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 empirical likelihood-type estimators when the estimating functions are well determined and the parameters are over identified. 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 empirical likelihood-type 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 given 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, which then overdetermines the model parameters. However, in some instances the instrumental variables may be only moderately or weakly correlated with the endogenous variables in the model, in which case while overdetermined, the parameters may not be strongly identified. 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 a moment based estimation and inference alternative to classical maximum likelihood methods (Owen, 1998, 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 linear structural equation estimators that are efficient, have small finite sample bias, and in terms of associated inference procedures, have accurate size, good power, and short confidence intervals with proper coverage, the purpose of this paper is to provide some empirical evidence relating to the finite sample performance of a range of empirical likelihood-type estimators in situations where the estimating functions overdetermine the model parameters, and the parameters are moderately well-identified.

Using Monte Carlo sampling procedures and a range of underlying data sampling processes relating to structural equations, we provide finite sample comparisons of 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, and a trio of empirical likelihood (EL) type estimators for recovering the unknown model parameters, including 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, and we suggest an algorithm that performed well in this regard.

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 traditional OptEF-2SLS-Optimal GMM and non-optimal GMM estimators exhibit substantial small sample bias?

ii) Do empirical likelihood (EL) type estimators exhibit substantial small sample bias?

iii) In terms of bias, are the EL-type estimators superior to traditional estimators?

iv) In terms of precision, are the EL-type estimators superior to traditional estimators?

v) Within the family of EL-type estimators, what is the relative performance of these estimators in terms of small sample bias and variance?

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

vii) What is the relative performance, in small samples, of the traditional and EL-type inference procedures relative to testing the moment restrictions?

viii) 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 and implications 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 can observe a vector of sample outcomes $\mathbf{y} = (y_1, y_2, \cdots, y_n)'$ associated with this linear model, $\mathbf{X}$ 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 $\mathrm{plim}\left[\mathrm{n}^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right] \neq \mathbf{0}$ and traditional Gauss-Markov based procedures such as the least squares (LS) estimator, or equivalently the method of moments (MOM)-extremum estimator $\hat{\boldsymbol{\beta}}_{\mathrm{mom}} = \arg_{\boldsymbol{\beta} \in \mathrm{B}}\left[\mathrm{n}^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}\right]$, are biased and inconsistent, with unconditional expectation and probability limit given by $\mathrm{E}\left[\hat{\boldsymbol{\beta}}\right] \neq \boldsymbol{\beta}$ and $\mathrm{plim}\left[\hat{\boldsymbol{\beta}}\right] \neq \boldsymbol{\beta}$.

2.1 Traditional Instrument-Based Estimators

Given a sampling process characterized by nonorthogonality of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$, and 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

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

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

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

If $m = k$, the vector of moment conditions just-determine the model parameters, then the sample moments (2.1) can be solved for the basic instrumental variable (IV) estimator $\hat{\boldsymbol{\beta}}_{\mathrm{iv}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Y}$. 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 1997; 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, $\boldsymbol{\beta}_{2sls} = (\mathbf{X}'\mathbf{P_z}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P_z}\mathbf{Y}$, where $\mathbf{P_z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ is the projection matrix for $\mathbf{Z}$. This estimator is equivalent to the estimator formed by applying the optimal estimating function (OptEF) transformation $n\left(\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}$ to the moment conditions in (2.2) (Godambe, 1960; 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 single estimation criterion based on a quadratic form in the sample moment information

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

The GMM estimator can be shown to have optimal asymptotic properties if the weighting matrix **W** is defined appropriately. 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, based on multinomial distributions 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. In our case, a solution is achieved by adjusting the expected counts (distance between the two sets of probabilities) by making use of a goodness-of-fit criterion subject to the moment constraints. As an estimating criterion, one possibility is to make use of the Cressie and Read (1984) power divergence family of statistics

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

where λ is an arbitrary unspecified parameter, and if the q_i's are to be interpreted as probabilities, then they satisfy $q_i \in (0,1)$, $\forall i$ and $\sum_{i=1}^{n} q_i = 1$. When $q_i = n^{-1}\ \forall i$, and 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 for $\boldsymbol{\beta}$:

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

Three main variants of $\mathrm{I}(\mathbf{w},\mathbf{q},\lambda)$ have emerged and received explicit attention in the literature. When $\lambda \to 0$, this leads to the traditional empirical log-likelihood objective function, $n^{-1}\sum_{i=1}^{n}\ln(\mathrm{w_i})$, and the maximum empirical likelihood (MEL) estimate of $\boldsymbol{\beta}$. When $\lambda \to -1$, this leads to the empirical exponential likelihood objective function, $-\sum_{i=1}^{n}\mathrm{w_i}\ln(\mathrm{w_i})$, and the Maximum Empirical Exponential Likelihood (MEEL) estimate of $\boldsymbol{\beta}$. 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 $\boldsymbol{\beta}$.

If the traditional Owen MEL criterion is used, the estimation objective involves finding the feasible weights $\hat{\mathbf{w}}$ that maximizes the joint empirical probability assigned to the observed set of sample observations, conditional on the moment constraints. In the sense of objective function analogies, the MEL approach is the closest to the classical maximum likelihood approach. The MEEL criterion of maximizing $-\sum_{i=1}^{n}\mathrm{w_i}\ln(\mathrm{w_i})$ is equivalent to defining an estimator by *minimizing* the Kullback-Leibler (KL) information criterion $\sum_{i=1}^{n}\mathrm{w_i}\ln(\mathrm{w_i}/n^{-1})$. 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, where the expectations are based on the $\hat{\mathbf{w}}$ distribution and the log-likelihood ratio has the restricted (by moment constraints) likelihood in the numerator and the unrestricted (i.e., uniform distribution) likelihood in the denominator. 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. Note that all of the preceding estimation objective functions achieve *unconstrained* (by moment constraints) optima when the empirical probability distribution is given by $\mathbf{w} = \mathbf{1}_n n^{-1}$. Note further that weights must be nonnegative valued in the MEEL and MEL cases, but negative weights are not ruled out by the MLEL specification.

The Lagrange function associated with the traditional MEL formulation is

$$L(\mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \eta) = n^{-1} \sum_{i=1}^{n} \ln(w_i) - \boldsymbol{\alpha}' \sum_{i=1}^{n} w_i \mathbf{z}'_{i.} (y_i - \mathbf{x}_{i.} \boldsymbol{\beta}) - \eta \left(\sum_{i=1}^{n} w_i - 1 \right) \qquad (2.6)$$

where $\boldsymbol{\alpha}$ and η are Lagrange multipliers. From the first order conditions relating to **w**, and noting that the optimal value of η is 1, constrained optimal w_i's can be expressed in terms of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ as

$$w_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \left[n \left(\boldsymbol{\alpha}' \mathbf{z}'_{i.} (y_i - \mathbf{x}_{i.} \boldsymbol{\beta}) + 1 \right) \right]^{-1}. \qquad (2.7)$$

The Lagrangian form of the MEEL formulation, expressed as a *minimization* problem, is

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

and the constrained optimal w_i's, expressed as a function of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, are

$$w_i(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{\exp\left(\boldsymbol{\alpha}' \mathbf{z}'_{i.} (y_i - \mathbf{x}_{i.} \boldsymbol{\beta}) \right)}{\sum_{j=1}^{n} \exp\left(\boldsymbol{\alpha}' \mathbf{z}'_{j.} (y_j - \mathbf{x}_{j.} \boldsymbol{\beta}) \right)}. \qquad (2.9)$$

Finally, the Lagrangian form of the MLEL formulation, also expressed as a *minimization* problem, is

$$L(\mathbf{w}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \eta) = n^{-1} \left(\sum_{i=1}^{n} \left(n^2 w_i^2 - 1 \right) \right) - \boldsymbol{\alpha}' \sum_{i=1}^{n} w_i \mathbf{z}'_{i.} (y_i - \mathbf{x}_{i.} \boldsymbol{\beta}) - \eta \left(\sum_{i=1}^{n} w_i - 1 \right) \qquad (2.10)$$

which leads to constrained optimal w_i 's, expressed as a function of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, given by

$$w_i(\boldsymbol{\beta},\boldsymbol{\alpha},\eta)=(2n)^{-1}\left(\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i-\mathbf{x}_{i.}\boldsymbol{\beta})+\eta\right). \tag{2.11a}$$

The Lagrange multiplier η in (2.11) can be eliminated by solving the adding up condition $\mathbf{1}'_n\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha})=1$ for η and then substituting the solved value of η into (2.11a) to obtain

$$w_i(\boldsymbol{\beta},\boldsymbol{\alpha})=(2n)^{-1}\left(\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i-\mathbf{x}_{i.}\boldsymbol{\beta})+2-n^{-1}\sum\nolimits_{i=1}^{n}\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i-\mathbf{x}_{i.}\boldsymbol{\beta})\right) \tag{2.11b}$$

where the value of η is given by

$$\eta(\boldsymbol{\alpha},\boldsymbol{\beta})=\left(2-n^{-1}\sum\nolimits_{i=1}^{n}\boldsymbol{\alpha}'\mathbf{z}'_{i.}(y_i-\mathbf{x}_{i.}\boldsymbol{\beta})\right). \tag{2.12}$$

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 (2.7), (2.9), or (2.11) with respect to $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and/or η are consistent asymptotically normally distributed, and asymptotically efficient relative to the optimal estimating function (OptEF) estimator, *given the set of estimating equations under consideration*. 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. Note further that for the typical application in which $q_i = n^{-1}\ \forall i$, *any* of the estimation objective functions contained in the Cressie-Read family achieve *unconstrained* (by moment equations) optima when the empirical probability distribution is given by $\mathbf{w} = \mathbf{1}_n n^{-1}$.

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, 1990) showed that an analog of Wilks' Theorem for likelihood ratios, -2ln(LR) $\overset{a}{\sim} \chi^2_j$, 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 the test of the hypothesis $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, or more generally for testing the linear combinations hypothesis $\mathbf{c}\boldsymbol{\beta} = \mathbf{r}$ when $rank(\mathbf{c}) = j$, is given for the MEL case by

$$LR_{EL}(\mathbf{y}) = \frac{\max_{\boldsymbol{\beta}}\left[\ell_E(\boldsymbol{\beta})\ s.t.\ \mathbf{c}\boldsymbol{\beta} = \mathbf{r}\right]}{\max_{\boldsymbol{\beta}} \ell_E(\boldsymbol{\beta})} \tag{2.13}$$

and

$$-2\ln\left(LR_{EL}(\mathbf{Y})\right) \overset{a}{\sim} Chisquare(j,0) \tag{2.14}$$

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

$\mathbf{c\beta} = \mathbf{r}$ on the Wald Criterion in the usual way by utilizing the inverse of the asymptotic covariance matrix of $\mathbf{c}\hat{\boldsymbol{\beta}}_{EL}$ as the weight matrix of a quadratic form in the vector $\mathbf{c}\hat{\boldsymbol{\beta}}_{EL} - \mathbf{r}$, or construct tests based on the Lagrange multipliers associated with the constraints $\mathbf{c\beta} = \mathbf{r}$ 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 also be assessed via a variation of the preceding testing methodology. We provide further details regarding the empirical implementation of test statistics and confidence region estimators in our discussion of the Monte Carlo experiments ahead.

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, referred to as Average Moment Tests by Imbens, Spady, and Johnson (1998), are calculated for all five estimators. The Wald test statistics are specified as

$$\text{Wald} = \left(\mathbf{1}_n'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)\right)'\left[\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)\right]^{-1}\left(\mathbf{1}_n'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\hat{\boldsymbol{\beta}}\right)\right)\right) \quad (2.15)$$

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 based on the same statistics utilized by Imbens, Spady, and Johnson and are given as follows:

$$LR_{EEL} = 2n\left(\mathbf{w}'\ln(\mathbf{w}) + \ln(n)\right) \tag{2.16}$$

$$LR_{EL} = -2\left(\mathbf{1}_n'\ln(\mathbf{w}) + n\ln(n)\right). \tag{2.17}$$

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

$$LR_{LEL} = n\left(1 - n^{-1}\mathbf{1}_n'\left[\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\left(\frac{\boldsymbol{\alpha}}{\eta}\right)\right]^2\right) = n\left(1 - \left(\frac{2}{\eta}\right)^2 n\sum_{i=1}^{n} w_i^2\right) \tag{2.18}$$

While perhaps not initially apparent from the functional definitions, it can be shown that $LR_{LEL} \equiv Wald_{LEL}$ and, thus, we later report on the performance of only one copy of this particular test. The $\mathbf{w}$ weights, $\boldsymbol{\beta}$ vector, and Lagrange multipliers $\boldsymbol{\alpha}$ and η appearing in the LR test statistics (2.16)-(2.18) 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 follow Imbens, Spady, and Johnson (1998, p. 341) and examine a quadratic form in the Lagrange multiplier vector that incorporates a robust estimator of the covariance matrix of the moment constraints, as

$$LM_{EEL} = n\alpha'\left[\left(\mathbf{h}(\beta)\odot\mathbf{w}\right)'\mathbf{h}(\beta)\right]\left[\left(\mathbf{h}(\beta)\odot\mathbf{w}\right)'\left(\mathbf{h}(\beta)\odot\mathbf{w}\right)\right]^{-1}\left[\left(\mathbf{h}(\beta)\odot\mathbf{w}\right)'\mathbf{h}(\beta)\right]\alpha \quad (2.19)$$

where $\mathbf{h}(\beta)\equiv\left(\mathbf{Z}\odot(\mathbf{Y}-\mathbf{X}\beta)\right)$ and $\mathbf{w}$, α and β 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

$$LM = n\alpha'\left(\Omega^{-1}-\Omega^{-1}\mathbf{G}'\mathbf{V}\mathbf{G}\Omega^{-1}\right)^{-}\alpha \quad (2.20)$$

where

$$\Omega \equiv n^{-1}\left(\mathbf{Z}\odot(\mathbf{Y}-\mathbf{X}\beta)\right)'\left(\mathbf{Z}\odot(\mathbf{Y}-\mathbf{X}\beta)\right),\ \ \mathbf{G}\equiv n^{-1}\mathbf{X}'\mathbf{Z},\ \ \mathbf{V}\equiv\left(\mathbf{G}\Omega^{-1}\mathbf{G}'\right)^{-1} \quad (2.21)$$

and the values of β and α are replaced by either MEEL or MLEL estimates. All of the LM tests are asymptotically Chisquare distributed under the null hypothesis, 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

$$Z = \frac{\hat{\beta}_i}{\widehat{std}\left(\hat{\beta}_i\right)} \overset{a}{\sim} N(0,1)\ under\ H_0 : \beta_i = 0, \quad (2.22)$$

with the associated confidence interval estimate given by

$$\left(\hat{\beta}_i - z_\tau \widehat{std}\left(\hat{\beta}_i\right), \hat{\beta}_i + z_\tau \widehat{std}\left(\hat{\beta}_i\right)\right) \quad (2.23)$$

where z_τ denotes the 100τ % quantile of the standard normal distribution, and $\hat{\beta}_i$ *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-OptGMM-OptEF, GMM(I), MEEL, MEL and MLEL estimators. The covariance matrices for the traditional estimators were defined as

$$AsyCov(\hat{\mathrm{B}}_{2sls}) = \hat{\sigma}^2 \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1} \tag{2.24}$$

and

$$AsyCov(\hat{\mathrm{B}}_{GMM(I)}) = \hat{\sigma}^2 \left(\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)\mathbf{Z}'\mathbf{X}\right)\left(\mathbf{X}'\mathbf{Z}\mathbf{Z}'\mathbf{X}\right)^{-1} \tag{2.25}$$

where $\hat{\sigma}^2$ is the usual consistent estimate of the equation noise variance. The common general form of the covariance matrices for the MEEL, MEL, and MLEL procedures is given by

$$AsyCov(\hat{\mathrm{B}}_{EL-type}) = \left[\left(\mathbf{X}'\left(\mathbf{Z}\odot\mathbf{w}\right)\right)\left[\left(\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\odot\mathbf{w}\right)'\left(\mathbf{Z}\odot\left(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}\right)\right)\right]^{-1}\left(\left(\mathbf{Z}\odot\mathbf{w}\right)'\mathbf{X}\right)\right]^{-1} \tag{2.26}$$

In empirical applications the $\mathbf{w}$ and $\boldsymbol{\beta}$ vectors are replaced by the appropriate estimates obtained from applications of the MEEL, MEL, or MLEL estimation procedures.

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 of being less than full rank. This occurs by design in these types of problems because the fundamental method by which the MEL, MEEL and MLEL methods (and in fact any method based on a Cressie-Read type estimation objective function) 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. Under these conditions, Imbens, Spady and Johnson found it advantageous in their particular EEL and EL simulation applications to utilize penalty function methods 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 MEL, MEEL, and MLEL optimization problems (2.6), (2.8), (2.10), 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, and so a numerical nonlinear equation solving procedure was employed. 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, the computational procedure involved the following two steps. In the first step relating to concentrating out the Lagrangians, the Lagrange multiplier vector $\boldsymbol{\alpha}$ was expressed as a function of $\boldsymbol{\beta}$ by utilizing the empirical moment conditions and the weight representation (2.7), (2.9), or (2.11b) for the vector $\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha})$, and solving

$$\boldsymbol{\alpha}(\boldsymbol{\beta}) \equiv \arg_{\boldsymbol{\alpha}}\left[\left(\mathbf{Z}\odot(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})\right)'\mathbf{w}(\boldsymbol{\beta},\boldsymbol{\alpha})=\mathbf{0}\right]. \tag{2.27}$$

The solution to (2.27) 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 (2.27) into the definition of η in (2.12), yielding

$$\eta(\boldsymbol{\beta}) \equiv \left(2 - n^{-1}\sum_{i=1}^{n}\boldsymbol{\alpha}(\boldsymbol{\beta})'\,\mathbf{z}_{i.}'\left(y_i - \mathbf{x}_{i.}\boldsymbol{\beta}\right)\right). \tag{2.28}$$

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

$$L_*(\boldsymbol{\beta}) \equiv L\left(\mathbf{w}\left(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta})\right),\boldsymbol{\beta},\ \boldsymbol{\alpha}(\boldsymbol{\beta}),\eta(\boldsymbol{\beta})\right)$$

$$\equiv \phi\left(\mathbf{w}\left(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta})\right)\right)-\boldsymbol{\alpha}(\boldsymbol{\beta})'\sum_{i=1}^{n} w_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta})\right)\ \mathbf{z}'_{i.}\left(y_i-\mathbf{x}_{i.}\boldsymbol{\beta}\right)-\eta(\boldsymbol{\beta})\left(\sum_{i=1}^{n} w_i\left(\boldsymbol{\beta},\boldsymbol{\alpha}(\boldsymbol{\beta})\right)-1\right) \tag{2.29}$$

The value of $L_*(\boldsymbol{\beta})$ is then optimized (in our characterizations (2.6),(2.8), and (2.10), maximized for MEL, minimized for MEEL and MLEL) with respect to the choice of $\boldsymbol{\beta}$, where $\phi(\bullet)$ can 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.

3. Design of Sampling Experiments

In terms of the EL-type formulations of Section 2, the solution for the optimal weights and parameter estimates cannot be expressed in closed form. Moreover, the finite sample probability distributions of the traditional 2SLS-GMM-OptEF estimators are also generally intractable. Consequently, the finite sample properties of these estimation and associated inference procedures cannot be derived from a direct evaluation of functional forms applied to distributions of random variables. We use a Monte Carlo sampling experiment to identify 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 Design

Consider a data sampling process of the following form:

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

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

where $\mathbf{X}_{i.} = (Z_{i1}, Y_{i2})$, and $i = 1, 2, ..., n$. The two-dimensional vector of unknown parameters, $\boldsymbol{\beta}$, in (3.1) is arbitrarily set equal to the vector $[-1, 2]'$. The outcomes of the (6×1) random vector $[Y_{i2}, \varepsilon_i, Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4}]$ 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, with $Z_{i5} \equiv 1, \forall i.$ 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 MC Scenario Characteristics

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 zeros, 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 of the scenarios are delineated in Table 3.1. 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 $\left(\mathbf{Z}_{.1}, \mathbf{Y}_2\right)$ is relatively high, being in the .84 to .95 range. All told, there were 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, and include 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) and between $\mathbf{y}$ with $\hat{\mathbf{y}}$ (measuring predictive risk). We also report on 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 of the estimates, $\text{Var}\left(\hat{\beta}_i\right)$.

Regarding inference performance, we compare the empirical size of ten alternative tests of moment equation validity with a nominal Chisquare based target size of .05, we examine the empirical coverage probability of confidence interval estimators based on each alternative estimation technique with a target coverage probability of .99, we calculate and compare the empirical expected lengths of confidence intervals, and we also report on the power of significance tests associated with the different estimation methods.

Five different estimators, and associated inference procedures, were examined including the GMM estimator based on the asymptotically optimal GMM weighting matrix (GMM-2SLS-OptEF), GMM based on an identity matrix weighting (GMM-I), and the three EL-type estimators, including the MEL, MEEL, and MLEL estimators.

Table 3.1 Monte Carlo Experiment Definitions, with $\beta = [-1, 2]'$, $\sigma_{\varepsilon_i} = \sigma_{Y_{2i}} = 5$, **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 $\hat{\mathbf{Y}}_2 = \mathbf{Z}\boldsymbol{\pi}$.

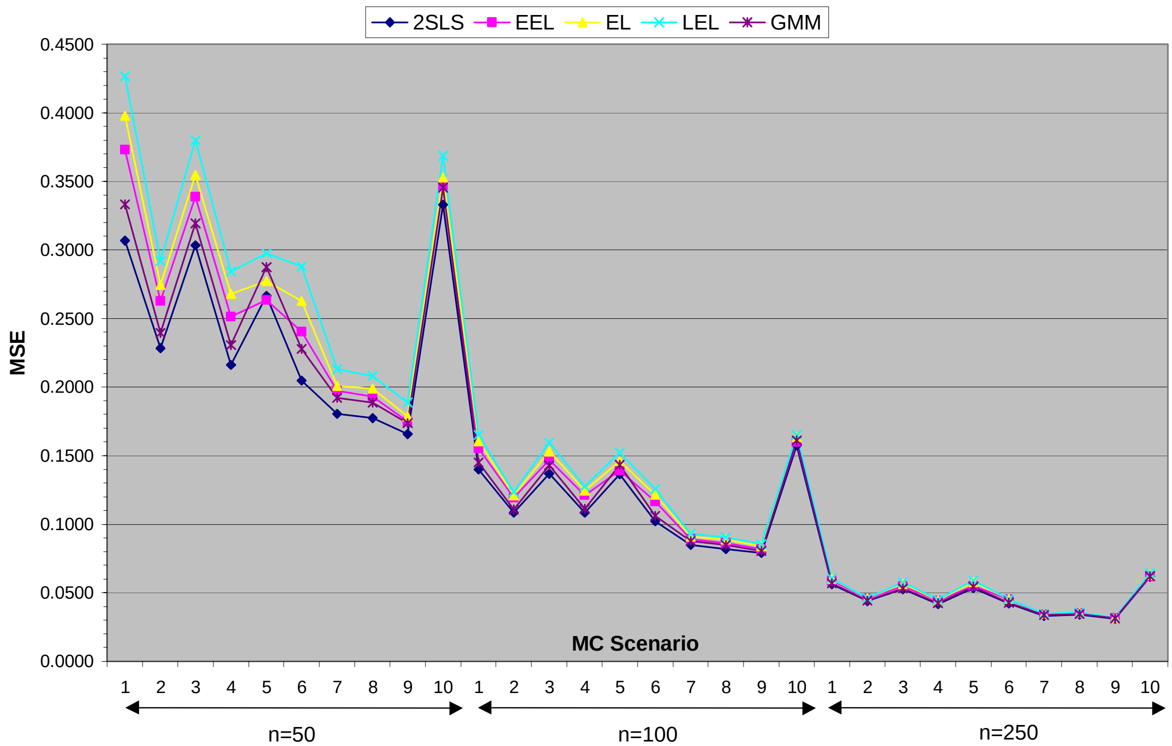
4. Monte Carlo Sampling Results

The results of the estimation and inference simulations are presented in this section. We limit our reporting of bias, variance, hypothesis tests and confidence region estimation performance relating to structural parameters to the β_2 coefficient and note that the results for the remaining structural parameter were qualitatively similar.

4.1 Estimator MSE Performance

The simulated mean squared errors associated with estimating the structural model parameters for each of the MC scenarios are presented in Appendix table A.1. A graph of the MSE results is presented in Figure 1. A number of general patterns are evident from the MC results. First of all, the OptEF-OptGMM-2SLS estimator 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 with MEEL 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, making the MSE performance ranking moot for larger sample sizes. 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).

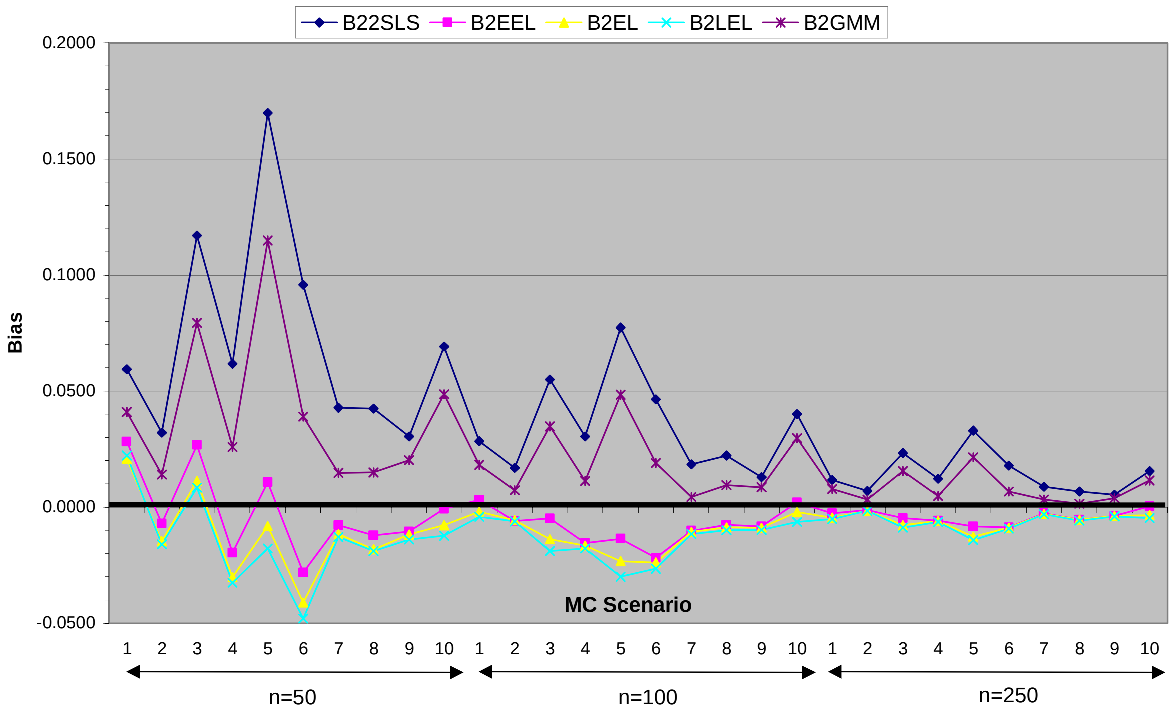
Figure 1. Parameter MSE Results



4.2. Bias and Variance

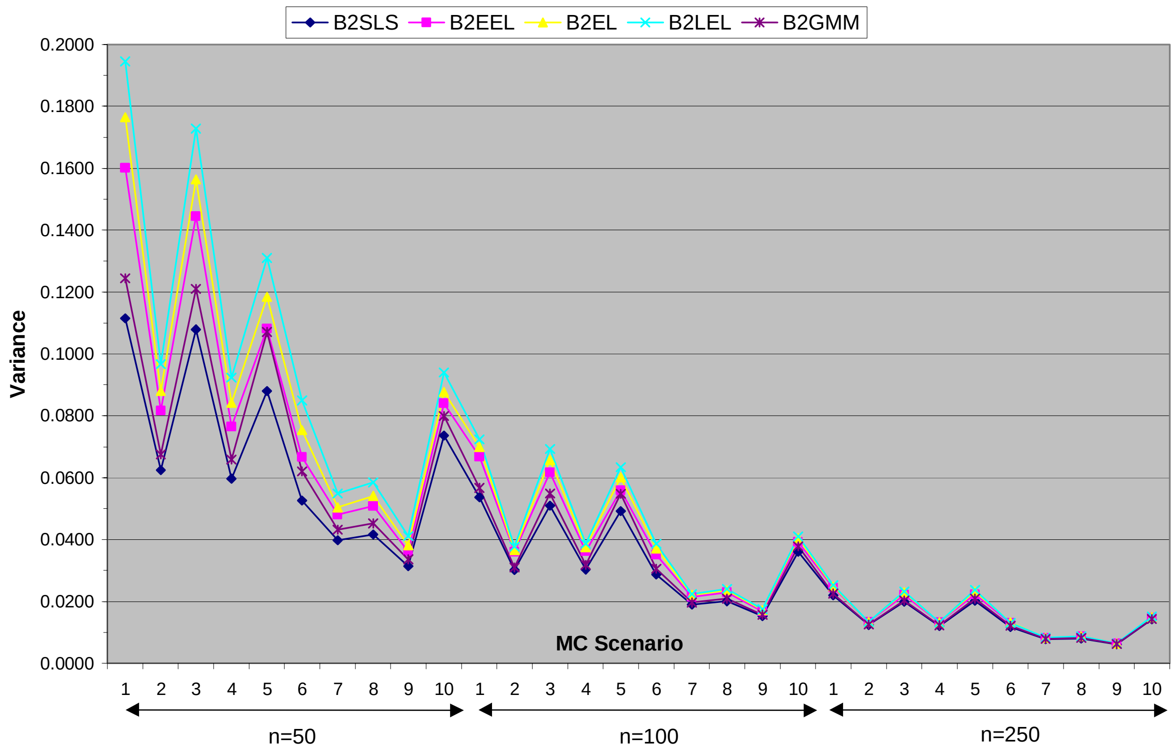
Empirical bias and variance results for the estimators of β_2 are presented in Figures 2 and 3, and Appendix tables A.2 and A.3. Again some general performance patterns emerge upon examining the figures and numbers. First of all, the EL-type estimators, as a group, generally tend to be less biased than 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).

Figure 2. Bias Results for B2 Parameter



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-estimator types, 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 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

Figure 3. Variance Results for B2 Parameter



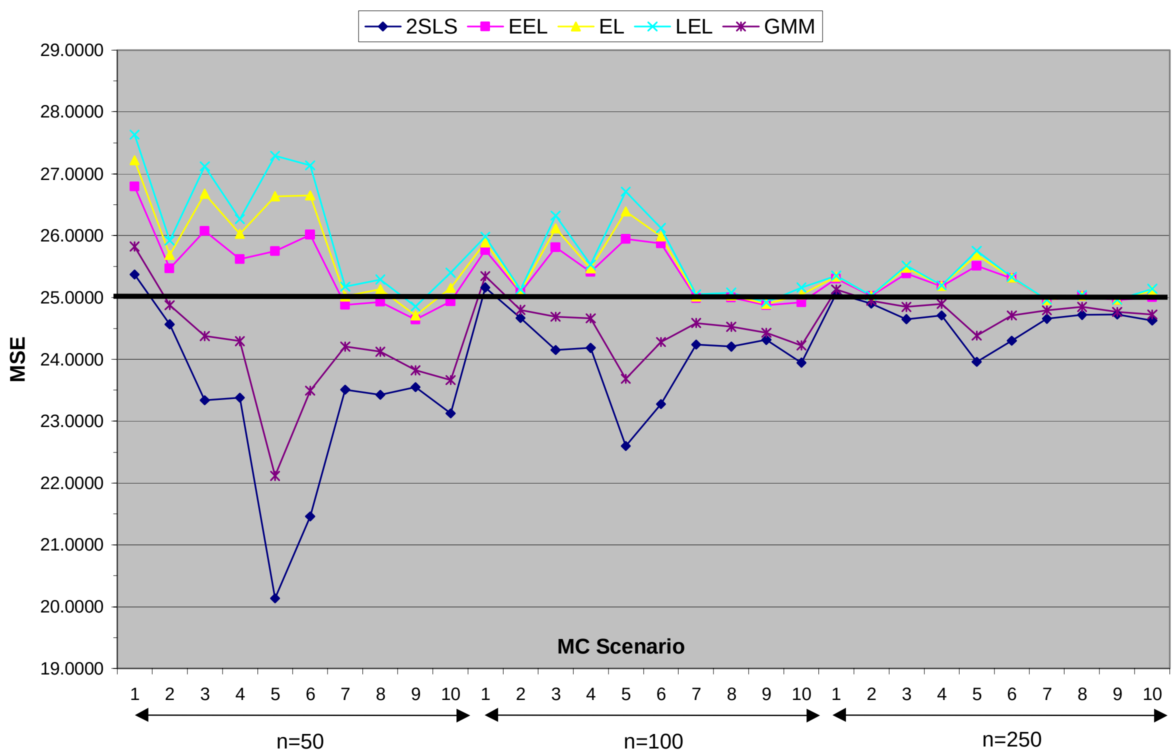
bias of all of the EL-type estimators tends to be small, bordering on inconsequential for most of the scenarios when sample sizes are n = 100 or larger. Fourth, for the largest sample size (n = 250), both bias and variance tends to be quite small for all of the estimators considered, although in a relative sense, the traditional estimators continued to have notably larger bias for most scenarios than any of the EL-type estimators.

4.3. Prediction MSE

Figure 4 and appendix table A.4 present results relating to the MSE in predicting the dependent variable of the structural equation. Judged in the context of generating predictions closest in expected Euclidean distance to actual dependent variable outcomes, it is clear that the 2SLS and GMM estimators are notably superior to the EL-type estimators across the majority of sampling scenarios, and in any case are never worse.

On the other hand, if one intends to use the estimated residuals from the model to generate an estimate of the model noise variance, then the EL-type methods exhibit MSE measures that are closer in proximity to the true noise variance, $\sigma_{\varepsilon}^{2}=25$, than do the traditional estimation methods. Among the EL-type methods, the general rank ordering of prediction MSE is MSE(MEEL) < MSE(MEL)< MSE(MLEL).

Figure 4. Prediction MSE Results

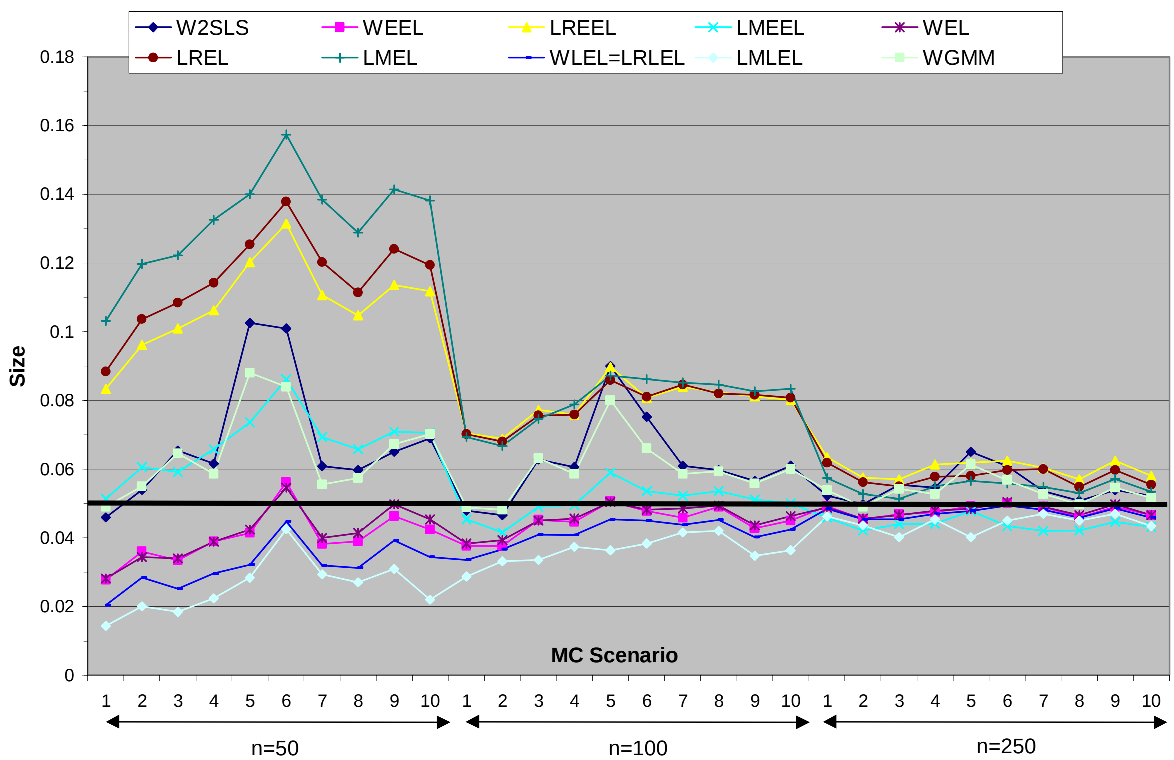


4.4. Size of Moment Validity Tests

Figure 5, and Appendix table A.5, present empirical sizes of the 10 different tests of moment validity decribed in section 2.3. The target size of the test was .05, and when n = 250 all of the test are generally within $\mp$ .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 lessor 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 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, although in the majority of cases when $n \geq 100$ both the 2SLS and GMM methods were within .01 of the test size target. It is interesting to note that GMM exhibited superior size performance to 2SLS in the majority of cases.

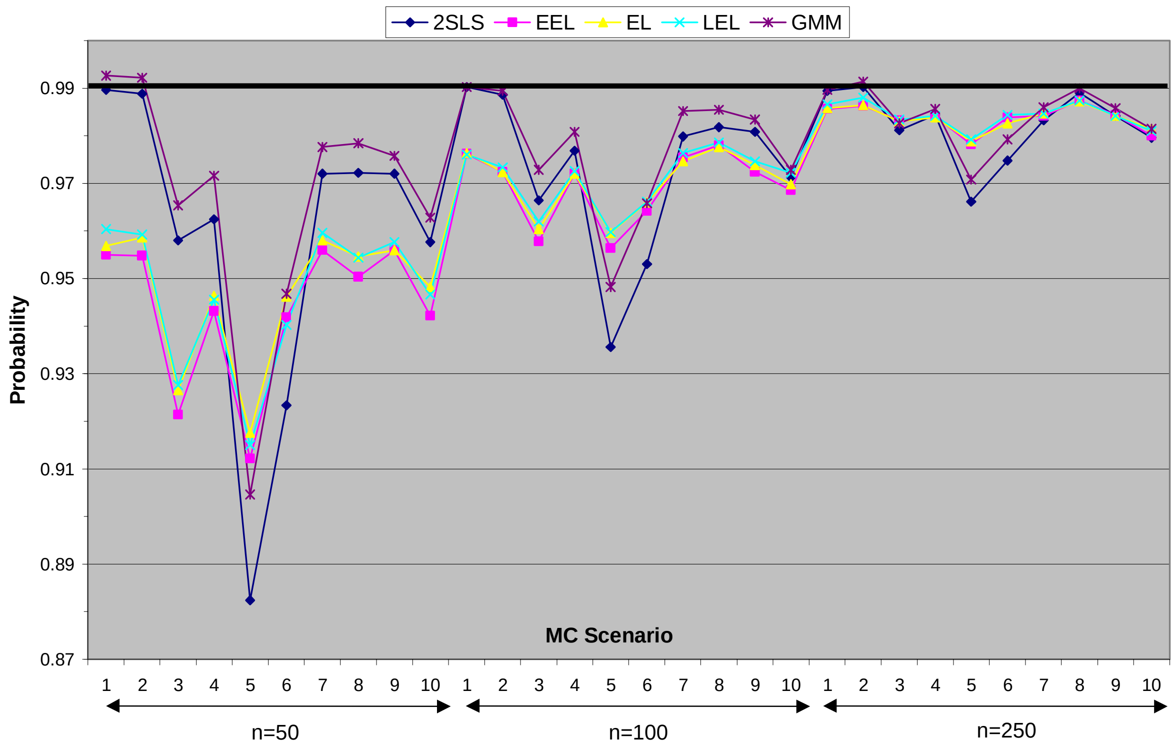
Figure 5. Empirical Size of Moment Validity Test, Target Size = .05



4.5 Confidence Interval Coverage and Size

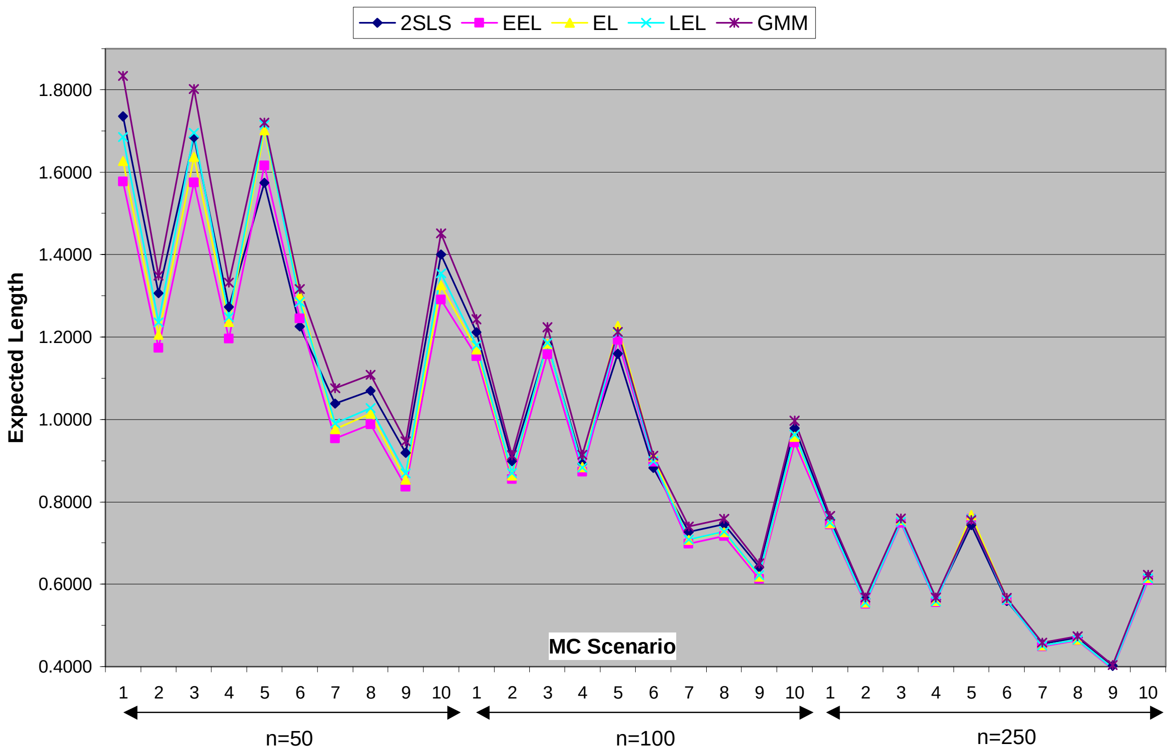
Figures 6 and 7, and Appendix Tables A.6 and A.7, display results relating to the empirical coverage probability of confidence intervals for the β_2 parameter, as well as the empirical expected length of the confidence intervals, where target coverage is .99.

Figure 6. Empirical Confidence Interval Coverage Probability for B2, Target = .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 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

Figure 7. Empirical Expected Confidence Interval Length for B2

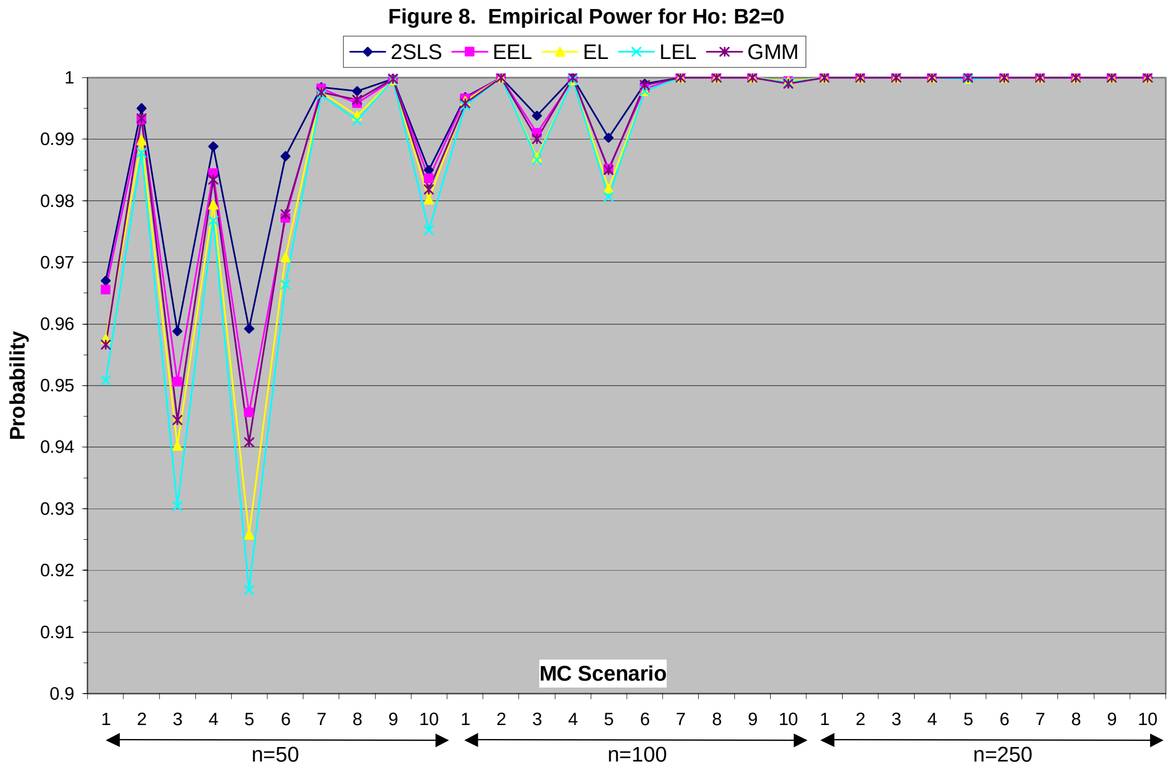


exhibited more volatility across MC scenarios than EL-methods. We note that the coverage results observed for the EL-methods is consistent with other observations in the literature that the EL-type approach consistently underachieves target coverage probability under the asymptotic Chisquare calibration (Baggerly, 2001). We add that, 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

Results relating to the power of the standard Z-test in testing the significance of the β_2 parameter are displayed in Figure 8 and in Appendix table A.8. All of the test procedures exhibited substantial power in correctly rejecting the 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 in the following order: P(MEEL) > P(MEL) > P(MLEL). When comparing power performance to the 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. It is also interesting to note that for the last 4 MC scenarios of the intermediate sample size n = 100 case, which are characterized by higher levels of correlation between the group of instruments and the endogenous explanatory variable of the structural model, there was also effectively no difference in power performance among any of the EL and traditional inference procedures.



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. The 2SLS estimator solves the problem of over-identification by taking a particular rank-k linear combination of the instruments. 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. Furthermore, in line with sampling processes often found in practice, there are questions concerning the possible estimation and inference impacts of IV's that are only weakly correlated with right hand side endogenous variables so that response parameters are only weakly identified or determined.

Given these questions and corresponding conjectures that appear in the literature, in this paper we attempt to provide some empirical evidence concerning 2SLS, GMM and EL type estimator performance by simulating a range of sampling processes and observing empirical repeated sampling behavior of the estimation procedures. While MC sampling results are never definitive, we feel that the base results presented in this paper provide important insights into the relative sampling performance of the different types of general moment based estimators for a range of data sampling processes. Some distinct and interesting patterns have emerged from the MC results in this study and may bear emphasizing here:

i) The entire suite of EL type estimators tend to exhibit less bias than the traditional estimators.

ii) The EL type estimators tend to exhibit more variance than the traditional estimators.

iii) In terms of MSE the 2SLS estimator wins almost all competitions. Around a sample size of 100 the estimators exhibit similar performances.

iv) 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.

v) 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.

vi) 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.

vii) A combination of concentrating out Lagrangian multipliers via numerical nonlinear equation solving algorithms, and then optimizing the concentrated optimization problem based on a 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.

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 can permit an escape from the convex hull of the moment data observations and possibly 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-types may tend to have improved performance because the moment information obtained from the non-symmetric and/ or improperly centered distributions may be better accommodated by the differential data weights available within the EL framework However, the answer is not clear because EL may generally obtain smaller levels of bias, but at the expense of increased variance.

Finally, in pursuit of achieving finite sample reductions in mean squared error, it is useful to consider, in a 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 utlizes EL-type moments, has been proposed by Mittelhammer and Judge (2000) and is currently under evaluation.

6. References

Akkeren, M., Judge, G. and Mittelhammer, R. (2001). "Generalized Moment Based Estimation and Inference." Journal of Econometrics, In Press.

Akkeren, M., Judge, G. and Mittelhammer, R. (2001). "Coordinate Based Empirical Likelihood-Like Estimation in Ill-Conditioned Inverse Problems," under review by Journal of American Statistical Association.

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

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.

Bera, A., Bilas, Y., 2000. MM, ME, EL, Ef and GMM approaches to estimation: a synthesis, working paper, University of Illinois.

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.

Choi, I., Philips, P.C.B., 1992. Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations, Journal of Econometrics 51, 113-150.

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.

Csiszar, I., 1998. Information theoretic methods in probability and statistics, IEEE Information Theory Society Newsletter 48, 21-30.

DiCiccio, T., Hall, P., Romano, J., 1991. Empirical likelihood is bartlett-correctable, The Annals of Statistics 19, 1053-1061.

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.

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

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.

Huber, P., 1981. Robust Statistics. New York: John Wiley and Sons.

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.

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

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

Kullback, S., Leibler, R. A., 1951. On information and sufficiency, The Annals of Mathematical Statistics 22, 79-86.

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

Maydalinos, M., Kandilorou, H., 2001. Specification Analysis in Equations with Stochastic Regressors, Journal of Business and Economic Statistics, 19:226-232.

Mittelhammer, R., Judge, G., 2000. Endogeneity and biased estimation under squared error loss. To appear in Handbook of Applied Econometric and Statistical Inference, Marcel-Dekker, Inc.

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

Mittelhammer, R. and Judge, G. (2000). Robust Empirical Likelihood Estimation of Models with Non-Orthogonal Noise Components, under review in The Economic Journal.

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

Murphy, A., Van Der Vaart, A., 2000. On profile likelihood, Journal of American Statistical Association 95, 449-485.

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

Nelson, C.R., Startz, R., Zivot, E., 1998. Valid confidence intervals and inference in the presence of weak instruments, International Economic Review 39, 1119-1144.

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., McFadden, D., 1994. Large sample estimation and hypothesis testing, in: Handbook of Econometrics, Vol. 4, 2111-2241.

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

O'Sullivan, F., 1986. A statistical perspective on ill-posed inverse problems, Statistical Science 1, 502-527.

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.

Pagan, A., Robertson, J., 1997. GMM and its problems. Manuscript, Australia National University.

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.

Shen, X., Shi, J., and Wong, W., 1999. Random Sieve Likelihood and General Regression Models, Journal of American Statistical Association 94:835-846.

Smith, R.J., 1997. Alternative semi-parametric approaches to generalized method of moments estimation, Economic Journal 107, 503-519.

Staiger, D., Stock, J.H., 1997. Instrumental variables regression with weak instruments, Econometrica 65, 557-586.

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

Tauchen, G., 1986. Statistical properties of generalized method-of-moments estimators of structural parameters obtained from financial market data, Journal of Business and Economic Statistics 4, 397-425.

Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426-482.

Zivot, E., 2001. Monte Carlo designs for IV regressions with weak instruments, working paper, University of Washington.

7. Appendix

Table A.1. Empirical Parameter Estimation MSE Results

MC	N	2SLS	MEEL	MEL	MLEL	GMM(I)
1	50	0.3069	0.3733	0.3978	0.4268	0.3332
2	50	0.2284	0.2629	0.2746	0.2920	0.2396
3	50	0.3034	0.3387	0.3546	0.3800	0.3193
4	50	0.2161	0.2514	0.2681	0.2843	0.2304
5	50	0.2664	0.2633	0.2773	0.2975	0.2874
6	50	0.2048	0.2403	0.2628	0.2879	0.2277
7	50	0.1805	0.1973	0.2006	0.2130	0.1921
8	50	0.1772	0.1931	0.1988	0.2082	0.1886
9	50	0.1658	0.1749	0.1787	0.1886	0.1738
10	50	0.3328	0.3467	0.3530	0.3685	0.3455
1	100	0.1400	0.1553	0.1606	0.1649	0.1452
2	100	0.1084	0.1193	0.1211	0.1236	0.1106
3	100	0.1367	0.1468	0.1530	0.1595	0.1431
4	100	0.1084	0.1212	0.1246	0.1275	0.1114
5	100	0.1366	0.1392	0.1461	0.1523	0.1432
6	100	0.1020	0.1165	0.1213	0.1256	0.1062
7	100	0.0847	0.0891	0.0910	0.0932	0.0877
8	100	0.0819	0.0863	0.0888	0.0905	0.0849
9	100	0.0789	0.0819	0.0838	0.0856	0.0805
10	100	0.1575	0.1597	0.1637	0.1653	0.1610
1	250	0.0561	0.0591	0.0602	0.0607	0.0570
2	250	0.0439	0.0457	0.0460	0.0460	0.0442
3	250	0.0525	0.0546	0.0565	0.0574	0.0532
4	250	0.0419	0.0444	0.0445	0.0447	0.0426
5	250	0.0536	0.0552	0.0577	0.0592	0.0545
6	250	0.0421	0.0451	0.0454	0.0455	0.0428
7	250	0.0333	0.0342	0.0346	0.0347	0.0337
8	250	0.0341	0.0351	0.0355	0.0357	0.0345
9	250	0.0309	0.0315	0.0319	0.0319	0.0313
10	250	0.0616	0.0619	0.0631	0.0641	0.0620

Table A.2. Empirical Bias Results for β_2

MC	N	2SLS	MEEL	MEL	MLEL	GMM(I)
1	50	0.0594	0.0282	0.0207	0.0223	0.0409
2	50	0.0321	-0.0071	-0.0147	-0.0159	0.0141
3	50	0.1171	0.0267	0.0114	0.0084	0.0794
4	50	0.0617	-0.0197	-0.0302	-0.0324	0.0258
5	50	0.1698	0.0107	-0.0082	-0.0178	0.1148
6	50	0.0958	-0.0281	-0.0411	-0.0480	0.0390
7	50	0.0428	-0.0078	-0.0116	-0.0128	0.0148
8	50	0.0424	-0.0121	-0.0181	-0.0188	0.0149
9	50	0.0304	-0.0105	-0.0119	-0.0140	0.0202
10	50	0.0691	-0.0008	-0.0078	-0.0123	0.0486
1	100	0.0284	0.0031	-0.0019	-0.0041	0.0183
2	100	0.0170	-0.0059	-0.0056	-0.0061	0.0073
3	100	0.0550	-0.0048	-0.0138	-0.0188	0.0347
4	100	0.0304	-0.0155	-0.0165	-0.0178	0.0114
5	100	0.0774	-0.0136	-0.0232	-0.0300	0.0485
6	100	0.0465	-0.0218	-0.0238	-0.0266	0.0190
7	100	0.0184	-0.0101	-0.0106	-0.0116	0.0045
8	100	0.0223	-0.0076	-0.0086	-0.0100	0.0094
9	100	0.0130	-0.0084	-0.0088	-0.0097	0.0086
10	100	0.0401	0.0021	-0.0021	-0.0063	0.0297
1	250	0.0118	-0.0027	-0.0045	-0.0051	0.0079
2	250	0.0069	-0.0012	-0.0016	-0.0016	0.0034
3	250	0.0233	-0.0047	-0.0078	-0.0090	0.0156
4	250	0.0122	-0.0058	-0.0060	-0.0063	0.0047
5	250	0.0330	-0.0084	-0.0126	-0.0141	0.0215
6	250	0.0179	-0.0088	-0.0088	-0.0093	0.0068
7	250	0.0088	-0.0028	-0.0029	-0.0030	0.0032
8	250	0.0068	-0.0054	-0.0055	-0.0058	0.0015
9	250	0.0054	-0.0038	-0.0038	-0.0039	0.0038
10	250	0.0154	0.0004	-0.0035	-0.0047	0.0115

Table A.3. Empirical Variance Results for β_2

MC	N	2SLS	MEEL	MEL	MLEL	GMM(I)
1	50	0.1115	0.1601	0.1765	0.1945	0.1245
2	50	0.0624	0.0817	0.0880	0.0967	0.0674
3	50	0.1079	0.1445	0.1564	0.1728	0.1210
4	50	0.0597	0.0765	0.0841	0.0923	0.0659
5	50	0.0880	0.1082	0.1184	0.1309	0.1071
6	50	0.0526	0.0667	0.0755	0.0850	0.0620
7	50	0.0397	0.0481	0.0504	0.0549	0.0432
8	50	0.0417	0.0509	0.0541	0.0585	0.0453
9	50	0.0314	0.0362	0.0381	0.0411	0.0336
10	50	0.0736	0.0840	0.0876	0.0940	0.0799
1	100	0.0537	0.0668	0.0699	0.0725	0.0567
2	100	0.0302	0.0360	0.0367	0.0378	0.0310
3	100	0.0509	0.0617	0.0657	0.0693	0.0549
4	100	0.0303	0.0362	0.0376	0.0387	0.0318
5	100	0.0491	0.0559	0.0601	0.0633	0.0547
6	100	0.0287	0.0351	0.0372	0.0387	0.0306
7	100	0.0189	0.0214	0.0221	0.0224	0.0197
8	100	0.0201	0.0230	0.0236	0.0240	0.0209
9	100	0.0153	0.0169	0.0173	0.0175	0.0157
10	100	0.0360	0.0393	0.0405	0.0411	0.0379
1	250	0.0220	0.0244	0.0249	0.0252	0.0225
2	250	0.0124	0.0134	0.0134	0.0134	0.0126
3	250	0.0199	0.0221	0.0229	0.0233	0.0204
4	250	0.0120	0.0132	0.0133	0.0133	0.0123
5	250	0.0202	0.0222	0.0232	0.0237	0.0209
6	250	0.0118	0.0131	0.0132	0.0131	0.0121
7	250	0.0078	0.0082	0.0083	0.0083	0.0079
8	250	0.0081	0.0087	0.0087	0.0088	0.0083
9	250	0.0061	0.0064	0.0065	0.0064	0.0062
10	250	0.0142	0.0147	0.0150	0.0152	0.0144

Table A.4. Empirical Prediction MSE Results

MC	N	2SLS	EEL	EL	LEL	GMM(I)
1	50	25.3711	26.7912	27.2195	27.6333	25.8216
2	50	24.5685	25.4681	25.6882	25.9214	24.8738
3	50	23.3399	26.0771	26.6747	27.1182	24.3737
4	50	23.3814	25.6191	26.0293	26.2682	24.2960
5	50	20.1322	25.7492	26.6360	27.2895	22.1142
6	50	21.4563	26.0143	26.6508	27.1324	23.4958
7	50	23.5054	24.8829	25.0167	25.1715	24.2043
8	50	23.4265	24.9287	25.1363	25.2913	24.1202
9	50	23.5492	24.6442	24.7104	24.8565	23.8231
10	50	23.1224	24.9378	25.1512	25.3976	23.6654
1	100	25.1621	25.7622	25.8901	25.9828	25.3478
2	100	24.6662	25.0789	25.0917	25.1266	24.8021
3	100	24.1506	25.8104	26.1145	26.3232	24.6914
4	100	24.1844	25.4131	25.4704	25.5314	24.6612
5	100	22.6019	25.9494	26.3881	26.7099	23.6837
6	100	23.2755	25.8698	25.9859	26.1217	24.2817
7	100	24.2388	24.9887	25.0155	25.0530	24.5864
8	100	24.2069	24.9946	25.0336	25.0823	24.5284
9	100	24.3158	24.8716	24.8928	24.9247	24.4327
10	100	23.9480	24.9240	25.0467	25.1615	24.2238
1	250	25.0682	25.3038	25.3380	25.3526	25.1257
2	250	24.9025	25.0238	25.0303	25.0309	24.9504
3	250	24.6480	25.3859	25.4801	25.5173	24.8473
4	250	24.7101	25.1814	25.1880	25.1967	24.8981
5	250	23.9566	25.5121	25.6899	25.7578	24.3817
6	250	24.3023	25.3108	25.3166	25.3314	24.7117
7	250	24.6558	24.9541	24.9595	24.9622	24.7952
8	250	24.7162	25.0270	25.0314	25.0393	24.8486
9	250	24.7242	24.9576	24.9606	24.9627	24.7660
10	250	24.6250	25.0057	25.1045	25.1404	24.7242

Table A.5. Empirical Size of Moment Validity Tests, Target Size = .05

MC	N	W2SLS	WEEL	LREEL	LMEEL	WEL	LREL	LMEL	WLEL=LRLEL	LMLEL	WGMM
1	50	0.046	0.0278	0.0834	0.0514	0.0282	0.0884	0.1032	0.0204	0.0144	0.049
2	50	0.054	0.036	0.0962	0.0606	0.0344	0.1036	0.1198	0.0284	0.02	0.055
3	50	0.0654	0.0334	0.101	0.0592	0.034	0.1084	0.1222	0.0252	0.0184	0.0646
4	50	0.0616	0.039	0.1062	0.0658	0.0388	0.1142	0.1326	0.0296	0.0224	0.0586
5	50	0.1026	0.0414	0.1202	0.0736	0.0424	0.1254	0.14	0.0322	0.0284	0.088
6	50	0.101	0.0562	0.1314	0.0862	0.0546	0.1378	0.1574	0.0448	0.0426	0.084
7	50	0.0608	0.0382	0.1106	0.0694	0.04	0.1202	0.1384	0.032	0.0294	0.0556
8	50	0.0598	0.039	0.1048	0.0658	0.0414	0.1114	0.1288	0.0312	0.027	0.0574
9	50	0.065	0.0464	0.1136	0.071	0.0498	0.124	0.1414	0.0392	0.031	0.0672
10	50	0.069	0.0424	0.1118	0.0706	0.0454	0.1194	0.1382	0.0344	0.022	0.0702
1	100	0.048	0.0376	0.0704	0.0454	0.0384	0.0702	0.0694	0.0336	0.0288	0.0488
2	100	0.0466	0.0376	0.0688	0.0416	0.0394	0.068	0.0668	0.0366	0.0332	0.0482
3	100	0.0628	0.0452	0.0772	0.049	0.045	0.0756	0.0746	0.041	0.0336	0.0632
4	100	0.0606	0.0446	0.0758	0.0496	0.0456	0.0758	0.0788	0.0408	0.0374	0.0586
5	100	0.09	0.0506	0.0898	0.059	0.0504	0.086	0.0872	0.0454	0.0364	0.08
6	100	0.0752	0.0478	0.0808	0.0536	0.0482	0.081	0.0862	0.045	0.0384	0.066
7	100	0.061	0.0458	0.084	0.0522	0.0486	0.0846	0.0852	0.0438	0.0416	0.0586
8	100	0.0598	0.049	0.0824	0.0536	0.0494	0.082	0.0846	0.0452	0.042	0.0594
9	100	0.0566	0.0428	0.0812	0.0512	0.0436	0.0816	0.0826	0.0402	0.0348	0.0558
10	100	0.061	0.045	0.0802	0.05	0.0464	0.0808	0.0834	0.0424	0.0364	0.06
1	250	0.0522	0.0492	0.0634	0.0458	0.0488	0.0618	0.0574	0.0484	0.0462	0.0538
2	250	0.0498	0.0454	0.0576	0.0422	0.0456	0.0562	0.0528	0.0454	0.0436	0.049
3	250	0.0554	0.0468	0.057	0.044	0.0466	0.055	0.0514	0.0454	0.0402	0.0542
4	250	0.0546	0.0474	0.0614	0.0442	0.048	0.0578	0.055	0.047	0.0454	0.0528
5	250	0.065	0.0492	0.0618	0.0478	0.0484	0.058	0.0566	0.0478	0.0402	0.0614
6	250	0.0608	0.0502	0.0624	0.0434	0.0504	0.0598	0.0558	0.0494	0.045	0.0568
7	250	0.0536	0.049	0.0604	0.042	0.049	0.06	0.0548	0.0482	0.047	0.0528
8	250	0.0508	0.046	0.057	0.0422	0.0466	0.0548	0.053	0.0458	0.045	0.0498
9	250	0.0538	0.049	0.0624	0.0446	0.0498	0.0598	0.0572	0.0486	0.047	0.0546
10	250	0.0522	0.0466	0.058	0.0432	0.0466	0.0554	0.0534	0.0458	0.0434	0.0516

A6. Empirical Confidence Interval Coverage Probability for B2, Target = .99

MC	N	2SLS	EEL	EL	LEL	GMM(I)
1	50	0.9896	0.955	0.9568	0.9604	0.9926
2	50	0.9888	0.9548	0.9586	0.9592	0.9922
3	50	0.958	0.9214	0.9266	0.9276	0.9654
4	50	0.9624	0.9432	0.9464	0.9456	0.9716
5	50	0.8824	0.9122	0.9176	0.915	0.9046
6	50	0.9234	0.9418	0.9462	0.9402	0.9468
7	50	0.972	0.956	0.958	0.9596	0.9776
8	50	0.9722	0.9504	0.9546	0.9544	0.9784
9	50	0.972	0.956	0.956	0.9576	0.9758
10	50	0.9576	0.9422	0.9482	0.9466	0.9628
1	100	0.9902	0.9762	0.9764	0.976	0.9902
2	100	0.9886	0.9724	0.9724	0.9734	0.9894
3	100	0.9664	0.9578	0.9604	0.962	0.9728
4	100	0.9768	0.972	0.972	0.9726	0.9808
5	100	0.9356	0.9564	0.9596	0.9598	0.9482
6	100	0.953	0.9642	0.966	0.9662	0.9658
7	100	0.9798	0.9754	0.9746	0.9764	0.9852
8	100	0.9818	0.978	0.9776	0.9786	0.9854
9	100	0.9808	0.9724	0.9738	0.9746	0.9834
10	100	0.9712	0.9686	0.9698	0.9722	0.9728
1	250	0.9894	0.9856	0.9858	0.9866	0.9896
2	250	0.9902	0.9862	0.9864	0.988	0.9914
3	250	0.9812	0.9832	0.983	0.9834	0.9826
4	250	0.9844	0.984	0.9838	0.9842	0.9856
5	250	0.9662	0.9782	0.9788	0.9792	0.9708
6	250	0.9748	0.9838	0.9826	0.9844	0.9792
7	250	0.9832	0.9842	0.9846	0.9846	0.986
8	250	0.989	0.9874	0.9872	0.9874	0.99
9	250	0.9842	0.9844	0.9842	0.9842	0.9858
10	250	0.9796	0.98	0.9814	0.981	0.9814

Table A.7. Empirical Expected Confidence Interval Length for B2

MC	N	2SLS	EEL	EL	LEL	GMM(I)
1	50	1.7359	1.5771	1.6270	1.6857	1.8332
2	50	1.3059	1.1732	1.2051	1.2362	1.3490
3	50	1.6847	1.5748	1.6372	1.6954	1.8023
4	50	1.2722	1.1956	1.2365	1.2479	1.3319
5	50	1.5741	1.6160	1.7016	1.7150	1.7205
6	50	1.2255	1.2453	1.2990	1.2848	1.3170
7	50	1.0390	0.9536	0.9760	0.9906	1.0756
8	50	1.0689	0.9875	1.0141	1.0277	1.1081
9	50	0.9195	0.8364	0.8539	0.8691	0.9463
10	50	1.4002	1.2912	1.3257	1.3548	1.4514
1	100	1.2108	1.1526	1.1694	1.1798	1.2428
2	100	0.8988	0.8552	0.8643	0.8704	0.9130
3	100	1.1828	1.1583	1.1836	1.1859	1.2237
4	100	0.8928	0.8730	0.8842	0.8834	0.9135
5	100	1.1592	1.1934	1.2264	1.2105	1.2124
6	100	0.8818	0.8969	0.9103	0.8988	0.9123
7	100	0.7274	0.6987	0.7072	0.7086	0.7398
8	100	0.7459	0.7169	0.7256	0.7279	0.7586
9	100	0.6413	0.6121	0.6189	0.6219	0.6505
10	100	0.9785	0.9436	0.9581	0.9627	0.9969
1	250	0.7580	0.7441	0.7487	0.7494	0.7656
2	250	0.5645	0.5523	0.5552	0.5558	0.5679
3	250	0.7506	0.7487	0.7552	0.7518	0.7601
4	250	0.5622	0.5567	0.5596	0.5583	0.5672
5	250	0.7431	0.7583	0.7674	0.7577	0.7560
6	250	0.5590	0.5628	0.5657	0.5615	0.5665
7	250	0.4555	0.4490	0.4513	0.4509	0.4585
8	250	0.4705	0.4631	0.4655	0.4652	0.4738
9	250	0.4016	0.3942	0.3962	0.3961	0.4039
10	250	0.6182	0.6098	0.6152	0.6143	0.6227

A.8. Empirical Power for Testing Ho: B2 = 0

MC	N	2SLS	EEL	EL	LEL	GMM(I)
1	50	0.967	0.9656	0.9576	0.9508	0.9566
2	50	0.995	0.9932	0.9898	0.9878	0.9934
3	50	0.9588	0.9506	0.9402	0.9304	0.9444
4	50	0.9888	0.9844	0.9794	0.9768	0.9834
5	50	0.9592	0.9456	0.9258	0.9168	0.9408
6	50	0.9872	0.9772	0.9708	0.9664	0.9778
7	50	0.9984	0.9982	0.9976	0.9972	0.9976
8	50	0.9978	0.9958	0.9938	0.993	0.9964
9	50	0.9998	0.9996	0.9996	0.9996	0.9998
10	50	0.985	0.9836	0.9802	0.9752	0.9818
1	100	0.9968	0.9966	0.9962	0.9954	0.9958
2	100	1	1	1	1	1
3	100	0.9938	0.991	0.987	0.9866	0.99
4	100	1	0.9996	0.9996	0.9996	1
5	100	0.9902	0.9852	0.982	0.9806	0.985
6	100	0.999	0.9982	0.9978	0.9978	0.9988
7	100	1	1	1	1	1
8	100	1	1	1	1	1
9	100	1	1	1	1	1
10	100	0.9992	0.9994	0.9994	0.9992	0.999
1	250	1	1	1	1	1
2	250	1	1	1	1	1
3	250	1	1	1	1	1
4	250	1	1	1	1	1
5	250	1	0.9998	0.9998	0.9998	1
6	250	1	1	1	1	1
7	250	1	1	1	1	1
8	250	1	1	1	1	1
9	250	1	1	1	1	1
10	250	1	1	1	1	1