Econometric Reviews
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597248

Finite Sample Evidence Suggesting a Heavy Tail Problem of the Generalized Empirical Likelihood Estimator

Patrik Guggenberger [a]
[a] Department of Economics, U.C.L.A., Los Angeles, California, USA

Online Publication Date: 01 July 2008

To cite this Article: Guggenberger, Patrik (2008) 'Finite Sample Evidence Suggesting a Heavy Tail Problem of the Generalized Empirical Likelihood Estimator', Econometric Reviews, 27:4, 526 — 541

To link to this article: DOI: 10.1080/07474930801960410
URL: http://dx.doi.org/10.1080/07474930801960410

Econometric Reviews, 26(4–6):526–541, 2008

ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474930801960410

FINITE SAMPLE EVIDENCE SUGGESTING A HEAVY TAIL PROBLEM OF THE GENERALIZED EMPIRICAL LIKELIHOOD ESTIMATOR

Patrik Guggenberger

Department of Economics, U.C.L.A., Los Angeles, California, USA

□ *Comprehensive Monte Carlo evidence is provided that compares the finite sample properties of generalized empirical likelihood (GEL) estimators to the ones of k-class estimators in the linear instrumental variables (IV) model. We focus on sample median, mean, mean squared error, and on the coverage probability and length of confidence intervals obtained from inverting a t-statistic based on the various estimators. The results indicate that in terms of the above criteria, all the GEL estimators and the limited information maximum likelihood (LIML) estimator behave very similarly. This suggests that GEL estimators might also share the "no-moment" problem of LIML. At sample sizes as in our Monte Carlo study, there is no systematic bias advantage of GEL estimators over k-class estimators. On the other hand, the standard deviation of GEL estimators is pronouncedly higher than for some of the k-class estimators. Therefore, if mean squared error is used as the underlying loss function, our study suggests the use of computationally simple estimators, such as two-stage least squares, in the linear IV model rather than GEL. Based on the properties of confidence intervals, we cannot recommend the use of GEL estimators either in the linear IV model.*

Keywords Generalized empirical likelihood estimator; Generalized method of moments; Monte Carlo simulation; No-moment problem.

JEL Classification C13; C15; C30.

1. INTRODUCTION

The class of generalized empirical likelihood (GEL) (Smith, 1997) estimators enjoys some desirable theoretical asymptotic properties relative to generalized method of moments (GMM) (Hansen, 1982) estimators.[1]

Received April 2004; Accepted December 2006
Address correspondence to Patrik Guggenberger, Bunche Hall 8385, Department of Economics UCLA, Box 951477, Los Angeles, CA 90095-1477; E-mail: guggenbe@econ.ucla.edu

[1]Particular examples of GEL estimators include the continuous updating (CU) estimator of Hansen et al. (1996), the empirical likelihood (EL) estimator of Owen (1988) and Qin and Lawless (1994), and the exponential tilting (ET) estimator, see, e.g., Imbens (1997) and Kitamura and Stutzer (1997).

For example, Newey and Smith (2004) (NS henceforth) show that for many models the dominant asymptotic bias of the empirical likelihood (EL) estimator does not grow with the number of moment restrictions, while that of the GMM estimator grows without bound.

On the other hand, very little is known about the theoretical finite sample properties of the GEL estimator and it is unclear whether the higher order advantages of GEL over GMM estimators translate into improved finite sample performance at empirically relevant sample sizes.[2]

The goal of this article then is to shed more light on the finite sample properties of GEL estimators via Monte Carlo simulation. More precisely, we study the properties of the continuous updating (CU), EL, and exponential tilting (ET) estimator and compare them to k-class estimators, namely, two-stage least squares (2SLS), limited information maximum likelihood (LIML) and Fuller's (1977) estimator with respect to several loss functions all in the context of a linear instrumental variables model. We are interested in how different GEL estimators perform compared to each other and compared to k-class estimators. We investigate whether the higher order bias advantages of GEL estimators are reflected in their finite sample performance. How does the mean and median bias compare to the one of classical k-class estimators, how is the bias affected by an increase in the number of instruments and do the finite sample results indicate a "no-moment problem" of the GEL estimator? How do the estimators perform with respect to a mean squared error (MSE) loss function and how reliable and precise are inference methods based on these estimators?

The main findings are summarized as follows. The finite sample properties of CU, EL, and ET are virtually identical with respect to all the above criteria under investigation in the Monte Carlo study. They are overall very similar to the properties of LIML. This suggests that

[2]With respect to moments, Mariano and Sawa (1972), Fuller (1977), and Kinal (1980) show that the limited information maximum likelihood (LIML) estimator $\hat{\beta}_{LIML}$ does not have any moments finite, i.e. $E\|\hat{\beta}_{LIML}\| = \infty$, the two-stage least squares (2SLS) estimator has as many moments finite as there are overidentifying restrictions and Fuller's estimator has at least two moments finite, $E\|\hat{\beta}_{Fuller}\|^2 < \infty$. To obtain a corresponding result for GEL estimators, $\hat{\beta}_{GEL}$, in the general case, is very difficult. GEL estimators are not given in closed form which makes theoretical finite sample analysis very complicated. In a linear IV model, it is easily shown that the CU estimator coincides with LIML and 2SLS in the just-identified case (under weak conditions that guarantee that the weighting matrix in the CU criterion function is positive definite) and therefore does not have moments in this case. However, even for the simplest overidentified case, where the parameter of interest is of dimension 1 and there are two instruments, a theoretical investigation into the moment question becomes extremely difficult. Kunitomo and Matsushita (2003) attempts a proof for the nonexistence of moments of the EL estimator for the overidentified linear model. However, in their proof they do not really consider the EL estimator because they artificially bound the empirical probabilities (p_i in their notation) away from zero (see their condition $0 < \epsilon \leq p_i < 1$ on p. 4, line 7↑ in their article for some sufficiently small $\epsilon > 0$). It is therefore still an unresolved task to analytically prove the conjectured nonexistence of moments of GEL estimators.

GEL estimators may have a "no-moment problem," $E\|\hat{\beta}_{GEL}\| = \infty$, just like LIML, which corroborates findings in Guggenberger and Hahn (2005).[3] In accordance with NS, our Monte Carlo results suggest that the finite sample bias of the GEL estimators is not negatively affected in a systematic manner by an increase in the number of moment conditions. While there is no clear ranking of GEL and k-class estimators with respect to bias properties, the standard deviation (Std) of GEL often exceeds the Std of 2SLS and Fuller, oftentimes by several magnitudes. Therefore, if MSE is the relevant loss function, we cannot recommend the use of GEL estimators in the linear instrumental variables model based on our simulation results. If interest focuses on the coverage probability of a confidence interval (CI) based on inverting a t-statistic based on a particular estimator, we find that, in an overall sense, Fuller-based CIs perform best, but they are not uniformly better than those based on 2SLS or GEL. As an important extension, one should study the finite sample performance of GEL estimators in several nonlinear models, because in nonlinear models they may have bias advantages over k-class estimators that potentially offset their higher Std, see NS.

Unfortunately, the definition of GEL estimators as the solution of a (potentially high-dimensional) saddle point problem (the $\arg\min - \sup$ problem in (2.3)) also makes the finite sample investigation of GEL estimators via Monte Carlo study extremely difficult. While the "inner maximization" problem is globally concave and one can rely on derivative-based maximization routines such as the Newton–Raphson algorithm, solving the "outer minimization" routine is much more challenging because several local minima may exist. Instead of relying on a minimization routine, such as Nelder–Meade, we evaluate the criterion function of the GEL estimators for all values in the parameter space on a very fine grid. This approach guarantees numerical stability and works successfully even in very weakly identified models to which we pay particular attention in our study. Obviously, this approach is very time consuming and would be too expensive in terms of computation time for Monte Carlo setups where the structural parameter vector is of dimension bigger than one and therefore we restrict attention to the scalar case. While grid search routines have been suggested elsewhere for isolated problem cases where minimization routines did not converge, see, e.g., Hansen et al. (1996) and Imbens et al. (1998) for further

[3]If an estimator does not have any finite moments, then its finite sample risk is infinite for many loss functions, including mean squared error loss. However, as discussed in a very interesting article by Zaman (1981), the use of estimators with no moments can be appropriate for other loss functions. For example, if the goal is to estimate $\theta = \beta^{-1}$ given iid data from a $N(\beta, \sigma^2)$ distribution with known variance σ^2, (Zaman, 1981, p. 291) proves that "asymptotically, for any interval around the true value, the maximum likelihood estimator has greater probability of being in that interval than any other asymptotically median unbiased estimator"—a very nice optimality result. However, the MLE estimator in this example does not have any finite moments.

discussion, to my knowledge, this is the first large scale systematic Monte Carlo study of the finite sample properties of GEL estimators that is based on this grid search approach. When implementing a grid search approach, the set over which the search is performed, necessarily, has to be restricted to a bounded set and therefore, strictly speaking, our Monte Carlo study investigates the properties of the *truncated version* of the GEL estimators. We choose the relatively large set $[-25, 25]$ for our search, where the true value of the structural parameter coefficient is 0. To simplify notation, rather than talking about the truncated GEL estimator, we will simply talk about the GEL estimator from now on, even though, strictly speaking this is wrong.

An important question that has intrigued finite sample theorists concerns the interpretation of sample moments of random variables whose moments are infinite, see Phillips (2003, p. 508), for a more detailed discussion of this issue. Even if an estimator has infinite moments, finite pseudomoments of the estimator, also called approximations, (Nagar, 1959), can be defined as the corresponding moments of an Edgeworth expansion of the estimator. Sargan (1988) collects several articles by Sargan that give conditions under which these moment approximations are valid. For the case when an estimator has infinite moments, Sargan (1982) shows that Monte Carlo sample moments are in fact estimating these moments of the Edgeworth approximation. See also Maasoumi (1977) for additional work on this subject matter.

We now briefly discuss various other recent articles that are concerned with Monte Carlo simulation of the finite sample properties of GEL in the linear instrumental variables model, namely, Mittelhammer et al. (2005), Kunitomo and Matsushita (2003), Anderson et al. (2005), Guggenberger (2005), and Guggenberger and Hahn (2005). Guggenberger and Hahn (2005) investigate the finite sample properties of EL indirectly by studying the finite sample properties of its two-step version that has a closed form solution and the same relevant higher order bias properties as EL. They find strong evidence for a "no-moment problem" of the EL estimator. Analogous findings are obtained in the small simulation study Guggenberger (2005) for the CU estimator. Mittelhammer et al. (2005) calculate GEL estimators using a combination of minimization routines including Nelder–Meade. They consider only relatively strongly identified scenarios (with $R^2 > 0.25$) and while evidence for higher variance of GEL estimators compared to 2SLS is reported, based on the strongly identified models in their Monte Carlo setup, no "no-moment problem" of the GEL estimators is suspected. If R^2 is very small, the criterion function for the GEL estimator is typically very flat and any of the standard minimization routines is known to run into problems. Either the routine does not converge or if it does, suspicion remains whether the true minimizer has been found. Kunitomo and Matsushita (2003) perform Monte Carlo simulations for an EL type and GMM estimator in the linear instrumental variables model with one included exogenous and one included endogenous variable.[4] They find that the EL type estimator is approximately median unbiased but—consistent with our findings—has a fat tailed distribution. Anderson et al. (2005) provide further evidence on these findings. Besides normal errors they allow for a variety of other error distributions and also consider conditional heteroskedasticity. They advocate the use of LIML because they find that it shares the nice median bias properties of EL but has tighter tails and is computationally simpler than EL.

2. MONTE CARLO EXPERIMENT

2.1. Model and Theoretical Properties

We investigate the finite sample properties of the EL, ET, and CU estimator applied to a simple linear instrumental variables model as in, for example, Hahn and Hausman (2002)

$$\begin{aligned} y_i &= x_i\beta_0 + \varepsilon_i, \\ x_i &= z_i'\pi + u_i, \qquad i = 1, \ldots, n, \end{aligned} \tag{2.1}$$

where y_i, x_i, and z_i denote the dependent variable, an (endogenous) scalar regressor and the vector of instrumental variables and n is the sample size. The value of β_0 is set equal to 0. Let z_i and $(\varepsilon_i, u_i)'$ be independent and IID $\mathcal{N}(0, I_K)$ and $\mathcal{N}(0, \Omega)$, respectively, where I_K is the K-dimensional identity matrix and Ω is a 2×2-matrix with diagonal and off diagonal elements 1 and ρ, respectively. Let

$$\mathbb{R}^2 \equiv E[(\pi' z_i)^2]/(E[(\pi' z_i)^2] + E[v_i^2]) = \pi'\pi/(\pi'\pi + 1)$$

denote the theoretical R^2 of the first stage regression. Assume $\pi = (\eta, \eta, \ldots, \eta)'$ and thus

$$\mathbb{R}^2 = \frac{K \cdot \eta^2}{K \cdot \eta^2 + 1}. \tag{2.2}$$

We simulate data for all the possible parameter combinations of

$$\begin{aligned} n &= 100, 200, \\ K &= 1, 5, 20, \end{aligned}$$

[4]As explained in an earlier footnote, a version of the EL estimator is used in which the empirical probabilities are bounded away from zero. It is not specified how the calculation of the EL estimator is implemented. In their case, the outer minimization problem in (2.3) is two-dimensional which probably makes a grid search routine too costly. Other numerical procedures face the just described problems of numerical stability especially in weakly identified scenarios.

$$\mathbb{R}^2 = .001, .01, .1, \qquad \text{and}$$
$$\rho = 0, .3, .5, .9.$$

Different choices of K allow us to investigate how the degree of over-identification affects the performance of the estimators. The parameter ρ determines the degree of endogeneity that ranges from none ($\rho = 0$) to very strong ($\rho = 0.9$). Finally, equation (2.2) implies $\eta = (\mathbb{R}^2/(K(1 - \mathbb{R}^2)))^{1/2}$ and therefore $\mathbb{R}^2$ (together with K) pins down the strength of the instruments.

Like GMM, the GEL estimators exploit the moment condition $E\varphi_i(\beta_0) = 0$, where $\varphi_i(\beta) \equiv (y_i - x_i\beta)z_i$. The general definition of GEL estimators is as follows. Let ϑ be a real-valued function $Q \to R$, where Q is an open interval of the real line that contains 0 and $\widehat{\Lambda}_n(\beta) := \{\lambda \in R^K : \lambda'\varphi_i(\beta) \in Q \text{ for } i = 1, \ldots, n\}$. Set $\vartheta_j(v) := (\partial^j \vartheta/\partial v^j)(v)$ and $\vartheta_j := \vartheta_j(0)$ for non-negative integers j. The GEL estimator is the solution to the saddle point problem

$$\hat{\beta}_\vartheta := \arg\min_{\beta \in B} \sup_{\lambda \in \widehat{\Lambda}_n(\beta)} \widehat{P}_\vartheta(\beta, \lambda), \tag{2.3}$$

where

$$\widehat{P}_\vartheta(\beta, \lambda) := \left(2\sum_{i=1}^{n} \vartheta(\lambda'\varphi_i(\beta))/n\right) - 2\vartheta_0 \tag{2.4}$$

and B is the parameter space for β. The CU, EL, and ET estimators correspond to $\vartheta(v) = -(1 + v)^2/2$, $\vartheta(v) = \ln(1 - v)$ and $\vartheta(v) = -\exp v$, respectively. In case of CU, one can explicitly solve for $\lambda(\beta)$, the λ that for a given β solves the inner maximization problem in (2.3). For other GEL estimators, it can be shown that $\lambda(\beta)$ exists with probability approaching 1, see NS and Guggenberger and Smith (2005). For CU it follows that[5]

$$\hat{\beta}_{CU} \equiv \arg\min_{\beta \in B} Q_n(\beta), \qquad \text{for}$$
$$Q_n(\beta) \equiv \left[\sum_{i=1}^{n} \varphi_i(\beta)\right]' \left[\sum_{i=1}^{n} \varphi_i(\beta)\varphi_i(\beta)'\right]^{-1} \left[\sum_{i=1}^{n} \varphi_i(\beta)\right].$$

Under technical conditions satisfied in our simulations, NS (Corollary 4.3, p. 229) show that the dominant higher order bias of the EL estimator equals the asymptotic bias for the (infeasible) GMM estimator that uses

[5]NS (2004, p. 222) show that this definition of the CUE $\hat{\beta}_{CU}$ coincides with the one given in Hansen et al. (1996), where in $Q_n(\beta)$ the term $\sum_{i=1}^{n} \varphi_i(\beta)\varphi_i(\beta)'$ is replaced by $\sum_{i=1}^{n} \varphi_i(\beta)\varphi_i(\beta)' - \sum_{i=1}^{n} \varphi_i(\beta) \sum_{i=1}^{n} \varphi_i(\beta)'$.

the optimal (asymptotic variance minimizing) linear combination.[6] On the other hand, in general, there are additional bias terms for GMM that stem from the estimation of the Jacobian and second moment matrices and from the use of a preliminary estimator needed to implement the two-step GMM estimator. Based on these theoretical asymptotic findings, one would expect the bias of EL to be smaller than the bias of GMM; this is to be expected especially in nonlinear models, where the estimation of the nonconstant Jacobian matrix constitutes an important source of bias.[7] NS (Theorem 4.5, p. 230) also show that the bias of EL is bounded as a function of the number of overidentifying restrictions, while—at least in certain heteroskedastic models—the bias of GMM grows linearly in K. It would therefore be instructive to also consider heteroskedastic designs and/or to study nonlinear models to see if this leads to a different bias behavior of GMM and EL in finite samples.

2.2. Implementation and Results

In the simulations, when implementing the grid search, it is obviously impossible to allow for the unbounded search interval $B = R$ for β. We choose $B = [-c, c]$ for $c = 25$ which results in a truncated version of the GEL estimator. In a "brute force" approach, we calculate $\widehat{P}_\vartheta(\beta, \lambda(\beta))$ for all $\beta \in B$ on a grid with stepsize 0.01 and pick $\hat{\beta}_\vartheta$ as the β that minimizes $\widehat{P}_\vartheta(\beta, \lambda(\beta))$ on this grid. Sensitivity analysis with respect to the stepsize shows that there is virtually no change in the results by using even smaller stepsizes than 0.01.

While for CU, $\widehat{P}_\vartheta(\beta, \lambda(\beta))$ is given in closed form, for EL and ET the maximization problem $\max_{\lambda \in \widehat{\Lambda}_n(\theta)} \widehat{P}_\vartheta(\theta, \lambda)$ has to be solved numerically for each $\beta \in B$. However, in this inner maximization problem in (2.3), the function to be maximized is globally concave. We implement a variant of the Newton–Raphson algorithm with implemented step size control, see Hamilton (1994, Chapter 5) for a reference. The maximization algorithm is initialized by setting λ equal to the zero K-vector and then at each iteration the algorithm tries several shrinking stepsizes in the search direction and accepts the first one that increases the value of the criterion function compared to the previous value for λ. This procedure enforces an "uphill climbing" feature of the algorithm and turns out to be very stable and reliable.

[6]Denote by $\varphi_{ij}(\beta)$ the jth component of $\varphi_i(\beta)$. Under the third moment condition $E\varphi_i(\beta_0)\varphi_i'(\beta_0)\varphi_{ij}(\beta_0) = 0$ the higher order asymptotic bias of all GEL estimators coincides, see NS (Corollary 4.4, p. 229).

[7]But note that even in the linear model with symmetric erros, GEL estimators have a theoretical higher order bias advantage over GMM if the preliminary estimator for two-step GMM is inefficient (see NS, p. 230).

For GEL estimators, the criterion function $\widehat{P}_\vartheta(\beta, \lambda)$ can be very flat and/or have multiple local extrema and therefore optimization routines (for the outer minimization in (2.3)), such as Newton–Raphson or Nelder–Meade, can be unstable and can often lead to wrong results. The heavy computational burden and unreliability of minimization routines for GEL estimators has been widely recognized, see, e.g., Hansen et al. (1996), Imbens et al. (1998), and Mittelhammer et al. (2005). That is why in our simulations we employ the stable but extremely time consuming "brute force" approach of grid evaluation, that has been suggested also by Hansen et al. (1996) for certain problem scenarios but, to our knowledge, has never been used in a comprehensive Monte Carlo study to investigate the finite sample properties of GEL. In scenarios where instruments are only weakly correlated with the endogenous regressor, the saddle point problem of the GEL estimator becomes particularly intractable. However, the grid search approach used here, finds the correct minimizer within the specified compact set. Of course, we do not recommend such a time consuming algorithm in practice when β is of high dimension. It is implemented here, because for scalar β, it is still feasible.

All the results reported below are based on 5,000 simulation repetitions. For each repetition, we calculate the EL, ET, CU estimators and, for comparison, the LIML, 2SLS, and Fuller's estimator. (We choose $\alpha = 1$ in the definition of the Fuller estimator, see, Fuller, 1977, p. 942). We restrict the latter estimators to B as well, that is, if, for example, the LIML estimator is bigger (smaller) than c $(-c)$ for a given data sample, it is set equal to c $(-c)$.

We then calculate the sample mean, median, Std, and root mean squared error (RMSE) of the estimators over the 5,000 samples. To gain insight into how binding the restriction onto the interval $[-c, c]$ is, we also calculate the percentage of times that the estimators fall into the "10%-tail" $[-c, -0.9c] \cup [0.9c, c]$ of the interval $[-c, c]$. We call this probability the "10%-tail-probability." Finally, we report actual coverage probability and median length of nominal 95% two-sided symmetric CIs based on inverting a t-statistic based on the various estimators. The 95% two-sided symmetric CI based on an estimator $\hat{\beta}$ is given by

$$\left\{\beta \in R : |\hat{\beta} - \beta| < 1.96\left[n^{-1}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2/(x'z(z'z)^{-1}z'x)\right]^{1/2}\right\}$$

and its length is given by $2 \times 1.96[n^{-1}\sum_{i=1}^{n}(y_i - x_i\hat{\beta})^2/(x'z(z'z)^{-1}z'x)]^{1/2}$, where $x = (x_1, \ldots, x_n)'$ and $z = (z_1, \ldots, z_n)'$.

Tables 1–4 below contain a subset of the simulation results, namely, all combinations of $n = (100, 200)$, $\mathbb{R}^2 = (0.001, 0.1)$, $K = (1, 5, 20)$ and

TABLE 1(a) $\mathbb{R}^2 = 0.001$, $\rho = 0.3$, $n = 100$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	0.36	0.30	5.24	5.25	1.00	7.37	0.03
LIML	0.36	0.30	5.24	5.25	1.00	7.37	0.03
2SLS	0.36	0.30	5.24	5.25	1.00	7.37	0.03
Fuller	0.29	0.29	0.38	0.48	1.00	5.73	0.00
				$K = 5$			
EL	0.26	0.27	5.11	5.12	0.82	2.55	0.02
LIML	0.21	0.28	5.25	5.26	0.82	2.55	0.02
2SLS	0.29	0.29	0.55	0.63	0.95	1.90	0.00
Fuller	0.29	0.28	0.63	0.69	0.90	2.05	0.00
				$K = 20$			
EL	0.29	0.30	5.35	5.36	0.27	1.18	0.03
LIML	0.19	0.30	5.31	5.32	0.27	1.17	0.02
2SLS	0.30	0.30	0.22	0.37	0.72	0.85	0.00
Fuller	0.29	0.30	0.86	0.91	0.39	1.02	0.00

$\rho = (0.3, 0.9)$. Tables 1(a)–4(a) contain the results for $n = 100$, Tables 1(b)–4(b) those for $n = 200$. The remaining results are available from the author upon request. The columns of each table report the sample mean, median, Std, RMSE, actual coverage probability, and median length of nominal 95% CIs and the 10%-tail-probabilities for the estimators EL, LIML, 2SLS, and Fuller. To save space, we do not explicitly report results for ET and CU because they coincide numerically (with very few exceptions) to the ones of EL for the median, the 10%-tail

TABLE 1(b) $\mathbb{R}^2 = 0.001$, $\rho = 0.3$, $n = 200$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	0.23	0.22	5.18	5.18	1.00	7.19	0.02
LIML	0.24	0.22	5.18	5.18	1.00	7.20	0.02
2SLS	0.24	0.22	5.18	5.18	1.00	7.20	0.02
Fuller	0.28	0.28	0.39	0.48	1.00	5.55	0.00
				$K = 5$			
EL	0.13	0.24	5.70	5.70	0.82	2.50	0.03
LIML	0.15	0.25	5.48	5.48	0.83	2.48	0.03
2SLS	0.29	0.29	0.56	0.63	0.95	1.87	0.00
Fuller	0.28	0.27	0.62	0.68	0.90	2.04	0.00
				$K = 20$			
EL	0.21	0.28	5.44	5.45	0.27	1.20	0.03
LIML	0.25	0.29	5.30	5.30	0.27	1.17	0.03
2SLS	0.30	0.30	0.22	0.37	0.71	0.86	0.00
Fuller	0.30	0.29	0.86	0.91	0.39	1.02	0.00

TABLE 2(a) $\mathbb{R}^2 = 0.1, \ \rho = 0.3, \ n = 100$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	−0.03	0.01	0.71	0.71	0.97	1.20	0.00
LIML	−0.03	0.01	0.71	0.71	0.97	1.20	0.00
2SLS	−0.03	0.01	0.71	0.71	0.97	1.20	0.00
Fuller	0.02	0.03	0.29	0.29	0.97	1.18	0.00
				$K = 5$			
EL	−0.05	0.01	1.68	1.68	0.86	1.04	0.00
LIML	−0.06	0.00	1.59	1.59	0.88	1.04	0.00
2SLS	0.07	0.09	0.28	0.29	0.95	1.00	0.00
Fuller	0.02	0.03	0.37	0.37	0.92	1.03	0.00
				$K = 20$			
EL	−0.04	0.03	3.10	3.10	0.46	0.78	0.01
LIML	−0.05	0.03	2.81	2.81	0.52	0.75	0.01
2SLS	0.19	0.19	0.18	0.26	0.79	0.69	0.00
Fuller	0.04	0.05	0.56	0.56	0.58	0.74	0.00

probability, the coverage probability, and median length of CIs and are qualitatively identical to the ones of EL for the other measures.

We now discuss the relative performance of each estimator with respect to the various different measures under investigation using the data from the tables and including additional insight from the simulation results not contained in the tables. We also discuss the effect of n, $\mathbb{R}^2$, K, and ρ on the performance of the estimators.

TABLE 2(b) $\mathbb{R}^2 = 0.1, \ \rho = 0.3, \ n = 200$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	−0.02	0.00	0.23	0.24	0.96	0.84	0.00
LIML	−0.02	0.00	0.23	0.24	0.96	0.84	0.00
2SLS	−0.02	0.00	0.23	0.24	0.96	0.84	0.00
Fuller	0.00	0.02	0.22	0.22	0.97	0.83	0.00
				$K = 5$			
EL	−0.02	0.00	0.50	0.50	0.91	0.77	0.00
LIML	−0.02	0.00	0.32	0.32	0.92	0.77	0.00
2SLS	0.04	0.05	0.20	0.21	0.95	0.76	0.00
Fuller	0.00	0.01	0.24	0.24	0.93	0.77	0.00
				$K = 20$			
EL	−0.09	0.01	1.31	1.31	0.65	0.63	0.00
LIML	−0.06	0.00	1.27	1.27	0.71	0.62	0.00
2SLS	0.14	0.14	0.16	0.21	0.83	0.59	0.00
Fuller	0.00	0.02	0.35	0.35	0.73	0.62	0.00

TABLE 3(a) $\mathbb{R}^2 = 0.001$, $\rho = .9$, $n = 100$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	0.69	0.79	3.90	3.96	0.75	3.75	0.01
LIML	0.72	0.79	3.90	3.96	0.75	3.75	0.01
2SLS	0.72	0.79	3.90	3.96	0.75	3.75	0.01
Fuller	0.87	0.87	0.20	0.89	0.72	2.70	0.00
				$K = 5$			
EL	0.80	0.85	3.65	3.74	0.36	1.22	0.01
LIML	0.91	0.85	3.96	4.06	0.36	1.23	0.01
2SLS	0.88	0.88	0.26	0.92	0.13	0.87	0.00
Fuller	0.88	0.88	0.30	0.93	0.19	0.95	0.00
				$K = 20$			
EL	0.77	0.88	3.86	3.94	0.14	0.55	0.01
LIML	0.82	0.88	3.76	3.85	0.14	0.56	0.01
2SLS	0.90	0.90	0.10	0.90	0.00	0.39	0.00
Fuller	0.89	0.88	0.41	0.98	0.10	0.47	0.00

Mean Bias: While oftentimes the bias of EL is smaller than the bias of the *k*-class estimators (see, e.g., Table 3(a)), this advantage is not systematic and there are scenarios, where the bias of *k*-class estimators is smaller than the one of EL. E.g., for $n = 100$, $\mathbb{R}^2 = 0.001$, $K = 1$, $\rho = 0.3$ the bias of EL is 0.36 while the one of Fuller is 0.29. The bias of 2SLS can be substantially larger than the biases of the other estimators in relatively strongly identified scenarios with many instruments, see Tables 2 and 4

TABLE 3(b) $\mathbb{R}^2 = 0.001$, $\rho = 0.9$, $n = 200$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	0.62	0.72	4.10	4.14	0.76	3.79	0.01
LIML	0.64	0.72	4.09	4.14	0.76	3.80	0.01
2SLS	0.64	0.72	4.09	4.14	0.76	3.80	0.01
Fuller	0.84	0.85	0.23	0.87	0.72	2.68	0.00
				$K = 5$			
EL	0.87	0.80	3.85	3.94	0.37	1.25	0.01
LIML	0.80	0.79	4.15	4.22	0.37	1.24	0.02
2SLS	0.87	0.87	0.26	0.91	0.13	0.86	0.00
Fuller	0.86	0.84	0.31	0.91	0.20	0.96	0.00
				$K = 20$			
EL	0.83	0.85	3.95	4.04	0.15	0.58	0.02
LIML	0.78	0.84	3.87	3.95	0.16	0.57	0.01
2SLS	0.89	0.89	0.10	0.90	0.00	0.39	0.00
Fuller	0.87	0.85	0.42	0.96	0.11	0.48	0.00

TABLE 4(a) $\mathbb{R}^2 = 0.1,\ \rho = 0.9,\ n = 100$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	−0.11	0.00	0.67	0.68	0.92	1.17	0.00
LIML	−0.11	0.00	0.67	0.68	0.92	1.17	0.00
2SLS	−0.11	0.00	0.67	0.68	0.92	1.17	0.00
Fuller	0.04	0.08	0.26	0.26	0.89	1.09	0.00
				$K = 5$			
EL	−0.18	0.00	1.06	1.07	0.88	1.00	0.00
LIML	−0.14	0.00	1.01	1.02	0.89	1.00	0.00
2SLS	0.22	0.25	0.23	0.31	0.68	0.79	0.00
Fuller	0.03	0.08	0.27	0.27	0.87	0.94	0.00
				$K = 20$			
EL	−0.18	0.03	2.87	2.87	0.59	0.72	0.01
LIML	−0.17	0.00	1.82	1.82	0.68	0.72	0.00
2SLS	0.57	0.57	0.12	0.58	0.02	0.41	0.00
Fuller	0.04	0.08	0.31	0.31	0.74	0.66	0.00

with $K = 20$; e.g., for $n = 100$, $\mathbb{R}^2 = 0.1$, $K = 20$, $\rho = 0.9$ the bias of 2SLS and Fuller is 0.57 and 0.04, respectively.

As to be expected from the theoretical results in NS, for EL there is no systematic increase in the bias with K increasing. Fuller's estimator seems to be affected even less by the number of instruments and the bias of LIML sometimes even decreases in K, see Table 1(a). On the other hand, the mean bias for 2SLS typically increases with the number of instruments, e.g.,

TABLE 4(b) $\mathbb{R}^2 = 0.1,\ \rho = 0.9,\ n = 200$

	Mean	Med	Std	RMSE	CovPr	Length	Tail prob
				$K = 1$			
EL	−0.05	0.00	0.26	0.26	0.94	0.83	0.00
LIML	−0.05	0.00	0.26	0.26	0.94	0.83	0.00
2SLS	−0.05	0.00	0.26	0.26	0.94	0.83	0.00
Fuller	0.01	0.04	0.21	0.21	0.92	0.80	0.00
				$K = 5$			
EL	−0.06	-0.01	0.30	0.30	0.92	0.77	0.00
LIML	−0.06	-0.01	0.28	0.29	0.93	0.77	0.00
2SLS	0.11	0.14	0.18	0.21	0.81	0.67	0.00
Fuller	0.00	0.03	0.22	0.22	0.91	0.74	0.00
				$K = 20$			
EL	−0.09	0.00	0.52	0.53	0.78	0.61	0.00
LIML	−0.06	0.00	0.30	0.31	0.83	0.61	0.00
2SLS	0.42	0.42	0.11	0.43	0.08	0.40	0.00
Fuller	0.00	0.04	0.23	0.23	0.83	0.59	0.00

in Table 3(a) the bias increases from 0.72 to 0.88 to 0.90 when the number of instruments increases from 1 to 5 to 20.

Typically, but not always, for all estimators, the bias increases as a function of ρ (compare Tables 1 with 3 and 2 with 4) and decreases as a function of $\mathbb{R}^2$ (compare Tables 1 with 2 and 3 with 4) and n. As one exception, compare Tables 3(a) and 3(b) for the case of $K = 5$ and 20. Here, the mean bias of EL increases from 0.80 to 0.87 and from 0.77 to 0.83 for $K = 5$ and 20, respectively, when n is increased from 100 to 200.

Median Bias: Typically, the median bias of all the estimators is very similar with the exception of 2SLS which can be substantially more biased than the other estimators in overidentified scenarios, especially when there is high endogeneity. For example, see Table 4 for $K = 20$, where the bias of 2SLS is 0.57 and 0.42 when $n = 100$ and 200, respectively, while the bias of the other estimators is negligible.

Again, for EL, there does not seem to be a systematic increase in the bias as a function of K—however, for some cases, we note a slight growth in the bias when K increases. For example, see Table 3. The same observation is true for LIML and Fuller but not for 2SLS whose bias increases in K in the strongly identified case, see Tables 2 and 4. As for the mean bias, we typically observe an increase/decrease/decrease in the median bias as a function of $\rho/\mathbb{R}^2/$ and n, respectively.

Std: The Std of EL and LIML are qualitatively very similar across all parameter combinations; sometimes the Std of EL is considerably larger than the one of LIML, see, e.g., Table 4(a) when $K = 20$. Furthermore, the Std of EL and LIML are considerably larger than the corresponding Std of 2SLS (if $K = 5$ or 20) and Fuller (sometimes by several magnitudes!, see for example Tables 1 and 3) and are very similar to 2SLS when $K = 1$. Since LIML has no moments finite, 2SLS has no moments finite when $K = 1$ and >2 moments finite in the overidentified situations considered here and Fuller always has two moments finite, our simulations show that if an estimator has no moments its Std is considerably larger relative to estimators with moments. These findings then also suggest that the GEL estimators suffer from a no-moment problem.

Typically, Std is a decreasing function of ρ and $\mathbb{R}^2$. The dependence on n is more complicated. While in the strongly identified case, Std decreases with increasing n, the opposite is typically true in the weakly identified case.

RMSE: In general, the findings for Std remain valid for the RMSE of the estimators as well, because in most cases, the Std dominates the bias. Therefore, in terms of RMSE, Fuller, and in overidentified models also 2SLS are to be recommended over GEL; in many scenarios, their RMSEs are smaller by several magnitudes.

Coverage Probability and Median Length of CIs: The coverage probability and median length of CIs are virtually identical for EL and LIML. CIs based on 2SLS have higher coverage probability for small ρ (and significantly so when K is large, see Tables 1 and 2 for $K = 20$) but much lower coverage probability when ρ is large and K is big (see Tables 3 and 4). CIs based on Fuller's estimator are arguably the best ones in an overall sense in terms of coverage probability. They have lower coverage probability in weakly identified and strongly endogenous situations than EL- and LIML-based CIs (see Table 3) but typically dominate those CIs otherwise. In terms of median length, Fuller clearly dominates EL and LIML. 2SLS is best in terms of median length in overidentified situations.

Typically, both coverage probability and median length are decreasing functions of K. Typically, but not always, coverage probability increases in $\mathbb{R}^2$ while median length decreases in $\mathbb{R}^2$. Median length decreases in n in strongly identified cases, coverage probability increases in n in strongly identified cases for large K, and otherwise there is not much dependence on $n = 100$ or $n = 200$. Median length typically decreases in ρ. Finally, coverage probability decreases in ρ in weakly identified situations (and in all situations for 2SLS).

10%-Tail-Probabilities: In analogy to the results for Std, we find that estimators without moments have relatively high-tail probabilities. For example, for the parameter combinations in Table 1(a) we find that for $K = 1$, in about 3% of the cases, the EL, LIML, and 2SLS estimator, take on values that are larger than 22.5 in absolute value. The corresponding number for Fuller is 0%. The dependence of the tail probabilities on the parameters ρ, $\mathbb{R}^2$, and n, is as for Std.

3. CONCLUSION

The similarity of GEL and LIML with regards to their Std and outlier probabilities suggest a "no-moment problem" of the GEL estimator. The Std and RMSE of GEL and LIML are very similar and considerably larger than the one of 2SLS (in the overidentified case) and Fuller. In agreement with the theoretical findings in NS, we find that the bias of GEL is not much affected by the number of moment conditions; the same is true for LIML and Fuller. We do not find a systematic bias advantage of GEL estimators over k-class estimators.

If the relevant loss function is MSE, then our results indicate that GEL estimators do not have an advantage over computationally simpler estimators, such as 2SLS, *in the linear model*, on the contrary they perform much worse. Based on the properties of CIs we cannot recommend the use of GEL estimators either, because Fuller-based CIs have overall better properties.

In sum, based on our study, GEL estimators should not be used for the linear model. In nonlinear models however, they may have a bias advantage that might offset (or maybe even more than that) their poorer performance in terms of Std. It would be important to adapt this study and to provide a comprehensive investigation for several nonlinear models in order to assess the comparative dis/advantage of GEL estimators over GMM estimators in these models.

Other criteria may be relevant for the choice of the estimator. For example, it is known that the finite sample distribution of LIML approximates its asymptotic normal distribution much faster than 2SLS, see Anderson et al. (1982, 2005). A similar finding was obtained in the Monte Carlo studies of Kunitomo and Matsushita (2003) for the EL estimator compared to the GMM estimator.

ACKNOWLEDGMENTS

I would like to thank Amos Golan, Essie Maasoumi, and two anonymous referees for very helpful suggestions that led to a much improved version of the article. I would like to thank Donald Andrews, Jinyong Hahn, and Ron Mittelhammer for very helpful discussion and suggestions.

REFERENCES

Anderson, T. W., Kunitomo, N., Sawa, T. (1982). Evaluation of the distribution function of the limited information maximum likelihood estimator. *Econometrica* 50:1009–1027.

Anderson, T. W., Kunitomo, N., Matsushita, Y. (2005). A new light from old wisdoms: Alternative estimation methods of simultaneous equations and microeconometric models. Unpublished Discussion Paper CIRJE-F-321.

Fuller, W. A. (1977). Some properties of a modification of the limited information estimator. *Econometrica* 45(4):939–953.

Guggenberger, P. (2005). Monte-Carlo evidence suggesting a no-moment problem of the continuous updating estimator. *Economics Bulletin* 3(13):1–6.

Guggenberger, P., Hahn, J. (2005). Finite sample properties of the 2-step empirical likelihood estimator. *Econometric Reviews* 24(3):247–263.

Guggenberger, P., Smith, R. J. (2005). Generalized empirical likelihood estimators and tests under partial, weak and strong identification. *Econometric Theory* 21(4):667–709.

Hahn, J., Hausman, J. (2002). A new specification test for the validity of instrumental variables. *Econometrica* 70:163–189.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.

Hansen, L. P. (1982). Large sample properties of generalized method of moments 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.

Imbens, G. (1997). One-step estimators for overidentified generalized method of moments models. *Review of Economic Studies* 64:359–383.

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

Kinal, T. W. (1980). The existence of moments of *k*-class estimators. *Econometrica* 48(1):241–249.

Kitamura, Y., Stutzer, M. (1997). An information-theoretic alternative to generalized method of moments estimation. *Econometrica* 65(4):861–874.

Kunitomo, N., Matsushita, Y. (2003). Finite sample distributions of the empirical likelihood estimator and the GMM estimator. CIRJE Discussion paper F-200.

Maasoumi, E. (1977). A study of improved methods of estimating reduced form coefficients based upon 3SLS estimators. Unpublished Ph.D. thesis, London School of Economics, London, UK.

Mariano, R. S., Sawa, T. (1972). The exact finite-sample distribution of the limited information maximum likelihood estimator in the case of two included endogenous variables. *Journal of the American Statistical Association* 67:159–163.

Mittelhammer, R. C., Judge, G. G., Schoenberg, R. (2005). Empirical evidence concerning the finite sample performance of EL-type structural equation estimation and inference methods. In: Andrews, D. W. K., Stock, J. H., eds. *Honor of Thomas Rothenberg*. Cambridge: Cambridge University Press, pp. 282–305.

Nagar, A. L. (1959). The bias and moments matrix of the general k-class estimators of the parameters in structural equations. *Econometrica* 27(4):575–595.

Newey, W. K., Smith, R. J. (2004). Higher order properties of GMM and Generalized Empirical Likelihood estimators. *Econometrica* 72(1):219–255.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75(2):237–249.

Phillips, P. C. B. (2003). Vision and influence in econometrics: John Denis Sargan. *Econometric Theory* 19(3):495–511.

Qin, J., Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* 22(1):300–325.

Sargan, J. D. (1982). On Monte Carlo estimate of moments that are infinite. In: Basman, R. L., Rhodes, G. F., (Jr), eds. *Advances in Econometrics*, 1. Greenwhich, CT: JAI Press, pp. 267–299.

Sargan, J. D. (1988). *Contributions to Econometrics*. In: Maasoumi, E., ed. Vol. 2. Cambridge: Cambridge University Press.

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

Zaman, A. (1981). Estimators without moments. The case of the reciprocal of a normal mean. *Journal of Econometrics* 15:289–298.