satisfy the saddlepoint equation (m equations in m unknowns)

$$\int \psi(z,\alpha) \exp\left\{\frac{\tau'}{n}\psi(z,\alpha)\right\} dF(z) = 0. \tag{2}$$

Denote the solution to the saddlepoint equation as $\tau(\alpha)$. An Edgeworth expansion is calculated for the conjugate distribution defined by $\tau(\alpha)$, i.e. $dH_{n,\tau(\alpha)}(z)$. 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 structure of the saddlepoint density is recorded in Theorem 2 from Almudevar, Field and Robinson (2000). Under sufficient conditions, the density for the location of solutions to the estimation equations (1) is given by

$$\begin{aligned} f_n(\alpha) &= \left(\frac{n}{2\pi}\right)^{\frac{m}{2}} \left| E\left[\partial\psi(z,\alpha)/\partial\alpha'\right]\right| \left| E\left[\psi(z,\alpha)\psi(z,\alpha)'\right]\right|^{-1/2} \\ &\qquad \times \exp\left\{n\kappa_n(\hat{\tau}_n(\alpha),\alpha)\right\}\left(1+\mathcal{O}\left(n^{-1}\right)\right) \end{aligned}$$

where $\hat{\tau}_n(\alpha)$ solves the saddlepoint equation $\int \psi(z,\alpha)\exp\left\{\frac{\tau'}{n}\psi(z,\alpha)\right\} dF(z) = 0$, the expectations are with respect to the conjugate distribution $dH_{n,\tau_n(\alpha)}(z)$, and $\kappa_n(\tau,\alpha) = \ln\left(\int \exp\left\{\tau'\psi(z,\alpha)\right\} dF(z)\right)$. This shows how the saddlepoint approximation is calculated. The saddlepoint approximation is nonnegative and gives a relative and faster rate of convergence than the asymptotic normal approximation.

The calculation of the saddlepoint density requires knowledge of the distribution $F(z)$ but in most economic applications this is unknown. Replacing the distribution with the empirical distribution results in the empirical saddlepoint approximation. The empirical saddlepoint density gives the distribution of the parameter values that solve the system of equations when the data are drawn from the empirical distribution. The basic structure of the empirical saddlepoint density is recorded in the theorem in Ronchetti and Welsh (1994). Under sufficient conditions, the density for the location of solutions to the estimation equations (1) is given by

$$\begin{aligned}\hat{f}_n(\alpha) &= \left(\frac{n}{2\pi}\right)^{\frac{m}{2}} \left|\sum_i \frac{\partial\psi(z_i,\alpha)'}{\partial\alpha}\hat{w}_i(\alpha)\right| \left|\sum_i \psi(z_i,\alpha)\psi(z_i,\alpha)'\hat{w}_i(\alpha)\right|^{-1/2} \\ &\quad \times \exp\left\{n\ln\left(\frac{1}{n}\sum_i \exp\left\{\tau_n(\alpha)'\psi(z_i,\alpha)\right\}\right)\right\}\left(1+\mathcal{O}\left(n^{-1/2}\right)\right),\end{aligned}$$

where $\hat{w}_i(\alpha) = \frac{\exp\{\tau_n(\alpha)'\psi(z_i,\alpha)\}}{\sum_j \exp\{\tau_n(\alpha)'\psi(z_j,\alpha)\}}$ and $\tau_n(\alpha)$ solves $n^{-1}\sum_i \psi(z_i,\alpha)\exp\{\tau'\psi(z_i,\alpha)\} = 0$. To simplify notation and calculations, the sample size scaling has been absorbed into the τ_n parameter. Using the empirical distribution gives a nonparametric procedure but results in a reduction in accuracy from a relative error of n^{-1} to a relative error of $n^{-1/2}$.

The saddlepoint approximation generalizes the asymptotic normal approximation. If the estimation equations are nonlinear, the saddlepoint approximation will incorporate the global structure of the estimation equations. The resulting approximation may be asymmetric and does not force the tail behavior associated with the normal approximation. The empirical saddlepoint approximation can have multiple modes. Consistency implies that the mass of the sampling distribution converges to a shrinking neighborhood of the population parameter value. In this neighborhood, the estimation equations will be nearly linear. Hence the saddlepoint approximation will converge to the normal approximation.

2.2 ESPL estimator defined

To simplify the notation, we make the dependence of all quantities on α implicit and use the following notation.

Definition 2.1. *Let* $\hat{w}_i = \hat{w}_i(\alpha)$, $\tau_n = \tau_n(\alpha)$, $\psi_i = \psi(z_i,\alpha)$, $\Psi_n = n^{-1}\sum_i \psi_i$, $M_\psi = E\left[\frac{\partial\psi(z_i,\alpha)}{\partial\alpha'}\right]$, $\hat{M}_\psi = n^{-1}\sum_i \frac{\partial\psi(z_i,\alpha)}{\partial\alpha'}$, $\hat{\Sigma}_\psi = n^{-1}\sum_i \psi(z_i,\alpha)\psi(z_i,\alpha)'$ *and* $\Sigma_\psi = E\left[\psi(z_i,\alpha)\psi(z_i,\alpha)'\right]$, *where expectations are with respect to* $F(z)$. *Quantities evaluated at* $\alpha = \alpha_0$ *are denoted with a subscript of* 0, *eg.* $M_{\psi 0} = E\left[\frac{\partial\psi(z_i,\alpha_0)}{\partial\alpha'}\right]$. *Let* $\psi_{(j)i}$ *denote the* j^{th} *element of the vector* $\psi(z_i,\alpha)$.

The ESPL estimator can now be defined. The likelihood function is created by considering the empirical saddlepoint density as a function of the parameters of interest

conditional on the observed sample. The objective function for the ESPL estimator is the log of the saddlepoint density normalized by the sample size with the constant removed.

Definition 2.2. *(The Empirical Saddlepoint Likelihood Estimator)*

$$\hat{\alpha}_{espl} \equiv \underset{\alpha}{\operatorname{argmax}} \quad \mathcal{L}_n(\alpha, \tau_n)$$

where

$$\begin{aligned}
&\mathcal{L}_n(\alpha, \tau_n) \\
&\quad = \; -\frac{1}{2n} \ln\left(\left|\sum_i \hat{w}_i \psi_i \psi_i'\right|\right) + \frac{1}{n} \ln\left(\left|\sum_i \hat{w}_i \frac{\partial \psi_i}{\partial \alpha'}\right|\right) + \ln\left(n^{-1} \sum_i \exp\left\{\tau_n' \psi_i\right\}\right),
\end{aligned}$$

$\hat{w}_i$ the solution to

$$\min_{\{w_i\}_{i=1}^n} \quad \sum_i w_i \ln(w_i)$$

subject to $\sum_i w_i \psi_i = 0$ and $\sum_i w_i = 1$, and τ_n is the Lagrange multiplier at the optimal value for the associated Lagrangian equation

$$\ell(w_1, \ldots, w_n, \tau, \mu) = \sum_i w_i \ln(w_i) - \tau' \left(\sum_i w_i \psi_i\right) + \mu \left(\sum_i w_i - 1\right).$$

The Lagrangian is the same one that occurs with the ET estimator and defines the constrained optimization problem of finding the multinomial density with the highest entropy subject to the density satisfying the estimation equations. Because the estimation equations are just-identified, entropy will be maximized by setting τ to zero and hence $w_i = 1/n$ for every solution to the estimation equations (1).

Intuition for the formal results presented below can be developed by noting the similarity between the ESPL objective function and the ET objective function. Note that

$$\begin{aligned}
&\mathcal{L}_n(\alpha, \tau_n) \\
&\quad = \; -\frac{1}{2n} \ln\left(\left|\sum_i \hat{w}_i \psi_i \psi_i'\right|\right) + \frac{1}{n} \ln\left(\left|\sum_i \hat{w}_i \frac{\partial \psi_i}{\partial \alpha'}\right|\right) + \ln\left(n^{-1} \sum_i \exp\left\{\tau_n' \psi_i\right\}\right) \\
&\quad = \; \ln\left(\left|\left[\sum_i \hat{w}_i \frac{\partial \psi_i'}{\partial \alpha}\right]\left[\sum_i \hat{w}_i \psi_i \psi_i'\right]^{-1}\left[\sum_i \hat{w}_i \frac{\partial \psi_i}{\partial \alpha'}\right]\right|^{\frac{1}{2n}} n^{-1} \sum_i \exp\left\{\tau_n' \psi_i\right\}\right)
\end{aligned}$$

which implies

$$\begin{aligned}\exp\left\{\mathcal{L}_n(\alpha,\tau_n)\right\} &= \left|\left[\sum_i \hat{w}_i \frac{\partial\psi_i'}{\partial\alpha}\right]\left[\sum_i \hat{w}_i\psi_i\psi_i'\right]^{-1}\left[\sum_i \hat{w}_i \frac{\partial\psi_i}{\partial\alpha'}\right]\right|^{\frac{1}{2n}} n^{-1}\sum_i \exp\left\{\tau_n'\psi_i\right\} \\ &= n^{-1}\sum_i \exp\left\{\tau_n'\psi_i\right\} + \mathcal{O}_p\left(n^{-1}\right)\end{aligned}$$

where the last equality occurs because in a neighborhood of $x = 0$, $|Z|^{\delta x} = 1 + \ln(|Z|)\,\delta x + \mathcal{O}(x^2)$. This shows that the ESPL objective function is equivalent to the ET objective function except for an $\mathcal{O}_p\left(n^{-1}\right)$ term. This term dies out fast enough that the ESPL estimator has the same consistency and asymptotic normality results as ET, however, it also dies out slow enough to contribute to the higher order bias.

The formal calculation of the estimator's properties will be eased by considering alternative ways of defining and characterizing the estimator.

Theorem 2.1. *The ESPL estimator $\hat{\alpha}_{espl}$ maximizes the objective function*

$$\begin{aligned}\mathcal{L}_n(\alpha,\tau_n(\alpha)) &= -\frac{1}{2n}\ln\left(\left|\frac{1}{n}\sum_i \exp\left\{\tau_n'\psi_i\right\}\psi_i\psi_i'\right|\right) + \frac{1}{n}\ln\left(\left|\frac{1}{n}\sum_i \exp\left\{\tau_n'\psi_i\right\}\frac{\partial\psi_i}{\partial\alpha'}\right|\right) \\ &\quad + \left(1-\frac{m}{2n}\right)\ \ln\left(\frac{1}{n}\sum_i \exp\left\{\tau_n'\psi_i\right\}\right),\end{aligned}$$

where τ_n is the solution to

$$S_n(\alpha,\tau) \equiv n^{-1}\sum_i \psi_i \exp\left\{\tau'\psi_i\right\} = 0. \tag{3}$$

The first order conditions can be written as either the system of $2m$ equations in $2m$ unknowns

$$\frac{\partial\mathcal{L}_n(\alpha,\tau)}{\partial\alpha} = 0 \tag{4}$$

$$n^{-1}\sum_i \psi_i \exp\left\{\tau'\psi_i\right\} = 0 \tag{5}$$

or the system of m equations in m unknowns

$$\frac{d\mathcal{L}_n(\alpha,\tau_n)}{d\alpha} = 0 \tag{6}$$

where the total derivative is used to indicate that τ_n is allowed to vary with α.

2.3 Asymptotic distribution and testing

The sampling distribution for the ESPL estimator will require fairly standard regularity conditions.

Assumption 2.1. *(Regularity Conditions)*

1. $\{z_i\}_{i=1}^{\infty}$ *forms an iid sequence.*
2. $\alpha_0 \in \text{int}(\mathcal{A})$ *is the unique solution to* $E\left[\psi(z_i, \alpha)\right] = 0$*, where* $\mathcal{A}$ *is a compact subset of* R^m*.*
3. $\psi(z_i, \alpha)$ *is continuous in* α *at each* $\alpha \in \mathcal{A}$ *with probability one.*
4. $E\left[\sup_{\alpha\in\mathcal{A}} \|\psi_i\|^{2+\delta}\right] < \infty$ *for some* $\delta > 0$ *and* $E\left[\sup_{\alpha\in\mathcal{N}} \left\|\frac{\partial\psi_i}{\partial\alpha}\right\|\right] < \infty$ *where* $\mathcal{N}$ *is an open neighborhood of* α_0*.*
5. $\Sigma_{\psi,0}$ *is nonsingular and finite and has rank* m*.*
6. $\psi(z_i, \alpha)$ *is twice continuously differentiable in* α *in a neighborhood* $\mathcal{N}$ *of* α_0*.*
7. $\text{rank}\left(M_{\psi,0}\right) = m$*.*
8. *(i)* $E\left[\frac{\partial\psi_{(j_1)}(z_i,\alpha_0)}{\partial\alpha_\ell}\psi_{(j_2)}(z_i, \alpha_0)\right]$ *is finite for* $j_1, j_2, \ell = 1, \ldots, m$*.*
   (ii) $E\left[\psi_{(j_1)}(z_i, \alpha_0)\psi_{(j_2)}(z_i, \alpha_0)\psi_{(j_3)}(z_i, \alpha_0)\right]$ *is finite for* $j_1, j_2, j_3 = 1, \ldots, m$*.*
   (iii) $E\left[\frac{\partial\psi_{(j1)}(z_i,\alpha_0)}{\partial\alpha_\ell}\frac{\partial\psi_{(j2)}(z_i,\alpha_0)}{\partial\alpha_\kappa}\right]$ *is finite for* $j1, j2, \ell, \kappa = 1, \ldots, m$*.*

These assumptions are slightly stronger than the assumptions needed for the first order asymptotics of the GEL estimator (Newey and Smith (2004)). In particular, Assumption 8 requires the existence of higher order moment terms and cross terms between the estimation equations and their derivatives. These enter the determinant term in the saddlepoint density and need to be bounded so that the determinant does not contribute to the first order asymptotic distribution but does affect the higher order bias. This is demonstrated in the following theorems recording the first order asymptotic behavior of the ESPL estimator and its higher order bias.

Theorem 2.2. *(ESPL: First order properties)*

Under Assumption 2.1, (i) *the ESPL estimator and the tilting parameter have the first order asymptotic distribution*

$$\sqrt{n}\begin{pmatrix} \hat{\alpha}_{espl} - \alpha_0 \\ \hat{\tau}_{espl} \end{pmatrix} \sim^a N\left(0, \begin{pmatrix} \left(M'_{\psi 0}\Sigma_{\psi 0}^{-1}M_{\psi 0}\right)^{-1} & 0 \\ 0 & 0 \end{pmatrix}\right)$$

and (ii) *alternatively, confidence intervals for the parameters can be created using the likelihood-ratio statistic*

$$2n\left(\mathcal{L}_n(\hat{\alpha}_{espl}) - \mathcal{L}_n(\alpha)\right) \sim^a \chi^2_m.$$

The ESPL permits multiple ways to test restrictions on the parameters. Consider the null hypothesis that the parameters satisfy the nonlinear restrictions, for $q \leq m$,

$$H_0 : r(\alpha) = \underset{q\times 1}{0} \tag{7}$$

and let $R(\alpha)_{m\times q} \equiv \frac{\partial r(\alpha)'}{\partial \alpha}$ with $R_0 = R(\alpha_0)$. One approach is to calculate the conditional ESPL (CESPL) estimator using the Lagrangian

$$\mathcal{L}_n(\alpha, \tau) + r(\alpha)'\gamma \tag{8}$$

subject to τ satisfying

$$n^{-1} \sum_i \psi_i \exp\{\tau'\psi_i\} = 0 \tag{9}$$

where γ is the Lagrange multiplier.

Theorem 2.3. *(Conditional ESPL: First order properties)*

Under Assumption 2.1 and when the parameter restriction (7) is true with R_0 full rank, the asymptotic distribution for the conditional parameter estimates defined by (8) and (9) is

$$\sqrt{n} \begin{bmatrix} \hat{\alpha}_{cespl} - \alpha_0 \\ \hat{\tau}_{ceslp} \\ \hat{\gamma} \end{bmatrix} \sim^a$$

$$N\left(0, \begin{pmatrix} M_{\psi 0}^{-1}\Sigma_{\psi 0}^{1/2\prime} P_\Gamma^\perp \Sigma_{\psi 0}^{1/2} \left(M_{\psi 0}^{-1}\right)' & 0 & 0 \\ 0 & \left(\Sigma_{\psi 0}^{1/2}\right)^{-1} P_\Gamma \left(\Sigma_{\psi 0}^{1/2\prime}\right)^{-1} & -\left(M_{\psi 0}^{-1}\right)' R_0 \left(\Gamma'\Gamma\right)^{-1} \\ 0 & -\left(\Gamma'\Gamma\right)^{-1} R_0' M_{\psi 0}^{-1} & \left(\Gamma'\Gamma_0\right)^{-1} \end{pmatrix}\right).$$

where $\Gamma = \Sigma_{\psi 0}^{1/2} \left(M_{\psi 0}^{-1}\right)' R_0$.

This gives two ways to test the null hypothesis. One approach is to use the tilting parameter with the test statistic

$$T_1 = n\hat{\tau}_{cespl}' \Sigma_{\psi 0}^{1/2\prime} (P_\Gamma)^{-g} \Sigma_{\psi 0}^{1/2} \hat{\tau}_{cespl} \tag{10}$$

which is asymptotically equivalent to

$$T_2 = n\hat{\tau}_{cespl}' \Sigma_{\psi 0} \hat{\tau}_{cespl}. \tag{11}$$

The other test statistic uses the Lagrange multiplier

$$LM = n\hat{\gamma}'\Gamma'\Gamma\hat{\gamma}. \tag{12}$$

Both of these statistics will have a chi-square distribution with q degrees of freedom

when the null hypothesis is true.

The Lagrange multiplier statistic can also be written in terms of the objective function for the constrained estimation problem. The first derivative of the Lagrangian with respect to α (equations (31) in the appendix) can be solved for $\hat{\gamma}$ as

$$\hat{\gamma} = -\left(\Gamma'\Gamma\right)^{-1}\Gamma'\Sigma\left(M_{\psi 0}^{-1}\right)'\frac{\partial \mathcal{L}_n(\hat{\alpha}_{cespl}, \hat{\tau}_{cespl})}{\partial \alpha}.$$

Substituting into (12) gives the score statistic version of the test statistic

$$n\frac{\partial \mathcal{L}_n(\hat{\alpha}_{cespl}, \hat{\tau}_{cespl})}{\partial \alpha'}M_{\psi 0}^{-1}\Sigma_{\psi 0}^{-1/2\prime}\Gamma\left(\Gamma'\Gamma\ \right)^{-1}\Gamma'\Sigma_{\psi 0}^{-1/2}\left(M_{\psi 0}^{-1}\right)'\frac{\partial \mathcal{L}_n(\hat{\alpha}_{cespl}, \hat{\tau}_{cespl})}{\partial \alpha}.$$

The test statistics can be made feasible by replacing the unknown terms with consistent estimates.

Consider the special case of testing if an $(m-k)$ subset of the parameters is zero. Partition the parameters $\alpha = \left[\begin{array}{cc} \theta'_{1\times k} & \lambda'_{1\times(m-k)} \end{array}\right]'$ and[3] consider the null hypothesis

$$H_0 : \lambda = 0. \tag{13}$$

For this hypothesis $R_0 = \left[\begin{array}{c} 0 \\ I_{m-k} \end{array}\right]$, and let $\overline{R}_0 = \left[\begin{array}{c} I_k \\ 0 \end{array}\right]$. This parameter restriction can be substituted directly into the estimation problem without the need for the Lagrange multiplier. The estimation problem then becomes

$$\hat{\theta}_{cespl} = \underset{\theta\in\Theta}{\operatorname{argmax}} \quad \mathcal{L}_n(\theta, 0, \hat{\tau}_{cespl}) \tag{14}$$

where $\hat{\tau}_{cespl}$ is selected to solve

$$S_n(\theta, 0, \tau) \equiv n^{-1}\sum_i \psi_i(\theta, 0)\exp\left\{\tau'\psi_i(\theta, 0)\right\} = 0. \tag{15}$$

Theorem 2.4. *(Conditional ESPL: First order properties Alternative Form)*

Under Assumption 2.1 and when the parameter restriction (13) is true, the asymptotic distribution for the conditional parameter estimate defined by (14) and (15) is

$$\sqrt{n}\left[\begin{array}{c} \hat{\theta}_{cespl} - \theta_0 \\ \hat{\tau}_{cespl} \end{array}\right] \sim^a N\left(0, \left(\begin{array}{cc} \left(\overline{R}_0' M_{\psi 0}' \Sigma_{\psi 0}^{-1} M_{\psi 0}\overline{R}_0\right)^{-1} & 0 \\ 0 & \left(\Sigma_{\psi 0}^{-1/2}\right)' P^{\perp}_{\Sigma_{\psi 0}^{-1/2} M_{\psi 0}\overline{R}_0}\Sigma_{\psi 0}^{-1/2} \end{array}\right)\right).$$

This has the more familiar form given in the EL, ET, GEL and ETEL literature for testing the overidentifying restrictions with the tilting parameter, see Imbens, Spady and Johnson (1998), Newey and Smith (2004), Schennach (2007).

[3]Functions previously dependent on α will now be written as functions of θ and λ, e.g. $\mathcal{L}_n(\alpha, \tau(\alpha)) = \mathcal{L}_n(\theta, \lambda, \tau(\theta, \lambda))$.

2.4 Higher order bias

The ESPL objective function is similar to the ET objective function. The difference is the determinant that includes efficient estimates of the covariance and the expectation of the first derivative. These terms converge to zero fast enough so as to not affect the first order asymptotic distribution but the convergence is slow enough that the terms do contribute to the higher order bias. The result is that the ESPL can be viewed as automatically performing partial bias correction relative to the one-step estimators. This is different from the analytical approach to bias correction proposed in Newey and Smith (2004). The ESPL does not require the analytic calculation of the higher order bias.

The calculation of the higher order bias requires additional restrictions on the estimation equations and the distribution of the observed data.

Assumption 2.2. *(Higher order Regularity Conditions) There exists a function $b(z_i)$ with $E[b(z_i)^6] < \infty$ such that, in a neighborhood $\mathcal{N}$ of α_0, all partial derivatives of $\psi(z_i, \alpha)$ with respect to α up to order four exist, are bounded by $b(z_i)$ and are Lipschitz in α with prefactor $b(z_i)$.*

This will imply enough moments to ensure the existence of the higher order bias. The higher order bias for the m-estimator that solves equation (1) is now presented.

Theorem 2.5. *(Higher order bias: m-estimator)*
If Assumptions 2.1 and 2.2 are satisfied, then the m-estimator's $\mathcal{O}(n^{-1})$ bias is

$$n^{-1} M_{\psi 0}^{-1} \quad \left(-a + E\left[\frac{\partial \psi_{i0}}{\partial \alpha'} M_{\psi 0}^{-1} \psi_{i0} \right] \right) \tag{16}$$

where a is a vector with elements $a_j = \operatorname{tr}\left(\left(M_{\psi 0}' \Sigma_{\psi 0}^{-1} M_{\psi 0} \right)^{-1} E\left[\partial^2 \psi_{(j)i0} / \partial \alpha \partial \alpha' \right] \right) / 2$.

This is the same as the higher-order bias for a GEL estimator when a just-identified system of moment conditions is used.[4] The higher order bias of the ESPL estimator includes two additional terms that are contributed by the determinant in the saddlepoint density.

Theorem 2.6. *(Higher order bias: ESPL estimator)*
If Assumptions 2.1 and 2.2 are satisfied, then ESPL estimator's $\mathcal{O}(n^{-1})$ bias is

$$\begin{aligned} n^{-1} M_{\psi 0}^{-1} \quad \Bigg(& -a + E\left[\frac{\partial \psi_{i0}}{\partial \alpha'} M_{\psi 0}^{-1} \psi_{i0} \right] \\ & + \Sigma_{\psi 0}^{-1} \left(M_{\psi 0}' \right)^{-1} c - \Sigma_{\psi 0}^{-1} \left(M_{\psi 0}' \right)^{-1} E\left[\frac{\partial \psi_{i0}'}{\partial \alpha} \Sigma_{\psi 0}^{-1} \psi_{i0} \right] \Bigg) \end{aligned} \tag{17}$$

[4] This is given in Theorem 4.2 of Newey and Smith (2004). Using the notation in Newey and Smith (2004), the matrix P is zero for a just-identified system of moment conditions.

where a is a vector with elements $a_j = \operatorname{tr}\left(\left(M'_{\psi 0}\Sigma^{-1}_{\psi 0}M_{\psi 0}\right)^{-1} E\left[\partial^2\psi_{(j)i0}/\partial\alpha\partial\alpha'\right]\right)/2$ *and c is a vector with elements* $c_j = \operatorname{tr}\left(M^{-1}_{\psi 0}E\left[\partial^2\psi_{i0}/\partial\alpha_j\partial\alpha'\right]\right)$.

2.5 Special Case: Approximations to the scores

In general, the higher order bias of the ESPL estimator can be larger or smaller than the higher order bias from the m-estimator. There are cases where the higher order bias of the ESPL will contain fewer terms.

Corollary 2.1. *(Higher order bias: special case)*
If Assumptions 2.1 and 2.2 are satisfied and

1. $E\left[\frac{\partial\psi_{i0}}{\partial\alpha'}M^{-1}_{\psi 0}\psi_{i0}\right] = \Sigma^{-1}_{\psi 0}\left(M'_{\psi 0}\right)^{-1}E\left[\frac{\partial\psi'_{i0}}{\partial\alpha}\Sigma^{-1}_{\psi 0}\psi_{i0}\right] + o_p(n^{-1/2})$ *and*
2. $a = \Sigma^{-1}_{\psi 0}\left(M'_{\psi 0}\right)^{-1}c/2 + o_p(n^{-1/2})$

then ESPL's $\mathcal{O}(n^{-1})$ bias is

$$n^{-1}M^{-1}_{\psi 0}a.$$

A leading case satisfied by the assumptions in **Corollary 2.1** is maximum likelihood in which case there would be equalities with no error terms. More generally the assumptions would be satisfied by estimation equations obtained by setting to zero the slowest converging term of the score function from a likelihood function. In this case $\frac{\partial\psi_{i0}}{\partial\alpha'} = \frac{\partial\psi'_{i0}}{\partial\alpha} + o_p(n^{-1/2})$, $\Sigma_{\psi 0} = -M_{\psi 0} + o_p(n^{-1/2})$ and $\partial^2\psi_{(j)i0}/\partial\alpha'\partial\alpha = \partial^2\psi_{i0}/\partial\alpha_j\partial\alpha' + o_p(n^{-1/2})$.

There are several other cases where the assumptions are expected to hold exactly. Even if the assumptions do not hold exactly, the higher order bias will be expected to be smaller the closer these assumptions are to being satisfied. When the assumptions of **Corollary 2.1** are satisfied, the determinant terms in the ESPL objective function result in automatic partial bias correction. This partial bias correction can be made complete. How the determinant terms in the ESPL objective function contribute to the higher order bias suggests an alternative estimator with zero higher order bias. The estimator defined by maximizing

$$\begin{aligned}\tilde{\mathcal{L}}_n(\alpha,\tau) &= -\frac{1}{2n}\ln\left(\left|\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\psi_i\psi'_i\right|\right) + \frac{1}{2n}\ln\left(\left|\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\frac{\partial\psi_i}{\partial\alpha'}\right|\right)\\ &\quad + \ln\left(\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\right) \qquad (18)\end{aligned}$$

will have zero higher order bias under the conditions in **Corollary 2.1**. The coefficient on the second term is now 1/2 and the coefficient on the last term is set to one. Future research will investigate this estimator.

2.6 Applications of ESPL

This section has introduced the ESPL estimator, presented it properties and showed how hypothesis can be tested. The next section presents an application of the ESPL estimator using the first order conditions from efficient two-step GMM. This will help build the connection between the widely used two-step GMM framework and the ESPL. However, it is important to note that the ESPL can be applied to a much larger class of problems. Parameters that are estimated by a just-identified system of equations that satisfy the regularity conditions of presented in Assumptions 2.1 can be estimated by the ESPL. For example in a separate paper, the first order conditions from the GEL estimators (Newey and Smith (2004)) are used to create a just-identified system of equations that are appropriate for ESPL estimation and testing.

3 ESPL for Overidentified Two-step GMM

As an application of the ESPL estimator and tests, a standard econometric model[5] will be investigated. The ESPL estimator requires a set of just-identified estimation equations. However, the standard econometric model is an overidentified system of moment conditions. This section shows how to create a just-identified system of estimation equations from an overidentified system of moment conditions. New equations and parameters are introduced to test the overidentifying restrictions. The new equations with the original moment conditions create a just-identified system of estimation equations. The original parameters of interest and the new parameter can be estimated jointly as ESPL estimates. The results from the previous section will lead to new tests for the validity of the overidentifying restrictions.

3.1 A standard econometric model

Before deriving the estimation equations, the notation for a standard econometrics model will be introduced. Consider an $m-$dimensional set of moment conditions $g(z_i, \theta)$ where θ is a $k-$dimensional set of parameters with $k < m$. The economic theory implies that the moment conditions have expectation zero at the population parameter value, i.e. $E\left[g(z_i, \theta_0)\right] = 0$. An iid sample of n observations is used to create the sample analog of the moment conditions $G_n(\theta) = \frac{1}{n}\sum_i g(z_i, \theta)$ and its first derivative $M_n(\theta) = \frac{\partial G_n(\theta)}{\partial \theta'}$. It is also assumed that the sample moment evaluated at θ_0 satisfy the central limit theorem $\sqrt{n}G_n(\theta_0) \sim^a N(0, \Sigma_g)$. The two-step GMM parameter estimate is defined as the parameter values that minimize the GMM objective function

$$\hat{\theta}_n = \underset{\theta \in \Theta}{\operatorname{argmin}}\, G_n(\theta)' W_n G_n(\theta) \tag{19}$$

where W_n is a symmetric positive definite weighting matrix that converges to Σ_g^{-1}. This weighting matrix is the inverse of a consistent estimate for Σ_g and is calculated by a consistent estimate of θ that is obtained in the first step of estimation.

Standard regularity conditions ensure that the GMM estimator is $\sqrt{n}-$consistent and asymptotically distributed as

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \sim^a N\left(0, \left(M_{g0}' \Sigma_g^{-1} M_{g0}\right)^{-1}\right)$$

where $M_{g0} = E\left[\frac{\partial g(z_i, \theta_0)}{\partial \theta'}\right]$. The economic theory implies all m moment conditions should equal zero. The first order conditions $M_n(\hat{\theta}_n)' W_n G_n(\hat{\theta}_n) = 0$ set k linear combinations of the sample moments to zero and the remaining $(m-k)$ overidentifying dimensions of the moments can be used to test the economic theory with the statistic

$$J = nG_n(\hat{\theta}_n)' W_n G_n(\hat{\theta}_n)$$

[5] Introductions to this type of econometric model are available in Mátyás (1999) and Hall (2005).

which is asymptotically distributed χ^2_{m-k} when the null hypothesis of the economic theory is correct.

3.2 From overidentified moment conditions to just-identified estimation equations.

This above notation will now be used to derive a just-identified system of estimation equations that will be appropriate for the calculation of the ESPL estimators. For each value of θ, the sample moment conditions $G_n(\theta)$ form 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[6]. This decomposition is a generalization of the decomposition used in Sowell (1996) where the tangent space at $\hat{\theta}_n$ 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 value[7] of θ, not only at $\hat{\theta}_n$.

For each value of θ, let $\overline{M}_n(\theta)$ denote the derivative of $G_n(\theta)$ 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 are $\overline{M}_n(\hat{\theta}_n)'W_n^{1/2}G_n(\hat{\theta}_n) = 0$. The columns of $\overline{M}_n(\hat{\theta}_n)$ define the k linear combinations used to identify and estimate θ.

The orthogonal complement of the space spanned by the columns of $\overline{M}_n(\hat{\theta}_n)$ is the $(m-k)-$dimensional space used to test the validity of the overidentifying restrictions and will be spanned by a new set of parameters denoted λ. The projection matrix for the space spanned by $\overline{M}_n(\theta)$, $P_{\overline{M}_n(\theta)}$, is a real symmetric positive semidefinite matrix, which is also idempotent and hence has a spectral decomposition. Denote a spectral decomposition[8]

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

where $C_n(\theta)'C_n(\theta) = I_m$. For each θ, the columns of $C_n(\theta)$ form an orthonormal basis. The basis elements will be selected so that they are differentiable in a neighborhood of θ. The derivatives of these basis elements are presented in Sowell (2007). The column span of $C_{1,n}(\theta)$ is the same as the column span of $\overline{M}_n(\theta)$, and the column span of $C_{2,n}(\theta)$ is the orthogonal complement of the column span of $\overline{M}_n(\theta)$. Hence, for each value of θ, the $m-$dimensional space containing $G_n(\theta)$ can be locally parameterized

[6] At $\hat{\theta}_n$ this has been called the space of overidentifying restrictions. In statistics, for other values of θ this has been called the ancillary space.

[7] When attention is restricted to the empirical saddlepoint density, then the decomposition only needs to exist for parameters in neighborhoods of the local minima.

[8] The spectral decomposition is not unique, raising a potential concern. However, the invariance of inference with respect to alternative spectral decompositions is documented in Sowell (2007).

by

$$\begin{bmatrix} C_{1,n}(\theta)'W_n^{1/2}G_n(\theta) \\ \lambda - C_{2,n}(\theta)'W_n^{1/2}G_n(\theta) \end{bmatrix}. \tag{20}$$

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 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.

The column span of $C_{1,n}(\theta)$ and $\overline{M}_n(\theta)$ are the same. Therefore, the system of equations $C_{1,n}(\hat{\theta}_n)'W_n^{1/2}G_n(\hat{\theta}_n) = 0$ and $\overline{M}_n(\hat{\theta}_n)'W_n^{1/2}G_n(\hat{\theta}_n) = 0$ are equivalent. Both define the same parameter estimates $\hat{\theta}_n$ and they can be solved independently of λ. The system of equations $\hat{\lambda}_n - C_{2,n}(\hat{\theta}_n)'W_n^{1/2}G_n(\hat{\theta}_n) = 0$ can then be used to calculate the estimate $\hat{\lambda}_n = C_{2,n}(\hat{\theta}_n)'W_n^{1/2}G_n(\hat{\theta}_n)$.

The overidentifying restrictions are tested with

$$\begin{aligned} n\hat{\lambda}_n'\hat{\lambda}_n &= nG_n(\hat{\theta}_n)'W_n^{1/2\,\prime}C_{2,n}(\hat{\theta}_n)C_{2,n}(\hat{\theta}_n)'W_n^{1/2}G_n(\hat{\theta}_n) \\ &= nG_n(\hat{\theta}_n)'W_n^{1/2\,\prime}P^{\perp}_{\overline{M}(\hat{\theta}_n)}W_n^{1/2}G_n(\hat{\theta}_n) \\ &= nG_n(\hat{\theta}_n)'W_n^{1/2\,\prime}W_n^{1/2}G_n(\hat{\theta}_n) \\ &= nG_n(\hat{\theta}_n)'W_nG_n(\hat{\theta}_n) \\ &= J. \end{aligned}$$

Premultiply (20) by $\begin{bmatrix} C_{1,n}(\theta) & -C_{2,n}(\theta) \end{bmatrix}$ and set it to zero to obtain the equivalent system of equations

$$\Psi_n(\hat{\alpha}_n) = W_n^{1/2}G_n(\hat{\theta}_n) - C_{2,n}(\hat{\theta}_n)\hat{\lambda}_n = 0 \tag{21}$$

where $\alpha = \begin{bmatrix} \theta' & \lambda' \end{bmatrix}'$. This is the just-identified system of equations that will be used to calculate the empirical saddlepoint density. The equations can also be written $\Psi_n(\alpha) = \frac{1}{n}\sum_{i=1}^n \psi(z_i, \alpha)$ where

$$\psi(z_i, \alpha) = W_n^{1/2}g(z_i, \theta) - C_{2,n}(\theta)\lambda.$$

These estimation equations give a just-identified system of m equations in m unknowns and simultaneously summarize both the first order conditions for GMM estimation and the statistics that test the overidentifying restrictions.

In (21), $C_{2,n}(\theta)$ changes with n. Formally (21) is created by solving the first $m(m+1)/2 + mk$ equations in terms of θ and substituting into the last m equations

of the system

$$n^{-1}\sum_i \begin{bmatrix} \text{vec}(W_n) - \text{vec}\left(g(z_i,\theta^{(1)})g(z_i,\theta^{(1)})'\right) \\ \text{vec}(M) - \text{vec}\left(\frac{\partial g(z_i,\theta)}{\partial \theta'}\right) \\ W_n^{1/2} g(z_i,\theta) - h(M,W_n)\lambda \end{bmatrix} = 0 \tag{22}$$

where $\theta^{(1)}$ is a consistent (first round) estimate of θ_0, vec is the operator that takes the unique elements from a matrix and maps to a column vector and $h(\cdot,\cdot)$ is the continuous and differentiable function that maps from M and W_n to $C_{2,n}$. The key issue is that the functions that create the individual elements of (22) do not change with the sample size n. The law of large numbers will imply no loss of generality by restricting attention to (21) to obtain first order asymptotic results for α.

3.3 Asymptotic distribution and testing

To reduce notation the dependence on θ will now be dropped.

Definition 3.1. *Let* $g_i = g(z_i,\theta)$, $G(\theta) = n^{-1}\sum_i g_i$, $M_g = E\left[\frac{\partial g(z_i,\theta)}{\partial \theta'}\right]$, $\hat{M}_g = n^{-1}\sum_i \frac{\partial g(z_i,\theta)}{\partial \theta'}$, $\hat{\Sigma}_g = n^{-1}\sum_i g(z_i,\theta)g(z_i,\theta)'$, *and* $\Sigma_g = E\left[g(z_i,\theta)g(z_i,\theta)'\right]$. *Quantities evaluated at* $\theta = \theta_0$ *are denoted with a subscript of* 0. *Let* $g_{(j)}(z_i,\theta)$ *denote the* j^{th} *element of the vector* $g(z_i,\theta)$.

To apply the theorems from the previous section requires sufficient conditions on the standard economic model. These are given in the next assumption. The assumptions below map directly to **Assumption 2.1**. The difference is that to move from overidentified moment conditions to the just-identified estimation equations required the introduction of the λ parameters. These are defined in terms of the first derivative of the moment conditions. Hence, the moment conditions will require an additional order of differentiability and the existence of moments involving higher order derivatives. The differentiability of the spectral decomposition presented in Sowell (2007) ensures a direct mapping between the moment conditions and the estimation equations.

Assumption 3.1. *(Regularity Conditions)*

1. $\{z_i\}_{i=1}^{\infty}$ *forms an iid sequence.*
2. $\theta_0 \in \text{int}(\Theta)$ *is the unique solution to* $E\left[g(z_i,\theta)\right] = 0$, *where* Θ *is a compact subset of* R^k.
3. $g(z_i,\theta)$ *and* $\frac{\partial g(z_i,\theta)}{\partial \theta'}$ *are continuous in* θ *at each* $\theta \in \Theta$ *with probability one.*
4. $E\left[\sup_{\theta\in\Theta} \|g(z_i,\theta)\|^{2+\delta}\right] < \infty$ *and*

$E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(z_i,\theta)}{\partial\theta_j}\right\|^{2+\delta}\right] < \infty$ *for* $j = 1,\ldots,k$ *and some* $\delta > 0$ *and*

$E\left[\sup_{\theta\in\mathcal{N}}\left\|\frac{\partial^2 g(z_i,\theta)}{\partial\theta_j\partial\theta}\right\|\right] < \infty$ *for* $j = 1,\ldots,k$ *where* $\mathcal{N}$ *is an open neighborhood of* θ_0.

5. $\Sigma_{g,0}$ *is nonsingular and finite and has rank* m.

6. $g(z_i,\theta)$ *is three times continuously differentiable in* θ *in a neighborhood* $\mathcal{N}$ *of* θ_0.

7. $\text{rank}\,(M_{g0}) = k$.

8. *(i)* $E\left[h1_{(j_1)}(z_i,\theta_0)h2_{(j_2)}(z_i,\theta_0)\right]$ *is finite for* $j_1, j_2 = 1,\ldots,k$ *where* $h1_{(j)}$ *can take the functions* $\frac{\partial g_{(j)}}{\partial\theta_{\ell_3}}$ *and* $\frac{\partial^2 g_{(j)}}{\partial\theta_{\ell_2}\partial\theta_{\ell_1}}$ *for* $\ell_1,\ell_2,\ell_3 = 1,\ldots,k$ *and* $h2_{(j)}$ *can take the functions* $g_{(j)}$ *and* $\frac{\partial g_{(j)}}{\partial\theta_{\ell_4}}$ *for* $\ell_4 = 1,\ldots,k$.

   (ii) $E\left[h_{(j_1)}(z_i,\theta_0)h_{(j_2)}(z_i,\theta_0)h_{(j_3)}(z_i,\theta_0)\right]$ *is finite for* $j_1,j_2,j_3 = 1,\ldots,k$ *where each* $h_{(j)}$ *can take the functions* $g_{(j)}$ *and* $\frac{\partial g_{(j)}}{\partial\theta_\ell}$ *for* $\ell = 1,\ldots,k$.

   (iii) $E\left[h1_{(j_1)}(z_i,\theta_0)h2_{(j_2)}(z_i,\theta_0)\right]$ *is finite for* $j_1,j_2 = 1,\ldots,k$ *where* $h1_{(j)}$ *can take the functions* $\frac{\partial^2 g_{(j)}}{\partial\theta_{\ell_2}\partial\theta_{\ell_1}}$ *for* $\ell_1,\ell_2 = 1,\ldots,k$ *and* $h2_{(j)}$ *can take the functions* $\frac{\partial^2 g_{(j)}}{\partial\theta_{\ell_3}\partial\theta_{\ell_4}}$ *for* $\ell_3,\ell_4 = 1,\ldots,k$.

9. $W_n \to^p \Sigma_g^{-1}$.

The new assumptions needed for the application of the saddlepoint approximation versus the GEL estimator are the existence of higher order moments for the estimation equations and their derivatives. The assumptions applied to $g(z,\theta)$ are stronger than for $\psi(z,\alpha)$. The difference is that $g(z,\theta)$ is allowed to be an overidentified system. To achieve a just-identified system of estimation equations requires spanning the $(m-k)-$dimensional space that spans the overidentifying restrictions. This is defined in terms of the derivative of the moment equations. Because the derivative of g is needed to obtain ψ, the restrictions on the derivatives of ψ result in an additional order of differentiability for $g(z,\theta)$.

A solution to (21) can be associated with either a local maximum or a local minimum of the original GMM objective function given in equations (19). Attention must be focused on solutions associated with the local minima of the original GMM objective function. Assumption 3.1 point 2 implies that 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. there may be multiple local minima. The saddlepoint density approximates the sampling density for the location of solutions to the estimation equations. These include both local maxima and local minima. For the empirical saddlepoint density, attention is focused on the local minima by setting the saddlepoint density to zero if the original GMM objective function is concave at the θ value in α.

Another restriction on the saddlepoint density is required for most nonlinear estimation equations. The problem occurs when the empirical saddlepoint equation does not have a solution. In this case the saddlepoint density is also set equal to zero. The lack of a solution to the empirical saddlepoint equation means the observed data are inconsistent with the selected parameter value.[9]

The estimation equations for the two-step GMM problem, given in equation (21), allow the theoretical results to be specialized in more familiar and useful forms.

The results for these estimation equations are immediate applications of the general theorems presented for the ESPL estimator. Hence these will be quoted as corollaries below. The special structure implied by these estimation equations is given by $\Sigma_{\psi 0} = I_m$,

$$M_{\psi 0} = \begin{bmatrix} \Sigma_g^{-1/2} M_{g0} & -C_2(\theta_0) \end{bmatrix} \text{ and } M_{\psi 0}^{-1} = \begin{bmatrix} \left(M_{g0}' \Sigma_{g0}^{-1} M_{g0}\right)^{-1} M_{g0}' \Sigma_{g0}^{-1/2\prime} \\ -C_2(\theta_0)' \end{bmatrix}.$$

Corollary 3.1. *(First order properties) Under Assumption 3.1,* (i) *the ESPL estimator and the tilting parameter have the distribution*

$$\sqrt{n} \begin{pmatrix} \hat{\theta}_{espl} - \theta_0 \\ \hat{\lambda}_{espl} \\ \hat{\tau}_{espl} \end{pmatrix} \sim^a N \left(0, \begin{pmatrix} \left(M_{g0}' \Sigma_{g0}^{-1} M_{g0}\right)^{-1} & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{pmatrix} \right).$$

and (ii) *confidence intervals for the parameters can be created using the likelihood-ratio statistic*

$$2n \left(\mathcal{L}_n(\hat{\theta}_{espl}, \hat{\lambda}_{espl}, \hat{\tau}_{espl}) - \mathcal{L}_n(\theta, \lambda, \hat{\tau}_{espl}) \right) \sim^a \chi_m^2.$$

Confidence intervals created by the likelihood ratio statistic are dramatically different from previously results in the literature. The confidence intervals are jointly created for both the original GMM parameter θ and the parameters that test the overidentifying restrictions λ. Because the ESPL does not impose the independence of these estimates, the confidence intervals can account for the dependence in the estimators that can occur in finite samples. Of course, the other result in the Corollary shows that asymptotic distribution of θ and λ are independent.

The just-identified estimation equations simultaneously estimate the parameters of interest and the parameters that test the overidentifying restrictions. This permits a new conditional estimator of the parameters of interest conditional on the overidentifying restrictions being true. This is the CESPL estimator for the hypothesis $H_0 : \lambda = 0$.

[9]This type of restriction has occurred recently in the statistics and econometrics literature, e.g. 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.

Corollary 3.2. *(Asymptotic Distribution: Conditional)*

Under Assumption 3.1 when the parameter restriction $\lambda = 0$ is true the asymptotic distribution for the conditional parameter estimates is

$$\sqrt{n}\begin{pmatrix} \hat{\theta}_{cespl}(0) - \theta_0 \\ \hat{\tau}\left(\hat{\theta}_{cespl}(0), 0\right) \\ \hat{\gamma} \end{pmatrix} \sim^a N\left(0, \begin{pmatrix} \left(M_{g0}'\Sigma_{g0}^{-1}M_{g0}\right)^{-1} & 0 & 0 \\ 0 & \Sigma_{\psi 0}^{-1/2}P^{\perp}_{\Sigma_{g0}^{-1/2}M_{g0}}\Sigma_{\psi 0}^{-1/2\prime} & C_2(\theta_0) \\ 0 & C_2(\theta_0)' & I_{(m-k)} \end{pmatrix}\right).$$

These lead to four natural tests for the overidentifying restrictions.

1. Wald $\quad n\hat{\lambda}_{espl}'\hat{\lambda}_{espl}$
2. LR $\quad 2n\left(\mathcal{L}_n(\hat{\theta}_{espl}, \hat{\lambda}_{espl}, \hat{\tau}_{espl}) - \mathcal{L}_n(\hat{\theta}_{cespl}(0), 0, \hat{\tau}_{cespl})\right)$
3. LM/score $\quad n\hat{\gamma}'\hat{\gamma}$ or $n\frac{\partial \mathcal{L}_n(\hat{\theta}_{cespl}(0),0,\hat{\tau}_{cespl})}{\partial \lambda'}\frac{\partial \mathcal{L}_n(\hat{\theta}_{cespl}(0),0,\hat{\tau}_{cespl})}{\partial \lambda}$
4. tilting $\quad n\hat{\tau}(\hat{\theta}_{cespl}(0), 0)'\Sigma_{\psi 0}^{1/2\prime}\left(P^{\perp}_{\Sigma_{g0}^{-1/2}M_{g0}}\right)^{-g}\Sigma_{\psi 0}^{1/2}\hat{\tau}(\hat{\theta}_{cespl}(0), 0)$

   or $\quad n\hat{\tau}(\hat{\theta}_{cespl}(0), 0)'\Sigma_{\psi 0}\hat{\tau}(\hat{\theta}_{cespl}(0), 0).$

Under the null hypothesis that the moments are correctly specified each of these statistics is distributed $\chi^2_{(m-k)}$.

The inference for the parameters of interest and the validity of the overidentifying restrictions can be built on several different asymptotically equivalent covariance estimators. Following the insights provided in Imbens, Spady and Johnson (1998), the tilting parameter test of the overidentifying restrictions will use the robust estimate of the covariance $\hat{\Sigma}_{\psi 0} = V_1 V_2^{-1} V_1$ where

$$V_1 = \sum_i \hat{w}_i \psi_i(\alpha_{cespl})\psi_i(\alpha_{cespl})' \text{ and } V_2 = \sum_i n\hat{w}_i^2 \psi_i(\alpha_{cespl})\psi_i(\alpha_{cespl})'$$

with

$$\hat{w}_i = \frac{\exp\left\{\hat{\tau}_{cespl}'\psi_i(\alpha_{cespl})\right\}}{\sum_j \exp\left\{\hat{\tau}_{cespl}'\psi_j(\alpha_{cespl})\right\}}.$$

The just-identified estimation equations can also be viewed as moment conditions as in two-step GMM. This suggests another test of the overidentifying restrictions, the analogue of the J statistic using the robust estimate of the covariance matrix. The statistic $\mathbf{J}_r = n\Psi_n(\hat{\theta}_{cespl}(0), 0)'V_1V_2^{-1}V_1\Psi_n(\hat{\theta}_{cespl}(0), 0)$ uses the robust estimate of the covariance matrix and the estimation equations evaluated at the CESPL estimates, $\Psi_n(\hat{\theta}_{cespl}(0), 0)$.

4 Simulations

This section reports simulations that demonstrate the small sample performance of the **ESPL** and **CESPL** estimators relative to currently available estimators: empirical likelihood (EL), exponential tilting (ET) and exponentially tilted empirical likelihood (ETEL).

These simulations pulling together two different but related literatures on the one-step GMM estimators. One literature concerns the bias of the estimators. Theoretical results include the calculation of the higher order bias, Newey and Smith (2004) and Schennach (2007). Empirical results include the bias from simulated models, Schennach (2007). The conclusions are that the smallest higher order bias are associated with the EL and ETEL estimators. The ET estimator appears to have a larger bias. The other literature concerns testing overidentifying restrictions. The theoretical work include the presentation of different tests that all have the same asymptotic distribution under the null hypothesis, Imbens, Spady and Johnson (1998), Newey and Smith (2004), Schennach (2007). The empirical results include the small sample performance of different tests statistics for different models, Imbens, Spady and Johnson (1998). The basic conclusion is that the best agreement with the asymptotic results occurs with a test statistic that is built on the ET estimator.

There is a tension in the literature because the lowest bias is associated with estimators that do not produce desirable tests for the overidentifying restrictions. Alternatively, the best test for the overidentifying restrictions is associated with a parameter estimated that tends to have higher bias. The simulations reported below remove this tension by demonstrating that the ESPL and the CESPL estimators have smaller bias than the one-step estimators and that tests build on the CESPL estimator have comparable or better performance than currently available tests.

4.1 The model

The model was first presented in Hall and Horowitz (1996) to demonstrate the superior performance of the bootstrap and has been used in Imbens, Spady and Johnson (1998), Kitamura (2001) and Schennach (2007) to judge the performance of different estimators and tests of overidentifying restrictions. This model can be interpreted as a simplified asset pricing model (Gregory, Lamarche and Smith (2002)). Schennach (2007) expanded the model to allow for an arbitrary number of moment conditions. The one parameter model has the moments

$$g_i(\theta) = \begin{bmatrix} \exp\{\mu - \theta(x_i + y_i) + 3y_i\} - 1 \\ y_i\left(\exp\{\mu - \theta(x_i + y_i) + 3y_i\} - 1\right) \\ (z_{i3}^2 - 1)\left(\exp\{\mu - \theta(x_i + y_i) + 3y_i\} - 1\right) \\ \vdots \\ (z_{im}^2 - 1)\left(\exp\{\mu - \theta(x_i + y_i) + 3y_i\} - 1\right) \end{bmatrix}$$

where $\theta_0 = 3$, x_i and y_i are iid from $N(0, .16)$, z_{ij} are iid $N(0, 1)$ for $j = 3, \ldots, m$ and $\mu = -.72$ is known. Implementation details concerning how the **ESPL** estimator was calculated are provided in the appendix.

4.2 Bias

Table 1 reports the bias for different sample sizes and different numbers of moment conditions for 10000 simulated samples. **The ESPL** dominates the other estimators with its bias at least 66% smaller and often an order of magnitude smaller.

Model		Methods				
m	n	**ESPL**	**CESPL**	EL	ETEL	ET
2	50	-.012	0.061	0.113	0.112	0.155
2	100	-.007	0.020	0.057	0.056	0.073
2	200	-.009	0.007	0.029	0.029	0.037
4	200	-.004	0.048	0.059	0.055	0.098
10	200	-.036	0.125	0.137	0.108	0.236

Table 1: The bias of the **ESPL**, **CESPL**, EL, ETEL and ET estimators for the Hall-Horowitz model. The sample size is denoted n and the number of moment conditions is m. When $m = 2$ only the first two moment conditions are used.

For each simulated model, the standard errors for the different estimates are all comparable. The reduction in bias did not result in increased variability.

Representative c.d.f. of the **ESPL**, **CESPL**, EL, ETEL and ET estimators of θ are presented in Figure 1. These models were selected so that the plots would be comparable to Figure 1 in Schennach (2007).

The most striking feature of the table and the figure is the lower sensitivity of the **ESPL** bias as the sample size decreases and the number of moment conditions increases. In both dimensions, the magnitude of the deterioration is much more dramatic for the other estimators.

4.3 Tests of overidentifying restrictions

For each simulated sample seven different tests of the overidentifying restrictions were calculated.

1. ET CF. The ET criteria function test.

2. $\mathrm{ET_r}$ The ET tilting parameter tests using the robust covariance estimator. In Imbens, Spady and Johnson (1998) this test was denoted $T^{LM}_{et(r)}$.

3. **Wald**. The Wald statistic from the unconditional ESPL estimation problem.

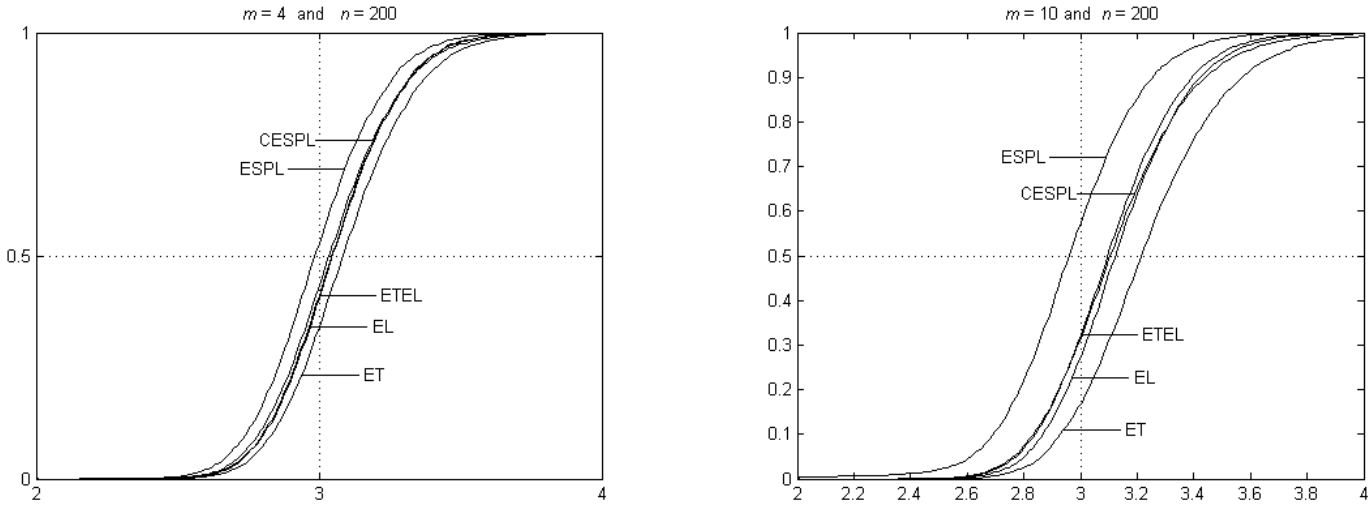

Figure 1: Cumulative distribution functions for the **ESPL**, **CESPL**, EL, ETEL and ET estimators of θ in the Hall and Horowitz model. The sample size is denoted by n and the number of moment conditions is m. These c.d.f.'s were calculated using 10000 simulated samples.

4. **LR**. The likelihood ratio statistic using the optimal objective function values from the unconditional and conditional ESPL estimation problems.

5. **Score**. The score/LM statistic from the CESPL estimation problem.

6. $\mathbf{Tilt}_r$. The tilting parameter test statistic from the CESPL estimation problem using the robust estimate of the covariance estimator.

7. $\mathbf{J}_r$. The J statistic using the just-identified estimation equations evaluated at the CESPL estimates with the robust estimate of the covariance matrix.

The ET statistics are presented so that the results will be comparable to the results reported in Imbens, Spady and Johnson (1998) where the $\mathrm{ET_r}$ was shown to have desirable performance. For each test the empirical size was calculated, and these are recorded in Table 2. In the table, an underline denotes the test with empirical size closest to the nominal size.

The table confirms that the new tests have comparable or better performance relative to currently available tests. The performance of the $\mathbf{Tilt}_r$ statistic is almost indistinguishable from the $\mathrm{ET_r}$ statistic advocated in Imbens, Spady and Johnson (1998). Of course the advantage of the $\mathbf{Tilt}_r$ statistic is that is built on the CESPL estimator which has smaller bias than the ET estimator.

A striking feature is the performance of the $\mathbf{J}_\mathrm{r}$ statistic across all the models. It is the best statistic for most models and size levels. It is always one of the top three

Model	Statistics	Size						
		.200	.100	.050	.025	.010	.005	.001
	ET CF	.302	.198	.136	.099	.069	.053	.028
	ET_r	.300	.173	.100	.059	.031	.021	.007
$n = 50$	**Wald**	.262	.177	.122	.086	.059	.040	.019
$m = 2$	**LR**	.281	.190	.133	.097	.066	.049	.027
	Score	.289	.192	.133	.097	.068	.053	.033
	$\mathbf{Tilt}_r$	.300	.168	.098	.059	.031	.022	.009
	$\mathbf{J}_r$	.225	.144	.088	.048	.021	.014	.008
	ET CF	.275	.173	.116	.081	.054	.040	.020
	ET_r	.272	.152	.084	.048	.024	.015	.004
$n = 100$	**Wald**	.239	.157	.110	.080	.051	.036	.016
$m = 2$	**LR**	.251	.160	.110	.077	.050	.036	.017
	Score	.258	.164	.108	.076	.047	.034	.016
	$\mathbf{Tilt}_r$	.272	.146	.082	.047	.023	.014	.004
	$\mathbf{J}_r$	.210	.135	.093	.062	.034	.021	.006
	ET CF	.249	.145	.088	.057	.034	.025	.012
	ET_r	.248	.128	.067	.035	.016	.009	.002
$n = 200$	**Wald**	.229	.142	.094	.064	.041	.031	.014
$m = 2$	**LR**	.237	.141	.090	.056	.035	.025	.012
	Score	.239	.138	.080	.052	.029	.020	.008
	$\mathbf{Tilt}_r$	.246	.125	.065	.034	.015	.008	.002
	$\mathbf{J}_r$	.201	.116	.072	.046	.027	.018	.006
	ET CF	.374	.257	.178	.126	.082	.060	.031
	ET_r	.341	.207	.126	.077	.042	.026	.010
$n = 200$	**Wald**	.328	.222	.152	.105	.066	.044	.018
$m = 4$	**LR**	.351	.241	.168	.124	.079	.058	.030
	Score	.191	.129	.092	.067	.047	.036	.021
	$\mathbf{Tilt}_r$	.329	.197	.120	.073	.039	.025	.009
	$\mathbf{J}_r$	.236	.130	.071	.037	.017	.009	.003
	ET CF	.617	.486	.379	.293	.222	.177	.102
	ET_r	.532	.376	.253	.180	.109	.075	.029
$n = 200$	**Wald**	.432	.305	.209	.144	.088	.059	.024
$m = 10$	**LR**	.569	.449	.347	.280	.208	.170	.099
	Score	.320	.254	.204	.168	.137	.115	.082
	$\mathbf{Tilt}_r$	.516	.358	.244	.173	.105	.071	.028
	$\mathbf{J}_r$	.129	.049	.024	.011	.006	.003	.002

Table 2: The empirical size for tests of the overidentifying restrictions of the Hall and Horowitz model. The sample size is denoted by n and the number of moment conditions is m.

statistics. The other striking feature is the robust performance of the $\mathbf{J}_{\mathrm{r}}$ statistic as the number of overidentifying restrictions increases. The other statistics' performance decays dramatically. Note that they are oversized for the $m = 10$ model. However, the $\mathbf{J}_{\mathrm{r}}$ statistic is undersized but is in much closer agreement with the nominal size.

Figure 2 summarized the performance of three test statistics the ET_{r} statistic, the $\mathbf{Tilt}_{\mathrm{r}}$ statistic and the $\mathbf{J}_{\mathrm{r}}$ statistic. These statistics are compared using QQ-plots (Quantile-Quantile plots) that record the quantiles for the simulated test statistics versus the quantiles for the asymptotic distribution of the test statistic under the null. These models and statistics were selected to be comparable to Figure 2 and Figure 3 in Imbens, Spady and Johnson (1998). The plots more clearly show the performance of the $\mathbf{Tilt}_r$ statistic is almost indistinguishable from the ET_{r}. In addition, the plots demonstrate the robustness of the $\mathbf{J}_{\mathrm{r}}$ statistic as the number of overidentifying restrictions increase. This is in contrast to the performance of the ET_{r} which deteriorates as the number of overidentifying restrictions increases.

The new tests for the overidentifying restrictions are at least comparable or better than currently available statistics. It is reassuring that dramatic reduction in the bias for the parameters of interest did not result in a reduction in the performance of these statistics.

4.4 Lessons from simulations

These simulations suggest that the ESPL is the prefered method to estimate and test. The smallest bias for the parameter estimates are obtained with the unconditional ESPL estimator. The most accurate test for the validity of the overidentifying restrictions is obtained with the CESPL estimator with the $\mathbf{Tilt}_{\mathrm{r}}$ and the $\mathbf{J}_{\mathrm{r}}$ statistics. The properties of the ET_{r} statistics is almost indistinguishable from the $\mathbf{Tilt}_{\mathrm{r}}$ statistic, however this statistic is built on the ET estimate that is always dominated by the ESPL and the CESPL estimates.

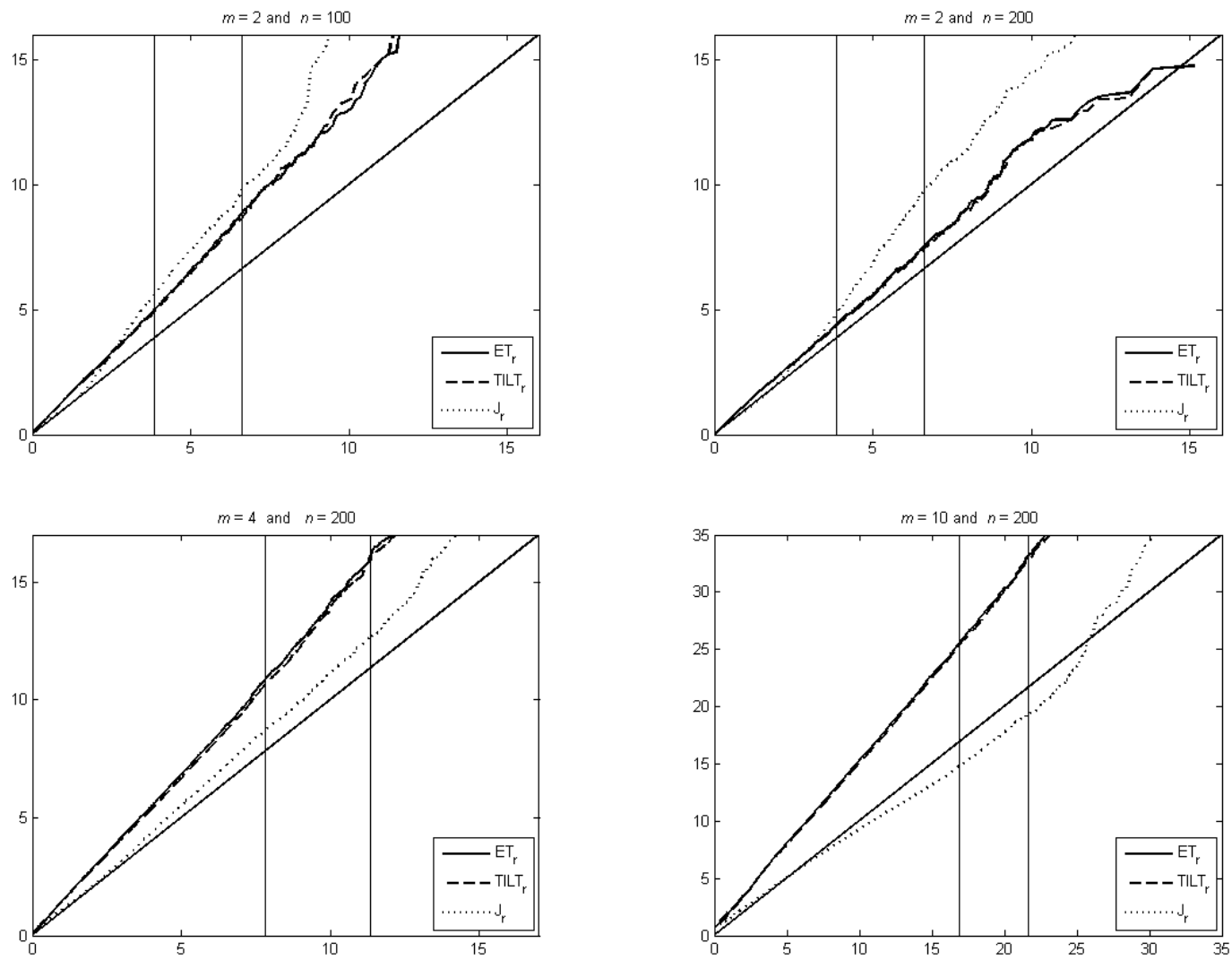

Figure 2: QQplots for tests of the overidentifying restrictions for the Hall and Horowitz model. The sample size is denoted by n and the number of moment conditions is m. These were calculated using 10000 simulated samples. Vertical lines show the nominal .95 and .99 levels, and the 45 degree line would represent perfect agreement.

5 Conclusion

This ESPL estimator has been introduced and the first order sampling properties are shown to be equivalent to efficient two-step GMM. The higher order bias is shown to be different from the higher order bias of the GEL estimators. For many commonly studied models, the higher order bias of the ESPL estimator will be smaller. The ESPL estimator is a maximum likelihood estimator and hence leads to the four commonly used test statistics: the likelihood ratio, the Wald, the score and the tilting.

As an application, the ESPL approach is used to investigate a generic econometric model that is specified as a set of overidentified moment conditions. New parameters are added so that the overidentified system of moment conditions is nested in a just identified system of estimation equations. Hypothesis tests of the new parameters become new statistics that test the overidentifying restrictions.

The theoretical results and the practicality of the estimators and tests are demonstrated by simulating the commonly studied Hall and Horowitz model. To facilitate comparisons, the model and parameter values are selected to exactly match previously published results. The estimation equations are created from the first order conditions from efficient two-step GMM. The results are that the ESPL and CESPL estimators have smaller bias than the currently available one-step GMM estimators, EL, ETEL and ET. In addition, the simulations show that new tests for the overidentifying restrictions have empirical size that is comparable or better than currently available tests statistics.

The ESPL can be thought of as a natural extension of three different literatures: the saddlepoint density, nonparametric maximum likelihood and information theoretic estimation. Hence this paper simultaneously extends and brings together these literatures. Previous researchers have noted some similarities, eg. tests using terms from the empirical saddlepoint and tests using the LR from the EL are compared in Monti and Ronchetti (1993). The definition of the ESPL estimator and method of proof used in this paper have further revealed these relationships.

Additional work remains. The ESPL can be calculated for alternative sets of moment conditions. The simulations demonstrate the improvement achieved by using the ESPL estimator applied to the moment conditions from two-step GMM. It is natural to expect some additional improvement if the first order conditions from a GEL estimator is used. The behavior of the EPSL estimator relative to other estimators should be investigated for other models through simulation. To increase its usefulness of the ESPL estimator should be extended to situations where the data have time dependence.

6 Appendix

6.1 Proofs

Proof of **Theorem 2.1**.

The first order conditions for the Lagrangian

$$\ell(w_1, \ldots, w_n, \tau, \mu) = \sum_i w_i \ln(w_i) - \tau' \left(\sum_i w_i \psi_i \right) + \mu \left(\sum_i w_i - 1 \right)$$

are

$$\frac{\partial \ell(\hat{w}_1, \ldots, \hat{w}_n, \hat{\tau}_n, \hat{\mu})}{\partial w_i} = 1 + \ln(\hat{w}_i) - \hat{\tau}_n' \psi_i + \hat{\mu} = \underset{1\times 1}{0} \text{ for } i = 1, \ldots, n \tag{23}$$

$$\frac{\partial \ell(\hat{w}_1, \ldots, \hat{w}_n, \hat{\tau}_n, \hat{\mu})}{\partial \mu} = \sum_i \hat{w}_i - 1 = \underset{1\times 1}{0} \tag{24}$$

$$\frac{\partial \ell(\hat{w}_1, \ldots, \hat{w}_n, \hat{\tau}_n, \hat{\mu})}{\partial \tau} = -\sum_i \hat{w}_i \psi_i = \underset{m\times 1}{0}. \tag{25}$$

For each i equations (23) can be solved for $\hat{w}_i$. Substitution into (24) gives a single equation that can be solved for $\hat{\mu}$. Substitute back into the function for $\hat{w}_i$ gives the weights as function of α and $\hat{\tau}_n$, $\hat{w}_i = \frac{\exp\{\hat{\tau}_n' \psi_i\}}{\sum_j \exp\{\hat{\tau}_n' \psi_j\}}$. To reduce notation the hat will now be dropped from $\hat{\tau}_n$. Substitute $\hat{w}_i$ into equation (25) to obtain m equations that can be equivalently written as the saddlepoint equation

$$\sum_i \psi_i \exp\{\tau_n' \psi_i\} = 0. \tag{26}$$

Substitute into the normalized log of the saddlepoint density to obtain

$$
\begin{aligned}
\mathcal{L}_n(\alpha,\tau_n) &= -\frac{1}{2n}\ln\left(\left|\sum_i \frac{\exp\{\tau_n'\psi_i\}}{\sum_j \exp\{\tau_n'\psi_j\}}\psi_i\psi_i'\right|\right) + \frac{1}{n}\ln\left(\left|\sum_i \frac{\exp\{\tau_n'\psi_i\}}{\sum_j \exp\{\tau_n'\psi_j\}}\frac{\partial\psi_i}{\partial\alpha'}\right|\right) \\
&\quad + \ln\left(\frac{1}{n}\sum_i \exp\{\tau_n'\psi_i\}\right) \\
&= -\frac{1}{2n}\ln\left(\left(\frac{1}{\frac{1}{n}\sum_j \exp\{\tau_n'\psi_j\}}\right)^m \left|\frac{1}{n}\sum_i \exp\{\tau_n'\psi_i\}\,\psi_i\psi_i'\right|\right) \\
&\quad + \frac{1}{n}\ln\left(\left(\frac{1}{\frac{1}{n}\sum_j \exp\{\tau_n'\psi_j\}}\right)^m \left|\frac{1}{n}\sum_i \exp\{\tau_n'\psi_i\}\,\frac{\partial\psi_i}{\partial\alpha'}\right|\right) \\
&\quad + \ln\left(\frac{1}{n}\sum_i \exp\{\tau_n'\psi_i\}\right) \\
&= -\frac{1}{2n}\ln\left(\left|\frac{1}{n}\sum_i \exp\{\tau_n'\psi_i\}\,\psi_i\psi_i'\right|\right) + \frac{1}{n}\ln\left(\left|\frac{1}{n}\sum_i \exp\{\tau_n'\psi_i\}\,\frac{\partial\psi_i}{\partial\alpha'}\right|\right) \\
&\quad + \left(1-\frac{m}{2n}\right)\ \ln\left(\frac{1}{n}\sum_i \exp\{\tau_n'\psi_i\}\right). \qquad (27)
\end{aligned}
$$

Hence maximizing equation (27) with respect to α subject to (26) is an alternative way to define the ESPL estimator.

The saddlepoint equation (26) can be viewed as $m-$equations in $2m$ unknowns, α and τ. The derivative of (27) with respect to α set equal to zero gives a different system of $m-$equations in the $2m-$unknowns. Together the first order conditions that define the ESPL estimator are given by the system $2m$ equations.

Alternatively, the implicit function theorem can be used to solve equation (26) for τ as a function of α. This implicit function will be denoted $\tau_n(\alpha)$. Substituting the implicit function into the derivative of (27) with respect to α set equal to zero gives the ESPL as the solution to a system of $m-$equations. ■

The implicit function $\tau_n(\alpha)$ will appear in future proofs. Its first and second derivatives evaluated at the population parameter values will be needed. The next Lemma gives the needed results.

Lemma 6.1. *Under Assumption 2.1 the derivatives of the implicit function defined by the saddlepoint equations satisfy*

$$
\frac{\partial\tau_n(\alpha_0)}{\partial\alpha'} = -\hat{\Sigma}_{\psi_0}^{-1}\hat{M}_{\psi 0} = \mathcal{O}_p(1)
$$

and

$$\begin{aligned}\frac{\partial^2\tau_n(\alpha_0)}{\partial\alpha\partial\alpha'} &= -\hat{\Sigma}_{\psi 0}^{-1}\frac{1}{n}\sum_i\left(\frac{\partial^2\psi_{i0}}{\partial\alpha\partial\alpha'}-\psi_{i0}\hat{\Sigma}_{\psi 0}^{-1}\hat{M}_{\psi 0}\frac{\psi_{i0}'}{\partial\alpha}\right)\\ &\quad +\hat{\Sigma}_{\psi 0}^{-1}\frac{1}{n}\sum_i\left(\frac{\partial\psi_{i0}}{\partial\alpha}\psi_{i0}'+\psi_{i0}\frac{\partial\psi_{i0}'}{\partial\alpha}\right)\hat{\Sigma}_{\psi 0}^{-1}\hat{M}_{\psi 0}\\ &= \mathcal{O}_p(1).\end{aligned}$$

Proof.

The implicit function theorem implies that in a neighborhood of α

$$\frac{\partial\tau_n(\alpha)}{\partial\alpha'} = -\left[\frac{1}{n}\sum_i\exp\left\{\tau'\psi_i\right\}\psi_i\psi_i'\right]^{-1}\left[\frac{1}{n}\sum_i\exp\left\{\tau'\psi_i\right\}\left(\frac{\partial\psi_i}{\partial\alpha'}+\psi_i\tau'\frac{\partial\psi_i}{\partial\alpha'}\right)\right]$$

which evaluated at α_0 and $\tau_0 = 0$ gives

$$\frac{\partial\tau_n(\alpha_0)}{\partial\alpha'} = -\hat{\Sigma}_{\psi_0}^{-1}\hat{M}_{\psi 0}$$

which converges to the finite vector $-\Sigma_{\psi_0}^{-1}M_{\psi 0}$. Differentiation of the first derivative with similar reasoning gives the second result. ■

Proof of **Theorem 2.2** .

To show the similarity with the literature, the same method of proof used in Schennach (2007) will be followed. The asymptotic properties of the ESPL estimator will be equivalent to the asymptotic properties of the EL estimator. This will be established by showing that the ESPL objective function is equivalent to the EL objective function in a neighborhood of the population parameter value. Formally it will be shown that (i) in a $\mathcal{O}(n^{-1/2})$ neighborhood of $\alpha = \alpha_0$ the expansion for the ESPL objective function $\mathcal{L}_n(\alpha, \tau_n(\alpha))$ is the same as the expansion of the EL objective function for terms that converge slower than $\mathcal{O}_p\left(n^{-1}\right)$ and (ii) that the expansion of the first order conditions for α and τ around $\alpha = \alpha_0$ and $\tau = \tau_0 = 0$ is identical to that for the EL estimator for terms that converge slower than $\mathcal{O}_p(n^{-1/2})$ in a $\mathcal{O}(n^{-1/2})$ neighborhood of $\alpha = \alpha_0$ and $\tau = 0$.

These two results will mean that asymptotically the EL estimator will solve the ESPL first order conditions. Since the Lagrange multiplier in the EL optimization problem converges in probability to zero and the EL and ESPL objective functions asymptotically converge to their maximum possible values when the Lagrange multiplier and τ are zero respectively, in a neighborhood of α_0 there can only be one solution to the first order conditions. Thus, the ESPL estimator inherits the first order properties of EL as presented in Owen (1990) and Newey and Smith (2004).

The expansion of the objective function $\mathcal{L}_n(\alpha, \tau_n(\alpha))$.

The proof proceeds by expanding the objective function in a three term Taylor series about the population parameter α_0. To simplify the presentation, the objective function can be written as the sum of three functions

$$\mathcal{L}_n(\alpha, \tau_n(\alpha)) = T_{1,n}(\alpha) + T_{2,n}(\alpha) + T_{3,n}(\alpha)$$

and each function will be expanded individually. The expansion requires these functions and their first two derivatives evaluated at α_0. First the needed terms will be calculated and then combined to obtain the needed expansion.

The first function is

$$T_{1,n}(\alpha) = -\frac{1}{2n} \ln \left(\left| \frac{1}{n} \sum_i \exp\left\{\hat{\tau}_n' \psi_i\right\} \psi_i \psi_i' \right| \right)$$

which evaluated at α_0 is

$$T_{1,n}(\alpha_0) \;=\; -\frac{1}{2n} \ln \left(\left| \frac{1}{n} \sum_i \exp\left\{\tau_0' \psi_{i0}\right\} \psi_{i0} \psi_{i0}' \right| \right) = \mathcal{O}_p(n^{-1}).$$

The first derivative of the first function is

$$\begin{aligned} \frac{\partial T_{1,n}(\alpha)}{\partial \alpha_j} \;=\; & -\frac{1}{2n} \mathrm{tr} \left\{ \left[\frac{1}{n} \sum_i \exp\left\{\hat{\tau}_n' \psi_i\right\} \psi_i \psi_i' \right]^{-1} \right. \\ & \times \left[\frac{1}{n} \sum_i \exp\left\{\hat{\tau}_n' \psi_i\right\} \left\{ \frac{\partial \psi_i}{\partial \alpha_j} \psi_i' + \psi_i \frac{\partial \psi_i'}{\partial \alpha_j} \right\} \right. \\ & \left. \left. + \frac{1}{n} \sum_i \exp\left\{\hat{\tau}_n \psi_i\right\} \left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_j} \psi_i + \hat{\tau}_n' \frac{\partial \psi_i}{\partial \alpha_j} \right) \psi_i \psi_i' \right] \right\} \end{aligned}$$

which evaluated at α_0 and using **Lemma 6.1** is

$$\begin{aligned} \frac{\partial T_{1,n}(\alpha_0)}{\partial \alpha_j} \;=\; & -\frac{1}{2n} \mathrm{tr} \left\{ \left[\frac{1}{n} \sum_i \psi_{i0} \psi_{i0}' \right]^{-1} \right. \\ & \left. \times \left[\frac{1}{n} \sum_i \left\{ \frac{\partial \psi_{i0}}{\partial \alpha_j} \psi_{i0}' + \psi_{i0} \frac{\partial \psi_{i0}'}{\partial \alpha_j} \right\} + \frac{1}{n} \sum_i \left(\frac{\partial \hat{\tau}_{n0}'}{\partial \alpha_j} \psi_{i0} \right) \psi_{i0} \psi_{i0}' \right] \right\} \\ =\; & \mathcal{O}_p(n^{-1}). \end{aligned}$$

The second derivative of the first function is

$$
\begin{aligned}
\frac{\partial^2 T_{1,n}(\alpha)}{\partial\alpha_\ell\partial\alpha_j} = -\frac{1}{2n}\mathrm{tr}\Bigg\{ & -\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\psi_i\psi_i'\right]^{-1} \\
& \times\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\psi_i}{\partial\alpha_\ell}\psi_i' + \psi_i\frac{\partial\psi_i'}{\partial\alpha_\ell}\right) + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_\ell}\psi_i + \hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_\ell}\right)\psi_i\psi_i'\right] \\
& \times\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\psi_i\psi_i'\right]^{-1} \\
& \quad\times\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\psi_i}{\partial\alpha_j}\psi_i' + \psi_i\frac{\partial\psi_i'}{\partial\alpha_j}\right) + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_j}\right)\psi_i\psi_i'\right] \\
& + \left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n\psi_i\right\}\psi_i\psi_i'\right]^{-1} \\
& \quad\times\Bigg[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial^2\psi_i}{\partial\alpha_\ell\partial\alpha_j}\psi_i' + \frac{\partial\psi_i}{\partial\alpha_j}\frac{\partial\psi_i'}{\partial\alpha_\ell} + \frac{\partial\psi_i}{\partial\alpha_\ell}\frac{\partial\psi_i'}{\partial\alpha_j} + \psi_i\frac{\partial^2\psi_i'}{\partial\alpha_\ell\partial\alpha_j}\right) \\
& \qquad + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_\ell}\psi_i + \hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_\ell}\right)\left(\frac{\partial\psi_i}{\partial\alpha_j}\psi_i' + \psi_i\frac{\partial\psi_i'}{\partial\alpha_j}\right) \\
& \qquad + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_j}\right)\left(\frac{\partial\psi_i}{\partial\alpha_\ell}\psi_i' + \psi_i\frac{\partial\psi_i'}{\partial\alpha_\ell}\right) \\
& \qquad + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_j}\frac{\partial\psi_i}{\partial\alpha_\ell} + \frac{\partial^2\hat{\tau}_n'}{\partial\alpha_\ell\partial\alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial^2\psi_i}{\partial\alpha_\ell\partial\alpha_j} + \frac{\partial\hat{\tau}_n'}{\partial\alpha_\ell}\frac{\partial\psi_i}{\partial\alpha_j}\right)\psi_i\psi_i' \\
& \qquad + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_\ell}\psi_i + \hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_\ell}\right)\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_j}\right)\psi_i\psi_i'\Bigg]\Bigg\}
\end{aligned}
$$

which evaluated at α_0 and using **Lemma 6.1** is

$$
\begin{aligned}
\frac{\partial^2 T_{1,n}(\alpha_0)}{\partial\alpha_\ell\partial\alpha_j} &= -\frac{1}{2n}\mathrm{tr}\left\{-\left[\frac{1}{n}\sum_i \psi_{i0}\psi_{i0}'\right]^{-1}\right. \\
&\times\left[\frac{1}{n}\sum_i\left(\frac{\partial\psi_{i0}}{\partial\alpha_\ell}\psi_{i0}'+\psi_{i0}\frac{\partial\psi_{i0}'}{\partial\alpha_\ell}\right)+\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_\ell}\psi_{i0}\psi_{i0}\psi_{i0}'\right] \\
&\times\left[\frac{1}{n}\sum_i\psi_{i0}\psi_{i0}'\right]^{-1}\left[\frac{1}{n}\sum_i\left(\frac{\partial\psi_{i0}}{\partial\alpha_j}\psi_{i0}'+\psi_{i0}\frac{\partial\psi_{i0}'}{\partial\alpha_j}\right)+\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\psi_{i0}\psi_{i0}\psi_{i0}'\right] \\
&+\left[\frac{1}{n}\sum_i\psi_{i0}\psi_{i0}'\right]^{-1} \\
&\quad\times\left[\frac{1}{n}\sum_i\left(\frac{\partial^2\psi_{i0}}{\partial\alpha_\ell\partial\alpha_j}\psi_{i0}'+\frac{\partial\psi_{i0}}{\partial\alpha_j}\frac{\partial\psi_{i0}'}{\partial\alpha_\ell}+\frac{\partial\psi_{i0}}{\partial\alpha_\ell}\frac{\partial\psi_{i0}'}{\partial\alpha_j}+\psi_{i0}\frac{\partial^2\psi_{i0}'}{\partial\alpha_j\partial\alpha_\ell}\right)\right. \\
&\qquad+\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_\ell}\psi_{i0}\left(\frac{\partial\psi_{i0}}{\partial\alpha_j}\psi_{i0}'+\psi_{i0}\frac{\partial\psi_{i0}'}{\partial\alpha_j}\right) \\
&\qquad+\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\psi_{i0}\left(\frac{\partial\psi_{i0}}{\partial\alpha_\ell}\psi_{i0}'+\psi_{i0}\frac{\partial\psi_{i0}'}{\partial\alpha_\ell}\right) \\
&\qquad+\frac{1}{n}\sum_i\left(\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\frac{\partial\psi_{i0}}{\partial\alpha_\ell}+\frac{\partial^2\hat{\tau}_{n0}'}{\partial\alpha_\ell\partial\alpha_j}\psi_{i0}+\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_\ell}\frac{\partial\psi_{i0}}{\partial\alpha_j}\right)\psi_{i0}\psi_{i0}' \\
&\qquad\left.\left.+\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_\ell}\psi_{i0}\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\psi_{i0}\psi_{i0}\psi_{i0}'\right]\right\} \\
&= \mathcal{O}_p\left(n^{-1}\right).
\end{aligned}
$$

The second function is

$$
T_{2,n}(\alpha)=\frac{1}{n}\ln\left(\left|\frac{1}{n}\sum_i\exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial\psi_i}{\partial\alpha}\right|\right)
$$

which evaluated at α_0 is

$$
T_{2,n}(\alpha_0) = \frac{1}{n}\ln\left(\left|\frac{1}{n}\sum_i\frac{\partial\psi_{i0}}{\partial\alpha'}\right|\right)=\mathcal{O}_p(n^{-1}).
$$

The first derivative of the second function is

$$\frac{\partial T_{2,n}(\alpha)}{\partial \alpha_j} = \frac{1}{n}\mathrm{tr}\left\{\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial \psi_i}{\partial \alpha}\right]^{-1}\right.$$
$$\left.\times\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial^2 \psi_i}{\partial \alpha_j \partial \alpha} + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial \psi_i}{\partial \alpha_j}\right)\frac{\partial \psi_i}{\partial \alpha}\right]\right\}$$

which evaluated at α_0 and using **Lemma 6.1** is

$$\begin{aligned}\frac{\partial T_{2,n}(\alpha_0)}{\partial \alpha_j} &= \frac{1}{n}\mathrm{tr}\left\{\left[\frac{1}{n}\sum_i \frac{\partial \psi_{i0}}{\partial \alpha'}\right]^{-1}\left[\frac{1}{n}\sum_i \frac{\partial^2 \psi_{i0}}{\partial \alpha_j \partial \alpha'} + \frac{1}{n}\sum_i \left(\frac{\partial \hat{\tau}_{n0}'}{\partial \alpha_j}\psi_i\right)\frac{\partial \psi_{i0}}{\partial \alpha'}\right]\right\} \\ &= \mathcal{O}_p(n^{-1}).\end{aligned}$$

The second derivative of the second function is

$$\frac{\partial^2 T_{2,n}(\alpha)}{\partial \alpha_\ell \partial \alpha_j} = \frac{1}{n}\mathrm{tr}\left\{-\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial \psi_i}{\partial \alpha}\right]^{-1}\right.$$
$$\left\{\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial^2 \psi_i}{\partial \alpha_\ell \partial \alpha} + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_\ell}\psi_i + \hat{\tau}_n'\frac{\partial \psi_i}{\partial \alpha_\ell}\right)\frac{\partial \psi_i}{\partial \alpha}\right\}$$
$$\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial \psi_i}{\partial \alpha}\right]^{-1}$$
$$\times\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial^2 \psi_i}{\partial \alpha_j \partial \alpha} + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial \psi_i}{\partial \alpha_j}\right)\frac{\partial \psi_i}{\partial \alpha}\right]$$
$$+\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial \psi_i}{\partial \alpha}\right]^{-1}$$
$$\times\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\frac{\partial^3 \psi_i}{\partial \alpha_\ell \partial \alpha_j \partial \alpha} + \frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_\ell}\psi_i + \hat{\tau}_n'\frac{\partial \psi_i}{\partial \alpha_\ell}\right)\frac{\partial^2 \psi_i}{\partial \alpha_j \partial \alpha}\right.$$
$$+\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial \psi_i}{\partial \alpha_j}\right)\frac{\partial^2 \psi_i}{\partial \alpha_\ell \partial \alpha}$$
$$+\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_j}\frac{\partial \psi_i}{\partial \alpha_\ell} + \frac{\partial^2 \hat{\tau}_n'}{\partial \alpha_\ell \partial \alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial^2 \psi_i}{\partial \alpha_\ell \partial \alpha_j} + \frac{\partial \hat{\tau}_n'}{\partial \alpha_\ell}\frac{\partial \psi_i}{\partial \alpha_j}\right)\frac{\partial \psi_i}{\partial \alpha}$$
$$\left.\left.+\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_\ell}\psi_i + \hat{\tau}_n'\frac{\partial \psi_i}{\partial \alpha_\ell}\right)\left(\frac{\partial \hat{\tau}_n'}{\partial \alpha_j}\psi_i + \hat{\tau}_n'\frac{\partial \psi_i}{\partial \alpha_j}\right)\frac{\partial \psi_i}{\partial \alpha}\right]\right\}$$

which evaluated at α_0 and using **Lemma 6.1** is

$$\begin{aligned}
\frac{\partial^2 T_{2,n}(\alpha_0)}{\partial\alpha_\ell\partial\alpha_j} &= \frac{1}{n}\mathrm{tr}\left\{-\left[\frac{1}{n}\sum_i\frac{\partial\psi_{i0}}{\partial\alpha}\right]^{-1}\left\{\frac{1}{n}\sum_i\frac{\partial^2\psi_{i0}}{\partial\alpha_\ell\partial\alpha}+\frac{1}{n}\sum_i\frac{\partial\hat\tau'_{n0}}{\partial\alpha_\ell}\psi_{i0}\frac{\partial\psi_{i0}}{\partial\alpha}\right\}\right. \\
&\quad\times\left[\frac{1}{n}\sum_i\frac{\partial\psi_{i0}}{\partial\alpha}\right]^{-1}\left[\frac{1}{n}\sum_i\frac{\partial^2\psi_{i0}}{\partial\alpha_j\partial\alpha}+\frac{1}{n}\sum_i\frac{\partial\hat\tau'_{n0}}{\partial\alpha_j}\psi_{i0}\frac{\partial\psi_{i0}}{\partial\alpha}\right] \\
&\quad+\left[\frac{1}{n}\sum_i\frac{\partial\psi_{i0}}{\partial\alpha}\right]^{-1}\left[\frac{1}{n}\sum_i\frac{\partial^3\psi_{i0}}{\partial\alpha_\ell\partial\alpha_j\partial\alpha}+\frac{1}{n}\sum_i\frac{\partial\hat\tau'_{n0}}{\partial\alpha_\ell}\psi_{i0}\frac{\partial^2\psi_{i0}}{\partial\alpha_j\partial\alpha}\right. \\
&\quad+\frac{1}{n}\sum_i\frac{\partial\hat\tau'_{n0}}{\partial\alpha_j}\psi_{i0}\frac{\partial^2\psi_{i0}}{\partial\alpha_\ell\partial\alpha} \\
&\quad+\frac{1}{n}\sum_i\left(\frac{\partial\hat\tau'_{n0}}{\partial\alpha_j}\frac{\partial\psi_{i0}}{\partial\alpha_\ell}+\frac{\partial^2\hat\tau'_{n0}}{\partial\alpha_\ell\partial\alpha_j}\psi_{i0}+\frac{\partial\hat\tau'_{n0}}{\partial\alpha_\ell}\frac{\partial\psi_{i0}}{\partial\alpha_j}\right)\frac{\partial\psi_{i0}}{\partial\alpha} \\
&\quad\left.\left.+\frac{1}{n}\sum_i\frac{\partial\hat\tau'_{n0}}{\partial\alpha_\ell}\psi_{i0}\frac{\partial\hat\tau'_{n0}}{\partial\alpha_j}\psi_{i0}\frac{\partial\psi_{i0}}{\partial\alpha}\right]\right\} \\
&= \mathcal{O}_p\left(n^{-1}\right).
\end{aligned}$$

The third function is

$$T_{3,n}(\alpha)=\left(1-\frac{m}{2n}\right)\ln\left(\frac{1}{n}\sum_i\exp\left\{\hat\tau'_n\psi_i\right\}\right)$$

which evaluated at α_0 is

$$T_{3,n}(\alpha_0) = \left(1-\frac{m}{2n}\right)\ln\left(\frac{1}{n}\sum_i\exp\left\{\tau'_0\psi_{i0}\right\}\right)=0.$$

The first derivative of the third function is

$$\frac{\partial T_{3,n}(\alpha)}{\partial\alpha_j}=\left(1-\frac{m}{2n}\right)\frac{\frac{1}{n}\sum_i\exp\left\{\hat\tau'_n\psi_i\right\}\left(\frac{\partial\hat\tau'_n}{\partial\alpha_j}\psi_i+\hat\tau'_n\frac{\partial\psi_i}{\partial\alpha_j}\right)}{\frac{1}{n}\sum_i\exp\left\{\hat\tau'_n\psi_i\right\}}.$$

which evaluated at α_0 is

$$\begin{aligned}
\frac{\partial T_{3,n}(\alpha_0)}{\partial\alpha_j} &= \left(1-\frac{m}{2n}\right)\frac{1}{n}\sum_i\left(\frac{\partial\hat\tau'_{n0}}{\partial\alpha_j}\psi_{i0}\right) \\
&= \left(\frac{\partial\hat\tau'_{n0}}{\partial\alpha_j}\frac{1}{n}\sum_i\psi_{i0}\right)+\mathcal{O}_p(n^{-1}).
\end{aligned}$$

The first derivative is in the form of a fraction. The denominator of the second derivative of the third function is

$$\left(\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\right)^2$$

and when evaluated at α_0 the denominator equals one. Hence attention can focus solely on the numerator of the second derivative of the third function:

$$\begin{aligned}
&\left(1-\frac{m}{2n}\right)\left(\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\right)\left[\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_j}\psi_i+\hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_j}\right)\right.\\
&\times\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_\ell}\psi_i+\hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_\ell}\right)\\
&\left.+\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_j}\frac{\partial\psi_i}{\partial\alpha_\ell}+\frac{\partial^2\hat{\tau}_n'}{\partial\alpha_\ell\partial\alpha_j}\psi_i+\hat{\tau}_n'\frac{\partial^2\psi_i}{\partial\alpha_\ell\partial\alpha_j}+\frac{\partial\hat{\tau}_n'}{\partial\alpha_\ell}\frac{\partial\psi_i}{\partial\alpha_j}\right)\right]\\
&-\left(\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_j}\psi_i+\hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_j}\right)\right)\\
&\times\left(\frac{1}{n}\sum_i \exp\left\{\hat{\tau}_n'\psi_i\right\}\left(\frac{\partial\hat{\tau}_n'}{\partial\alpha_\ell}\psi_i+\hat{\tau}_n'\frac{\partial\psi_i}{\partial\alpha_\ell}\right)\right)
\end{aligned}$$

which evaluated at α_0 gives

$$\begin{aligned}
&\frac{\partial^2 T_{3,n}(\alpha_0)}{\partial\alpha_\ell\partial\alpha_j}\\
=\;&\left(1-\frac{m}{2n}\right)\left[\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\psi_{i0}\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_\ell}\psi_{i0}+\frac{1}{n}\sum_i\left(\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\frac{\partial\psi_{i0}}{\partial\alpha_\ell}+\frac{\partial^2\hat{\tau}_{n0}'}{\partial\alpha_\ell\partial\alpha_j}\psi_{i0}+\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_\ell}\frac{\partial\psi_{i0}}{\partial\alpha_j}\right)\right.\\
&\left.-\left(\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\psi_i(\alpha_0)\right)\left(\frac{1}{n}\sum_i\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_\ell}\psi_i(\alpha_0)\right)\right]\\
=\;&\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\left[\frac{1}{n}\sum_i\psi_{i0}\psi_{i0}'\right]\frac{\partial\hat{\tau}_{n0}}{\partial\alpha_\ell}+\frac{\partial\hat{\tau}_{n0}'}{\partial\alpha_j}\left[\frac{1}{n}\sum_i\frac{\partial\psi_{i0}}{\partial\alpha_\ell}\right]+\left[\frac{1}{n}\sum_i\frac{\partial\psi_{i0}'}{\partial\alpha_j}\right]\frac{\partial\hat{\tau}_{n0}}{\partial\alpha_\ell}+\mathcal{O}_p\left(n^{-1/2}\right).
\end{aligned}$$

Stack these and use **Lemma 6.1** to obtain

$$\begin{aligned}
\frac{\partial^2 T_{3,n}(\alpha_0)}{\partial\alpha'\partial\alpha} &= -\left[\frac{1}{n}\sum_i\frac{\partial\psi_{i0}'}{\partial\alpha}\right]\left[\frac{1}{n}\sum_i\psi_{i0}\psi_{i0}'\right]^{-1}\left[\frac{1}{n}\sum_i\frac{\partial\psi_{i0}}{\partial\alpha'}\right]+\mathcal{O}_p\left(n^{-1/2}\right)\\
&= -M_{\psi0}'\Sigma_{\psi0}^{-1}M_{\psi0}+\mathcal{O}_p\left(n^{-1/2}\right). \qquad (28)
\end{aligned}$$

Now construct the expansion of the objective function

$$\begin{aligned}\mathcal{L}_n(\alpha,\tau_n(\alpha)) &= \mathcal{L}_n(\alpha_0,\tau_n(\alpha_0)) + \frac{\partial \mathcal{L}_n(\alpha_0,\tau_n(\alpha_0))}{\partial \alpha'}(\alpha-\alpha_0) \\ &\quad + \frac{1}{2}(\alpha-\alpha_0)' \frac{\partial^2 \mathcal{L}_n(\alpha_0,\tau_n(\alpha_0))}{\partial \alpha \partial \alpha'}(\alpha-\alpha_0) + o_p(\delta^2)\end{aligned}$$

where $\delta = |\max(\alpha-\alpha_0)|$. The first term of the expansion only contributes $\mathcal{O}_p\left(n^{-1}\right)$ because each of the three functions that compose the objective function are $\mathcal{O}_p(n^{-1})$ when evaluated at α_0.

Similarly the first two functions that create the objective function will only contribute $\mathcal{O}_p\left(n^{-1}\right)$ to the expansion. The first derivative of the third function evaluated at α_0 will be the only contribution to the linear term of the expansion. For $j = 1,\ldots,m$

$$\frac{\partial \mathcal{L}_n(\alpha_0,\tau_n(\alpha_0))}{\partial \alpha_j} = \left(\frac{\partial \hat{\tau}'_{n0}}{\partial \alpha_j}\frac{1}{n}\sum_i \psi_{i0}\right) + \mathcal{O}_p(n^{-1}).$$

The terms can be stacked and apply **Lemma 6.1** to obtain

$$\begin{aligned}\frac{\partial \mathcal{L}_n(\alpha_0,\tau_n(\alpha_0))}{\partial \alpha'} &= \frac{1}{n}\sum_i \psi'_{i0}\frac{\partial \hat{\tau}_{n0}}{\partial \alpha'} + \mathcal{O}_p(n^{-1}) \\ &= -\Psi'_{n0}\hat{\Sigma}^{-1}_{\psi 0}\hat{M}_{\psi 0} + \mathcal{O}_p(n^{-1}). \qquad (29)\end{aligned}$$

Expanding $\Psi_n(\alpha)$ about the population parameter value and evaluating at the estimated value

$$\Psi_n(\hat{\alpha}) = \Psi_{n0} + \frac{\partial \Psi_{n0}}{\partial \alpha'}(\hat{\alpha}-\alpha_0) + \mathcal{O}_p(n^{-1}).$$

The first order conditions imply the LHS is zero, and solving for moment conditions evaluated at the population parameter values gives

$$\Psi'_{n0} = -(\hat{\alpha}-\alpha_0)'\hat{M}'_{\psi 0} + \mathcal{O}_p(n^{-1}).$$

Substituting into (29) gives the linear term as

$$\frac{\partial \mathcal{L}_n(\alpha_0,\tau_n(\alpha_0))}{\partial \alpha'}(\hat{\alpha}-\alpha_0) = (\hat{\alpha}-\alpha_0)'\hat{M}'_{\psi 0}\hat{\Sigma}^{-1}_{\psi 0}\hat{M}_{\psi 0}(\hat{\alpha}-\alpha_0) + \mathcal{O}_p\left(n^{-3/2}\right). \qquad (30)$$

As with the linear term, the first two functions that create the objective function will not contribute to the quadratic term in the expansion. The second derivative of the third function evaluated at α_0 will be the only contribution to the quadratic term.

Equations (30) and (28) can be used to give the expansion of the objective function in terms of α as

$$\mathcal{L}_n(\alpha,\tau_n(\alpha)) = \frac{1}{2}(\hat{\alpha}-\alpha_0)'\hat{M}'_{\psi 0}\hat{\Sigma}^{-1}_{\psi 0}\hat{M}_{\psi 0}(\hat{\alpha}-\alpha_0) + \mathcal{O}_p(n^{-1})$$

which is the negative of the expansion for EL. Hence in a $\mathcal{O}(n^{-1/2})$ neighborhood of $\alpha = \alpha_0$ the expansion for the ESPL objective function $\mathcal{L}_n(\alpha, \tau_n(\alpha))$ is the same as the expansion of the EL objective function for terms that converge slower than $\mathcal{O}_p(n^{-1})$ and the ESPL estimates will have the same $\sqrt{n}$ consistency as EL.

The expansion of the first order conditions (asymptotic normality).

The above expansion of the objective function is written as only a function of α, because the saddlepoint equation had been used to write τ as an implicit function of α. However, the first order conditions are written as a function of the $2m$ variables α and τ. Instead of using the implicit function $\tau_n(\alpha)$, now τ is an estimated parameter and hence the saddlepoint equation is included in the expansion.

The first order asymptotic distribution will be derived using the traditional approach of expanding the first order conditions and solving for the centered and normalized parameters.

The parameters α and τ are selected simultaneously as the solution to the values that set the first derivative of the log of the EL function to zero[10]:

$$\begin{aligned}
\frac{\partial \mathcal{L}_n(\alpha,\tau)}{\partial \alpha_j} &= -\frac{1}{2n}\mathrm{tr}\left\{\left[\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\,\psi_i\psi_i'\right]^{-1}\right. \\
&\quad \left.\times\left[\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\left\{\frac{\partial\psi_i}{\partial\alpha_j}\psi_i' + \psi_i\frac{\partial\psi_i'}{\partial\alpha_j}\right\} + \frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\,\tau'\frac{\partial\psi_i}{\partial\alpha_j}\psi_i\psi_i'\right]\right\} \\
&\quad +\frac{1}{n}\mathrm{tr}\left\{\left[\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\,\frac{\partial\psi_i}{\partial\alpha}\right]^{-1}\right. \\
&\quad \left.\times\left[\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\,\frac{\partial^2\psi_i}{\partial\alpha_j\partial\alpha} + \frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\,\tau'\frac{\partial\psi_i}{\partial\alpha_j}\frac{\partial\psi_i}{\partial\alpha}\right]\right\} \\
&\quad +\left(1-\frac{m}{2n}\right)\frac{\frac{1}{n}\sum_i \exp\{\tau'\psi_i\}\,\tau'\frac{\partial\psi_i}{\partial\alpha_j}}{\frac{1}{n}\sum_j \exp\{\tau'\psi_j\}}
\end{aligned}$$

and the saddlepoint equation

$$S_n(\alpha,\tau) = \frac{1}{n}\sum_i \psi_i \exp\{\tau'\psi_i\} = 0.$$

These define the ESPL estimates. Expanding about the population parameter values and solving for the parameters gives:

[10]note the difference with the derivatives calculated for the expansion of the objective function. now τ is not an implicit function of α and hence $\partial\tau/\partial\alpha = 0$.

$$\begin{bmatrix} \frac{\partial \mathcal{L}_n(\hat{\alpha}_{espl},\hat{\tau}_{espl})}{\partial \alpha} \\ S_n(\hat{\alpha}_{espl},\hat{\tau}_{espl}) \end{bmatrix} = \begin{bmatrix} \frac{\partial \mathcal{L}_n(\alpha_0,\tau_0)}{\partial \alpha} \\ S_n(\alpha_0,\tau_0) \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 \mathcal{L}_n(\alpha_0,\tau_0)}{\partial \alpha' \partial \alpha} & \frac{\partial^2 \mathcal{L}_n(\alpha_0,\tau_0)}{\partial \tau' \partial \alpha} \\ \frac{\partial S_n(\alpha_0,\tau_0)}{\partial \alpha'} & \frac{\partial S_n(\alpha_0,\tau_0)}{\partial \tau'} \end{bmatrix} \begin{bmatrix} (\hat{\alpha}_{espl} - \alpha_0) \\ \hat{\tau}_{espl} \end{bmatrix} + \mathcal{O}_p\left(n^{-1}\right).$$

The estimation FOC's imply the LHS is zero. Solving for the parameters gives

$$\begin{aligned}
\sqrt{n}\begin{bmatrix} (\hat{\alpha}_{espl} - \alpha_0) \\ \hat{\tau}_{espl} \end{bmatrix} &= -\begin{bmatrix} E\left[\frac{\partial^2 \mathcal{L}_n(\alpha_0,\tau_0)}{\partial \alpha' \partial \alpha}\right] & E\left[\frac{\partial^2 \mathcal{L}_n(\alpha_0,\tau_0)}{\partial \tau' \partial \alpha}\right] \\ E\left[\frac{\partial S_n(\alpha_0,\tau_0)}{\partial \alpha'}\right] & E\left[\frac{\partial S_n(\alpha_0,\tau_0)}{\partial \tau'}\right] \end{bmatrix}^{-1} \sqrt{n}\begin{bmatrix} \frac{\partial \mathcal{L}_n(\alpha_0,\tau_0)}{\partial \alpha} \\ S_n(\alpha_0,\tau_0) \end{bmatrix} + \mathcal{O}_p\left(n^{-1/2}\right) \\
&= -\begin{bmatrix} 0 & M'_{\psi 0} \\ M_{\psi 0} & \Sigma_{\psi 0} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \sqrt{n}\Psi_n(\alpha_0) \end{bmatrix} + \mathcal{O}_p\left(n^{-1/2}\right) \\
&= -\begin{bmatrix} -M_{\psi 0}^{-1}\Sigma_{\psi 0}\left(M'_{\psi 0}\right)^{-1} & M_{\psi 0}^{-1} \\ \left(M'_{\psi 0}\right)^{-1} & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \sqrt{n}\Psi_n(\alpha_0) \end{bmatrix} + \mathcal{O}_p\left(n^{-1/2}\right) \\
&\sim^a N\left(0, \operatorname{diag}\left(\begin{array}{cc} M_{\psi 0}^{-1}\Sigma_{\psi 0}\left(M'_{\psi 0}\right)^{-1} & 0 \end{array} \right)\right).
\end{aligned}$$

■

Proof of **Theorem 2.3** .

The first order conditions for the constrained estimation problem are

$$\begin{aligned}
\frac{\partial \mathcal{L}_n(\hat{\alpha}_{cespl},\hat{\tau}_{cespl})}{\partial \alpha} + R(\hat{\alpha}_{cespl})\hat{\gamma} &= 0 \\
n^{-1}\sum_i \psi_i(\hat{\alpha}_{cespl}) \exp\left\{\hat{\tau}'_{cespl}\psi_i(\hat{\alpha}_{cespl})\right\} &= 0 \\
r(\hat{\alpha}_{cespl}) &= 0.
\end{aligned} \tag{31}$$

Expand these first order conditions about the population parameter values

$$\begin{aligned}
&\begin{bmatrix} \frac{\partial \mathcal{L}_n(\hat{\alpha}_{cespl},\hat{\tau}_{cespl})}{\partial \alpha} + R(\hat{\alpha}_{cespl})\hat{\gamma} \\ n^{-1}\sum_i \psi_i(\hat{\alpha}_{cespl}) \exp\left\{\hat{\tau}'_{cespl}\psi_i(\hat{\alpha}_{cespl})\right\} \\ r(\hat{\alpha}_{cespl}) \end{bmatrix} \\
= &\begin{bmatrix} 0 \\ \Psi_{n0} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & \frac{\partial \Psi'_{n0}}{\partial \alpha} & R_0 \\ \frac{\partial \Psi_{n0}}{\partial \alpha'} & \Sigma_{\psi 0} & 0 \\ R'_0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\alpha}_{cespl} - \alpha_0 \\ \hat{\tau}_{cespl} \\ \hat{\gamma} \end{bmatrix} + \mathcal{O}_p\left(n^{-1}\right) \\
= &\begin{bmatrix} 0 \\ \Psi_{n0} \\ 0 \end{bmatrix} + \begin{bmatrix} 0 & M'_{\psi 0} & R_0 \\ M_{\psi 0} & \Sigma_{\psi 0} & 0 \\ R'_0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{\alpha}_{cespl} - \alpha_0 \\ \hat{\tau}_{cespl} \\ \hat{\gamma} \end{bmatrix} + \mathcal{O}_p\left(n^{-1}\right).
\end{aligned}$$

The LSH is zero because of the first order conditions from equations (31). Solving for the parameters gives

$$\begin{bmatrix} \hat{\alpha}_{cespl} - \alpha_0 \\ \hat{\tau}_{cespl} \\ \hat{\gamma} \end{bmatrix} = -\begin{bmatrix} 0 & M'_{\psi 0} & R_0 \\ M_{\psi 0} & \Sigma_{\psi 0} & 0 \\ R'_0 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \Psi_n(\alpha_0) \\ 0 \end{bmatrix} + \mathcal{O}_p\left(n^{-1}\right).$$

now the partitioned inverse is

$$\begin{aligned} \begin{bmatrix} A & B \\ B' & 0 \end{bmatrix}^{-1} &= \begin{bmatrix} A^{-1} - A^{-1}B\left(B'A^{-1}B\right)^{-1}B'A^{-1} & A^{-1}B\left(B'A^{-1}B\right)^{-1} \\ \left(B'A^{-1}B\right)^{-1}B'A^{-1} & -\left(B'A^{-1}B\right)^{-1} \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} A^{-1}B \\ -I \end{bmatrix}\left(B'A^{-1}B\right)^{-1}\begin{bmatrix} B'A^{-1} & -I \end{bmatrix}. \end{aligned}$$

This gives the inverse as

$$\begin{aligned} \begin{bmatrix} 0 & M'_{\psi 0} & R_0 \\ M_{\psi 0} & \Sigma_{\psi 0} & 0 \\ R'_0 & 0 & 0 \end{bmatrix}^{-1} = &\begin{bmatrix} M_{\psi 0}^{-1}\Sigma_{\psi 0}\left(M'_{\psi 0}\right)^{-1} & M_{\psi 0}^{-1} & 0 \\ \left(M'_{\psi 0}\right)^{-1} & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \\ &- \begin{bmatrix} M_{\psi 0}^{-1}\Sigma_{\psi 0}\left(M'_{\psi 0}\right)^{-1}R_0 \\ \left(M'_{\psi 0}\right)^{-1}R_0 \\ -I \end{bmatrix} \\ &\times \left(R'_0 M_{\psi 0}^{-1}\Sigma_{\psi 0}\left(M'_{\psi 0}\right)^{-1}R_0\right)^{-1} \\ &\times \begin{bmatrix} R'_0 M_{\psi 0}^{-1}\Sigma_{\psi 0}\left(M'_{\psi 0}\right)^{-1} & R'_0 M_{\psi 0}^{-1} & -I \end{bmatrix}. \end{aligned}$$

Substitute in the inverse to obtain

$$\begin{aligned} &\sqrt{n}\begin{bmatrix} \hat{\alpha}_{cespl} - \alpha_0 \\ \hat{\tau}_{cespl} \\ \hat{\gamma} \end{bmatrix} \\ = &-\begin{bmatrix} M_{\psi 0}^{-1} - M_{\psi 0}^{-1}\Sigma_{\psi 0}^{1/2\prime}\Gamma\left(\Gamma'\Gamma\right)^{-1}R'_0 M_{\psi 0}^{-1} \\ -\left(M'_{\psi 0}\right)^{-1}R_0\left(\Gamma'\Gamma\right)^{-1}R'_0 M_{\psi 0}^{-1} \\ \left(\Gamma'\Gamma\right)^{-1}R'_0 M_{\psi 0}^{-1} \end{bmatrix}\sqrt{n}\Psi_n(\alpha_0) + \mathcal{O}_p\left(n^{-1/2}\right) \\ \sim^a &N\left(0, \begin{pmatrix} M_{\psi 0}^{-1}\Sigma_{\psi 0}^{1/2\prime}P_\Gamma^\perp\Sigma_{\psi 0}^{1/2}\left(M_{\psi 0}^{-1}\right)' & 0 & 0 \\ 0 & \left(\Sigma_{\psi 0}^{1/2}\right)^{-1}P_\Gamma\left(\Sigma_{\psi 0}^{1/2\prime}\right)^{-1} & -\left(M_{\psi 0}^{-1}\right)'R_0\left(\Gamma'\Gamma\right)^{-1} \\ 0 & -\left(\Gamma'\Gamma\right)^{-1}R'_0 M_{\psi 0}^{-1} & \left(\Gamma'\Gamma\right)^{-1} \end{pmatrix}\right) \end{aligned}$$

where $\Gamma = \Sigma_{\psi 0}^{1/2}\left(M_{\psi 0}^{-1}\right)'R_0$. ■

Proof of **Theorem 2.4** .

The first order conditions for this problem are

$$\begin{aligned} \frac{\partial \mathcal{L}_n(\hat{\theta}_{cespl}, 0, \hat{\tau}_{cespl})}{\partial \theta} &= 0 \\ S_n(\hat{\theta}_{cespl}, 0, \hat{\tau}_{cespl}) = n^{-1} \sum_i \psi_i(\hat{\theta}_{cespl}, 0) \exp\left\{ \hat{\tau}'_{cespl} \psi_i(\hat{\theta}_{cespl}, 0) \right\} &= 0. \end{aligned}$$

These define the CESPL estimates. The asymptotic distribution will be determined by expanding about the population parameter values and solving for the parameters.

$$\begin{aligned} & \begin{bmatrix} \frac{\partial \mathcal{L}_n(\hat{\theta}_{cespl}, 0, \hat{\tau}_{cespl})}{\partial \theta} \\ S_n(\hat{\theta}_{cespl}, 0, \hat{\tau}_{cespl}) \end{bmatrix} \\ = & \begin{bmatrix} \overline{R}_0' \frac{\partial \mathcal{L}_n(\alpha_0, \tau_0)}{\partial \alpha} \\ S_n(\theta_0, 0, \tau_0) \end{bmatrix} + \begin{bmatrix} \overline{R}_0' \frac{\partial^2 \mathcal{L}_n(\alpha_0, \tau_0)}{\partial \alpha' \partial \alpha} \overline{R}_0 & \overline{R}_0' \frac{\partial^2 \mathcal{L}_n(\alpha_0, \tau_0)}{\partial \tau' \partial \alpha} \\ \frac{\partial S_n(\alpha_0, \tau_0)}{\partial \alpha'} \overline{R}_0 & \frac{\partial S_n(\alpha_0, \tau_0)}{\partial \tau'} \end{bmatrix} \begin{bmatrix} \hat{\theta}_{cespl} - \theta_0 \\ \hat{\tau}_{cespl} \end{bmatrix} + \mathcal{O}_p\left(n^{-1}\right). \end{aligned}$$

The FOC of estimation implies that the LHS is zero. To reduce notation let $\overline{\Gamma} = \Sigma_{\psi 0}^{-1/2} M_{\psi 0} \overline{R}_0$. Solving for the parameters gives

$$\begin{aligned} & \sqrt{n} \begin{bmatrix} \hat{\theta}_{cespl} - \alpha_0 \\ \hat{\tau}_{cespl} \end{bmatrix} \\ = & -\begin{bmatrix} \overline{R}_0' \frac{\partial^2 \mathcal{L}_n(\alpha_0, \tau_0)}{\partial \alpha' \partial \alpha} \overline{R}_0 & \overline{R}_0' \frac{\partial^2 \mathcal{L}_n(\alpha_0, \tau_0)}{\partial \tau' \partial \alpha} \\ \frac{\partial S_n(\alpha_0, \tau_0)}{\partial \alpha'} \overline{R}_0 & \frac{\partial S_n(\alpha_0, \tau_0)}{\partial \tau'} \end{bmatrix}^{-1} \sqrt{n} \begin{bmatrix} \overline{R}_0' \frac{\partial \mathcal{L}_n(\alpha_0, \tau_0)}{\partial \alpha} \\ S_n(\alpha_0, \tau_0) \end{bmatrix} + \mathcal{O}_p\left(n^{-1/2}\right) \\ = & -\begin{bmatrix} 0 & \overline{R}_0' M_{\psi 0}' \\ M_{\psi 0} \overline{R}_0 & \Sigma_{\psi 0} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \sqrt{n} \Psi_n(\alpha_0) \end{bmatrix} + \mathcal{O}_p\left(n^{-1/2}\right) \\ = & -\begin{bmatrix} -\left(\overline{\Gamma}'\overline{\Gamma}\right)^{-1} & \left(\overline{\Gamma}'\overline{\Gamma}\right)^{-1} \overline{\Gamma}' \Sigma_{\psi 0}^{-1/2} \\ \Sigma_{\psi 0}^{-1/2\prime} \overline{\Gamma} \left(\overline{\Gamma}'\overline{\Gamma}\right)^{-1} & \Sigma_{\psi 0}^{-1/2\prime} P_{\overline{\Gamma}}^{\perp} \Sigma_{\psi 0}^{-1/2} \end{bmatrix} \begin{bmatrix} 0 \\ \sqrt{n} \Psi_n(\alpha_0) \end{bmatrix} + \mathcal{O}_p\left(n^{-1/2}\right) \\ \sim^a & N\left(0, \operatorname{diag}\begin{bmatrix} \left(\overline{\Gamma}'\overline{\Gamma}\right)^{-1} & \Sigma_{\psi 0}^{-1/2\prime} P_{\overline{\Gamma}}^{\perp} \Sigma_{\psi 0}^{-1/2} \end{bmatrix}\right). \end{aligned}$$

■

Proof of **Theorem 2.5**.

The two term expansion of the system of equations gives

$$(\alpha_n - \alpha_0) = -\left[E \frac{\partial \Psi_{n0}}{\partial \alpha'}\right]^{-1} \Psi_{n0} + \mathcal{O}_p(n^{-1}) \tag{32}$$

A three term expansion of the system gives

$$(\alpha_n - \alpha_0) = -\left[E\frac{\partial \Psi_{n0}}{\partial \alpha'}\right]^{-1} \left\{ \Psi_{n0} + \left[\frac{\partial \Psi_{n0}}{\partial \alpha'} - E\frac{\partial \Psi_{n0}}{\partial \alpha'}\right](\alpha_n - \alpha_0) \right.$$
$$\left. + \frac{1}{2}\left[(\alpha_n - \alpha_0)' \left[E\frac{\partial^2 \Psi_{(j)n0}}{\partial \alpha \partial \alpha'}\right](\alpha_n - \alpha_0)\right]_{j=1,\ldots,m} \right\} + \mathcal{O}_p(n^{-3/2}).$$

Use equation (32) to substitute for the first set of parameters on the RHS and then take expectations to obtain

$$\begin{aligned} E(\alpha_n - \alpha_0) &= -\left[E\frac{\partial \Psi_{n0}}{\partial \alpha'}\right]^{-1} \left\{ -E\left[\frac{\partial \Psi_{n0}}{\partial \alpha'}\left[E\frac{\partial \Psi_{n0}}{\partial \alpha'}\right]^{-1} \Psi_{n0}\right] \right. \\ &\quad \left. + \frac{1}{2}E\left[(\alpha_n - \alpha_0)' \left[E\frac{\partial^2 \Psi_{(j)n0}}{\partial \alpha \partial \alpha'}\right](\alpha_n - \alpha_0)\right]_{j=1,\ldots,m} \right\} + \mathcal{O}(n^{-3/2}) \\ &= \frac{1}{n} M_{\psi 0}^{-1} \left\{ E\left[\frac{\partial \psi_{i0}}{\partial \alpha'} M_{\psi 0}^{-1} \psi_{i0}\right] - a \right\} + \mathcal{O}(n^{-3/2}) \end{aligned}$$

where a is a vector with elements $a_j = \operatorname{tr}\left(\left(M'_{\psi 0}\Sigma_{\psi 0}^{-1} M_{\psi 0}\right)^{-1} E\left[\partial^2 \psi_{(j)i0}/\partial \alpha \partial \alpha'\right]\right)/2$. ■

Proof of **Theorem 2.6** .

Let $A_n(\alpha, \tau) = n\left(\frac{\partial T_{1,n}(\alpha)}{\partial \alpha} + \frac{\partial T_{2,n}(\alpha)}{\partial \alpha}\right)$ and $B_n(\alpha, \tau) = \frac{\partial T_{3,n}(\alpha)}{\partial \alpha}$ and note that it was shown in the proof of **Theorem 2.2** that both $A_n(\alpha_0, \tau_0)$ and $B_n(\alpha_0, \tau_0)$ are $\mathcal{O}_p(1)$. With this notation the ESPL estimator's first order conditions (4) and (5) can be written

$$\begin{aligned} \frac{1}{n} A_n(\hat{\alpha}_{espl}, \hat{\tau}_{espl}) + B_n(\hat{\alpha}_{espl}, \hat{\tau}_{espl}) &= 0 \\ S_n(\hat{\alpha}_{espl}, \hat{\tau}_{espl}) &= 0. \end{aligned}$$

Denote this system $H_n(\hat{\alpha}_{espl}, \hat{\tau}_{espl}) = n^{-1}\sum_i h_i(\hat{\alpha}_{espl}, \hat{\tau}_{espl}) = 0$ and the asymptotic distribution of the parameters $\sqrt{n}\left(\ (\hat{\alpha}_{espl} - \alpha_0)' \quad \hat{\tau}'_{espl}\ \right)' \sim N(0, \Omega)$. Let ∂ denote the derivative with respect to all the parameters of the function. With this notation,

the expectation of the centered estimates can be written

$$
\begin{aligned}
& E\begin{pmatrix} \hat{\alpha}_{espl} - \alpha_0 \\ \hat{\tau}_{espl} \end{pmatrix} \\
= & -\left[E\partial H_n(\alpha_0,\tau_0)\right]^{-1}\Big\{EH_n(\alpha_0,\tau_0) - E\Big[\partial H_n(\alpha_0,\tau_0)\left[E\partial H_n(\alpha_0,\tau_0)\right]^{-1}H_n(\alpha_0,\tau_0)\Big] \\
& +\frac{1}{2}\left[\operatorname{tr}\left\{E\partial^2 H_{(j)n}(\alpha_0,\tau_0)\frac{\Omega}{n}\right\}\right]_{j=1,\ldots,2m}\Bigg\} + \mathcal{O}_p(n^{-3/2}) \\
= & \frac{1}{n}\left[E\partial h_{i0}\right]^{-1}\Big\{-Eh_{i0} + E\Big[\partial h_{i0}\left[E\partial h_{i0}\right]^{-1}h_{i0}\Big] \\
& \qquad -\frac{1}{2}\left[\operatorname{tr}\left\{E\partial^2 h_{(j)i0}\Omega\right\}\right]_{j=1,\ldots,2m}\Bigg\} + \mathcal{O}_p(n^{-3/2}).
\end{aligned}
$$

Substitute for the terms to obtain

$$
\begin{aligned}
& E\begin{pmatrix} \hat{\alpha}_{espl} - \alpha_0 \\ \hat{\tau}_{espl} \end{pmatrix} \\
= & \frac{1}{n}\begin{bmatrix} -M_{\psi0}^{-1}\Sigma_{\psi0}\left(M'_{\psi0}\right)^{-1} & M_{\psi0}^{-1} \\ M_{\psi0}^{-1\prime} & 0 \end{bmatrix} \\
& \times\Bigg\{-\begin{bmatrix} E\left[A_n(\alpha_0)\right] \\ 0 \end{bmatrix} + E\left[\begin{bmatrix} 0 & \frac{\partial\psi'_{i0}}{\partial\alpha} \\ \frac{\partial\psi_{i0}}{\partial\alpha'} & \psi_{i0}\psi'_{i0} \end{bmatrix}\begin{bmatrix} -M_{\psi0}^{-1}\Sigma_{\psi0}\left(M'_{\psi0}\right)^{-1} & M_{\psi0}^{-1} \\ M_{\psi0}^{-1\prime} & 0 \end{bmatrix}\begin{bmatrix} 0 \\ \psi_{i0} \end{bmatrix}\right] \\
& -\frac{1}{2}\begin{bmatrix} \left[\operatorname{tr}\left\{0_{m\times m}\right\}\right]_{j1=1,\ldots,m} \\ \left[\operatorname{tr}\left\{E\frac{\partial^2\psi_{(j2)i0}}{\partial\alpha\partial\alpha'}M_{\psi0}^{-1}\Sigma_{\psi0}\left(M'_{\psi0}\right)^{-1}\right\}\right]_{j2=1,\ldots,m} \end{bmatrix}\Bigg\} + \mathcal{O}(n^{-3/2}) \\
= & \frac{1}{n}\begin{bmatrix} -M_{\psi0}^{-1}\Sigma_{\psi0}\left(M'_{\psi0}\right)^{-1} & M_{\psi0}^{-1} \\ M_{\psi0}^{-1\prime} & 0 \end{bmatrix}\Bigg\{-\begin{bmatrix} E\left[A_n(\alpha_0)\right] \\ 0 \end{bmatrix} + E\begin{bmatrix} 0 \\ \frac{\partial\psi_{i0}}{\partial\alpha'}M_{\psi0}^{-1}\psi_{i0} \end{bmatrix} \\
& -\frac{1}{2}\begin{bmatrix} 0_{m\times1} \\ \left[\operatorname{tr}\left\{E\frac{\partial^2\psi_{(j2)i0}}{\partial\alpha\partial\alpha'}M_{\psi0}^{-1}\Sigma_{\psi0}\left(M'_{\psi0}\right)^{-1}\right\}\right]_{j2=1,\ldots,m} \end{bmatrix}\Bigg\} + \mathcal{O}(n^{-3/2}) \\
= & \frac{1}{n}\begin{bmatrix} M_{\psi0}^{-1}\Sigma_{\psi0}\left(M'_{\psi0}\right)^{-1}E\left[A_n(\alpha_0)\right] + M_{\psi0}^{-1}E\left[\frac{\partial\psi_{i0}}{\partial\alpha'}M_{\psi0}^{-1}\psi_{i0}\right] - M_{\psi0}^{-1}a \\ -M_{\psi0}^{-1\prime}E\left[A_n(\alpha_0)\right] \end{bmatrix} + \mathcal{O}(n^{-3/2}).
\end{aligned}
$$

The terms that form $A_n(\alpha_0)$ were derived in the proof of **Theorem 2.2**, i.e. the

first derivative of $T_{1,n}$ and $T_{2,n}$ evaluated at α_0:

$$\begin{aligned}
EA_n(\alpha_0)_{(j)} &= -\frac{1}{2}\text{tr}\left\{\Sigma_{\psi 0}^{-1}E\left\{\frac{\partial\psi_{i0}}{\partial\alpha_j}\psi_{i0}' + \psi_{i0}\frac{\partial\psi_{i0}'}{\partial\alpha_j}\right\}\right\} + \text{tr}\left\{M_{\psi 0}^{-1}E\left[\frac{\partial^2\psi_{i0}}{\partial\alpha_j\partial\alpha'}\right]\right\} + \mathcal{O}(n^{-1/2}) \\
&= -\text{tr}\left\{\Sigma_{\psi 0}^{-1}E\left[\psi_{i0}\frac{\partial\psi_{i0}'}{\partial\alpha_j}\right]\right\} + \text{tr}\left\{M_{\psi 0}^{-1}E\left[\frac{\partial^2\psi_{i0}}{\partial\alpha_j\partial\alpha'}\right]\right\} + \mathcal{O}(n^{-1/2}) \\
&= -E\left[\frac{\partial\psi_{i0}'}{\partial\alpha_j}\Sigma_{\psi 0}^{-1}\psi_{i0}\right] + \text{tr}\left\{M_{\psi 0}^{-1}E\left[\frac{\partial^2\psi_{i0}}{\partial\alpha_j\partial\alpha'}\right]\right\} + \mathcal{O}(n^{-1/2}).
\end{aligned}$$

Stacking this gives

$$EA_n(\alpha_0) = -E\left[\frac{\partial\psi_{i0}'}{\partial\alpha}\Sigma_{\psi 0}^{-1}\psi_{i0}\right] + c + \mathcal{O}(n^{-1/2})$$

where c is a vector with elements $c_j = \text{tr}\left(M_{\psi 0}^{-1}E\left[\partial^2\psi_{i0}/\partial\alpha_j\partial\alpha'\right]\right)$. Substitute to obtain the higher order bias for the ESPL estimator

$$\begin{aligned}
E(\alpha_{espl} - \alpha_0) &= \frac{1}{n}\Bigg(M_{\psi 0}^{-1}\Sigma_{\psi 0}\left(M_{\psi 0}'\right)^{-1}\left(-E\left[\frac{\partial\psi_{i0}'}{\partial\alpha}\Sigma_{\psi 0}^{-1}\psi_{i0}\right] + c\right) \\
&\qquad + M_{\psi 0}^{-1}E\left[\frac{\partial\psi_{i0}}{\partial\alpha'}M_{\psi 0}^{-1}\psi_{i0}\right] - M_{\psi 0}^{-1}a\Bigg) + \mathcal{O}(n^{-3/2}).
\end{aligned}$$

■

6.2 Implementation for two-step GMM estimation equations

This section summarizes how the simulations were performed.

- Starting values. For the EL, ET, ETEL and CESPL a grid of 61 evenly spaced values spaces between [-1, 5] was evaluated. The parameter where the objective function obtained its extreme value was then used as the starting value for a nonlinear search routine.

  For the ESPL estimate the starting values for λ were set to alternate between .001 and -.001. The ESPL objective function was optimized twice: once starting θ at the EL estimate and the other starting at the CESPL estimate.[11]

- Weighting matrix for ESPL and the CESPL. To remove the degree of freedom in the selection of the first round weighting matrix, the EL parameter estimate was used to calculate the moment conditions and then the weighting matrix. The weights $1/n$ were used to calculate the weighting matrix.

- Parameter restrictions. The nonlinear optimization routine was bounded not to search further than ± 20 from its start values.

  For ESPL and CESPL the numerical optimization routine was restricted to selecting $\theta \in [-10, 10]$ and $\lambda_j \in [-10/\sqrt{n}, 10/\sqrt{n}]$. The restriction on λ_j is equivalent to ensuring that each t-statistic that tests an overidentifying restriction is less than 10 in absolute value.

- Starting values for the saddlepoint equations. For both the ESPL and the CESPL the starting values for the solution to the saddlepoint equation were set to alternate between .001 and -.001.

- To account for the parameter values where the empirical saddlepoint density is zero, the ESPL objective function was set to $-10.0E-30$ when:

  1. The inner product of the saddlepoint equation is not below .02. This indicates parameter values that are inconsistent with the observed data.
  2. The GMM objective function is concave. This would mean that the parameter value is associated with a local maximum of the GMM objective function instead of a local minimum.
  3. Any of the weights $w_i(\alpha)$ are zero.
  4. The covariance matrix using the weights $w_i(\alpha)$ is not positive semi-definite. This would imply that the conjugate density is not well defined.
  5. The first derivative of the estimation equations using the weights $w_i(\alpha)$ is singular. This would imply that the conjugate density is not well defined.

[11] For the simulated samples of size $n = 50$ better starting values were required. An evenly spaced grid was used to obtain a starting value for the numerical optimization routine. For this sample size, to complete all 10,000 simulations required less than 40 minutes on a notebook computer.

7 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.

   vol. 73(3, pp. 983-1002.

2. Butler, Ronald W. (2007) *Saddlepoint Approximations with Applications* Cambridge University Press.

3. Daniels, H.E. (1954) "Saddlepoint approximations in statistics" *Annals of Mathematical Statistics* vol. 25, pp. 631-650.

4. Field, C. A.(1982) "Small Sample Asymptotics for Multivariate M-Estimates" *Annals of Statistics*, vol. 10, pp. 672-689.

5. Field, C. A. and E. Ronchetti (1990) *Small Sample Asymptotics* Hayward, CA. IMS Monograph Series, vol 13.

6. Goutis, Constantin and George Casella (1999) "Explaining the Saddlepoint Approximation" *The American Statistician* vol. 53, no. 3. pp. 216-224.

7. Gregory, A. W., et al. (2002). "Information-theoretic estimation of preference parameters: macroeconomic applications and simulation evidence" *Journal of Econometrics* 107(1-2): 213-233.

8. Hall, A. R. H. (2005). *Generalized Method of Moments*, Oxford, UK, Oxford University Press.

9. 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.

10. 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.

11. Huzurbazar, S. (1999) "Practical Saddlepoint Approximations" *The American Statistician*, vol. 53, no. 3, pp. 225-232.

12. Imbens, Guido W. (1997) "One-Step Estimators for Over-Identified Generalized Method of Moments Models" *The Review of Economic Studies* vol. 64, no. 3. pp. 359-383.

13. Imbens, Guido W., Richard H. Spady and Phillip Johnson (1998) "Information Theoretic Approaches to Inference in Moment Condition Models" *Econometrica* vol. 66, no. 2, pp. 333-357.

14. Jensen, J. L. (1995) *Saddlepoint Approximations*, Oxford University Press, Oxford.

15. Jensen, J. L. and A. T. A. Wood (1998). "Large Deviation and Other Results for Minimum Contrast Estimators." *Annals of the Institute of Statistical Mathematics* vol. 50, no. 4, pp. 673-695.

16. Kitamura, Yuichi and Michael Stutzer (1997) "An Information-Theoretic Alternative to Generalized Method of Moments Estimation" *Econometrica* vol. 65, no. 4, pp. 861-874.

17. Kolassa, J.E. (1997) *Series Approximation Methods in Statistics* 2nd edition, new York, Springer-Verlag, Lecture notes in Statistics, 88.

18. Mátyás, L. (1999). *Generalized Method of Moments Estimation*, Cambridge, UK, Cambridge University Press.

19. Monti, A. C. and E. Ronchetti (1993) "On the Relationship Between Empirical Likelihood and Empirical Saddlepoint Approximation for Multivariate M-Estimators." *Biometrika* vol. 80 no. 2, pp. 329-338.

20. Newey, W. K., and D. L. McFadden (1994) "Large Sample Estimation and Hypothesis Testing." in *Handbook of Econometrics. Vol. 4.* Edited by R. F. Engle and D. McFadden. Amsterdam, The Netherlands: Elsevier Science, 1999, pp. 2113-2245.

21. 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.

22. Reid, N. (1988) "Saddlepoint Methods and Statistical Inference" *Statistical Science* vol 3, no. 2, pp. 213-227.

23. Ronchetti, 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.

24. Schennach, S. M. (2007) "Point Estimation with Exponentially Tilted Empirical Likelihood" *The Annals of Statistics* vol. 35 no. 2 pp. 634-672.

25. Skovgaard, I. M. (1990) "On the Density of minimum contrast estimators", *Annals of Statistics*, vol. 18, pp. 779-789.

26. Sowell, F. B. (1996) "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework" *Econometrica* vol. 64, no. 5. pp. 1085-1107.

27. Sowell, F. B. (2007) "The Empirical Saddlepoint Approximation for GMM Estimators" working paper, Tepper School of Business, Carnegie Mellon University.