ELSEVIER

Journal of Econometrics 107 (2002) 87–98

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Confidence intervals in generalized method of moments models

Guido W. Imbens[a,*], Richard Spady[b]

[a]Department of Economics and Department of Agricultural and Resource Economics, 549 Evans Hall # 3880, UC Berkeley, Berkeley, CA, 94720-3880, USA, and NBER Bunche Hall, University of California at Los Angeles, 405 Hilgard Avenue, Los Angeles, CA 90095, USA

[b]Department of Economics, Northwestern University, Evanston, IL, USA

Abstract

We consider the construction of confidence intervals for parameters characterized by moment restrictions. In the standard approach to generalized method of moments (GMM) estimation, confidence intervals are based on the normal approximation to the sampling distribution of the parameters. There is often considerable disagreement between the nominal and actual coverage rates of these intervals, especially in cases with a large degree of overidentification. We consider alternative confidence intervals based on empirical likelihood methods which exploit the normal approximation to the Lagrange multipliers calculated as a byproduct in empirical likelihood estimation. In large samples such confidence intervals are identical to the standard GMM ones, but in finite samples their properties can be substantially different. In some of the examples we consider, the proposed confidence intervals have coverage rates much closer to the nominal coverage rates than the corresponding GMM intervals.

JEL classification: C12; C13; C14; C33

Keywords: Empirical likelihood; Generalized method of moments; Overidentification; Hypothesis testing; Confidence intervals

1. Introduction

We consider the construction of confidence intervals for parameters characterized by moment restrictions. In the standard approach to generalized method of moments (GMM, Hansen, 1982) estimation, confidence intervals are based on the normal approximation to the sampling distribution of the parameters. There is often

* Corresponding author.

E-mail address: imbens@econ.berkeley.edu (G.W. Imbens).

0304-4076/02/$ - see front matter
PII: S0304-4076(01)00114-2

considerable disagreement between the nominal and actual coverage rates of these intervals, especially in cases with a large degree of overidentification (Robertson and Pagan, 1997; Altonji and Segal, 1995; Hall and Horowitz, 1996; Hansen et al., 1996). We consider alternative confidence intervals based on empirical likelihood methods (Owen, 1988; Back and Brown, 1990; Qin and Lawless, 1994; Imbens, 1997; Smith, 1997; Kitamura and Stutzer, 1997; Imbens et al., 1998; Smith, 1999; Newey and Smith, 2000). Our alternatives exploit the normal approximation to the Lagrange multipliers calculated as a byproduct in empirical likelihood estimation (e.g., Mykland, 1995; Smith, 2000). We show that in large samples such confidence intervals are identical to those constructed using the standard GMM approach. In finite samples, however, the properties of these confidence intervals can be substantially different. In some of the examples we consider, the proposed confidence intervals have coverage rates much closer to the nominal coverage rates than the corresponding GMM intervals.

2. Standard GMM estimation and asymptotic properties

First we review the standard (e.g., Hansen, 1982) approach to inference in models characterized by moment restrictions. Let $\psi(Z,\theta)$ be a moment function of a random variable Z and an unknown parameter θ, with $\mathrm{E}[\psi(Z,\theta)]=0$ for a unique value of the parameter, denoted by θ^*. We have a random sample of size N from the distribution of Z, denoted by $z_1,\ldots,z_N$. Point estimates for θ are obtained by choosing the value of θ such that the average moment, $\sum\psi(z_i,\theta)/N$ is as close as possible to zero. When the dimension of the moment function, denoted by M, is larger than K, the dimension of the parameter θ, it is generally not possible to set the average moment exactly equal to zero. The standard Hansen approach to GMM is to estimate θ by minimizing

$$Q_{N,C}(\theta) \equiv \left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi(Z_i,\theta)\right]' C^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\psi(Z_i,\theta)\right],$$

for some positive definite $M\times M$ matrix C. The optimal choice for the weight matrix is the inverse of the expectation of the outer product of the moments,

$$\Delta \equiv \mathrm{E}[\psi(Z,\theta^*)\psi(Z,\theta^*)'].$$

This expectation is typically not known to the researcher so that it is infeasible to minimize $Q_{N,\Delta^{-1}}(\theta)$. Instead researchers typically estimate the inverse of the optimal weight matrix Δ using a consistent, but possibly inefficient, estimator $\tilde{\theta}$:

$$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N}\psi(z_,\tilde{\theta})\psi(z_i,\tilde{\theta})'.$$

In the second step one then estimate θ by minimizing the quadratic form with the estimated Δ:

$$\hat{\theta}^{\text{gmm}} = \operatorname*{argmin}_{\theta} Q_{N,\hat{\Delta}}(\theta).$$

In large samples the distribution of $\hat{\theta}^{\text{gmm}}$ is approximately normal,

$$\sqrt{N}(\hat{\theta}^{\text{gmm}}-\theta^*) \xrightarrow{\text{d}} \mathcal{N}(0,(\Gamma'\Delta^{-1}\Gamma)^{-1}),$$

where

$$\Gamma \equiv \mathrm{E}\frac{\partial\psi}{\partial\theta'}(Z,\theta^*).$$

It is important to note that the large sample distribution of $\hat{\theta}^{\text{gmm}}$ is not affected by the estimation error in the weight matrix Δ^{-1}.

For testing purposes it is also important that the value of the objective function $Q_{N,\Delta}(\theta)$ has a known distribution, both when evaluated at the true value of the parameter and when evaluated at the estimated value:

$$Q_{N,\hat{\Delta}}(\theta^*) \xrightarrow{\mathrm{d}} \mathscr{X}^2(M),$$

and

$$Q_{N,\hat{\Delta}}(\hat{\theta}^{\text{gmm}}) \xrightarrow{\mathrm{d}} \mathscr{X}^2(M-K).$$

Like the normalized limiting distribution of the estimators for θ, these limiting chi-squared distributions are not affected by the estimation of Δ, or more formally, $Q_{N,\hat{\Delta}}(\theta^*) - Q_{N,\Delta}(\theta^*) = \mathrm{o_p}(1)$, and $Q_{N,\hat{\Delta}}(\hat{\theta}^{\text{gmm}}) - Q_{N,\Delta}(\hat{\theta}^{\text{gmm}}) = \mathrm{o_p}(1)$.

We focus on confidence intervals for θ_1, the first element of θ, with θ_2 denoting the remaining part of θ; $\theta = (\theta_1, \theta_2')'$. The first confidence interval we consider is based on the normal approximation to the sampling distribution of $\hat{\theta}_1^{\text{gmm}}$. Let V_{11} be the (1,1) element of $(\Gamma'\Delta^{-1}\Gamma)^{-1}$, the covariance matrix for $\hat{\theta}^{\text{gmm}}$. Then, the standard $100\alpha\%$ confidence interval for θ_1 is

$$\mathrm{CI}_\alpha^{\text{gmm}} = (\hat{\theta}^{\text{gmm}} - z_{1-\alpha/2}\sqrt{V}_{11}, \hat{\theta}^{\text{gmm}} + z_{\alpha+(1-\alpha/2)}\sqrt{V}_{11}), \tag{1}$$

where z_α is the α quantile of the standard normal distribution. Because V is unknown, we have to use an estimate for V, and we use the standard estimate

$$\hat{V} = (\hat{\Gamma}'\hat{\Delta}^{-1}\hat{\Gamma})^{-1},$$

where

$$\hat{\Delta} = \frac{1}{N}\sum_{i=1}^{N}\psi(Z_i,\hat{\theta}^{\text{gmm}})\psi(Z_i,\hat{\theta}^{\text{gmm}})', \quad \text{and} \quad \hat{\Gamma} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial\psi}{\partial\theta'}(Z_i,\hat{\theta}^{\text{gmm}}),$$

are the estimated matrix of expected outer product and derivatives, respectively. This use of estimates does not affect the asymptotic validity of the confidence intervals.

In practice, however, the sampling variation in $\hat{\Delta}$ and $\hat{\Gamma}$ can have important consequences for the accuracy of the asymptotic approximations in finite samples. Altonji and Segal (1996) consider in a simulation study biases arising from sampling variation in $\hat{\Delta}$, and Newey and Smith (2000) document substantial higher order biases arising from correlations between the estimated derivative $\hat{\Gamma}$ and the average moments $\sum\psi(Z_i,\theta^*)/N$.

3. Alternative estimators

The lack of accuracy of asymptotic approximations in finite samples for standard GMM methods motivates the search for alternative methods. To develop such alternatives to the standard GMM estimators for θ we first consider estimators for the distribution function of Z. The idea behind empirical likelihood approaches is that without restrictions on the joint distribution of the observables the empirical distribution function is an optimal estimator for the distribution function. In the presence of restrictions this is no longer true, because the empirical distribution function does not necessarily satisfy these restrictions. We therefore consider estimators for the distribution function that are closest to the empirical distribution function, among those distribution functions that obey all (moment) restrictions. The measure of closeness we use is the Cressie–Read power-divergence statistics (Cressie and Read, 1984; Read and Cressie, 1988; Corcoran, 1995; Imbens et al., 1998). See Smith (1999) for an alternative class of empirical likelihood type estimators. Consider two discrete distributions with common support $s_1,\dots,s_L$ and probabilities $p_1,\dots,p_L$ and $q_1,\dots,q_L$. The power-divergence statistic, indexed by a parameter λ, is

$$I_\lambda(p,q)=\frac{1}{\lambda(1+\lambda)}\sum_{i=1}^{L}p_i\left[\left(\frac{p_i}{q_i}\right)^{\lambda}-1\right].$$

If $p_i=q_i$ for all $i=1,\dots,K$, then the statistic $I_\lambda(p,q)$ equals zero. If some of the probabilities differ, that is, $p_i\neq q_i$ for some i, the power-divergence statistic is positive. We consider this statistic with the support equal to the sample support, $z_1,\dots,z_N$, the first distribution, $\{p_i\}$, equal to the empirical distribution, or $p_i=1/N$ for all i, and choose the distribution $\{q_i\}$ to be the distribution closest to $\{p_i\}$ among all distributions with the same support, that obey the moment restrictions for at least one value of θ. Formally, for a given choice of λ, and with ι an N-vector of ones, we solve

$$\min_{\theta,\pi} I_\lambda(\iota/N,\pi)\quad\text{subject to}\sum_{i=1}^{N}\psi(z_i,\theta)\pi_i=0,\quad\text{and}\sum_{i=1}^{N}\pi_i=1.$$

For any choice of λ the estimator for θ is consistent with a limiting normal distribution that is identical to that of the efficient GMM estimator

$$\sqrt{N}(\hat{\theta}^{\,\mathrm{CR}(\lambda)}-\theta^*)\xrightarrow{\mathrm{d}}\mathcal{N}(0,(\Gamma'\Delta^{-1}\Gamma)^{-1}).$$

The choice of λ does not matter for the asymptotic distribution. It does matter for higher order approximations to the distribution. Newey and Smith (2000) show that all members in this class have, in the case with many over-identifying restrictions, lower bias than the standard GMM estimator.

Three choices for λ have received most of the attention in the literature. Taking the limit $\lambda\to 0$ leads to the empirical likelihood (EL) estimator. If the distribution of Z is in fact discrete, this estimator corresponds to the multinomial maximum likelihood estimator. In that case the estimates for the probabilities, q_i, are also equal to the maximum entropy estimates (e.g., Golan et al., 1996a, b), derived from a different perspective. Newey and Smith (2000) show that this empirical likelihood estimator particularly attractive bias properties when there are many over-identifying restrictions. Fixing $\lambda=-2$ leads to the log Euclidean likelihood (LEL) estimator. This estimator is very similar to standard GMM estimators. It has in fact been developed through a different approach by Hansen et al. (1996) as the continuously updated GMM estimator, obtained by maximizing the quadratic form $Q_{N,\hat{\Delta}(\theta)}(\theta)$ over the θ in the average moments as well as the θ in the weight matrix. See Newey and Smith (2000) for the demonstration of this equivalence. Finally, taking the limit $\lambda \to -1$ gives the exponential tilting (ET) estimator. Although all estimators in this empirical likelihood class share a number of desirable features, we focus in this discussion on the ET estimator for a number of reasons. In simulations it appears more stable than some of the other estimators in the Cressie–Read class, as discussed in Imbens et al. (ISJ, 1998) and it is easier to compute.

With $\lambda \to -1$, the optimization program can be rewritten as maximizing the Kullback—Leibler information criterion, or equivalently, as minimizing Shannon's entropy measure (e.g., Golan et al., 1996b):

$$\min_{\theta,\pi} \sum_{i=1}^{N} \pi_i \ln \pi_i \quad \text{subject to} \sum_{i=1}^{N} \psi(z_i,\theta)\pi_i = 0, \quad \text{and} \sum_{i=1}^{N} \pi_i = 1.$$

This characterization of the estimator for θ is not particularly convenient because it requires minimization in a space with dimension larger than the sample size. In comparison, the standard two-step GMM procedures only requires two minimizations in K-dimensional spaces. One way to simplify this problem is to concentrate out the probabilities π, similar to the dual form in the maximum entropy approach (Golan et al., 1996a). This leads to a characterization of the estimator for θ as the solution to a saddle point problem involving both θ and the normalized Lagrange multipliers in the original optimization program:

$$\max_{\theta} \min_{t} K(t,\theta),$$

with t the normalized Lagrange multipliers, and where $K(t,\theta)$ is the empirical cumulant generating function,

$$K(t,\theta) = \ln\left[\frac{1}{N}\sum_{i=1}^{N} \exp(t'\psi(z_i,\theta))\right], \tag{2}$$

with derivatives $K_t(t,\theta) = (\partial K/\partial\theta)(t,\theta)$ and $K_\theta(t,\theta) = (\partial K/\partial\theta)(t,\theta)$. The implicit estimates of the probabilities π_i are

$$\hat{\pi}_i = \frac{\exp(t'\psi(z_i,\theta))}{\sum_{j=1}^{N}\exp(t'\psi(z_j,\theta))}.$$

For fixed θ, solving for t is straightforward through minimization of $\exp(K(t,\theta))$. This objective function is strictly convex, and with its second derivative easy to calculate, Newton–Raphson methods work very efficiently and fast. Given the implicit function $t(\theta)$ defined as

$$t(\theta) = \operatorname*{argmin}_{t} K(t,\theta),$$

one can then maximize $K(t(\theta),\theta))$ using standard algorithms such as Davidon–Fletcher–Powell (e.g., Gill et al., 1981). In various Monte Carlo experiments we have found this a reliable and fast method for calculating the empirical likelihood estimates.

4. Confidence intervals based on Lagrange multipliers

As the empirical likelihood estimators are in general, that is for all choices of λ in the Cressie–Read characterization, as efficient as the standard GMM estimator, one can construct confidence intervals using the symmetric interval around the estimator using the standard estimator for the variance, $(\Gamma'\Delta^{-1}\Gamma)^{-1}$. Here, we discuss some alternatives. The key to our proposed construction of confidence intervals are the normalized Lagrange multipliers $\hat{t}$ calculated in the empirical likelihood approach. In large samples these Lagrange multipliers have a normal distribution with mean zero and variance depending on variance and average derivative of the moments. Mykland (1995), ISJ, and Smith (2000) discuss tests based on these Lagrange multipliers. See also Diciccio and Romano (1990) and Efron (1981) for the construction of confidence intervals in related settings. Let $\hat{t}$ denoted the solution for t in the optimization problem (2). Then

$$\sqrt{N}\hat{t} \xrightarrow{\mathrm{d}} \mathcal{N}(0, V_t),$$

where

$$V_t = \Delta^{-1} - \Delta^{-1}\Gamma(\Gamma'\Delta^{-1}\Gamma)^{-1}\Gamma'\Delta^{-1}),$$

where, as before, Δ is the covariance matrix of the moments ψ and Γ the matrix of expected derivatives. This covariance matrix V_t is singular with rank equal to the number of over-identifying restrictions $M - K$. In particular, if the model is just-identified, all Lagrange multipliers are equal to zero and the rank of V_t is zero.

The construction of the confidence intervals follows that of tests for the over-identifying restrictions. In the standard approach to GMM such tests are based on the GMM objective function $Q_{N,\hat{\Delta}}(\hat{\theta})$. ISJ suggest alternative tests based on the limiting distribution of

$$\hat{t}'\Delta\hat{t} \xrightarrow{\mathrm{d}} \mathcal{X}^2(M - K).$$

The preferred test statistic in ISJ uses the sandwich estimator for covariance matrix:

$$N\hat{t}'AB^{-1}A\hat{t},$$

where

$$A = \sum_{i=1}^{N} \hat{\pi}_i \psi(z_i, \hat{\theta})\psi(z_i, \hat{\theta})', \quad \text{and} \quad B = \sum_{i=1}^{N} \hat{\pi}_i^2 \psi(z_i, \hat{\theta})\psi(z_i, \hat{\theta})'.$$

Just as the empirical likelihood estimators are first order equivalent to the GMM estimator, these tests for over-identifying restrictions are equivalent, up to first order, to the standard GMM test.

The ISJ tests for over-identifying restrictions are the basis for our proposed confidence intervals. Again we focus on construction of confidence intervals for the first element of the parameter vector θ where $\theta = (\theta_1, \theta_2')'$. Define the Lagrange multiplier as a function of the parameters:

$$t(\theta_1, \theta_2) = \operatorname*{argmin}_t K(t, \theta_1, \theta_2)$$

and the estimator for θ_2 as a function of θ_1:

$$\hat{\theta}_2(\theta_1) = \underset{\theta_2}{\operatorname{argmax}}\, K(t(\theta_1, \theta_2), \theta_1, \theta_2).$$

First consider the just-identified case with the number of moments equal to the number of over-identifying restrictions. With $M = K$ and no binding restrictions (other than the adding up of the probabilities) all the Lagrange multipliers in the program (2) are equal to zero,

$$t(\hat{\theta}_1, \hat{\theta}_2) = 0.$$

Now, consider imposing the restriction $\theta_1 = \theta_1^0$. The implied solutions for θ_2 and t are

$$\theta_2(\theta_1^0), \quad \text{and} \quad t(\theta_1^0, \theta_2(\theta_1^0)).$$

Using the results in ISJ, we have the following convergence in distribution:

$$T(\theta_1^0) = Nt(\theta_1^0, \theta_2(\theta_1^0))' \Delta t(\theta_1^0, \theta_2(\theta_1^0)) \to \mathrm{d}\mathscr{X}^2(1),$$

under the null hypothesis that $\theta_1 = \theta_1^0$. The proposed confidence interval for this case is then the set of θ_1 such that the implied test statistic $T(\theta_1)$ is less than or equal to the appropriate quantile for the chi-squared distribution with one degree of freedom. Formally, the proposed $100\alpha\%$ confidence interval is

$$\mathrm{CI}_\alpha^{\mathrm{et}} = \{\theta_1 | Nt(\theta_1, \theta_2(\theta_1))' \Delta t(\theta_1, \theta_2(\theta_1)) < \mathscr{X}_\alpha^2(1)\}, \tag{3}$$

where the α quantile of a chi-squared distribution with k degrees of freedom is denoted by $\mathscr{X}_\alpha^2(k)$. The estimate for Δ used in the construction is again the sandwich estimator $AB^{-1}A$ as described before.

The justification of this construction of confidence intervals for the just-identified case follows directly from the construction of tests for over-identifying case. This can be extended to construction of confidence intervals for the over-identified case in two ways. First, consider the Lagrange multiplier for the unrestricted estimator. The unrestricted estimator for θ_1 is $\hat{\theta}_1$, and the unrestricted estimator for θ_2 is $\theta_2(\hat{\theta}_1)$, leading to the Lagrange multiplier:

$$t_u = t(\hat{\theta}_1, \theta_2(\hat{\theta}_1)).$$

Unlike in the just-identified case discussed above, t_u will generally differ from zero when $M > K$. It has a limiting normal distribution with rank of the covariance matrix equal to the number of over-idenifying restrictions. For the restricted case,

$$t_r = t(\theta_1^0, \theta_2(\theta_1^0)).$$

The first statistic we consider is based on a second test for the hypothesis that $\theta_1 = \theta_1^0$, using the fact that

$$N(t_u - t_r)' \Delta (t_u - t_r) \tag{4}$$

has a chi-squared distribution with one degree of freedom. The corresponding $100\alpha\%$ confidence interval is

$$\mathrm{CI}_\alpha^{\mathrm{et1}} = \{\theta_1 | N(t(\theta_1, \theta_2(\theta_1)) - t(\hat{\theta}_1, \theta_2(\hat{\theta}_1)))' \Delta (t(\theta_1, \theta_2(\theta_1)) - t(\hat{\theta}_1, \theta_2(\hat{\theta}_1))) < \mathscr{X}_\alpha^2(1)\}.$$

The second statistic we consider as the basis of the confidence intervals is the difference in the over-identifying test statistics:

$$T_1 = Nt_r' \Delta t_r - Nt_u' \Delta t_u.$$

Under the hypothesis that $\theta_1 = \theta_1^0$, this test statistic has again a chi-squared distribution with one degree of freedom. The corresponding $100\alpha\%$ confidence interval is

$$\mathrm{CI}_\alpha^{\mathrm{et2}} = \{\theta_1 | Nt(\theta_1, \theta_2(\theta_1))' \Delta t(\theta_1, \theta_2(\theta_1)) - Nt(\hat{\theta}_1, \theta_2(\hat{\theta}_1))' \Delta t(\hat{\theta}_1, \theta_2(\hat{\theta}_1)) < \mathscr{X}_\alpha^2(1)\}.$$

This construction has a disadvantage. The test statistic is not necessarily positive, which can lead to awkward results for the corresponding confidence intervals.

The following argument shows that in large samples the two confidence intervals are very close (see also Smith, 1999):

$$N(t_u - t_r)' \Delta (t_u - t_r) - (Nt_r' \Delta t_r - Nt_u' \Delta t_u) = 2N(t_u - t_r)' \Delta t_u.$$

Use the Cholesky decomposition to define $\Delta^{1/2}$ such that $\Delta = \Delta^{1/2} \Delta^{1/2'}$, and define $\varepsilon = \Delta^{-1/2} \sum \psi(Z, \theta_0)/\sqrt{N}$. ISJ show that $\sqrt{N} t_u$ can be written as

$$\sqrt{N} t_u = (\mathscr{I} - \Delta^{-1} \Gamma (\Gamma' \Delta^{-1} \Gamma)^{-1} \Gamma') \Delta^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \psi(Z_i, \theta_0) + \mathrm{o_p}(1)$$

$$= (\Delta^{-1/2'} - \Delta^{-1} \Gamma (\Gamma' \Delta^{-1} \Gamma)^{-1} \Gamma' \Delta^{-1/2}) \varepsilon + \mathrm{o_p}(1).$$

Using R_X to denote the matrix calculating the deviation from the least squares projection on X, or $R_X = \mathscr{I} - X(X'X)^{-1}X'$, and defining $\tilde{\Gamma} = \Delta^{-1/2} \Gamma$, we can write this as

$$\sqrt{N} t_u = \Delta^{-1/2'} R_{\tilde{\Gamma}} \varepsilon, \quad \text{and similarly,} \quad \sqrt{N} t_r = \Delta^{-1/2'} R_{\tilde{\Gamma}_2} \varepsilon.$$

Now the difference in test-statistics can be written as

$$2N(t_u - t_r)' \Delta t_u = 2\varepsilon (R_{\tilde{\Gamma}} - R_{\tilde{\Gamma}_2}) \Delta^{-1/2} \Delta \Delta^{-1/2'} R_{\tilde{\Gamma}} \varepsilon + \mathrm{o_p}(1)$$

$$= 2\varepsilon (R_{\tilde{\Gamma}} - R_{\tilde{\Gamma}_2}) R_{\tilde{\Gamma}} \varepsilon + \mathrm{o_p}(1).$$

Because $\tilde{\Gamma} = (\tilde{\Gamma}_1 \tilde{\Gamma}_2)'$, the residual from the projection on $\tilde{\Gamma}$ is orthogonal to both $\tilde{\Gamma}$ and $\tilde{\Gamma}_2$, and hence $(R_{\tilde{\Gamma}} - R_{\tilde{\Gamma}_2}) R_{\tilde{\Gamma}}$ is equal to zero, and

$$N(t_u - t_r)' \Delta (t_u - t_r) - (Nt_r' \Delta t_r - Nt_u' \Delta t_u) = \mathrm{o_p}(1).$$

Hence, the two confidence intervals will be identical in large samples.

5. A Monte Carlo investigation

To compare the finite sample properties of the three methods of constructing confidence intervals we carry out a small Monte Carlo investigation. We focus on three examples where standard asymptotics are not necessarily a good guide to finite sample properties. In each example we simulate 10 000 data sets. For each data set we first estimate the unknown parameter using the standard two-step GMM estimator. The initial weight matrix used is the identity matrix. Given the GMM estimator an estimate

Table 1
Example 1, exponential distribution with two moments

	GMM	ET	Nominal coverage	Coverage rates					
				Wald		LM1		LM2	
				Reg	Boot	Reg	Boot	Reg	Boot
Mean	0.969	0.976	0.900	0.782	0.865	0.860	0.878	0.845	0.889
SD	0.105	0.105	0.950	0.845	0.921	0.918	0.932	0.906	0.937
0.025*	0.774	0.782	0.990	0.908	0.970	0.972	0.978	0.961	0.978
0.975*	1.183	1.189	0.999	0.947	0.987	0.991	0.992	0.985	0.991

*Quantile.

of the variance is obtained and the confidence intervals are constructed. Second, we calculated the exponential tilting estimator and the restricted and unrestricted Lagrange multipliers. We then check whether the true value of the parameter is in the confidence interval by calculating the two test-statistics.

In addition to the coverage rates obtained by the above procedures we report coverage rates from bootstrapped versions of each of these procedures. In each case we use the non-parametric bootstrap to get quantiles for the distribution of the test statistics which are then used to construct the confidence intervals. We do not correct the probabilities with which we sample in the bootstrap to take account of the overidentification, as described in Brown et al. (1997). The motivation is that because we can write the estimator as a solution to a set of moment equations, it can be interpreted as a just-identified GMM estimator, and therefore standard results on bootstrapping for GMM apply. We can then interpret the chi-squared statistics as pivotal, and the bootstrapping can result in improvements of the testing procedures.

5.1. Example 1: exponential distribution with two moments

In the first example there is a single random variable Z and a single parameter θ. There are two moment conditions, connecting the first and second moment of the random variable to the unknown parameter.

$$\psi(z,\theta) = \begin{pmatrix} z-\theta \\ z^2 - 2\theta^2 \end{pmatrix}.$$

In the simulations Z has a exponential distribution with mean $\theta = 1$. The number of observations is 100. The number of replications reported in Table 1 is 10 000.

The results in Table 1 report the summary statistics for the distribution of the two estimators, means and standard deviations over the 10 000 replications and 0.025 and 0.975 quantiles. It can be seen that the sampling distribution of especially the standard GMM estimator is very asymmetric, due to the extreme skewness of the second moment. This carries over into the properties of the confidence intervals and leads the coverage rate of the standard, symmetric, confidence interval to be below the nominal levels. The confidence intervals based on the Lagrange multipliers perform much better, especially the one based on the difference in Lagrange multipliers in (4).

Table 2
Example 2, Burnside–Eichenbaum

	GMM	ET	Nominal coverage	Coverage rates					
				Wald		LM1		LM2	
				Reg	Boot	Reg	Boot	Reg	Boot
Mean	0.965	0.977	0.900	0.720	0.887	0.783	0.904	0.801	0.930
SD	0.045	0.045	0.950	0.797	0.941	0.855	0.953	0.867	0.969
0.025	0.873	0.888	0.990	0.939	0.987	0.938	0.991	0.939	0.995
0.975	1.060	1.072	0.999	0.987	0.999	0.979	0.999	0.979	0.999

The bootstrapped confidence intervals all perform much better than the corresponding non-bootstrapped versions, with the bootstrapped Lagrange multiplier based intervals still superior to the bootstrapped standard GMM ones.

5.2. Example 2: Burnside–Eichenbaum

This example is related to one studied by Burnside and Eichenbaum (1996) and Altonji and Segal (1996). Let $Z_1,\ldots,Z_M$ be independent normal random variables with mean zero and variance θ. We consider estimation of θ through the M moments:

$$\psi(z_1,\ldots,z_M,\theta)=\begin{pmatrix} z_1^2-\theta \\ z_2^2-\theta \\ \vdots \\ z_M^2-\theta \end{pmatrix}. \tag{5}$$

In the simulations we use samples of size 100 and $M=10$ moments. The true value of θ is one.

The results for this data generating process are in Table 2. Again we calculate both the standard GMM and exponential tilting estimators. The same three confidence intervals as before are constructed and the frequency with which the confidence interval does not include the true value is reported. The exponential tilting estimator is slightly less biased than the standard GMM estimator. Its sampling distribution is also more symmetric. Altonji and Segal focus on the bias in the estimation of θ in this context and attribute it to estimation of the weight matrix in the GMM procedure.

The performance of the Lagrange-multiplier-based confidence intervals is again considerably better than those of the standard confidence interval. The interval based on the difference in Lagrange multipliers gets the undercoverage down to about half that of the standard interval, and the interval based on the difference in over-identifying test-statistics is even more accurate.

The bootstrapped intervals perform quite well here, with nominal and actual coverage rates very close.

Table 3
Example 3, normal distribution

	GMM	ET	Nominal coverage	Coverage rates					
				Wald		LM1		LM2	
				Reg	Boot	Reg	Boot	Reg	Boot
Mean	−0.003	−0.003	0.900	0.846	0.926	0.883	0.922	0.899	0.912
SD	0.032	0.032	0.950	0.900	0.965	0.935	0.956	0.952	0.969
0.025	−0.072	−0.069	0.990	0.970	0.996	0.985	0.993	0.991	0.993
0.975	0.068	0.066	0.999	0.991	1.000	0.997	1.000	0.999	1.000

5.3. *Example 3: normal distribution with five moments*

In the third example the data come from a normal distribution. We observe a normal random variable with mean θ and unit variance. The moments we consider are based on the first 5 cumulants:

$$\psi(z,\theta)=\begin{pmatrix} z-\theta \\ z^2-\theta^2-1 \\ z^3-\theta^3-3\theta \\ z^4-\theta^4-6\theta^2-3 \\ z^5-\theta^5-10\theta^3-15\theta \end{pmatrix}. \tag{6}$$

The number of observations is 1000, and the true value for θ is 0. In this example some of the moments have large kurtosis, suggesting that the estimates of the covariance matrix of the moments may be imprecise. Table 3 reports the results for this example. Again we find that the Lagrange multiplier based confidence intervals outperform the standard intervals by a wide margin.

In this example, the bootstrapped intervals improve the coverage rate of the standard intervals but do not significantly affect the Lagrange multiplier based intervals which are quite accurate already.

6. Conclusion

In this paper, we suggest alternatives ways of constructing confidence intervals in generalized method of moments settings. The confidence intervals apply both in the just and over-identified case and have the same first order asymptotic properties as the standard intervals. We study the finite sample properties of the intervals in a number of examples. In two of the examples the proposed confidence intervals perform considerably better with closely similar nominal and actual coverage rates. In the third example all confidence intervals perform poorly, with the proposed intervals performing best. Comparisons with bootstrapped versions of all intervals suggests that although bootstrapping generally improves the performance of all confidence intervals, after bootstrapping the empirical likelihood based intervals continue to behave better.

References

Altonji, J., Segal, L., 1996. Small sample bias in GMM estimation of covariance structures. Journal of Business and Economic Statistics 14, 353–366.

Back, K., Brown, D., 1990. Estimating distributions from moment restrictions. Working paper, Graduate School of Business, Indiana University.

Brown, B., Newey, W., May, S., 1997. Efficient bootstrapping for GMM, unpublished manuscript, department of economics, MIT.

Burnside, C., Eichenbaum, M., 1996. Small sample properties of generalized method of moments based Wald tests. Journal of Business and Economic Statistics 14, 294–308.

Corcoran, S., 1995. Empirical discrepancy theory, Manuscript, Department of Statistics, Oxford University.

Cressie, N., Read, T., 1984. Multinomial goodness-of-fit tests. Journal of the Royal Statistical Society, Series B 46, 440–464.

Diciccio, T., Romano, J., 1990. Nonparametric confidence limits by resampling methods and least favorable families. International Statistical Review 9&58, 59–76.

Efron, B., 1981. Nonparametric standard errors and confidence intervals (with discussion). Canadian Journal of Statistics 9, 139–172.

Gill, P., Murray, W., Wright, M., 1981. Practical Optimization. Academic Press, San Diego.

Golan, A., Judge, G., Miller, A., 1996a. Maximum Entropy Econonometrics, Wiley, New York.

Golan, A., Judge, G., Perloff, J., 1996b. A maximum entropy approach to recovering information from multinomial data. Journal of the American Statistical Association 91, 841–853.

Hansen, L.-P., 1982. Large sample properties of generalized method of moment estimators. Econometrica 50, 1029–1054.

Hansen, L.-P., Heaton, J., Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14 (3), 262–280.

Hall, P., Horowitz, J., 1996. Boostrap critical values for tests based on generalized method of moment estimators. Econometrica 64, 891–916.

Imbens, G.W., 1997. One-step estimators in overidentified generalized method of moments estimator. Review of Economic Studies 64, 359–383.

Imbens, G.W., Spady, R.H., Johnson, P., 1998. Information theoretic approaches to inference in moment condition models. Econometrica 66 (2), 333–357.

Kitamura, Y., Stutzer, M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.

Mykland, P., 1995. Dual likelihood. Annals of Statistics 23, 396–421.

Newey, W., Smith, R., 2000. Asymptotic bias and equivalence of GMM and GEL. Unpublished manuscript, Department of Economics, MIT, Cambridge, MA.

Owen, A., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.

Qin, J., Lawless, J., 1994. Generalized estimating equations. Annals of Statistics 22, 300–325.

Read, T., Cressie, N., 1988. Goodness-of-Fit Statistics for Discrete Multivariate Data. Springer, New York.

Robertson, J., Pagan, A., 1997. Exploring some studies of the finite sample performance of GMM. Mimeo, Department of Economics, Australian National University.

Smith, R., 1997. Alternative semi-parametric likelihood approaches to generalized method of moments estimation. Economic Journal 107 (441), 503–519.

Smith, R., 1999. Generalized empirical likelihood methods for moment condition models. Unpublished manuscript, Department of Economics, University of Bristol.

Smith, R., 2000. Empirical likelihood estimation and inference. In: Salmon, M., Marriott, D. (Eds.), Applications of Differential Geometry to Econometrics. Cambridge University Press, Cambridge (Chapter 4).