Title

Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods

Authors

Mittelhammer, Ron C
Judge, George G.
Schoenberg, Ron

Publication Date

2003

Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods

Ron C. Mittelhammer[1], George G. Judge[2], and Ron Schoenberg[3]

Washington State University, University of California, Berkeley, and Aptech Systems, Inc.

Abstract

This paper presents empirical evidence concerning the finite sample performance of conventional and generalized empirical likelihood-type estimators that utilize instruments in the context of linear structural models characterized by endogenous explanatory variables. There are suggestions in the literature that traditional and non-traditional asymptotically efficient estimators based on moment equations may, for the relatively small sample sizes usually encountered in econometric practice, have relatively large biases and/or variances and provide an inadequate basis for estimation and inference. Given this uncertainty we use a range of data sampling processes and Monte Carlo sampling procedures to accumulate finite sample empirical evidence concerning these questions for a family of generalized empirical likelihood-type estimators in comparison to conventional 2SLS and GMM estimators. Solutions to EL-type empirical moment-constrained optimization problems present formidable numerical challenges. We identify effective optimization algorithms for meeting these challenges.

Keywords: Unbiased moment based estimation and inference, empirical likelihood, empirical exponential likelihood, semiparametric models, conditional estimating equations, finite sample bias and precision, squared error loss, instrumental conditioning variables.

AMS 1991 Classifications: Primary 62E20
JEL Classifications: C10, C24

[1] 111E Hulbert Hall, Washington State University, Pullman, WA, 99164. Email:**mittelha@ WSU.edu**
[2] 207 Giannini Hall, University of California, Berkeley, CA, 94720. Email: **Judge@ARE.Berkeley.edu**
[3] Aptech Systems, Maple Valley, WA, Email: **Ron@cablespeed.com**

Empirical Evidence Concerning the Finite Sample Performance of EL-Type Structural Equation Estimation and Inference Methods

1. Introduction

It is known in the literature that a number of moment-based estimators for the linear structural model are asymptotically normally distributed and mutually asymptotically equivalent. There is also a growing body of evidence (see for example Newey and Smith (2000) and the references therein) that traditional asymptotically efficient moment-based estimators may exhibit large biases and/or variances when applied to the relatively small samples usually encountered in applied economic research.

Econometric models that specify a set of moment-orthogonality conditions relating to the underlying data sampling process, and involving parameters, data outcomes, and model noise, lead to a corresponding set of unbiased empirical estimating functions. These estimating functions often involve instrumental variables (IV) whose number exceeds the number of unknown parameters of interest, and overdetermines the model parameters. In some instances the instrumental variables may be only moderately or weakly correlated with the endogenous variables in the model. In this situation it is generally recognized that significant bias and/or variability problems may arise and that large sample normal approximations may provide a poor basis for evaluating finite sample performance (see for example Nelson and Startz (1990), Maddala and Jeong (1992), Bound, Jaeger and Baker (1995), and Stock and Wright 2000).

In an effort to avoid an explicit likelihood function specification, semi-parametric empirical likelihood (EL) type estimators have been proposed as moment based estimation and inference alternatives to classical maximum likelihood methods (Owen, 1988, 1991; Qin and Lawless, 1994; Imbens, et al. 1998; Corcoran, 2000 and Mittelhammer, Judge and Miller, 2000). Given this new class of estimators, and in line with the ongoing search for efficient linear structural equation estimators having small finite sample bias, and associated inference procedures with accurate size, good power, and short confidence intervals with proper coverage, we provide some empirical evidence relating to the finite sample performance of a trio of empirical likelihood-type estimators when estimating functions overdetermine the model parameters and parameters are moderately well-identified. The results are based on Monte Carlo sampling experiments applied to a range of underlying data sampling processes and to estimators that include the optimal estimating function (OptEF) and two stage least squares (2SLS) estimator, the generalized method of moments (GMM) estimator based on an identity weight matrix, as well as the empirical likelihood (EL), exponential empirical likelihood (EEL) and log Euclidean likelihood (LEL) estimators. As noted by Imbens, et. al. (1998), the computation of solutions to EL type moment-constrained optimization problems can present formidable numerical challenges. From both a theoretical and practical standpoint, reliable and efficient solution algorithms are critically needed. Toward this end, we suggest an algorithm that performs well.

In the context of finite sample situations where the instrumental variables (IV) are moderately well-correlated with the endogenous variables in question and the orthogonality condition between the IV and the structural equation noise holds, we seek information relative to the following questions:

i) Do empirical likelihood (EL) type estimators offer reductions in either small sample bias or variance relative to traditional OptEF-2SLS and non-optimal GMM estimators?

ii) In terms of a mean square measure of estimator performance, are any of the EL-type estimators superior to the traditional semiparametric estimators?

iii) In terms of inference in small samples, do EL-type testing procedures have, relative to traditional testing procedures, more accurate coverage, shorter confidence intervals, and/or test sizes that are closer to nominal target size?

iv) What is the relative small sample performance of the traditional and EL-type inference procedures relative to testing the moment restrictions?

v) What is the basis for a reliable and efficient solution algorithm for EL-type moment-constrained estimation problems?

The format of the paper is as follows: In Section 2 the linear structural model is defined and the competing semiparametric estimators and inference procedures are specified. In Section 3 the design of the sampling experiment is presented and the alternative data sampling processes are defined. Monte Carlo estimation results are presented and discussed in section 4. Conclusions, implications, and speculations are presented in section 5.

2. Statistical Models, Estimators, and Inference Procedures

Consider a single structural equation that is contained within a system of structural equations and that has the semiparametric linear statistical model form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. We observe a vector of sample outcomes $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ associated with this linear model, where $\mathbf{X}$ is a $(n \times k)$ matrix of stochastic explanatory variables, $\boldsymbol{\varepsilon}$ is an unobservable random noise vector with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 \mathbf{I}_n$, and $\boldsymbol{\beta} \in \mathbf{B}$ is a $(k \times 1)$ vector of unknown parameters. If one or more of the regressors is correlated with the equation noise, then $\mathrm{E}\left[\mathrm{n}^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] \neq \mathbf{0}$ or $\operatorname{plim}\left[\mathrm{n}^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] \neq \mathbf{0}$ and traditional Gauss-Markov based procedures such as the least squares (LS) estimator, or equivalently the method of moments (MOM)-extremum estimator defined by $\hat{\boldsymbol{\beta}}_{\mathrm{mom}} = \arg_{\boldsymbol{\beta}\in\mathrm{B}}\left[\mathrm{n}^{-1}\mathbf{X}'\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)=\mathbf{0}\right]$, are biased and inconsistent, with unconditional expectation and probability limit given by $\mathrm{E}\left[\hat{\boldsymbol{\beta}}\right] \neq \boldsymbol{\beta}$ and $\operatorname{plim}\left[\hat{\boldsymbol{\beta}}\right] \neq \boldsymbol{\beta}$.

2.1 Traditional Instrument-Based Estimators

Given a sampling process characterized by nonorthogonality of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, in order to avoid the use of strong distributional assumptions it is conventional to introduce additional information in the form of a $(n \times m)$, $m \geq k$, random matrix $\mathbf{Z}$ of instrumental variables whose elements are correlated with $\mathbf{X}$ but uncorrelated with $\boldsymbol{\varepsilon}$. This information is introduced into the statistical model by specifying the sample analog moment condition

$$\mathbf{h}\left(\mathbf{Y},\mathbf{X},\mathbf{Z};\boldsymbol{\beta}\right) = \mathrm{n}^{-1}\left[\mathbf{Z}'\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right] \xrightarrow{\mathrm{p}} \mathbf{0}, \tag{2.1}$$

relating to the underlying population moment condition derived from the orthogonality of instruments and model noise defined by

$$\mathrm{E}\left[\mathbf{Z}'\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right] = \mathbf{0} \quad . \tag{2.2}$$

If $m = k$, the vector of moment conditions just-determine the model parameters, and the sample moments (2.1) can be solved for the basic instrumental variable (IV) estimator $\hat{\boldsymbol{\beta}}_{\mathrm{iv}} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{Y}$. When the usual regularity conditions are fulfilled, this IV estimator is consistent, asymptotically normal distributed, and is an optimal estimating function

(OptEF) estimator (Godambe 1960; Heyde 1989; Mittelhammer, Judge, and Miller 2000).

For $m > k$, the vector of moment conditions overdetermine the model parameters and other IV-like estimation procedures are available, such as the well known two stage least squares (2SLS) estimator, $\boldsymbol{\beta}_{2sls} = \left(\mathbf{X}'\mathbf{P_z X}\right)^{-1}\mathbf{X}'\mathbf{P_z Y}$, where $\mathbf{P_z} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'$ is the projection matrix for $\mathbf{Z}$. This estimator is equivalent to the estimator formed by applying the optimal estimating function (OptEF) transformation $n\left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}$ to the moment conditions in (2.2) (Godambe, 1960; Judge, et. al., 1985; Heyde and Morton, 1998).

The GMM estimator (Hansen, 1982) is another estimator that makes use of the information in (2.2). The GMM estimators minimize a quadratic form in the sample moment information

$$\begin{aligned}\hat{\boldsymbol{\beta}}(\mathbf{W}) &= \arg\min_{\boldsymbol{\beta}\in \mathrm{B}}\left[\mathrm{Q_n}(\boldsymbol{\beta})\right] = \arg\min_{\boldsymbol{\beta}\in \mathrm{B}}\left[\left(\mathrm{n}^{-1}\mathbf{Z}'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right)'\mathbf{W}\left(\mathrm{n}^{-1}\mathbf{Z}'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right)\right] \\ &= \arg\min_{\boldsymbol{\beta}\in \mathrm{B}}\left[\mathrm{n}^{-2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{ZWZ}'(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right]\end{aligned}. \tag{2.3}$$

The GMM estimator can be shown to have optimal asymptotic properties if the weighting matrix **W** is appropriately defined. The optimal choice of **W** in the context of moment conditions (2.2) leads back to the definition of the 2SLS-OptEF estimator.

2.2 Empirical Likelihood (EL) Type Estimators

In contrast to traditional instrument-moment based estimators, the empirical likelihood approach (Owen, 1988, 1991, 2001; Qin and Lawless, 1994, Imbens, et. al. (1998), Corcoran, 2000, and Mittelhammer, Judge and Miller, 2000) allows the investigator to employ likelihood methods for model estimation and inference without having to choose a specific parametric family of probability densities on which to base the likelihood function. Under the EL concept, empirical likelihood weights supported on a sample of observed data outcomes are used to reduce the infinite dimensional problem of nonparametric likelihood estimation to a finite dimensional one.

2.2.1 Estimation

The constrained estimation problem underlying the EL approach is in many ways analogous to allocating probabilities in a contingency table where w_j and q_j are observed and expected probabilities. A solution is achieved by minimizing the divergence between the two sets of probabilities by optimizing a goodness-of-fit criterion subject to the moment constraints. One possible set of divergence measures is the power divergence family of statistics (Cressie and Read,1984; Read and Cressie,1988)

$$I(\mathbf{w},\mathbf{q},\lambda) = \frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{n} w_i\left[\left(\frac{w_i}{q_i}\right)^{\lambda} - 1\right], \tag{2.4}$$

where λ is an arbitrary unspecified parameter. In the limit as λ ranges from -1 to 1, several estimation and inference procedures emerge.

If in an instrumental variable context for the linear structural equation we use (2.4) as the goodness-of-fit criterion and (2.1) as the moment-estimating function information, the EL estimation problem can be formulated as the following extremum- type estimator:

$$\hat{\boldsymbol{\beta}} = \arg\max_{\boldsymbol{\beta}}\left[\ell_E(\boldsymbol{\beta}) = \max_{\mathbf{w}}\left\{-I(\mathbf{w},\mathbf{q},\lambda)\,|\,\sum_{i=1}^{n} w_i\mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}) = \mathbf{0}, \sum_{i=1}^{n} w_i = 1, w_i \ge 0\,\forall i, \boldsymbol{\beta}\in\mathbf{B}\right\}\right] \tag{2.5}$$

Three main variants of $I(\mathbf{w},\mathbf{q},\lambda)$ have received explicit attention in the literature. Letting $\lambda \to 0$ leads to the traditional empirical log-likelihood objective function,

$n^{-1}\sum_{i=1}^{n}\ln\left(\mathrm{w_i}\right)$, and the maximum empirical likelihood (MEL) estimate of $\boldsymbol{\beta}$. When $\lambda \to -1$, the empirical exponential likelihood objective function, $-\sum_{i=1}^{n}\mathrm{w_i}\ln\left(\mathrm{w_i}\right)$, is defined and the Maximum Empirical Exponential Likelihood (MEEL) estimate of $\boldsymbol{\beta}$ results. Finally, when $\lambda = 1$, the log Euclidean likelihood function $-n^{-1}\left(\sum_{i=1}^{n}\left(n^2 w_i^2 - 1\right)\right)$ is implied and leads to the Maximum Log Euclidean Likelihood (MLEL) estimate of $\boldsymbol{\beta}$.

In the sense of objective function analogies, the Owen MEL approach is the closest to the classical maximum likelihood approach. The MEEL criterion of maximizing $-\sum_{i=1}^{n}\mathrm{w_i}\ln\left(\mathrm{w_i}\right)$ is equivalent to defining an estimator by *minimizing* the Kullback-Leibler (KL) information criterion $\sum_{i=1}^{n}\mathrm{w_i}\ln\left(\mathrm{w_i}/n^{-1}\right)$ (Kullback, 1959; Golan, Judge, and Miller, 1996). Interpreted in the KL context, the MEEL estimation objective finds the feasible weights $\hat{\mathbf{w}}$ that define the minimum value of all possible expected log-likelihood ratios consistent with the structural moment constraints. The MLEL solution seeks feasible weights $\hat{\mathbf{w}}$ that minimize the Euclidean distance of $\mathbf{w}$ from the uniform probability distribution, the square of this Euclidean distance being $\left(\mathbf{w}-\mathbf{1}_{\mathrm{n}}n^{-1}\right)'\left(\mathbf{w}-\mathbf{1}_{\mathrm{n}}n^{-1}\right)$, where $\mathbf{1}_n$ denotes an $n \times 1$ vector of unit values. All of the preceding estimation objective functions achieve *unconstrained* (by moment constraints) optima when the empirical probability distribution is given by $\mathbf{w} = \mathbf{1}_n n^{-1}$.

If the optimization problem is cast in Lagrangian form, where $\boldsymbol{\alpha}$ and η are Lagrange multipliers for the moment and adding up conditions respectively, then the constrained optimal w_i's for the MEL estimator can be expressed as

$$w_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = \left[n\left(\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}) + 1\right)\right]^{-1}, \qquad (2.6)$$

and the constrained optimal w_i's for the MEEL estimator can be expressed as

$$w_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = \frac{\exp\left(\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta})\right)}{\sum_{j=1}^{n}\exp\left(\boldsymbol{\alpha}'\mathbf{z}'_{j.}(y_j - \mathbf{x}_{j.}\boldsymbol{\beta})\right)}. \qquad (2.7)$$

In the case of the MLEL estimator, the constrained optimal w_i's can be expressed as $w_i(\boldsymbol{\beta},\boldsymbol{\alpha},\eta) = (2n)^{-1}\left(\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}) + \eta\right)$. The Lagrange multiplier η can be eliminated by solving the adding up condition $\mathbf{1}'_{\mathrm{n}}\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha},\eta) = 1$ for η, yielding the expression $\eta(\boldsymbol{\alpha},\boldsymbol{\beta}) = \left(2 - n^{-1}\sum_{i=1}^{n}\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta})\right)$, and then substitution into $w_i(\boldsymbol{\beta},\boldsymbol{\alpha},\eta)$ yields

$$w_i(\boldsymbol{\beta},\boldsymbol{\alpha}) = (2n)^{-1}\left(\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}) + 2 - n^{-1}\sum_{i=1}^{n}\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta})\right). \qquad (2.8)$$

Under the usual regularity conditions assumed when establishing the asymptotics of traditional structural equation estimators, all of the preceding EL-type estimators of $\boldsymbol{\beta}$ obtained by optimizing the w_i's in (2.6), (2.7), or (2.8) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and/or η are, *given the set of estimating equations under consideration,* consistent, asymptotically normally distributed, and asymptotically efficient relative to the optimal estimating function (OptEF) estimator. Calculating the solution to the MEL, MEEL, or MLEL estimation problem will generally require that a computer-driven optimization algorithm be employed. When $m = k$, the solutions to all of the EL-type extremum problems lead back to the standard IV estimator $\hat{\boldsymbol{\beta}}_{iv}$ with $w_i = n^{-1}$. When $m \geq k$, the estimating equations overdetermine the unknown parameter values to be recovered and a nontrivial EL solution results. The solution to the constrained optimization problem (2.5) based on any of the members of the Cressie-Read family of estimation objective functions yields an optimal estimate, $\hat{\mathbf{w}}$ and $\hat{\boldsymbol{\beta}}$, that cannot, in general, be expressed in closed form and thus must be obtained using numerical methods.

2.2.2 Inference

EL-type inference methods, including hypothesis testing and confidence region estimation, bear a strong analogy to inference methods used in traditional ML and GMM approaches. Owen (1988, 1991) showed that an analog of Wilks' Theorem for likelihood ratios, -2ln(LR) $\overset{a}{\sim} \chi_j^2$, hold for the empirical likelihood (MEL) approach, where j denotes the number of functionally independent restrictions on the parameter space. Baggerly (1998) demonstrated that this calibration remains applicable when the likelihood is replaced with any properly scaled member of the Cressie-Read family of power divergence statistics (2.4). In this context, the empirical likelihood ratio (LR) for testing the linear hypothesis $\mathbf{c}\boldsymbol{\beta} = \mathbf{r}$ when $rank(\mathbf{c}) = j$, is given for the MEL case by

$$LR_{EL}(\mathbf{y}) = \frac{\max_{\boldsymbol{\beta}}\left[\ell_E(\boldsymbol{\beta})\ s.t.\ \mathbf{c}\boldsymbol{\beta} = \mathbf{r}\right]}{\max_{\boldsymbol{\beta}} \ell_E(\boldsymbol{\beta})} \tag{2.9}$$

where $-2\ln\left(LR_{EL}(\mathbf{Y})\right) \overset{a}{\sim} Chisquare(j,0)$ under H_o when $m \geq k$. An analogous pseudo-LR approach can be applied, mutatis mutandis, to other members of the Cressie-Read family. One can also base tests of $\mathbf{c}\boldsymbol{\beta} = \mathbf{r}$ on the Wald Criterion in the usual way by utilizing the inverse of the asymptotic covariance matrix of $\mathbf{c}\hat{\boldsymbol{\beta}}_{\mathrm{EL}}$ as the weight matrix of a quadratic form in the vector $\mathbf{c}\hat{\boldsymbol{\beta}}_{\mathrm{EL}} - \mathbf{r}$, or construct tests based on the Lagrange multipliers associated with the constraints $\mathbf{c}\boldsymbol{\beta} = \mathbf{r}$ imposed on the EL-type optimization problem. Confidence region estimates can be obtained from hypothesis test outcomes in the usual way based on duality. The validity of the moment conditions (2.1)-(2.2) can be assessed via a variation of the preceding testing methodology. We provide further details ahead regarding the empirical implementation of inference methods.

2.3. Test Statistics

Two different types of inference contexts are examined in this paper, including testing the validity of the moment constraints, and testing hypotheses and generating confidence intervals for parameters of the structural model.

2.3.1 Moment Validity Tests

Regarding the validity of the moment restrictions, Wald-type quadratic form tests, often referred to as Average Moment Tests, are calculated for all five estimators. The Wald test statistics are specified as

$$\text{Wald} = \left(\mathbf{1}_n'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)\right)'\left[\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)\right]^{-1}\left(\mathbf{1}_n'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)\right) \quad (2.10)$$

where $\hat{\boldsymbol{\beta}}$ is any one of the five different estimators of the $\boldsymbol{\beta}$ vector, and $\odot$ denotes the *generalized* Hadamard (elementwise) product operator. Under the null hypothesis of moment validity, the Wald statistic has an asymptotic Chisquare distribution with degrees of freedom equal to the degree of overidentification of the parameter vector, i.e., m-k.

Pseudo Likelihood Ratio (LR) -type tests of moment validity, referred to as Criterion Function Tests by Imbens, Spady, and Johnson (1998, p.342), are also calculated for the three EL-type procedures. The respective test statistics for the MEEL and MEL procedures are $LR_{EEL} = 2n\left(\mathbf{w}'\ln(\mathbf{w})+\ln(n)\right)$ and $LR_{EL} = -2\left(\mathbf{1}_n'\ln(\mathbf{w})+n\ln(n)\right)$.

In the case of MLEL, the pseudo-likelihood ratio statistic is derived as a special case of the generalized empirical likelihood (GEL) class of procedures identified by Newey and Smith (2000, p. 8) given by

$$LR_{LEL} = n\left(1 - n^{-1}\mathbf{1}_{\mathrm{n}}'\left[\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\left(\frac{\boldsymbol{\alpha}}{\eta}\right)\right]^2\right) = n\left(1-\left(\frac{2}{\eta}\right)^2 n\sum_{i=1}^{n} w_i^2\right) \tag{2.11}$$

Since $LR_{LEL} \equiv Wald_{LEL}$, we later report on the performance of only one version of this particular test. The $\mathbf{w}$ weights, $\boldsymbol{\beta}$ vector, and Lagrange multipliers $\boldsymbol{\alpha}$ and η appearing in the LR test statistics are replaced by the respective EL-type estimates. All of the pseudo LR -type test statistics follow the same asymptotic Chisquare distribution as for the Wald statistics of moment validity.

The final set of moment validity tests are based on the Lagrange multipliers of the moment constraints. In the case of the EEL-type test statistic, we examine the following quadratic form in the Lagrange multiplier vector that incorporates a robust estimator of the covariance matrix of the moment constraints,

$$LM_{EEL} = n\boldsymbol{\alpha}'\left[\left(\mathbf{h}(\boldsymbol{\beta})\odot\mathbf{w}\right)'\mathbf{h}(\boldsymbol{\beta})\right]\left[\left(\mathbf{h}(\boldsymbol{\beta})\odot\mathbf{w}\right)'\left(\mathbf{h}(\boldsymbol{\beta})\odot\mathbf{w}\right)\right]^{-1}\left[\left(\mathbf{h}(\boldsymbol{\beta})\odot\mathbf{w}\right)'\mathbf{h}(\boldsymbol{\beta})\right]\boldsymbol{\alpha} \tag{2.12}$$

where $\mathbf{h}(\boldsymbol{\beta}) \equiv \left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right)$ and $\mathbf{w}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are estimated on the basis of the MEEL method. In the case of the MEL and MLEL methods, we instead utilize LM tests that are based on equivalences with GEL tests implied by the asymptotic results of Newey and Smith(2000, p. 8). Both of these LM tests are based on the statistic

$$LM = n\boldsymbol{\alpha}'\left(\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}^{-1}\mathbf{G}'\mathbf{V}\mathbf{G}\boldsymbol{\Omega}^{-1}\right)^{-}\boldsymbol{\alpha} \tag{2.13}$$

where $\mathbf{\Omega} \equiv n^{-1}\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right)'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right)$, $\mathbf{G} \equiv n^{-1}\mathbf{X}'\mathbf{Z}$, $\mathbf{V} \equiv \left(\mathbf{G}\mathbf{\Omega}^{-1}\mathbf{G}'\right)^{-1}$ and the values of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ are replaced by either MEEL or MLEL estimates. Under the null hypothesis, all of the LM tests are asymptotically Chisquare distributed with degrees of freedom equal to m-k.

2.3.2 Tests of Parameter Restrictions

A test of the significance of the parameters of the structural model is conducted based on the usual asymptotic normally distributed Z-statistic and concomitantly, by duality, the accuracy of confidence region coverage of the parameters is examined. The test statistic for all of the estimation procedures examined has the familiar form

$$Z = \frac{\hat{\beta}_i}{\widehat{std}\left(\hat{\beta}_i\right)} \overset{a}{\sim} N(0,1) \textit{ under } H_0 : \beta_i = 0, \tag{2.14}$$

and the associated confidence interval estimate is $\left(\hat{\beta}_i - z_\tau \widehat{std}\left(\hat{\beta}_i\right), \hat{\beta}_i + z_\tau \widehat{std}\left(\hat{\beta}_i\right)\right)$ where z_τ denotes the 100τ % quantile of the standard normal distribution. In (2.14) $\hat{\beta}_i$ *and* $\widehat{std}\left(\hat{\beta}_i\right)$ are the appropriate estimates of the parameter and the estimated standard error of the estimate based on one of the five alternative estimation procedures. The respective estimates of the standard errors used in the test and confidence interval procedures were obtained as the square roots of the appropriate diagonal elements of the asymptotic covariance matrices of the B2SLS-OptEF, GMM(I), and the EL-type estimators defined respectively as

$$AsyCov(\hat{\mathbf{B}}_{2sls}) = \hat{\sigma}^2\left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}, \tag{2.15}$$

$$AsyCov(\hat{\mathbf{B}}_{GMM(I)}) = \hat{\sigma}^2\left(\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)\mathbf{Z}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X}\right)^{-1}, \tag{2.16}$$

and

$$AsyCov(\hat{\mathbf{B}}_{EL-type}) = \left[\left(\mathbf{X}'(\mathbf{Z} \odot \hat{\mathbf{w}})\right) \left[\left(\left(\mathbf{Z} \odot \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)\right) \odot \hat{\mathbf{w}} \right)' \left(\mathbf{Z} \odot \left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)\right) \right]^{-1} \left((\mathbf{Z} \odot \hat{\mathbf{w}})' \mathbf{X} \right) \right]^{-1} \tag{2.17}$$

where $\hat{\sigma}^2$ is the usual consistent estimate of the equation noise variance, and $\hat{\mathbf{w}}$ and $\hat{\mathbf{B}}$ are the appropriate estimates obtained from applications of the MEEL, MEL, or MLEL estimation procedure.

2.4 Computational Issues and Approach

As noted by Imbens, Spady, and Johnson (1998), the computation of solutions to EL-type constrained optimization problems can present formidable numerical challenges especially because, in the neighborhood of the solution to such problems, the gradient matrix associated with the moment constraints will approach an ill-conditioned state. This occurs by design in these types of problems because the fundamental method by which EL-type methods resolve the overdetermined nature of the empirical moment conditions, $\sum_{i=1}^{n} w_i \mathbf{z}'_{i.} \left(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}\right) = \mathbf{0}$, is to choose sample weights that ultimately transform the m moment equations into a functionally dependent, lower rank (k< m) system of equations capable of being solved uniquely for the parameters. This creates instability in gradient-based constrained optimization algorithms regarding the representation of the feasible spaces and feasible directions for such problems. Moreover, attempting to solve the optimization problems in primal form is complicated by the dimensionality of the problem, where there are as many w_i sample weights as there are sample observations, and requires that explicit constrained optimization methods be used to enforce the moment conditions and the convexity properties of the sample weights.

Given these complications, Imbens, Spady and Johnson found it advantageous in their EEL and EL simulations to utilize a dual penalty function method for enforcing the moment constraints, whereby a penalty-augmented objective function is optimized within the context of an unconstrained optimization problem. While their penalty-function approach appeared to perform well for the range of applications that were examined in their work, the algorithm failed (non-convergence) too frequently when applied to the IV-based moment constrained problems examined in this paper.

The computational approach utilized in this work for solving the EL-type problems consisted of concentrating out the Lagrange multiplier vector and scalar, $\boldsymbol{\alpha}$ and η, from the EL-type optimization problems, expressing $\boldsymbol{\alpha}$ and η as a function of the $\boldsymbol{\beta}$ vector (in the case of MEEL and MEL, the optimal η is simply the scalar 1). The actual process of concentrating out the Lagrange multipliers cannot be accomplished in closed form, requiring a numerical nonlinear equation solving procedure, but solving the system of equations proved to be quite stable and efficient . Then the resulting concentrated Lagrange representations of the EL-type estimation problems were optimized with respect to the choice of $\boldsymbol{\beta}$, leading to the parameter estimates.

More specifically, in the first step of the computational procedure the Lagrange multiplier vector $\boldsymbol{\alpha}$ was expressed as a function of $\boldsymbol{\beta}$ by utilizing the empirical moment conditions and the weight representation (2.6)-(2.8) for the vector $\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha})$ as

$$\boldsymbol{\alpha}(\boldsymbol{\beta}) \equiv \arg_{\boldsymbol{\alpha}}\left[\left(\mathbf{Z}\odot(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right)'\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha}) = \mathbf{0}\right]. \tag{2.18}$$

The solution to (2.18) was determined numerically using the NLSYS nonlinear equation solver in the GAUSS mathematical programming language (Aptech Systems, Maple Valley, Washington, Version 3.6). Regarding the Lagrange multiplier η, the first order conditions for either the MEL or MEEL estimation problems imply that $\eta(\boldsymbol{\beta}) \equiv 1$. In the case of the MLEL problem, $\eta(\boldsymbol{\beta})$ can be defined by substituting the value of $\boldsymbol{\alpha}(\boldsymbol{\beta})$ obtained from (2.18) into the definition of $\eta(\boldsymbol{\alpha},\boldsymbol{\beta})$ that precedes (2.8), yielding

$$\eta(\boldsymbol{\beta}) \equiv \left(2 - n^{-1}\sum_{i=1}^{n} \boldsymbol{\alpha}(\boldsymbol{\beta})' \mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta})\right). \tag{2.19}$$

In the second step relating to optimization, the concentrated Lagrange function can be represented as

$$\begin{aligned} L_*(\boldsymbol{\beta}) &\equiv L\left(\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta})),\boldsymbol{\beta},\, \boldsymbol{\alpha}(\boldsymbol{\beta}),\eta(\boldsymbol{\beta})\right) \\ &\equiv \phi\left(\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta}))\right) - \boldsymbol{\alpha}(\boldsymbol{\beta})' \sum_{i=1}^{n} w_i(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta}))\, \mathbf{z}'_{i.}(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}) - \eta(\boldsymbol{\beta})\left(\sum_{i=1}^{n} w_i(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta})) - 1\right) \end{aligned} \tag{2.20}$$

The value of $L_*(\boldsymbol{\beta})$ is then optimized (maximized for MEL, minimized for MEEL and MLEL) with respect to the choice of $\boldsymbol{\beta}$, where $\phi(\bullet)$ can also denote any of the estimation objective functions in the Cressie-Read family. The algorithm used to accomplish the optimization step was based on a Nelder-Meade polytope-type direct search procedure written by the authors and implemented in the GAUSS programming language (Nelder and Mead,1965; Jacoby, Kowalik, and Pizzo,1972; and Bertsekas,1995) using the values .5, .5, and 1.1, respectively, for the reflection, contraction, and expansion coefficients. The Nelder-Meade approach is especially well-suited for this problem because it requires only that the function itself be evaluated at trial values of the $\boldsymbol{\beta}$ vector, and does not require calculation of the numerical derivatives of the first or second order used by gradient-based search algorithms, which were inaccurate and unstable in the current context.

3. Design of Sampling Experiments

The finite sample properties of the EL-type estimators and associated inference procedures delineated in section 2 cannot be derived from a direct evaluation of closed functional forms applied to distributions of random variables. Moreover, the finite sample probability distributions of the traditional 2SLS and GMM estimators are also generally intractable. Consequently, we use Monte Carlo sampling experiments to examine and compare the finite sample performance of competing estimators and inference methods. While these results are specific to the collection of particular Monte Carlo experiments analyzed, the wide ranging sampling evidence reported does provide an indication of the types of relative performance that can occur over a range of scenarios for which the unknown parameters of a model are moderately well-identified.

3.1 Experimental Sampling Design

Consider a data sampling process of the following form:

$$Y_{i1} = Z_{i1}\beta_1 + Y_{i2}\beta_2 + e_i = \mathbf{X}_{i.}\boldsymbol{\beta} + \varepsilon_i \tag{3.1}$$

$$Y_{i2} = \sum\nolimits_{j=1}^{5} \pi_j Z_{ij} + v_i = \mathbf{Z}_{i.}\boldsymbol{\pi} + v_i \tag{3.2}$$

where $\mathbf{X}_{i.} = \left(Z_{i1}, Y_{i2}\right)$, and $i = 1,2,...,n$. In the sampling experiment, the two-dimensional vector of unknown parameters, $\boldsymbol{\beta}$, in (3.1) is arbitrarily set equal to the vector $\left[-1,2\right]'$. The outcomes of the (6×1) random vector $\left[Y_{i2}, \varepsilon_i, Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}\right]$ are generated *iid* from a multivariate normal distribution with a zero mean vector and standard deviations uniformly set to 5 for the first two random variables, and 2 for the remaining random variables and $Z_{i5} \equiv 1, \forall i$. Also various other conditions relating to the correlations among the six scalar random variables were assumed. The values of the π_j's in (3.2) are determined by the regression function between Y_{i2} and $[Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5}]$, which is itself a function of the covariance specification relating to the marginal normal distribution associated with the (5×1) random vector $[Y_{i2,} Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}]$. Thus the π_j's generally change as the scenario postulated for the correlation matrix of the sampling process changes. In this sampling design, the outcomes of $[Y_{i1}, V_i]$ are then calculated by applying the equations (3.1-3.2) to the outcomes of $[Y_{i2,} Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}, Z_{i5}]$.

3.2 Sample Characteristics and Outcome Basis

Regarding the details of the sampling scenarios simulated for these Monte Carlo experiments, sample sizes of n = 50, 100 and 250 were examined. The outcomes of ε_i were generated independently of the vector $[Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}]$ so that the correlations between ε_i and the Z_{ij}'s were zero, thus fulfilling a fundamental condition for $[Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}]$ to be considered a set of valid instrumental variables for estimating the unknown parameters in (3.1). Regarding the degree of nonorthogonality and identifiability in (3.1), correlations of .25, .50, and .75 between the random variables Y_{i2} and ε_i were utilized to simulate moderately, to relatively strongly correlated-nonorthogonality relationships between the explanatory variable Y_{i2} and the equation noise ε_i.

For each sample size, alternative scenarios were examined relating to both the degree of correlation existing between each of the random instruments in the matrix **Z** and the $\mathbf{Y}_2$ variable, and the levels of collinearity existing among the instrumental variables themselves. By varying the degrees of intercorrelation among the variables, the overall correlation of the instrumental variables with $\mathbf{Y}_2$ is effected, and contributes to determining the overall effectiveness of the set of instruments in predicting values of the endogenous $\mathbf{Y}_2$. The joint correlation between $\mathbf{Y}_2$ and the set of instruments range from a relatively low .25 to a relatively strong .68.

The major characteristics of each sampling scenario are delineated in Table 3.1.

Table 3.1 Monte Carlo Experiment Definitions, with $\boldsymbol{\beta} = [-1, 2]'$, $\sigma_{\varepsilon_i} = \sigma_{Y_{2i}} = 5$, **and** $\sigma_{Z_{ij}} = 2$, $\forall$ **i and** $\mathbf{j = 1, ..., 5}$.

Experiment Number	$\rho_{y_{2i},\varepsilon_i}$	$\rho_{y_{2i},z_{i,1}}$	$\rho_{y_{2i},z_{ij}\,;\,j>1}$	$\rho_{z_{ij},z_{ik}}$	$R^2_{Y_1,\hat{Y}_1}$	$R^2_{Y_2,\hat{Y}_2}$
1	.25	.25	.25	0	.84	.25
2	.25	-.25	.25	.5	.86	.40
3	.50	.25	.25	0	.89	.25
4	.50	-.25	.25	.5	.90	.40
5	.75	.25	.25	0	.95	.25
6	.75	-.25	.25	.5	.94	.40
7	.50	.1	.5	.25	.89	.53
8	.50	.1	.5	.5	.89	.50
9	.50	.1	.5	.75	.89	.68
10	.50	.5	.1	.75	.89	.53

Note: $\rho_{y_{2i},\varepsilon_i}$ denotes the correlation between Y_{2i} and e_i, and measures the degree of nonorthogonality; $\rho_{y_{2i},z_{ij}}$ denotes the common correlation between Y_{2i} and each of the four random instrumental variables, the Z_{ij}'s; $\rho_{z_{ij},z_{ik}}$ denotes the common correlation between the four random instrumental variables; $R^2_{Y_1,\hat{Y}_1}$ denotes the population squared correlation between $\mathbf{Y}_1$ and $\hat{\mathbf{Y}}_1 = \mathbf{X}\boldsymbol{\beta}$; and $R^2_{Y_2,\hat{Y}_2}$ denotes the population squared correlation between $\mathbf{Y}_2$ and $\hat{\mathbf{Y}}_2 = \mathbf{Z}\boldsymbol{\pi}$.

In general, the scenarios range from relatively weak but independent instruments to stronger but more highly multicollinear instruments. All models have a relatively strong signal component in the sense that the squared correlation between the dependent variable $\mathbf{Y}_1$ and the explanatory variables $\left(\mathbf{Z}_{.1}, \mathbf{Y}_2\right)$ ranges between .84 and .95. In total there are 10 different MC experimental designs in combination with the three different sample sizes, resulting in 30 different sampling scenarios in which to observe estimator and inference behavior.

The sampling results, reported in section 4, are based on 5000 Monte Carlo repetitions, which was sufficient to produce stable estimates of the empirical mean squared error (MSE), expressed in terms of the mean of the empirical squared Euclidean distance between the true parameter vector $\boldsymbol{\beta}$ and $\hat{\boldsymbol{\beta}}$ (measuring parameter estimation risk), the MSE between $\mathbf{y}$ with $\hat{\mathbf{y}}$ (measuring predictive risk), the average estimated bias in the estimates, $\text{Bias}\left(\hat{\boldsymbol{\beta}}\right) = \text{E}[\hat{\boldsymbol{\beta}}]\text{-}\boldsymbol{\beta}$, and the average estimated variances, $\text{Var}\left(\hat{\beta}_i\right)$.

Regarding inference performance, we: *i)* compare the empirical size of ten alternative tests of moment equation validity with the typical nominal target size of .05, *ii)* examine the empirical coverage probability of confidence interval estimators based on a target coverage probability of .99, *iii)* compare the empirical expected lengths of confidence intervals, and *iv)* examine power of significance tests associated with the different estimation methods.

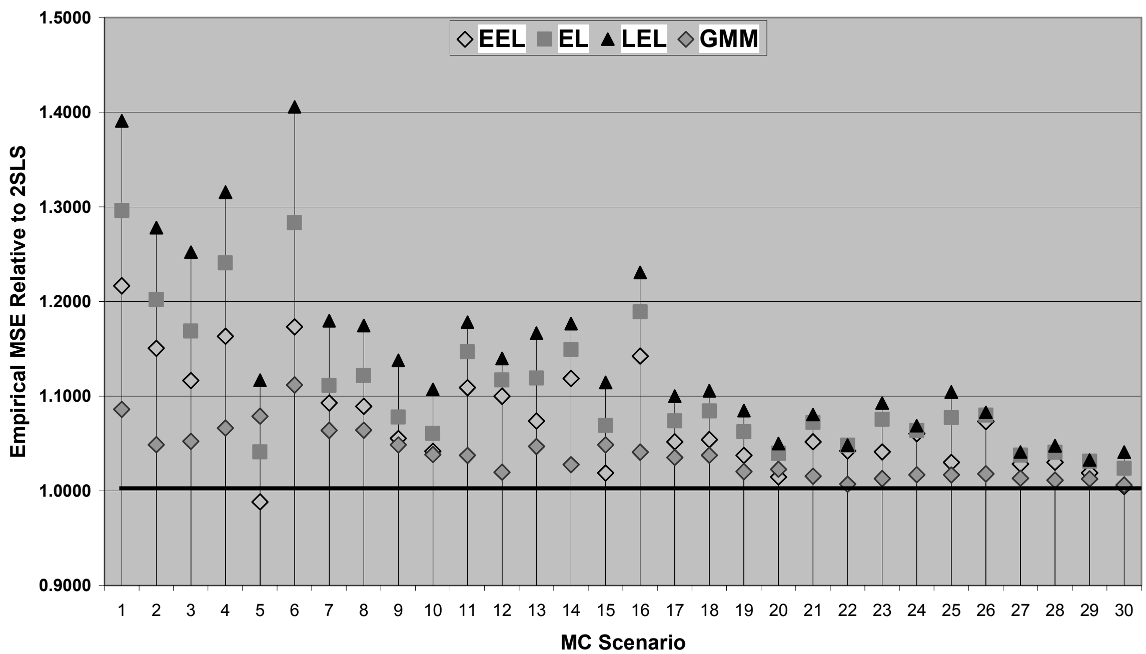
4. Monte Carlo Sampling Results

The results of the estimation and inference simulations are presented in this section. We report MSE results for the entire parameter vector $\boldsymbol{\beta}$, but limit our reporting of bias, variance, hypothesis tests and confidence region estimation performance to the structural parameter β_2 and note that the results for the remaining structural parameter were qualitatively similar. Tables containing the detailed simulation results are available from the authors.

4.1 Estimator MSE Performance

The simulated mean squared errors associated with estimating the $\boldsymbol{\beta}$ vector are presented in Figure 1, where results are expressed relative to the MSE of the 2SLS estimator and scenarios are numbered sequentially to repeatedly represent the 10 sampling scenarios in Table 3.1 for each of the sample sizes 50, 100, and 250. A number of general patterns are evident from the MC results. First of all, the 2SLS estimator

Figure 1. Parameter Vector Relative MSE



dominates the other four estimators in terms of parameter MSE, with the exception of the smallest sample size and scenario 5, in which case the MEEL estimator is marginally superior to all others. Second, the MSEs of the GMM(I) estimator are very close to the

MEEL estimator across all scenarios, but MEEL is actually MSE superior to GMM(I) in only a few cases. Third, there is a general order ranking of the MSEs of the EL-type estimators whereby generally MSE(MEEL) < MSE(MEL)< MSE(MLEL). However, differences in MSE performance among these estimators is small at n = 100 and practically indistinguishable at n = 250. Fourth, the MSE differences between *all* of the estimators dissipate as the sample size increases, with the differences being negligible at the largest sample size (n = 250).

4.2. Bias and Variance

Empirical bias and variance results for the estimators of β_2 are presented in Figures 2 and 3. Again some general estimator performance patterns emerge. First of all, the EL-type estimators, as a group, generally tend to be less biased than either the 2SLS or GMM estimators, but the EL estimators also tend to exhibit more variation than the traditional estimators. These performance patterns are especially evident for the small sample size (n = 50). Second, volatility in bias across MC scenarios is notably more pronounced for 2SLS and GMM than for the EL estimators, while just the opposite is true regarding volatility in variance measures across MC scenarios. Again this performance pattern is notably more pronounced at the smallest sample size than for the larger sample sizes. Third, regarding comparisons among EL-type estimators, the MEEL estimator tends to be the least variable among the three EL alternatives, with the ranking of variability tending to be in the order var(MEEL) < var(MEL) < var(MLEL). The ranking of relative bias performance among the EL estimators is less distinct, where especially for the smallest sample size, each of the EL–type estimators exhibits least bias for at least one MC

Figure 2. Bias in Estimating B2 = 2

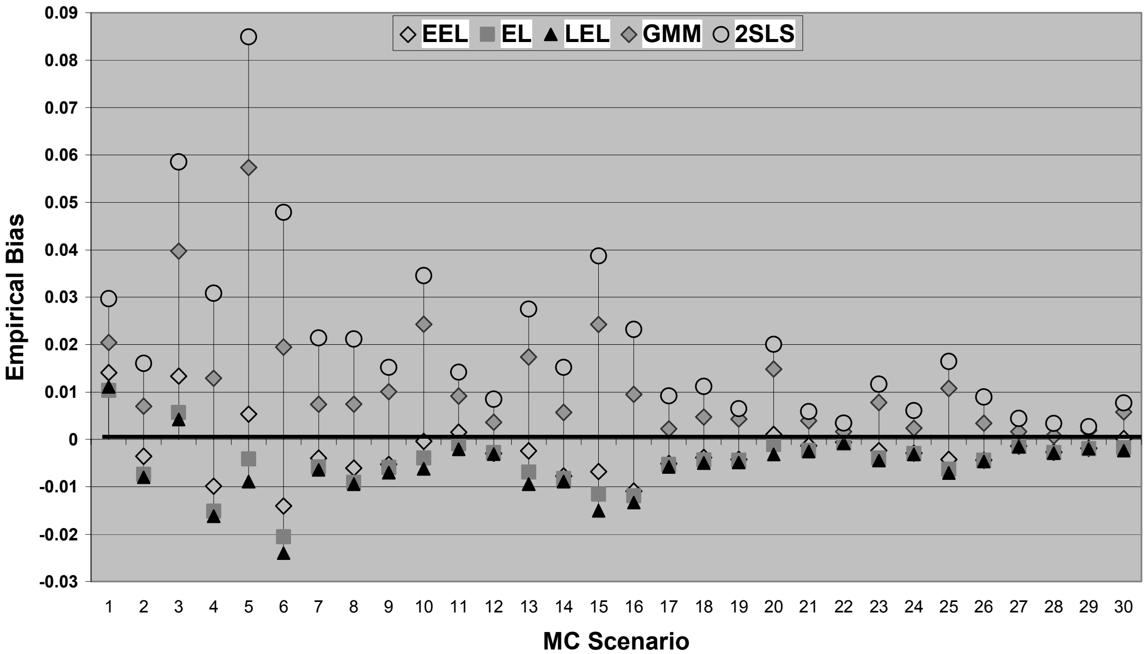
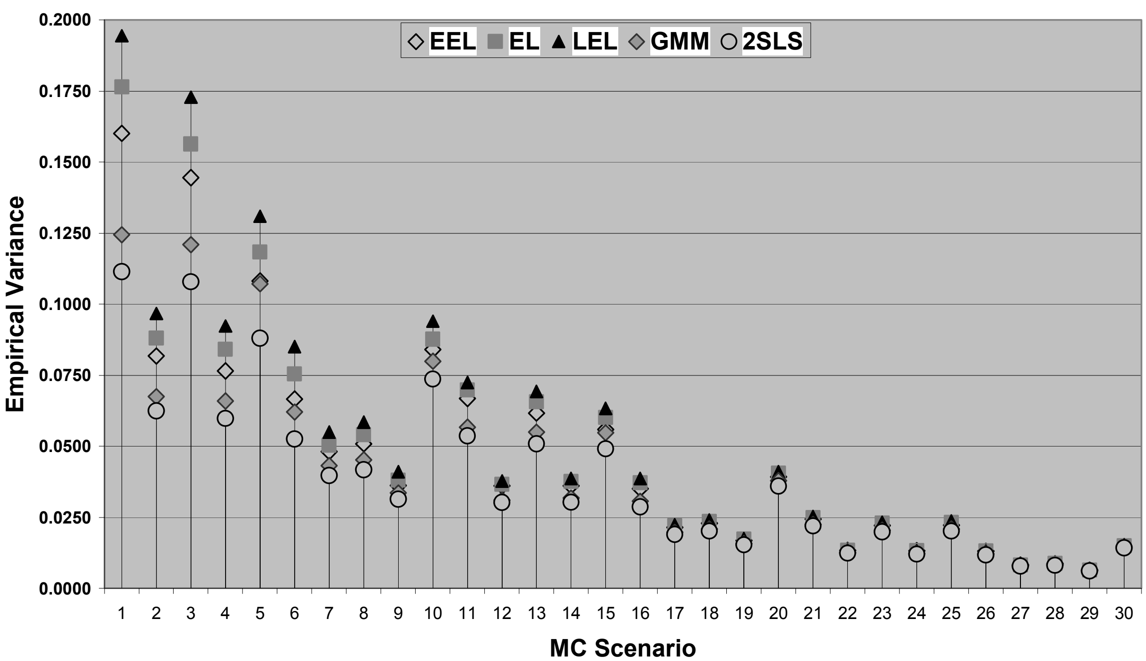


Figure 3. Variance in Estimating B2 = 2



scenario. For larger sample sizes the MEEL estimator more often than not has the smallest bias, but again there are exceptions for some scenarios, and in any case the bias of all of the EL-type estimators tends to be small, bordering on inconsequential for most of the scenarios when sample sizes are n = 100 or larger. Fourth, for the largest sample size (n = 250), both bias and variance tends to be quite small for all of the estimators considered, although in a relative sense, the traditional estimators continued to have notably larger bias for most scenarios than any of the EL-type estimators.

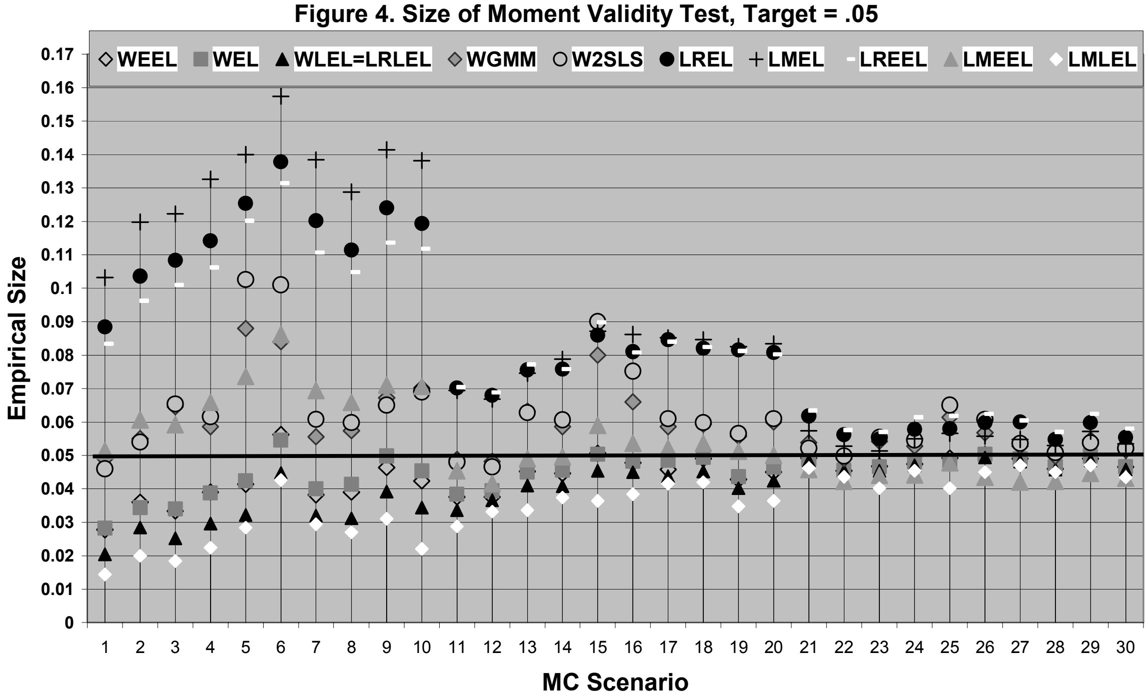
4.3. Prediction MSE

Judged in the context of generating predictions closest in expected Euclidean distance to actual dependent variable outcomes, the 2SLS and GMM estimators were notably superior to the EL-type estimators across the majority of sampling scenarios, and in any case were never worse. On the other hand, if one intended to use estimated residuals to generate an estimate of the model noise variance, the EL-type methods exhibited MSE measures that were closer in proximity to the true noise variance of $\sigma^2 = 25$. Among the EL-type methods, the general rank ordering of prediction MSE was MSE(MEEL) < MSE(MEL)< MSE(MLEL).

4.4. Size of Moment Validity Tests

Figure 4 presents empirical sizes of the 10 different tests of moment validity decribed in section 2.3. The target size of the test was set to the typical .05 level, and when n = 250 all of the test are generally within $\pm .01$ of this level across all MC scenarios. However, a number of the test procedures, most notably the LR tests for MEEL and MEL, the LM test for MEL, and to a lesser extent the Wald-Average Moment Test for 2SLS and GMM, are erratic and notably distant from the target test size when n = 50. The most consistent

Figure 4. Size of Moment Validity Test, Target = .05

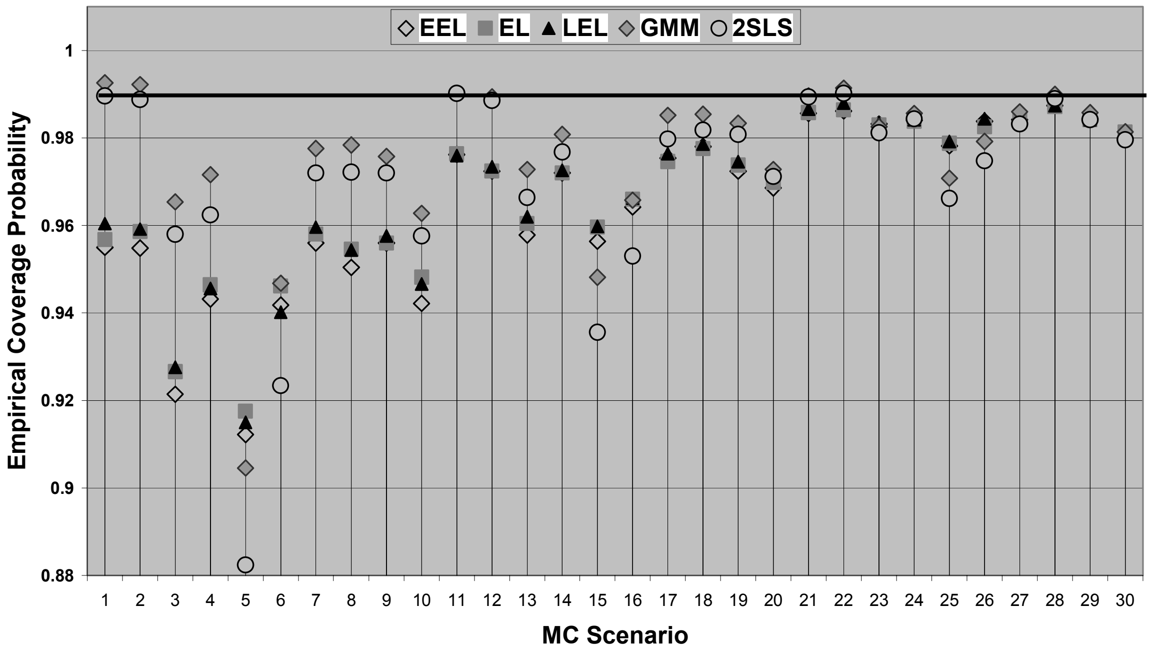


suite of tests in terms of average proximity to the true test size across MC scenarios were the Wald-Average Moment Tests for all three of the EL-type estimators. In addition the LM tests in the case of MEEL and MLEL was reasonably accurate when $n \geq 100$. As noted in the literature, for a subset of the scenarios, the size of the tests based on the traditional 2SLS and GMM methods were substantially distant from target size.

4.5 Confidence Interval Coverage and Expected Length

Figure 5 displays results relating to the empirical coverage probability of confidence intervals for the β_2 parameter, where target coverage is .99. Except for two scenarios involving the 2SLS and GMM methods, all of the confidence intervals are generally within .01 of the target coverage for the large sample size of n = 250. Again with the preceding two exceptions noted relating to the traditional estimators, coverage is

Figure 5. Confidence Interval Coverage Probability for B2, Target = .99



generally within .03 of target for the sample size of n = 100. Coverage degrades significantly for the small sample size n = 50, with the traditional estimators generally having better coverage, although they also exhibit demonstrably the worst coverage performance for two sampling scenarios. Moreover, the traditional methods exhibited more volatility across MC scenarios than EL-methods. We note that the coverage results observed for the EL-methods is consistent with other observations in the literature that the EL-type methods consistently underachieve target coverage probability under the asymptotic Chisquare calibration (Baggerly, 2001). In the large majority of cases, the traditional inference procedures also underachieved target coverage.

In the case of expected confidence interval length, a clearer relative performance pattern was apparent. In particular, the general relative ranking of CI length among the five alternative estimators was given by the following ordering of empirical average lengths: CI(MEEL) < CI(MEL) < CI(MLEL) < CI(2SLS) < CI(GMM). As expected, differences in length were most pronounced at the smallest sample size, in some cases exceeding 15%, but differences dissipated to effectively negligible levels when n = 250.

4.6. Test Power

All of the test procedures exhibited substantial power in rejecting the false null hypothesis $H_o : \beta_2 = 0$, where all rejection probabilities were in the range of .92 or higher. Among the EL-type methods, the relative power performance ranking was P(MEEL) > P(MEL) > P(MLEL). When comparing power performance to traditional methods, it was generally the case that 2SLS resulted in the most test power, followed by either MEEL or GMM, depending on the scenario, although the powers of the latter two procedures were in any case always very close to each other. The differences in power dissipated substantially for the higher sample sizes, and when n = 250, there was effectively no difference in power between any of the procedures, with all procedures achieving the ideal power of 1.

5. Some Final Remarks

In statistical models consisting of linear structural equations, the 2SLS and GMM estimators have long been the estimator of choice when the number of moment conditions-IV variables exceeded the number of unknown response parameters in the equation in question. Both the 2SLS and GMM estimators solve the problem of over-identification by defining particular rank-k linear combinations of the moment conditions. In contrast the nontraditional EL-type estimator transforms the overdetermined moments problem into a set of equations that is solvable for the model parameters by imposing a functional dependence on the moment equations through the choice of sample observation weights. Although both the traditional and EL-type estimators perform well in terms of first order asymptotics, questions persist as to their small sample bias and variance performance in estimation, and their coverage, interval width and power characteristics in terms of inference.

Given these questions and corresponding conjectures that appear in the literature, in this paper we provide some empirical evidence concerning the sampling performance of 2SLS, GMM and EL-type methods by simulating a range of sampling processes and observing empirical sampling behavior of the estimators and associated inference procedures. While MC sampling results are never definitive, the base results presented in this paper provide insights into the relative sampling performance of different types of general moment based estimators for a range of data sampling processes. Some distinct and interesting estimation and inference properties that we observed and did not know prior to our study are:

i) The EL-type estimators tend to exhibit less bias and more variance than the traditional estimators.

ii) In terms of MSE the 2SLS estimator wins almost all competitions. At a sample size of 100 or more, the estimators exhibit similar performances.

iii) In terms of accurate size of moment tests, the EL-type inference methods are superior, based on the average moment (or Wald) statistics, across all sample sizes. For sample sizes of 100 or more the LM tests also do reasonably well, especially in the case of MEEL and MLEL, and for a sample size of 250 all of the moment tests are in the neighborhood of the correct size.

iv) On confidence interval coverage, the traditional estimators perform somewhat erratically across differing data sampling processes until the highest sample size is reached. The EL-type methods are similar to each other in interval coverage performance, and exhibit a more orderly convergence to the correct coverage.

v) Test power for significance tests is very high for a sample size of 100 and is essentially 1 and ideal across all significance tests for sample size 250.

vi) A combination of concentrating out Lagrangian multipliers via numerical nonlinear equation solving algorithms, and then optimizing the concentrated optimization problem based on a non-gradient driven, direct search polytope (Nelder-Meade) type optimization algorithm appears to be a tractable and computationally efficient method for calculating solutions to EL-type problems in the IV-based moment constraint setting.

Many of the results appear reasonable and consistent with the limited amount of previous finite sample results (Mittelhammer and Judge, 2001a,b) and speculations in the literature relating to applications of EL-type estimators to structural equation estimation. The different pseudo-distance measures optimized by the trio of EL-type methods result in differing sampling performances for the varying estimator and test statistics, and those preferring a particular pseudo-distance measure will no doubt still be able to rationalize why their choice was not superior for a particular estimation or inference comparison. However, it is striking that none of the EL-type methods was found to be a compelling alternative to the ubiquitous 2SLS approach for parameter estimation, and there was only limited cases where the EL-type methods exhibited competitive inference properties.

Speculating further about the observed results, both the 2SLS and EL-type methods begin with the same ill-posed, over-identified set of moment conditions but transform them in differing ways into well-posed systems of equations that are solvable for the parameters. The 2SLS approach applies an optimal (in the optimal estimating function, OptEF, sense) linear transformation to the moment conditions that has a unique solution. This OptEF transformation can be derived analytically, its functional form is completely known, and it does not depend on any of the $\boldsymbol{\beta}$ or σ^2 to be estimated. Even though the unknown variance parameter σ^2 does appear in the explicit OptEF transformation, it is a scale factor that is redundant and can be eliminated when the optimal transform matrix is applied. On the other hand, the EL-type methods introduce *n* additional unknown parameters in order to resolve the overdetermined nature of the moment equations. These parameters must be estimated from the data, and act as slack variables that scale the sample observation components of the moment conditions to define a functionally dependent set of equations with rank equal to the dimension of the $\boldsymbol{\beta}$ parameter vector. The particular set of transformed moment conditions that is solved for $\boldsymbol{\beta}$ in EL-type methods is, in a sense, arbitrarily determined by an arbitrary choice of pseudo-distance measure (some member of the Cressie-Read family), and an optimal choice for finite samples, if it exists at all, is a measure zero set. Thus, it is to be expected that almost all EL-type methods are suboptimal in the class of estimating function-type estimators.

Looking towards future research, there are several ways to extend the empirical evidence concerning the performance of EL type estimators in recovering unknown response parameters in structural equations. We and others have noted that confidence regions generated by EL-type distance measures using χ^2 calibrations consistently under cover. Baggerly (2001) has suggested forming empirical regions through the use of a studentization of the moment constraints. Studentizing permits an escape from the convex hull of the moment data observations and may yield more accurate inferences in small samples.

It would be interesting to extend performance questions to data sampling processes that involve non-normal, non-symmetric distributions. Here the EL methods may exhibit improved performance because the moment information obtained from non-symmetric and/ or improperly centered distributions may be better accommodated by the flexible data weights available within the EL framework However, the answer is not clear because EL may attain smaller levels of bias, but at the expense of increased variance.

One interesting alternative data sampling process would be a statistical model in which Y is a discrete random variable. Based on preliminary work, we speculate that the use of EL- type estimators would perform well relative to semi-parametric alternatives in terms of quadratic loss.

Finally, in pursuit of achieving finite sample reductions in mean squared error, it is useful to consider, in a semi-parametric Stein-type of way, a mixture estimator that combines a consistent estimator having questionable finite properties, with an estimator that is inconsistent but has small finite sample variability. Such an estimator, which utilizes an EL-type moment formulation, has been proposed by Mittelhammer and Judge (2001c) and is currently under further evaluation.

6. References

Baggerly, K.A., 2001. Studentized Empirical Likelihood and Maximum Entropy: Empirical t, working paper, Dept. Of Statistics, Rice University.

Baggerly, K.A., 1998. Empirical likelihood as a goodness of fit measure, Biometrika 85, 535-547.

Bertsekas, D.P., 1995. Nonlinear Programming, Belmont: Athena Scientific.

Bound, J., Jaeger, D., Baker, R., 1995. Problems with instrumental variable estimation when the correlation between the instruments and the endogenous variables is weak, Journal of the American Statistical Association 90, 443-450.

Corcoran, S.A., 2000. Empirical exponential family likelihood using several moment conditions, Statistica Sinica 10, 545-557.

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

Godambe, V., 1960. An optimum property of regular maximum likelihood estimation, Annals of Mathematical Statistics 31, 1208-1212.

Golan, A., Judge, G. G., Miller, D., 1996. Maximum Entropy Econometrics. New York: John Wiley and Sons.

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

Heyde, C., 1989. Quasi-likelihood and optimality of estimating functions: some current and unifying themes, Bulletin of International Statistical Institute 1, 19-29.

Heyde, C., Morton, R., 1998. Multiple roots in general estimating equations, Biometrika 85 (4), 954-959.

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

Jacoby, S.L.S., Kowalik, J.S., and Pizzo, J.T. , 1972. Iterative Methods for Nonlinear Optimization Problems, New York: Prentice Hall, 1972.

Judge, G., Hill, R., Griffiths, W., Lutkepohl, H., Lee, T., 1985. The Theory and Practice of Econometrics. New York: John Wiley and Sons.

Kullback, S., 1959. Information Theory and Statistics. New York: John Wiley and Sons.

Maddala, G.S., Jeong, J., 1992. On the exact small sample distribution of the instrumental variable estimator, Econometrica 60, 181-183.

Mittelhammer, R. and Judge, G. (2001a). "Finite Sample Performance of the Empirical Likelihood Estimation under Endogeneity" in Computer Aided Econometrics, edited by David Giles, In Press.

Mittelhammer, R. and Judge, G. (2001b). "Endogeneity and Moment Based Estimation under Squared Error Loss," in Handbook of Applied Econometrics and Statistical Inference, edited by Alan Wan and Aman Ullah, Marcel Dekker Inc,, In Press.

Mittelhammer, R. and Judge, G. (2001c). "Robust Empirical Likelihood Estimation of Models with Non-Orthogonal Noise Components", Volume in Honor of Henri Theil, forthcoming in the Journal of Agriculture and Applied Economics, in press.

Mittelhammer, R., Judge, G., and Miller, D. (2000). Econometric Foundations, Cambridge University Press, 835 pages.

Nelder, J.A. and Mead, R., 1965. A Simplex Method for Function Minimization, Computer Journal 7, pp.308-313.

Nelson, C.R, Startz, R., 1990. Some further results on the exact small sample properties of the instrumental variable estimator, Econometrica 58, 967-976.

Newey, W. K., Smith, R.J., 2000. Asymptotic bias and equivalence of GMM and GEL estimators, MIT working paper.

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

Owen, A., 1991. Empirical likelihood for linear models, The Annals of Statistics 19 (4), 1725-1747.

Owen, A., 2001. Empirical Likelihood. New York: Chapman and Hall.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimating equations, The Annals of Statistics 22 (l), 300-325.

Read, T.R., Cressie, N.A., 1988. Goodness of fit statistics for discrete multivariate data. New York: Springer Verlag.

Stock, J. H., Wright, J. H., 2000. GMM With weak identification, Econometrica 68 (5),1055-1096.