EMPIRICAL LIKELIHOOD ESTIMATORS

OF THE

LINEAR SIMULTANEOUS EQUATIONS MODEL

Thomas L. Marsh[a], Ron C. Mittelhammer[b], and George G. Judge[c]

Selected Paper
American Agricultural Economics Association
Chicago, August 2001

[a]Assistant Professor, 342 Waters Hall, Kansas State University, Manhattan, KS, 66506, 785-532-4913, tlmarsh@agecon.ksu.edu.; [b]Professor, Washington State University, and [c]Professor, University of California, Berkeley.

Empirical Likelihood Estimators of the Linear Simultaneous Equations Model

Abstract: Information theoretic estimators are specified for a system of linear simultaneous equations, including maximum empirical likelihood, maximum empirical exponential likelihood, and maximum log Euclidean likelihood. Monte Carlo experiments are used to compare finite sample performance of these estimators to traditional generalized method of moments.

Keywords: endogeneity, mean square error, system of simultaneous equations, information theoretic estimators

1. Introduction

It is becoming increasingly evident that asymptotically justified estimators can have severe performance limitations in finite sample estimation, especially for economic data sampling processes that involve endogenously and simultaneously determined random variables (Zellner; van Akkeren, Judge, and Mittelhammer; Mittelhammer and Judge). For example the traditional two and three stage least squares estimators are consistent and asymptotically normally distributed, but have no optimality justification for small sample estimation except in very restrictive sampling contexts. Moreover, there is increasing evidence that traditional asymptotically efficient moment based estimators may have large biases for the relatively small sample sizes usually encountered in applied economic research (Newey and Smith 2000). In response to limitations of traditional approaches for small sample estimation, we investigate alternative empirical likelihood-type estimators of the linear simultaneous system of equations and their performance in relatively small finite samples.

Empirical likelihood-type estimators have been suggested in various forms as alternatives to traditional estimators [Owen, 1988, 1991, 2000; Qin and Lawless; Kitamura and Stutzer; Imbens, Spady, and Johnson; Mittelhammer, Judge and Miller]. Empirical likelihood estimators do not require specification of the specific parametric functional form of likelihood functions, but rather make mild assumptions concerning the existence of certain zero-valued moment conditions. To date, there has been only limited analysis of small- and medium-sized sample performance of these estimators. Imbens, Spady, and Johnson investigated the properties of point estimators and hypothesis testing procedures in the context of single parameter models. Mittelhammer and Judge examined single equations models when the orthogonality condition between explanatory variables and equation noise is not fulfilled. Finite sample properties of empirical likelihood-type estimators have yet to be analyzed rigorously within a simultaneous systems context.

In this paper we examine the performance of three different Empirical Likelihood (*EL*) estimators within a linear simultaneous systems framework. These include the Maximum Empirical Likelihood (*MEL*), Maximum Empirical Exponential Likelihood (*MEEL*), and Maximum Log Euclidean Likelihood (*MLEL*) estimators. To evaluate the performance of the *EL* type estimators over a range of finite sample sizes, Monte Carlo sampling experiments are performed for a system of three simultaneous equations. The mean square error between the true and estimated values of model parameters is used to compare the finite sample performance of the various *EL* estimators, as well as their performance relative to the generalized method of moments estimation procedure. In addition, Monte Carlo experiments are used to compare the size and power of asymptotic normal *Z* and asymptotic chi-square Wald tests. Although these results are specific to the collection of particular Monte Carlo experiments analyzed, the sampling evidence reported does provide an indication of the relative performance among the estimators.

2. Empirical Likelihood Estimators

Consider the ith equation of a system of q linear simultaneous equations

$$\mathbf{Y}_i = \mathbf{Y}_{(i)}\boldsymbol{\gamma}_i + \mathbf{X}_{(i)}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i = \mathbf{M}_{(i)}\boldsymbol{\delta}_i + \boldsymbol{\varepsilon}_i \text{ for } i = 1,\ldots,q$$

where $\mathbf{Y}_{\mathrm{i}}$ is a $n\times 1$ vector of endogenous variables, and $\mathbf{Y}_{(i)}$ and $\mathbf{X}_{(i)}$ represent the $(n\times q_i)$ matrix of endogenous and $(n\times k_i)$ matrix of predetermined explanatory variables, respectively. The $(n\times 1)$ vector $\boldsymbol{\varepsilon}_i$ represents the unobserved residuals for the ith equation. The parameters to be estimated include the $(q_i\times 1)$ vector $\boldsymbol{\gamma}_i$ associated with the right hand side endogenous variables and the $(k_i\times 1)$ vector $\boldsymbol{\beta}_i$ associated with the predetermined variables. The structural parameters are combined into the $((q_i+k_i)\times 1)$ vector $\boldsymbol{\delta}_i = [\boldsymbol{\gamma}_i{}'\,|\,\boldsymbol{\beta}_i{}']'$.

In the event one or more of the regressors is correlated with the equation noise, then $\mathrm{E}\left[\mathrm{n}^{-1}\mathbf{M}_{(\mathrm{i})}\boldsymbol{\varepsilon}_{\mathrm{i}}\right] \neq \mathbf{0}$ or $\mathrm{plim}\left[\mathrm{n}^{-1}\mathbf{M}_{(\mathrm{i})}\boldsymbol{\varepsilon}_{\mathrm{i}}\right] \neq \mathbf{0}$ and traditional Gauss-Markov procedures such as the least squares (LS) estimator, or equivalently the method of moments (*MOM*) -extremum estimator $\hat{\boldsymbol{\delta}}_{\mathrm{i,mom}} = \arg_{\beta\in \mathrm{B}}\left[\mathrm{n}^{-1}\mathbf{X}'\left(\mathbf{Y} - \mathbf{M}_{(\mathrm{i})}\boldsymbol{\delta}_{(\mathrm{i})}\right) = \mathbf{0}\right]$, are biased and inconsistent, with unconditional expectation and probability limit such that $\mathrm{E}\left[\hat{\boldsymbol{\delta}}_{\mathrm{i}}\right] \neq \boldsymbol{\delta}_{\mathrm{i}}$ and $\mathrm{plim}\left[\hat{\boldsymbol{\delta}}_{i}\right] \neq \boldsymbol{\delta}_{i}$. For a complete system of simultaneous equations a consistent generalized method of moments (*GMM*) estimator can be derived from

$$\left[\mathrm{n}^{-1}\left(\mathbf{I}_{\mathrm{q}} \otimes \mathbf{Z}'\right)\right]\left[\mathbf{Y}_{\mathrm{v}} - \mathbf{M}\boldsymbol{\delta}\right] = \mathbf{0}$$

where $\mathbf{Y}_{v} = vec\left(\mathbf{Y}_{1},...,\mathbf{Y}_{q}\right)$ is a $(nq\times 1)$ vector of vertically concatenated endogenous variables, $\mathbf{Z}$ is a $(n\times m)$ matrix of instrumental variables, $\mathbf{M}$ is a block diagonal matrix whose ith block is given by $\mathbf{M}_{(\mathrm{i})}$ and $\boldsymbol{\delta} = vec\left(\boldsymbol{\delta}_{1},...,\boldsymbol{\delta}_{q}\right)$ is a $(K\times 1)$ vector of structural parameters to be estimated. Here $K = \sum_{i=1}^{q}(q_i + k_i)$ is the total number of endogenous and predetermined structural parameters in the system. If **Z**=**X** then the *GMM* estimator is equivalent to three stage least squares (*3SLS*). Hansen and Hansen, Heaton, and Yaron provide details on large and finite properties of *GMM* estimators.

In contrast to the *GMM* approach, empirical moment conditions for *EL* type estimators are expressed in the form

$$\left[\mathbf{I}_{\mathrm{q}} \otimes \left(\mathbf{p} \odot \mathbf{Z}\right)'\right]\left[\mathbf{Y}_{\mathrm{v}} - \mathbf{M}\boldsymbol{\delta}\right] = \mathbf{0}$$

where the unknown $(n\times 1)$ vector $\mathbf{p}$ consists of an empirical probability distribution supported on the sample outcomes, and $\odot$ denotes the extended Hadamard (elementwise) product operator. Comparing the two moment conditions it is evident that the *GMM* approach restricts $p_i = 1/n$ for $i = 1,...,n$, while the *EL* approach treats the unknown $p_i's$ as parameters to be estimated. Note that although we are currently examining a linear system of equations , the nonlinear equivalent follows with only slight modifications in notation.

The extremum problem for information theoretic estimation can be formulated as

$$\max_{\mathbf{p},\boldsymbol{\delta}}\left\{\phi(\mathbf{p}) \text{ s.t. } \left[\mathbf{I}_q \otimes (\mathbf{p} \odot \mathbf{Z})'\right]\left[\mathbf{Y}_v - \mathbf{M}\boldsymbol{\delta}\right] = \mathbf{0},\ \sum_{i=1}^{n} p_i = 1, p_i \geq 0\, \forall i\right\}$$

which maximizes the objective function $\phi(\mathbf{p})$ subject to moment, normalization, and nonnegativity constraints. The different objective functions considered for the functional specification of $\phi(\mathbf{p})$ include the traditional empirical log-likelihood objective function $\sum_{i=1}^{n} \ln(p_i)$, the empirical exponential likelihood (or negative entropy) function $\sum_{i=1}^{n} p_i \ln(p_i)$, and the log Euclidean likelihood function $n^{-1}\left(\sum_{i=1}^{n}\left(n^2 p_i^2 - 1\right)\right)$. Each specification leads to a uniquely defined estimator of $\boldsymbol{\delta}$. These estimating criteria are nested within the Cressie-Read power divergence statistic that is based on the concept of closeness between estimated and empirical distributions relating to the choice of **p**-distributions. The Cressie-Read statistic is discussed further in Cressie and Read, Read and Cressie, and Baggerly.

The Lagrangian form of the extremum problem is given by

$$L(\mathbf{p},\boldsymbol{\delta},\boldsymbol{\lambda},\eta) = \phi(\mathbf{p}) - \sum_{i=1}^{q} \boldsymbol{\lambda}_i'\left[(\mathbf{p} \odot \mathbf{Z})'\left(\mathbf{Y}_{\cdot i} - \mathbf{M}_{(i)}\boldsymbol{\delta}_i\right)\right] - \eta\left(\sum_{i=1}^{n} p_i - 1\right)$$

where $\boldsymbol{\lambda} = vec(\boldsymbol{\lambda}_1,...,\boldsymbol{\lambda}_q)$ is a $(mq \times 1)$ vector and η is a (1×1) scalar set of Lagrange multipliers. First order conditions are given by

$$\frac{\partial L}{\partial p_j} = \frac{\partial \phi(\mathbf{p})}{\partial p_j} - \sum_{i=1}^{q} \boldsymbol{\lambda}_i' \left[\mathbf{Z}_{j\cdot}' \left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot] \boldsymbol{\delta}_i \right) \right] - \eta = 0$$

$$\frac{\partial L}{\partial \boldsymbol{\delta}_{i\ell}} = \boldsymbol{\lambda}_i' (\mathbf{p} \odot \mathbf{Z})' \mathbf{M}_{(i)}[.,\ell] = 0$$

$$\frac{\partial L}{\partial \boldsymbol{\lambda}_i} = -(\mathbf{p} \odot \mathbf{Z})' \left(\mathbf{Y}_{\cdot i} - \mathbf{M}_{(i)} \boldsymbol{\delta}_i \right) = [\mathbf{0}]$$

$$\frac{\partial L}{\partial \eta} = \sum_{i=1}^{n} p_i - 1 = 0$$

and $p_j \geq 0, \forall j$. The first set of equations links the unknown $p_i's$ to the other unknown parameters $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$ through the empirical moment conditions. The second and third sets of equations relax traditional orthogonality conditions required by two and three stage least squares. The fourth equation is the required normalization condition for the empirically estimated probability weights. Provided $f(\mathbf{p}) = \frac{\partial \phi(\mathbf{p})}{\partial p_j}$ is a monotonic function, then an inverse function, $f^{-1}(\cdot)$, exists and the general solution for $\mathbf{p}$ is

$$\mathbf{p} = f^{-1}\left[\sum_{i=1}^{q} \boldsymbol{\lambda}_i' \left[\mathbf{Z}_{j\cdot}' \left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot] \boldsymbol{\delta}_i \right) \right] + \eta \right]$$

For the three distinct objective functions identified above, three separate econometric estimators are derived below.

2.1 Maximum Empirical Likelihood

The empirical log-likelihood objective function, $\phi(\mathbf{p}) = \sum_{i=1}^{n} \ln(p_i)$, yields the Maximum Empirical Likelihood (*MEL*) estimate of $\boldsymbol{\delta}$. The first order condition with respect to p_j is given by

$$\frac{\partial L}{\partial p_j} = \frac{1}{n p_j} - \sum_{i=1}^{q} \boldsymbol{\lambda}_i' \left[\mathbf{Z}_{j\cdot}' \left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot] \boldsymbol{\delta}_i \right) \right] - \eta = 0$$

The optimal p_j can be expressed as (note it can be shown that $\eta = 1$ at the optimal solution)

$$p_j(\boldsymbol{\lambda},\boldsymbol{\delta}) = \left[n\sum_{i=1}^{q} \boldsymbol{\lambda}_i' \mathbf{Z}_{j\cdot}'\left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot]\boldsymbol{\delta}_i\right) + n \right]^{-1}.$$

Concentrating the objective function by substituting $p_j(\boldsymbol{\lambda},\boldsymbol{\delta})$ for p_j generates a system of $(K+mq)$ first order conditions and $(K+mq)$ unknowns represented by $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$. This leads to a conventional empirical likelihood estimator of the linear simultaneous equations model.

2.2 Maximum Empirical Exponential Likelihood

The empirical exponential likelihood function, $\phi(\mathbf{p}) = \sum_{i=1}^{n} p_i \ln(p_i)$, leads to the Maximum Empirical Exponential Likelihood (*MEEL*) estimate of $\boldsymbol{\delta}$. The first order condition with respect to p_j is given by

$$\frac{\partial L}{\partial p_j} = 1 + \ln(p_j) - \sum_{i=1}^{q} \boldsymbol{\lambda}_i' \left[\mathbf{Z}_{j\cdot}'\left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot]\boldsymbol{\delta}_i\right)\right] - \eta = 0$$

The optimal p_j can be expressed as

$$p_j(\boldsymbol{\lambda},\boldsymbol{\delta}) = \frac{\exp\left(\sum_{i=1}^{q} \boldsymbol{\lambda}_i' \left[\mathbf{Z}_{j\cdot}'\left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot]\boldsymbol{\delta}_i\right)\right]\right)}{\sum_{j=1}^{n} \exp\left(\sum_{i=1}^{q} \boldsymbol{\lambda}_i' \left[\mathbf{Z}_{j\cdot}'\left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot]\boldsymbol{\delta}_i\right)\right]\right)}$$

Concentrating the objective function by substituting $p_j(\boldsymbol{\lambda},\boldsymbol{\delta})$ for p_j yields a system of $(K+mq)$ first order conditions and $(K+mq)$ unknowns represented by $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$. For further insight into the *MEEL* estimator see Mittelhammer, Judge, and Miller (Chapter 17).

The *MEEL* estimator has similarities to generalized maximum entropy estimators proposed by Golan, Judge, and Miller in that it uses the same functional form of objective function. However, the *MEEL* estimator is fundamentally different from generalized maximum entropy estimators of the linear simultaneous equations model. *MEEL* does not utilize user supplied support spaces for the parameters and error terms as do generalized maximum entropy estimators, but rather recovers the unknown structural parameters $\boldsymbol{\delta}$ and empirically estimated probability weights **p** supported on the sample outcomes. See Marsh, Mittelhammer, and Cardell for a generalized maximum entropy analysis of the linear simultaneous equations model.

2.3 Maximum Log Euclidean Likelihood

The log Euclidean likelihood function $\phi(\mathbf{p}) = n^{-1}\left(\sum_{i=1}^{n}\left(n^2 \mathrm{p}_i^2 - 1\right)\right)$ yields the Maximum Log Euclidean Likelihood (*MLEL*) estimate of $\boldsymbol{\delta}$. The first order condition with respect to p_j is now

$$\frac{\partial \mathrm{L}}{\partial \mathrm{p}_{\mathrm{j}}} = -2\mathrm{np}_{\mathrm{j}} - \sum_{\mathrm{i}=1}^{\mathrm{q}} \boldsymbol{\lambda}_{\mathrm{i}}^{'}\left[\mathbf{Z}_{\mathrm{j}\cdot}{}'\left(\mathbf{Y}_{\mathrm{ji}} - \mathbf{M}_{(\mathrm{i})}[\mathrm{j},\cdot]\boldsymbol{\delta}_{\mathrm{i}}\right)\right] - \eta = 0\,.$$

The optimal $\mathrm{p_j}$ can be expressed as

$$\mathrm{p}_{\mathrm{j}}(\boldsymbol{\lambda},\boldsymbol{\delta}) = (2\mathrm{n})^{-1}\left[\boldsymbol{\lambda}_{\mathrm{i}}^{'}\mathbf{Z}_{\mathrm{j}\cdot}{}'\left(\mathbf{Y}_{\mathrm{ji}} - \mathbf{M}_{(\mathrm{i})}[\mathrm{j},\cdot]\boldsymbol{\delta}_{\mathrm{i}}\right) + \eta\right].$$

Again concentrating the objective function by substituting $\mathrm{p}_{\mathrm{j}}(\boldsymbol{\lambda},\boldsymbol{\delta})$ for p_j yields a system of $(K+mq)$ first order conditions and $(K+mq)$ unknowns represented by $\boldsymbol{\delta}$ and $\boldsymbol{\lambda}$. Of the three specifications considered in this study, the *MLEL* estimator has received the least attention in the econometrics and statistics literature.

3. Asymptotic Properties

The *MEL*, *MEEL*, and *MLEL* estimators are all consistent, asymptotically normally distributed, and asymptotically efficient relative to the optimal estimating function estimator (Imbens, Spady, and Johnson; Kitamura and Stutzer; Mittelhammer, Judge, and Miller, Chapter 17). The estimators are asymptotically distributed as $n^{1/2}\left(\boldsymbol{\delta}^{\ell} - \boldsymbol{\delta}\right) \xrightarrow{d} \mathrm{N}(\mathbf{0},\boldsymbol{\Omega})$ where the index ℓ represents the specific *EL* estimator $\ell \in \{MEL, MEEL, MLEL\}$. For iid sampling the asymptotic covariance matrix $\boldsymbol{\Omega}$ is defined as

$$\mathbf{\Omega} = \left(\text{plim}\left[n^{-1} \frac{\partial \mathbf{h}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} \right]' \text{plim}\left[n^{-1} \sum_{j=1}^{n} \mathbf{h}_j(\boldsymbol{\delta}) \mathbf{h}_j(\boldsymbol{\delta})' \right] \text{plim}\left[n^{-1} \frac{\partial \mathbf{h}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}'} \right] \right)^{-1}$$

In the expression above

$$\mathbf{h}_j(\boldsymbol{\delta}) = \left(\begin{bmatrix} Y_{j1} - \mathbf{M}_{(1)}[j,\cdot]\boldsymbol{\delta}_1 \\ \vdots \\ Y_{jq} - \mathbf{M}_{(q)}[j,\cdot]\boldsymbol{\delta}_q \end{bmatrix} \otimes \mathbf{Z}[j,.]' \right)$$

and

$$\mathbf{h}(\boldsymbol{\delta}) = \sum_{j=1}^{n} \mathbf{h}_j(\boldsymbol{\delta}) = \left[\mathbf{I}_q \otimes \mathbf{Z}' \right] \left[\mathbf{Y}_v - \mathbf{M}\boldsymbol{\delta} \right] .$$

See Imbens, Spady, and Johnson, as well as Kitamura and Stutzer, for underlying assumptions and proof of consistency and asymptotic normality. In the case of non-iid sampling, Kitamura and Stutzer extend the above covariance expression.

3.1 Hypothesis Testing

Since the *EL* estimators are consistent and asymptotically normally distributed, asymptotically valid normal and chi-square test statistics can be used to test hypotheses about $\boldsymbol{\delta}$. Consistent estimates of asymptotic covariance matrices can be constructed by using *MEL*, *MEEL*, and *MLEL* estimates of $\boldsymbol{\delta}$, respectively. The plim terms can be based on either sample averages or else expectations can be taken with respect to the estimated $\hat{\mathbf{p}}$ distributions, effectively replacing the n-divisors in the preceding expression defining $\mathbf{\Omega}$ with non-uniform $\hat{\mathrm{p}}_{\mathrm{j}}$ weights applied to each sample observation, respectively.

Because $Z = \dfrac{\hat{\delta}_{ij} - \delta_{ij}^0}{\sqrt{\hat{\Omega}_{ii}}}$ is asymptotically N(0,1) under the null hypothesis H_0: $\delta_{ij} = \delta_{ij}^0$, the statistic *Z* can be used to test hypothesis about the values of the δ_{ij}'s.

To define Wald tests on the elements of $\boldsymbol{\delta}$, let $\mathrm{H_0}\text{: } \mathbf{R}(\boldsymbol{\delta}) = [\mathbf{0}]$ be the null hypothesis to be tested. Here $\mathbf{R}(\boldsymbol{\delta})$ is a continuously differentiable L-dimensional vector function with $\operatorname{rank}\left(\frac{\partial \mathbf{R}(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}\right) = L \leq K$. In the special case of a linear null hypothesis, $\mathrm{H_0}\text{: } \mathbf{R}\boldsymbol{\delta} = \mathbf{r}$. Then the Wald test statistic has a χ^2 limiting distribution with L degrees of freedom under $\mathrm{H_0}$, where the statistic is defined by

$$W = \left(\mathbf{R}(\boldsymbol{\delta}) - \mathbf{r}\right)' \left[\frac{\partial \mathbf{R}}{\partial \boldsymbol{\delta}}' \, \boldsymbol{\Omega} \frac{\partial \mathbf{R}}{\partial \boldsymbol{\delta}}\right]^{-1} \left(\mathbf{R}(\boldsymbol{\delta}) - \mathbf{r}\right) \sim \chi^2_L .$$

Imbens, Spady, and Johnson and Kitamura and Stutzer provide further details on Lagrange multiplier and pseudo-likelihood ratio hypothesis testing procedures. Mittelhammer, Judge, and Miller discuss specification and application of empirical likelihood ratio tests.

4. Finite Sample Properties

The derivation of the finite sample properties of the *EL* estimators presented above is not tractable. Hence, Monte Carlo sampling experiments are used to identify and compare the repeated sampling properties of the estimators. In this study we attempt to focus on small-to-medium sample size performance of the *EL* estimators, and their relative performance to *3SLS*. To measure the performance of the estimators, we use the mean square error (*MSE*) between the true and estimated values of structural coefficients. Moreover, rejection probabilities of true and false hypothesis are used to estimate the size and power of statistical tests.

4.1 Monte Carlo Experiments

For the sampling experiments we specified an overdetermined simultaneous system with contemporaneously correlated errors that is similar to empirical models discussed in Cragg, Tsurumi, and Marsh, Mittelhammer, and Cardell. The structural parameters $\boldsymbol{\Gamma}$ and $\mathbf{B}$ of the system in (1) are given as

$$\boldsymbol{\Gamma} = \begin{pmatrix} -1 & .267 & .087 \\ .222 & -1 & 0 \\ 0 & .046 & -1 \end{pmatrix} \quad \mathbf{B}' = \begin{pmatrix} 6.2 & 0 & .7 & 0 & .96 & 0 & .06 \\ 4.4 & .74 & 0 & 0 & .13 & 0 & 0 \\ 4 & 0 & .53 & .11 & 0 & .56 & 0 \end{pmatrix} \quad .$$

The disturbance outcomes are drawn from a multivariate normal distribution with mean zero and covariance $\boldsymbol{\Sigma} \otimes \mathbf{I}$. The contemporaneous error covariance specification $\boldsymbol{\Sigma}$ is given by

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & -1 & -.125 \\ -1 & 4 & .0625 \\ -.125 & .0625 & 8 \end{pmatrix}.$$

The exogenous variable values are all drawn independently from a $N(0,1)$ distribution. Results for each estimator were obtained by solving by the respective first order conditions defined previously. In particular, the *EL* type solutions were calculated using the GAUSS constrained optimization application module provided by Aptech Systems, Maple Valley, WA.

4.2 Results: Point Estimates

Table 1 contains the mean values of the distribution of estimated $\boldsymbol{\Gamma}$ parameters based on 1000 Monte Carlo repetitions for sample sizes of 50, 100, and 200 observations per equation. From this information we can infer several implications as to the performance of the *EL* estimators. At 50 observations all three of the EL-type estimates are quite similar across all coefficients. As the observations increased from 50 to 200, the *MEEL*, *MEL* and *MLEL* estimators appear to be converging to the true parameter values. Hence, *MEEL*, *MEL*, *MLEL*, and *3SLS* all exhibit the property of consistency across the given sample sizes, as anticipated from asymptotic theory.

Table 2 contains standard errors of the distributions of estimated **Γ** parameters based on 1000 Monte Carlo repetitions for sample sizes of 50, 100, and 200 observations per equation. Across the sample sizes, *MEL* and *MEEL* exhibited smaller standard errors for the structural parameters than did *MLEL*. Overall, the standard errors from *3SLS* are larger than those for *MEL* and *MEEL* for three of the four parameters. As the observations increase from 50 to 200, standard errors for each estimator appear to be converging towards one another, which is expected for asymptotically equivalent estimators.

In Table 3 the mean square error between the true and estimated structural parameters **Γ** are reported based on 1000 Monte Carlo repetitions for sample sizes of 50, 100, and 200 observations per equation. *MEL* and *MEEL* exhibited smaller *MSE* for the structural parameters than did *MLEL*. The *MSE* from *3SLS* is larger than those for *MEL* and *MEEL* except for one parameter. As the sample size increases from 50 to 200, each estimator shows a decrease in *MSE* values for all four structural parameters, as expected.

In all, the *MEL* and *MEEL* estimators outperformed *3SLS* in *MSE*. Further, *MEL* and *MEEL* estimators outperformed *MLEL* in *MSE*. These results are for the most part encouraging and suggest need for additional finite sample analysis of the *EL* estimators considered in this study.

4.3. Results: Hypothesis Testing

To investigate the size of the asymptotically normal *Z* test, single scalar hypotheses of the form H_0: $\gamma_{ij} = \gamma_{ij}^0$ are tested with γ_{ij}^0 set equal to the true values of the structural parameters. Critical values of the tests are based on a normal distribution with a .05 level of significance. To complement this analysis, we investigated the size and power of a joint hypothesis H_0: $\gamma_{21} = \gamma_{21}^0, \gamma_{31} = \gamma_{31}^0$ using the Wald test with a .05 level of significance, where the critical value is based on a central chi-square distribution with 2 degrees of freedom. Again the scenarios are analyzed using 1000 Monte Carlo repetitions for sample sizes of 50, 100, and 200 per equation.

Table 4 contains rejection probabilities relating to true scalar null hypotheses for structural parameters and for the asymptotic normal *Z* tests based on the *MEEL*, *MEL*, *MLEL*, and *3SLS* estimators. These values estimate the size of the test statistic and should approach .05 with increasing sample size. For example, consider the scalar hypothesis test for the parameter γ_{21}=.222. At 50 observations the size for *3SLS* is 0.0809 and for MEEL it is 0.1260, while at 200 observations the size for *3SLS* is 0.0600 and for *MEEL* it is 0.0550. Overall, 3SLS appeared to estimate the size of the *Z* test statistic better than the other estimators at 50 observations and MEEL performed best at 200 observations.

Table 5 contains rejection probabilities for true and false joint hypotheses based on the asymptotic chi-square Wald test for the three EL-type estimators and the 3SLS estimator. For the joint hypothesis tests, the size and power values are similar across the estimators. The test sizes approached .05 and the values of the power of the tests approached 1 with increasing sample sizes. Overall, the results indicate that asymptotic test properties based on *EL* estimators do not dominate, nor are they dominated by, *3SLS*.

5. Summary and Conclusions

Three information theoretic estimators for the linear simultaneous equations model were specified, including Maximum Empirical Likelihood (*MEL*), Maximum Exponential Empirical Likelihood (*MEEL*), and Maximum Log Euclidean Likelihood (*MLEL*). Asymptotic properties and hypothesis testing techniques were identified and discussed for each estimator. To evaluate the performance of the empirical log-likelihood type estimators over a range of finite sample sizes, Monte Carlo sampling experiments were performed for a linear system of three simultaneous equations. Their relative performance was assessed, and also compared to the traditional asymptotically optimal generalized method of moment estimator (three stage least squares).

In the Monte Carlo experiments completed, the *MEL* and *MEEL* estimators outperformed *3SLS* in means square error (*MSE*) between the true and estimated structural coefficients of the endogenous variables for smaller sample sizes. The *MEEL* and *MEL* estimators outperformed *MLEL* across all sample sizes. The *MEL*, *MEEL*, *MLEL*, and *3SLS* estimators appeared to be converging with increasing sample sizes, exhibiting consistency and asymptotic efficiency. The Monte Carlo results also indicated that performance of asymptotic normal Z and chi-square Wald tests based on *EL* estimators do not dominate, nor are they dominated by, *3SLS*.

The findings of this study are encouraging regarding the use of alternative estimators to the traditional GMM procedure in finite samples when estimating the parameters of a system of simultaneous structural equations and suggest the need for additional finite sample analysis of *EL* –type estimators. Future areas of research include the performance of the estimators under alternative stochastic assumptions than normal iid sampling and the investigation of the finite sample performance of alternative test statistics, such as the Lagrange multiplier and pseudo-likelihood ratio statistics.

References

Baggerly, K.A., 1998. Empirical Likelihood as a Goodness of Fit Measure. Biometrika, 85: 535-547.

Cressie, N. and Read, T., 1984. Multinomial Goodness of Fit Tests. Journal of Royal Statistical Society of Series B 46:440-464.

Cragg, J. G. 1967. "On the Relative Small Sample Properties of Several Structural-Equation Estimators." Econometrica, 35.

Golan, A., Judge, G. G. and 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., and Yaron, A., 1996. Finite sample properties of some alternative GMM estimators. Journal of Business and Economic Statistics 14: 262-280.

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

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

Marsh, T. L., R. C. Mittelhammer, and N. S. Cardell. 2000. "A Generalized Maximum Entropy Estimator of the Simultaneous Linear Statistical Model." Working paper, Kansas State University.

Mittelhammer, R., and Judge, G. 2001. Robust Empirical Exponential Likelihood Estimation of Models with Non-Orthogonal Noise Components. Working paper, Washington State University.

Mittelhammer, R., Judge, G. and Miller, D., 2000. Econometric Foundations. New York: Cambridge University Press.

Newey, W. K. and 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., 2000. Empirical Likelihood. New York: Chapman and Hall.

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

Read, T.R. and Cressie, N.A., 1988. Goodness of Fit Statistics for Discrete Multivariate Data. New York: Springer Verlag.

Tsurumi, H. 1990. "Comparing Bayesian and Non-Bayesian Limited Information Estimators." S. Geisser, J.S. Hodges, S.J. Press, and A. Zellner (ed.), Bayesian and Likelihood Methods in Statistics and Econometrics, North Holland Publishing, Amsterdam.

van Akkeren, M., Judge, G. G. and Mittelhammer, R.C., 2000. Generalized moment based estimation and inference. To appear as a chapter in the Handbook of Applied Econometrics and Statistical Inference, ed. by Aman Ullah and Alan Wan.

Zellner, A. "The Finite Sample Properties of Simultaneous Equations' Estimates and Estimators Bayesian and Non-Bayesian Approaches." *Journal of Econometrics*, 83, p. 185-212, 1998

Table 1. Mean value based on a distribution of parameter estimates from 1000 Monte Carlo simulations using *3SLS*, *MEEL*, *MEL*, and *MLEL*.

Obs	*3SLS*	*MEEL*	*MEL*	*MLEL*
γ_{21} = .222				
50	0.2010	0.2003	0.1986	0.2024
100	0.2727	0.2428	0.2470	0.2405
200	0.2208	0.2185	0.2182	0.2216
γ_{12} =.267				
50	0.2629	0.2356	0.2371	0.2321
100	0.3091	0.2110	0.2874	0.2677
200	0.2653	0.2657	0.2624	0.2566
γ_{32} = .046				
50	0.0477	0.0548	0.0570	0.0581
100	0.0522	0.0673	0.0615	0.0679
200	0.0565	0.0491	0.0478	0.0537
γ_{13} = .087				
50	0.0783	0.0778	0.0774	0.0779
100	0.0880	0.0902	0.0871	0.0894
200	0.0876	0.0901	0.0892	0.0863

Table 2. Standard errors based on a distribution of parameter estimates from 1000 Monte Carlo simulations using *3SLS*, *MEEL*, *MEL*, and *MLEL*.

Obs	*3SLS*	*MEEL*	*MEL*	*MLEL*
γ_{21} = .222				
50	0.2193	0.2138	0.2084	0.2197
100	0.1596	0.1504	0.1527	0.1619
200	0.0952	0.0930	0.0928	0.0973
γ_{12} =.267				
50	0.7232	0.6388	0.6095	0.6444
100	0.4895	0.4220	0.4128	0.4274
200	0.3081	0.2914	0.2932	0.2988
γ_{32} = .046				
50	0.6211	0.5797	0.5447	0.6129
100	0.4613	0.3978	0.3800	0.4029
200	0.2962	0.2846	0.2858	0.2986
γ_{13} = .087				
50	0.3983	0.4061	0.4037	0.4074
100	0.2693	0.2727	0.2709	0.2727
200	0.1899	0.1909	0.1907	0.1920

Table 3. Mean square errors based on a distribution of parameter estimates from 1000 Monte Carlo simulations using *3SLS*, *MEEL*, *MEL*, and *MLEL*.

Obs	*3SLS*	*MEEL*	*MEL*	*MLEL*
γ_{21} = .222				
50	0.0485	0.0461	0.0439	0.0486
100	0.0254	0.0226	0.0233	0.0262
200	0.0091	0.0087	0.0086	0.0095
γ_{12} =.267				
50	0.5225	0.4086	0.3721	0.4161
100	0.2394	0.1785	0.1707	0.1832
200	0.0948	0.0848	0.0859	0.0893
γ_{32} = .046				
50	0.3853	0.3358	0.2965	0.3754
100	0.2126	0.1585	0.1445	0.1627
200	0.0877	0.0809	0.0816	0.0892
γ_{13} = .087				
50	0.1585	0.1649	0.1629	0.1659
100	0.0724	0.0743	0.0733	0.0743
200	0.0360	0.0364	0.0363	0.0368

Table 4. Rejection Probabilities for True Hypotheses with Asymptotic Normal *Z* test.

Single Hypotheses

3SLS

Obs	γ_{21}=.222	γ_{12}=.267	γ_{32}=.046	γ_{31}=.087
50	0.0890	0.0660	0.0790	0.0610
100	0.0730	0.0570	0.0760	0.0570
200	0.0600	0.0540	0.0260	0.0690

MEEL

Obs	γ_{21}=.222	γ_{12}=.267	γ_{32}=.046	γ_{31}=.087
50	0.1260	0.0690	0.0550	0.1100
100	0.0790	0.0420	0.0350	0.0760
200	0.0550	0.0510	0.0450	0.0510

MEL

Obs	γ_{21}=.222	γ_{12}=.267	γ_{32}=.046	γ_{31}=.087
50	0.1170	0.0660	0.0610	0.1020
100	0.0740	0.0400	0.0310	0.0770
200	0.0570	0.0510	0.0270	0.0650

MLEL

Obs	γ_{21}=.222	γ_{12}=.267	γ_{32}=.046	γ_{31}=.087
50	0.1310	0.0730	0.0620	0.1190
100	0.0830	0.0440	0.0310	0.0820
200	0.0630	0.0490	0.0260	0.0700

Table 5. Rejection Probabilities for True and False Hypotheses with Asymptotic Chi-Square Wald Test.

	Joint Hypotheses			
	3SLS		*MEEL*	
	γ_{21}=.222 γ_{32}=.046	γ_{21}=0 γ_{32}=0	γ_{21}=.222 γ_{32}=.046	γ_{21}=0 γ_{32}=0
50	0.1240	0.2700	0.1630	0.3040
100	0.0940	0.3740	0.0990	0.3710
200	0.0590	0.6470	0.0580	0.6430

	Joint Hypotheses			
	MEL		*MLEL*	
	γ_{21}=.222 γ_{32}=.046	γ_{21}=0 γ_{32}=0	γ_{21}=.222 γ_{32}=.046	γ_{21}=0 γ_{32}=0
50	0.1640	0.3120	0.1790	0.3150
100	0.0980	0.3670	0.1010	0.3720
200	0.0590	0.6420	0.0640	0.6590