INFORMATION THEORETIC ALTERNATIVES TO TRADITIONAL SIMULTANEOUS EQUATIONS ESTIMATORS IN THE PRESENCE OF HETEROSKEDASTICITY

Thomas L. Marsh[a] and Ron C. Mittelhammer[b]

Direct Correspondence to:

Thomas L. Marsh
Department of Agricultural Economics
342 Waters Hall
Kansas State University
Manhattan, KS 66506-4011
Phone 785-532-4913, Fax 785-532-6925
tlmarsh@agecon.ksu.edu

Selected Paper
American Agricultural Economics Association
Long Beach, 2002

[a]Assistant Professor, Kansas State University, Manhattan, KS, 66506, 785-532-4913, tlmarsh@agecon.ksu.edu and [b]Professor, Washington State University.

Information Theoretic Alternatives to Traditional Simultaneous Equations Estimators in the Presence of Heteroskedasticity

Abstract: Finite sampling properties of information theoretic estimators of the simultaneous equations model, including maximum empirical likelihood, maximum empirical exponential likelihood, and maximum log Euclidean likelihood, are examined in the presence of selected forms of heteroskedasticity. Extensive Monte Carlo experiments are used to compare finite sample performance of Wald, Likelihood ratio, and Lagrangian multiplier tests constructed from information theoretic estimators to those from traditional generalized method of moments.

Keywords: endogeneity, mean square error, simultaneous equations

1. Introduction

Heteroskedasticity arises when observations exhibit heterogeneity of variances across sample observations. Traditional Generalized Method of Moments (*GMM*) estimators generally remain unbiased and consistent in the face of heteroskedasticity of nondescript form, but the typical covariance matrix estimators derived from *GMM* methods are inconsistent in that context, the estimators themselves are generally inefficient, and inference procedures based on the estimators are biased and can behave poorly. Appropriate inference methods in the context of an unknown form of heteroskedasticity in simultaneous equations models exist asymptotically (White), but are not necessarily accurate or useful in finite sample situations. In contrast, information theoretic estimators, which are asymptotically efficient (Kitamura and Stutzer), can be applied directly to data that are heteroskedastic and offer a potentially robust estimation and inference alternative to standard estimation methods. This is particularly important in finite samples where the sampling properties of traditional estimators and testing procedures are often suspect.

Information theoretic 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]. Information theoretic estimators do not require specification of the specific parametric functional form of sampling distributions or likelihood functions, but rather make mild assumptions concerning the existence of zero-valued moment conditions that are no more stringent than used in GMM estimation. To date, there has been only limited analysis of the finite sample performance of these estimators. The most extensive Monte Carlo investigation currently available is the work of Imbens, Spady, and Johnson, who investigated the properties of point estimators and hypothesis testing procedures in the limited context of models having scalar parameters under the assumption of *iid* sample observations.

In this paper we examine the performance of three different Information Theoretic (*IT)* estimators as competitors to traditional least squares and maximum likelihood methods for estimating the parameters of linear simultaneous equations models. These *IT* alternatives include the Maximum Empirical Likelihood (*MEL*), Maximum Empirical Exponential Likelihood (*MEEL*), and Maximum Log Euclidean Likelihood (*MLEL*) estimators. The finite sample performance of the *IT* type estimators is investigated in the context of an extensive Monte Carlo analysis conducted over a range of finite sample sizes and over a variety of different forms of heteroskedasticity designed to typify sampling processes encountered in applied econometric work. The competing estimators are compared on the basis of both parameter estimation risk and dependent variable prediction risk. In addition, Monte Carlo simulations are used to compare the accuracy of the size and power of asymptotically normally distributed Z-tests and asymptotically chi-square distributed Wald, Likelihood Ratio, and Lagrangian Multiplier tests.

Although results of Monte Carlo analyses are specific to the collection of particular Monte Carlo experiments analyzed, the extensive Monte Carlo sampling results examined in this paper suggest that the alternative *IT* estimators exhibit some inherent robustness to heteroskedasticity both in terms of parameter estimator risk properties and in terms of the accuracy of test procedures derived from them. A principal implication for empirical application is the recommendation of using one of the *IT* alternatives, i.e. *MEEL, MEL*, or *MLEL*, in place of, or in addition to, the standard least squares or maximum likelihood methods when heteroskedasticity is suspected but the precise functional form of the heteroskedasticity is unknown.

2. Empirical Likelihood Estimators

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

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

where $\mathbf{Y}_i$ 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 explanatory 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 $\boldsymbol{\delta}_i = [\boldsymbol{\gamma}_i ' | \boldsymbol{\beta}_i ']'$.

For a complete system of simultaneous equations a consistent generalized method of moments (*GMM*) estimator can be derived from the empirical moments

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

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 structural parameters in the system. Setting (2) to zero generally produces an inconsistent system of equations, so in the estimated optimal *GMM* estimation approach, the estimator of $\boldsymbol{\delta}$ is instead defined to be

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

where $\hat{\mathbf{W}}_n$ is an estimate of the asymptotically optimal weight matrix. In effect the moments $\mathbf{h}(\boldsymbol{\delta})$ are driven to zero in weighted (by $\hat{\mathbf{W}}_{\mathrm{n}}$) Euclidean distance as closely as possible via a choice of parameter values. If $\hat{\mathbf{W}}_n = \left(\hat{\boldsymbol{\Sigma}} \otimes n^{-1}\mathbf{Z}'\mathbf{Z}\right)^{-1}$ and **Z**=**X** then the *GMM* estimator is equivalent to three stage least squares (*3SLS*).

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

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

where the unknown $(n\times 1)$ vector **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 $\mathrm{p}_i = 1/n$ for $i = 1,...,n$, while the *EL* approach treats the unknown $\mathrm{p}_{\mathrm{i}}'\mathrm{s}$ as parameters to be estimated. Note that although we are currently examining a linear system of equations, the single-equation equivalent follows for *q*=1.

The extremum problem for information theoretic estimation can be formulated as

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

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 $\sum_{i=1}^{n}\ln(\mathrm{p}_{\mathrm{i}})$, the empirical exponential likelihood (or negative entropy) function $\sum_{i=1}^{n}\mathrm{p}_{\mathrm{i}}\ln(\mathrm{p}_{\mathrm{i}})$, and the log Euclidean likelihood function $n^{-1}\left(\sum_{i=1}^{n}\left(n^2\mathrm{p}_i^2 - 1\right)\right)$. 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

$$\text{(6)} \qquad L(\mathbf{p},\boldsymbol{\delta},\boldsymbol{\lambda},\eta) = \phi(\mathbf{p}) - \sum_{i=1}^{q} \lambda_i' \left[(\mathbf{p} \odot \mathbf{Z})' \left(\mathbf{Y}_{\cdot i} - \mathbf{M}_{(i)} \boldsymbol{\delta}_i \right) \right] - \eta \left(\sum_{i=1}^{n} p_i - 1 \right)$$

where $\boldsymbol{\lambda} = vec(\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_q)$ 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 that $\mathbf{f}(\mathbf{p}) = \left(f_1(\mathbf{p}), \ldots, f_n(\mathbf{p}) \right)$, where $f_j(\mathbf{p}) = \frac{\partial \phi(\mathbf{p})}{\partial p_j} \ \forall j$, admits an inverse function, $\mathbf{f}^{-1}(\cdot)$, the general solution for **p** is

$$\text{(7)} \qquad \mathbf{p} = \mathbf{f}^{-1} \left(\sum_{i=1}^{q} \boldsymbol{\lambda}_i' \left[\mathbf{Z}_{j\cdot}' \left(\mathbf{Y}_{ji} - \mathbf{M}_{(i)}[j,\cdot] \boldsymbol{\delta}_i \right) \right] + \eta, \ \forall j \right)$$

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, $\phi(\mathbf{p}) = \sum_{i=1}^{n} \ln(\mathrm{p_i})$, yields the Maximum Empirical Likelihood (*MEL*) estimate of $\boldsymbol{\delta}$. The optimal $\mathrm{p_j}$ can be expressed as (note it can be shown that $\eta = 1$ at the optimal solution)

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

Concentrating the objective function by substituting $\mathrm{p_j}(\boldsymbol{\lambda}, \boldsymbol{\delta})$ 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, $\phi(\mathbf{p}) = \sum_{i=1}^{n} \mathrm{p_i} \ln(\mathrm{p_i})$, leads to the Maximum Empirical Exponential Likelihood (*MEEL*) estimate of $\boldsymbol{\delta}$. The optimal $\mathrm{p_j}$ can be expressed as

$$\text{(8)} \qquad \mathrm{p_j}(\boldsymbol{\lambda}, \boldsymbol{\delta}) = \frac{\exp\left(\sum_{\mathrm{i}=1}^{\mathrm{q}} \boldsymbol{\lambda}_{\mathrm{i}}^{'} \left[\mathbf{Z}_{\mathrm{j}\cdot}{}' \left(\mathbf{Y}_{\mathrm{ji}} - \mathbf{M}_{(\mathrm{i})}[\mathrm{j},\cdot] \boldsymbol{\delta}_{\mathrm{i}} \right) \right] \right)}{\sum_{\mathrm{j}=1}^{\mathrm{n}} \exp\left(\sum_{\mathrm{i}=1}^{\mathrm{q}} \boldsymbol{\lambda}_{\mathrm{i}}^{'} \left[\mathbf{Z}_{\mathrm{j}\cdot}{}' \left(\mathbf{Y}_{\mathrm{ji}} - \mathbf{M}_{(\mathrm{i})}[\mathrm{j},\cdot] \boldsymbol{\delta}_{\mathrm{i}} \right) \right] \right)}$$

Concentrating the objective function by substituting $\mathrm{p_j}(\boldsymbol{\lambda}, \boldsymbol{\delta})$ 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 is similar to the generalized maximum entropy estimators proposed by Golan, Judge, and Miller in the sense that it uses the same basic 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 $\mathbf{p}$ 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 optimal $\mathrm{p_j}$ can be expressed as

$$\text{(9)} \qquad \mathrm{p_j}(\boldsymbol{\lambda},\boldsymbol{\delta}) = (2\mathrm{n})^{-1}\left[\boldsymbol{\lambda}_{\mathrm{i}}'\mathbf{Z}_{\mathrm{j\cdot}}'\left(\mathbf{Y}_{\mathrm{ji}} - \mathbf{M}_{(\mathrm{i})}[\mathrm{j},\cdot]\boldsymbol{\delta}_{\mathrm{i}}\right) + \eta\right].$$

Again concentrating the objective function by substituting $\mathrm{p_j}(\boldsymbol{\lambda},\boldsymbol{\delta})$ 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 and Tests for IT Estimators

The *MEL*, *MEEL*, and *MLEL* estimators are all consistent, asymptotically normally distributed, and asymptotically efficient relative to the optimal estimating function estimator (See Mittelhammer, Judge, and Miller, Chapter 17, for further discussion). 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 $\boldsymbol{\Omega}$ can be defined as

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

In the expression above **h** represents the estimating function specified as

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

and note that

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

See Imbens, Spady, and Johnson, as well as Kitamura and Stutzer, for underlying assumptions and proof of consistency and asymptotic normality. The estimated asymptotic covariance matrix can be used in the usual way to form asymptotically valid hypothesis tests and confidence interval estimators for the parameters of the structural model. In the case of non-iid sampling, the above covariance expression can be extended (Kitamura and Stutzer).

3.1 Testing Moment Conditions

A statistical test of particular interest for *IT*-type estimators is the test of the validity of moment conditions

(13) $$H_0 : \mathbf{E}\left[\mathbf{h}\left(\mathbf{Y}, \boldsymbol{\delta} \right) \right] = \mathbf{0}$$

which are the moment conditions specified in the *IT* extremum problem. This evaluates the hypothesis that there is a value $\boldsymbol{\delta}^0$ that solves the above moment conditions and is equivalent to testing the unbiasedness of the estimating function $\mathbf{h}(\mathbf{Y}, \boldsymbol{\delta})$. Imbens, Spady, and Johnson, as well as Mittelhammer and Judge, examine properties of Wald (*W*), Psuedo-Likelihood Ratio (*LR*), and Lagrange Multiplier

(*LM*) test statistics. Here, we examine the performance of the *W*, *LR*, and *LM* tests in the presence of heteroskedasticity.

The Wald test statistic has a χ^2 limiting distribution with *m-k* degrees of freedom under H_0, where the statistic is defined by

$$(14) \qquad \mathbf{h}(\mathbf{Y},\boldsymbol{\delta})'\boldsymbol{\Omega}^{-1}\mathbf{h}(\mathbf{Y},\boldsymbol{\delta}) \xrightarrow{d} \chi^2_{m-k}.$$

An empirical likelihood ratio test statistic under H_0 can be defined by

$$(15) \qquad 2n\, \hat{\mathbf{p}}'\ln(\hat{\mathbf{p}}) + \ln(n) \xrightarrow{d} \chi^2_{m-k}$$

The appropriate *LM* test statistic under H_0 is defined by

$$(16) \qquad n\, \hat{\boldsymbol{\lambda}}'\mathbf{R}\, \hat{\boldsymbol{\lambda}} \xrightarrow{d} \chi^2_{m-k}$$

where the matrix **R** is defined by

$$\mathbf{R} = \left[\sum_{i=1}^{n} \hat{p}_i \mathbf{h}_i \mathbf{h}_i'\right]\left[\sum_{i=1}^{n} \hat{p}_{\boldsymbol{i}}^2\, \mathbf{h}_i \mathbf{h}_i'\right]^{-1}\left[\sum_{i=1}^{n} \hat{p}_i \mathbf{h}_i \mathbf{h}_i'\right]$$

and constitutes a robust estimator of the covariance matrix of the moment functions (Imbens, Spady, and Johnson).

4. Finite Sample Properties

Because the derivation of the finite sample properties of the *IT* estimators presented above are not tractable, 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 *IT* estimators, and their performance relative to *2SLS* and *GMM*. In the experiments below, the covariance matrices for *2SLS* and *GMM* are not corrected for heteroskedasticity because our objective is to benchmark the performance of *IT* estimators to traditional estimators. To measure the performance of the estimators, we use the squared error loss between the true and estimated values of structural coefficients and the prediction squared error between the actual and predicted dependent variable. Regarding performance in inference settings, simulated rejection probabilities of true and false hypothesis are used to evaluate the size and power of statistical tests.

4.1 Monte Carlo Experiments

To analyze the influence of heteroskedasticity, we consider a single-equation SEM with a single exogenous and single right hand side endogenous variable. The data sampling process has the following form

(17) $$\mathrm{Y}_{i1} = \mathrm{Y}_{i2}\gamma + Z_{i1}\beta + \boldsymbol{\varepsilon}_{i1} = \mathbf{M}_{i1}\boldsymbol{\delta} + \boldsymbol{\varepsilon}_{i1}$$

where $\mathbf{M}_1 = [\mathbf{Y}_2 \ \mathbf{Z}_1]$, $\boldsymbol{\delta} = [\gamma' \ \beta']'$, and

(18) $$\mathbf{Y}_{i2} = \pi_0 + \sum_{j=1}^{4} Z_{ij}\pi_j + v_i$$

The vector of unknown parameters, $\boldsymbol{\delta}$, is defined to be $[-1,2]'$. Outcomes of the random vector $[\mathbf{Y}_{i2}, \varepsilon_i, \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \mathbf{Z}_{i3}, \mathbf{Z}_{i4}]$ are generated *iid* from a multivariate normal distribution having a zero mean vector and standard deviation of five. To generate valid instruments, correlations between ε_i and the $\mathbf{Z}_{ij}$'s were set to zero. A correlation of 0.50 was specified between the random variables $\mathbf{Y}_{i2}$ and ε_i to represent moderate endogeneity. Likewise, degrees of correlation between the instruments and the y_{i2} variable and the levels of collinearity existent among the instrumental variables were set to 0.50.

Because heteroskedasticity arises when observations exhibit heterogeneity of variances across sample observations, one general functional representation of heteroskedasticity is $\sqrt{\mathrm{var}(\varepsilon_i)} = \sigma_i = g_i(z_{1i})$. For the Monte Carlo simulations below, four specific cases are examined: Case 1 represents the homoskedastic error model, or $\sigma_i = \sigma, \forall i$; Case 2 assumes the standard error of the SEM is proportional to the first exogenous variable for each observation, or $\sigma_i = (.5z_{i1})\sigma, \forall i$; Case 3 specifies that the standard error is proportional to the square of the first exogenous variable for each observation, or $\sigma_i = \left(.5 z_{i1}^2\right)\sigma, \forall i$; and Case 4 is a standard multiplicative heteroskedastic model, or $\sigma_i = \exp\left(.5 z_{1i}\right)\sigma, \forall i$. Estimates below are based on 1000 Monte Carlo repetitions for sample sizes of n = 50, 100, 500, and 1000.

4.2 Results: Parameter and Prediction Expected SEL

In performing comparisons, we examine the impacts that different forms of heteroskedasticity have on precision of parameter estimates and accuracy of predictive fit. Table 1 reports mean estimates of the squared error loss (*SEL*) between the true and estimated coefficient values of β for *MEEL, MEL, MLEL*, *2SLS*, and *GMM*. For each estimator the mean *SEL* values increased for the heteroskedastic error Cases 2-4 relative to the homoskedastic Case 1, with the largest SEL values being associated with Case 3, followed by Case 4 and Case 2 respectively. For n = 50, with the exception of Case 4 denoting multiplicative heteroskedasticity, *2SLS* had the smallest mean *SEL*. As the sample size increased from 50 observations the mean *SEL* values decreased in each experiment. Overall, for Cases 1-3, the estimators had very similar *SEL* values. For Case 4 with multiplicative heteroskedasticity the *MEEL*, *MEL* and *MLEL* estimators dramatically outperformed *2SLS* and *GMM* in *SEL*. These results underscore an advantage of *IT* estimators over traditional estimators in that they appear to be more robust across different forms of heteroskedasticity.

Table 2 reports mean prediction squared error (*PSE*) between the actual and predicted $\mathbf{Y}_{i1}$ values. Across the different forms of heteroskedasticity, and for all sample sizes, *2SLS* and *GMM* have smaller mean *PSE* than do *MEEL*, *MEL* or *MLE,* although for the most part the prediction performance of all of the estimators was quite similar. The relative difference of the mean *PSE* values between the estimators decreased with increasing observations. Overall, *2SLS*, followed by *GMM* and *MEEL* respectively, exhibited the smallest values of mean *PSE*.

4.3 Results: Standard Errors

Table 3 contains empirical measures of the bias of the estimated standard error of $\hat{\beta}$, calculated as the simulated standard error of the empirical distribution of the $\hat{\beta}$'s (measuring the true standard error) minus the mean asymptotic standard error of the $\hat{\beta}$. From this information we can infer several implications as to the performance of the *IT* estimators. At n = 50 all three of the *IT*-type estimates exhibit more bias than do *2SLS* and *GMM*, but by 100 observations the *IT*-type estimates exhibit less bias than *2SLS* and *GMM* for the heteroskedastic Cases 2-4. As the observations increase from 100 to 1000, the *MEEL*, *MEL* and *MLEL* asymptotic standard errors are converging to the simulated true standard errors. By 1000 observations, for all three cases of heteroskedasticity, the *MEEL*, *MEL*, and *MLEL* standard error estimates have dramatically smaller empirical biases that do *2SLS* and *GMM*.

4.4 Results: Coverage Probability and Power

Table 4 contains coverage probabilities for $H_0 : \beta = -1$ across the homoskedastic and heteroskedastic error models. The targeted true coverage probability is 0.99, which is the complement of the true size of the test, or (1-0.01)=0.99. Across all forms of heteroskedasticity, the *MEEL*, *MEL*, and *MLEL* are converging to 0.99 as the observations increase from 50 to 1000. In contrast, the *2SLS* and *GMM* estimators are not. This illustrates the inference robustness of *IT*-type estimators in the presence of heteroskedasticity.

Table 5 contains coverage probabilities for $H_0 : \beta = 0$, which represents an observation on the power of the test for this incorrect null hypothesis. Interestingly, all the *IT*-type estimators for heteroskedastic Case 2 and Case 4 appear to perform better than *2SLS* and *GMM*. However, none of the estimators perform well for Case 3, with *2SLS* and *GMM* performing better than the *IT* estimators. It is apparent that the power of the test for $H_0 : \beta = 0$ is sensitive to the type of heteroskedasticity existent in the error of the SEM. Nonetheless, these results suggest that *IT*-type estimators can be attractive alternatives to the standard 2SLS and GMM procedures in heteroskedastic situations.

4.5 Results: Moment Condition Tests

Figures 1-4 illustrate the size of Wald tests for *2SLS* and *GMM*, as well as Wald, LR, and LM tests for the three *IT* estimators. The targeted true size of the tests was 0.05. The figures lead to several observations regarding moment condition testing. First, across all experiments the test statistics for the *IT* estimators are converging to the true size 0.05 as the observations increase from 50 to 1000. Second, the LR and LM tests for the *MEL* estimator appear less robust than those of *MEEL* and *MLEL* for 50 and 100 observations, but become more alike for 500 to 1000 observations.

Interestingly, in comparing the results of all of the estimators, the Wald tests for 2*SLS* and *GMM* were not robust across the different forms of heteroskedasticity. In particular, the Wald tests for *2SLS* and *GMM* appear not to converge to correct size for Case 3 with multiplicative heteroskedasticity. Given the popularity of the Wald test and the likelihood of heteroskedasticity in cross-sectional data, these results have important implications in applied econometrics.

5. 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 information theoretic estimators in the presence of selected forms of heteroskedasticity over a range of finite sample sizes, Monte Carlo sampling experiments were performed for a single-equation simultaneous equations model. Their relative performance was assessed, and also compared to the traditional *2SLS* and *GMM* with covariance matrices not corrected for heteroskedasticity.

In the Monte Carlo experiments examined, the estimation performance of the *MEEL*, *MEL*, and *MLEL* estimators was quite competitive with *2SLS* and *GMM* across all sampling scenarios, and for multiplicative heteroskedasticity, the IT estimators performed dramatically better than the *2SLS* and *GMM* estimators. More specifically, the prediction performance of the *MEL*, *MEEL*, and *MLEL* estimators was also very close to the *2SLS* and *GMM* estimators and the latter pair of estimators dominated the IT estimators in prediction performance across all scenarios. The estimated IT standard errors converged to the simulated true standard errors, but *2SLS* and *GMM* did not. Wald, LR, and LM moment condition tests for the three *IT* estimators were robust in the presence of the different forms of heteroskedasticity, but *2SLS* and *GMM* were not. In this aspect of inference, the IT estimators were clearly better than the *2SLS* and *GMM* estimators. Finally the power of tests based on the *IT* estimators, as well as *2SLS* and *GMM* , were sensitive to the form of heteroskedasticity, and comparisons of the relative peformance of the *IT* and *2SLS/GMM* estimators was mixed.

These finding provide insights into the performance of *IT* –type estimators relative to traditional GMM procedures in finite samples when estimating the parameters of a simultaneous equations model. The results suggest that there may be an important role for *IT*-type estimators as alternatives to traditional GMM-type estimators in empirical contexts where heteroskedasticity is suspected, especially in the context of inference, but also for purposes of parameter estimation when certain forms of heteroskedasticity are present. Nonetheless, the results also suggest that the IT suite of estimators is not a panacea, with the traditional estimators still being the better choice in terms of parameter SEL and prediction SEL in many sampling contexts representative of empirical practice. Additional analysis of the finite sample performance of *IT*-type estimators is needed (i.e., comparing *IT*-estimators to White's heteroskedastic corrected *2SLS* and *GMM* estimators), and the provocative results of this study suggest that such efforts are warranted.

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.

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

Golan, A., Judge, G. G. and Miller, D., 1996. Maximum Entropy 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.

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

Mittelhammer, R. C. and G. Judge. 2002. Robust Empirical Exponential Likelihood Estimation of Models with Non-Orthogonal Noise Components, Forthcoming.

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.

White, H. 1982. “Instrumental Variables Regression with Independent Observations.” Econometrica, 50:483-500.

Table 1. Squared error loss (*SEL*) between true and estimated parameter values of β.

Obs	*2SLS*	*MEEL*	*MEL*	*MLEL*	*GMM*
			Case 1: $\sigma_i = \sigma$		
50	0.5178	0.5756	0.5913	0.6156	0.5982
100	0.2472	0.2478	0.2497	0.2578	0.2532
500	0.0473	0.0479	0.0489	0.0499	0.0477
1000	0.0252	0.0256	0.0259	0.0261	0.0252
			Case 2: $\sigma_i = \left(.5z_{1i}\right)\sigma$		
50	0.7587	0.8177	0.8803	0.8959	0.7932
100	0.3644	0.3758	0.3996	0.3947	0.3749
500	0.0742	0.0763	0.0790	0.0767	0.0741
1000	0.0387	0.0392	0.0400	0.0392	0.0389
			Case 3: $\sigma_i = \left(.5z_{1i}^2\right)\sigma$		
50	8.6367	9.0579	9.5703	9.4338	9.2679
100	4.9133	4.7792	5.0028	4.6950	5.0489
500	1.1684	1.1161	1.1393	1.0712	1.1754
1000	0.5870	0.5698	0.5775	0.5496	0.5900
			Case 4: $\sigma_i = \exp\left(.5z_{1i}\right)\sigma$		
50	4.6157	3.9375	4.0166	3.9356	4.8360
100	2.3104	1.5248	1.5454	1.3808	2.3684
500	0.6840	0.2565	0.2687	0.2617	0.6879
1000	0.3426	0.1173	0.1297	0.1282	0.3425

Table 2. Prediction squared error (*PSE*) between actual and predicted values of $\mathbf{Y}_{i1}$.

Obs	*2SLS*	*MEEL*	*MEL*	*MLEL*	*GMM*
			Case 1: $\sigma_i = \sigma$		
50	22.9731	25.8801	26.2404	26.6622	24.9978
100	23.5294	25.0039	25.2186	25.3949	24.4969
500	24.7011	25.0736	25.1552	25.1883	24.9111
1000	24.9235	25.1559	25.1793	25.1735	25.0287
			Case 2: $\sigma_i = (.5z_{1i})\sigma$		
50	24.6441	25.8179	26.1924	26.6100	24.9491
100	25.1713	25.5305	25.6737	25.7255	25.2433
500	25.1316	25.1464	25.1539	25.1481	25.1329
1000	24.9395	24.9464	24.9492	24.9473	24.9405
			Case 3: $\sigma_i = (.5z_{1i}^2)\sigma$		
50	272.7837	289.8926	293.6553	296.7647	282.6930
100	288.8104	296.4845	298.0896	298.7322	293.0778
500	293.9879	295.9929	296.3469	296.3198	294.8674
1000	300.3475	301.2506	301.3542	301.4237	300.7975
			Case 4: $\sigma_i = \exp(.5z_{1i})\sigma$		
50	148.1361	158.9957	160.0775	161.2424	153.1170
100	159.0306	164.8555	164.9043	165.8437	161.9508
500	180.8251	182.3342	182.2686	182.3376	181.3416
1000	187.3857	188.5066	188.3031	188.3390	187.6527

Table 3. Estimates of bias for standard error of β under homoskedastic errors and selected forms of heteroskastic errors.

Obs	*2SLS*	*MEEL*	*MEL*	*MLEL*	*GMM*
			Case 1: $\sigma_i = \sigma$		
50	0.0168	0.1023	0.0973	0.1077	0.0348
100	0.0130	0.0344	0.0293	0.0355	0.0110
500	0.0018	0.0047	0.0056	0.0074	0.0017
1000	0.0049	0.0057	0.0061	0.0068	0.0047
			Case 2: $\sigma_i = (.5z_{1i})\sigma$		
50	0.1991	0.2142	0.2216	0.2267	0.1929
100	0.1332	0.0925	0.0978	0.0963	0.1330
500	0.0632	0.0193	0.0221	0.0187	0.0624
1000	0.0494	0.0131	0.0142	0.0124	0.0496
			Case 3: $\sigma_i = (.5z_{1i}^2)\sigma$		
50	0.8361	0.9511	0.9406	0.9352	0.8401
100	0.6805	0.5303	0.5265	0.4669	0.6758
500	0.3820	0.1324	0.1277	0.0999	0.3826
1000	0.2670	0.0594	0.0561	0.0399	0.2678
			Case 4: $\sigma_i = \exp(.5z_{1i})\sigma$		
50	0.6167	0.7448	0.7092	0.6885	0.5982
100	0.3734	0.3088	0.2801	0.2181	0.3653
500	0.2794	0.0326	0.0310	0.0233	0.2788
1000	0.1898	-0.0151	-0.0084	-0.0093	0.1892

Table 4. Coverage probability for $H_0 : \beta = -1$ for homoskedastic and selected forms of heteroskastic errors. True coverage probability is 0.99, or equivalently its complement the true test size of 0.01.

Obs	*2SLS*	*MEEL*	*MEL*	*MLEL*	*GMM*
			Case 1: $\sigma_i = \sigma$		
50	0.957	0.932	0.937	0.943	0.957
100	0.980	0.973	0.981	0.976	0.983
500	0.989	0.987	0.988	0.988	0.990
1000	0.982	0.984	0.984	0.983	0.986
			Case 2: $\sigma_i = (.5z_{1i})\sigma$		
50	0.951	0.942	0.947	0.950	0.958
100	0.955	0.966	0.967	0.969	0.960
500	0.947	0.982	0.981	0.982	0.947
1000	0.949	0.985	0.985	0.985	0.949
			Case 3: $\sigma_i = (.5z_{1i}^2)\sigma$		
50	0.910	0.926	0.925	0.946	0.921
100	0.918	0.956	0.957	0.975	0.919
500	0.899	0.978	0.980	0.983	0.903
1000	0.903	0.981	0.983	0.986	0.901
			Case 4: $\sigma_i = \exp(.5z_{1i})\sigma$		
50	0.948	0.970	0.971	0.973	0.951
100	0.949	0.970	0.974	0.979	0.952
500	0.921	0.985	0.985	0.978	0.919
1000	0.920	0.986	0.992	0.993	0.922

Table 5. Coverage probability for $H_0 : \beta = 0$ (or power of test) for homoskedastic and selected forms of heteroskastic errors. True coverage probability is 0.99.

Obs	*2SLS*	*MEEL*	*MEL*	*MLEL*	*GMM*
			Case 1: $\sigma_i = \sigma$		
50	0.442	0.438	0.422	0.412	0.393
100	0.595	0.575	0.546	0.532	0.561
500	0.982	0.977	0.974	0.976	0.978
1000	1.000	1.000	1.000	1.000	1.000
			Case 2: $\sigma_i = (.5z_{1i})\sigma$		
50	0.276	0.291	0.281	0.250	0.266
100	0.481	0.431	0.424	0.404	0.471
500	0.978	0.931	0.923	0.920	0.973
1000	0.999	0.998	0.998	0.998	0.999
			Case 3: $\sigma_i = (.5z_{1i}^2)\sigma$		
50	0.145	0.149	0.148	0.121	0.138
100	0.165	0.110	0.111	0.086	0.157
500	0.313	0.127	0.118	0.105	0.309
1000	0.441	0.200	0.194	0.171	0.432
			Case 4: $\sigma_i = \exp(.5z_{1i})\sigma$		
50	0.169	0.222	0.214	0.191	0.155
100	0.195	0.232	0.231	0.195	0.188
500	0.398	0.564	0.537	0.494	0.393
1000	0.634	0.834	0.828	0.782	0.634

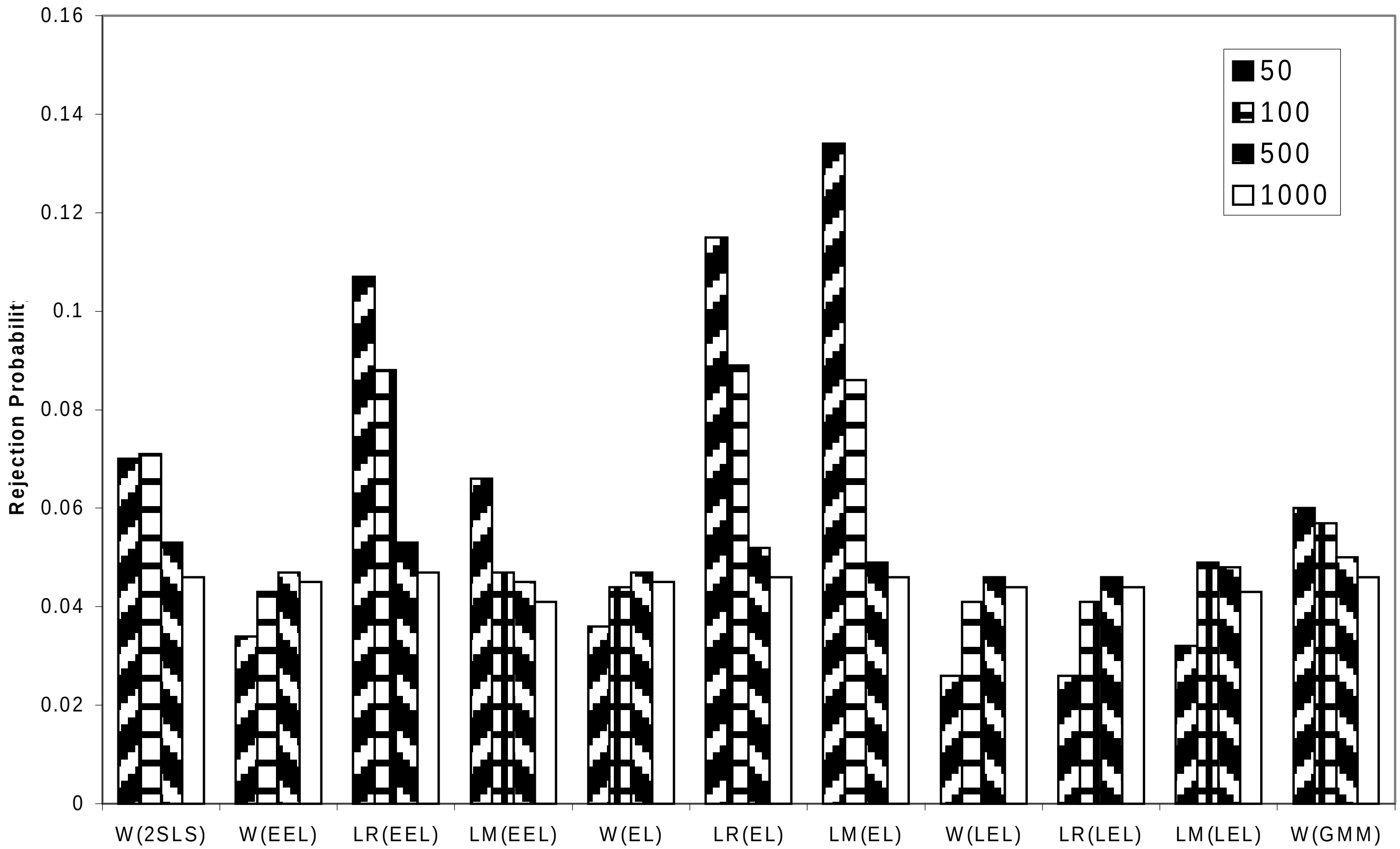

Figure 1. Testing the validity of moment conditions. Case 1: homoskedastic errors, or $\sigma_i = \sigma$.

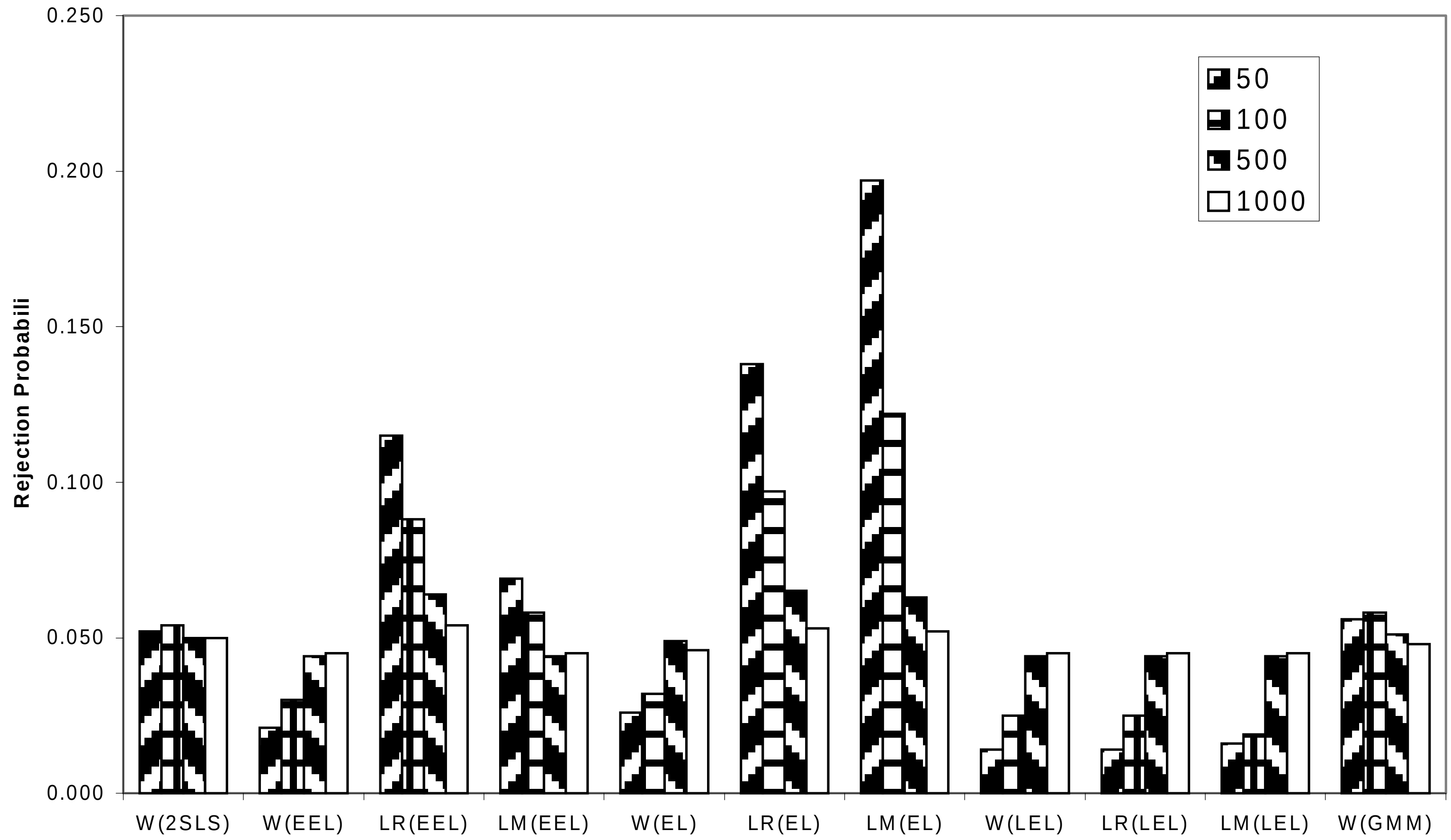

Figure 2. Testing the validity of moment conditions. Case 2: sqrt(variance) is proportional to the first exogenous variable for each observation, or $\sigma_i = (.5z_{i1})\sigma$.

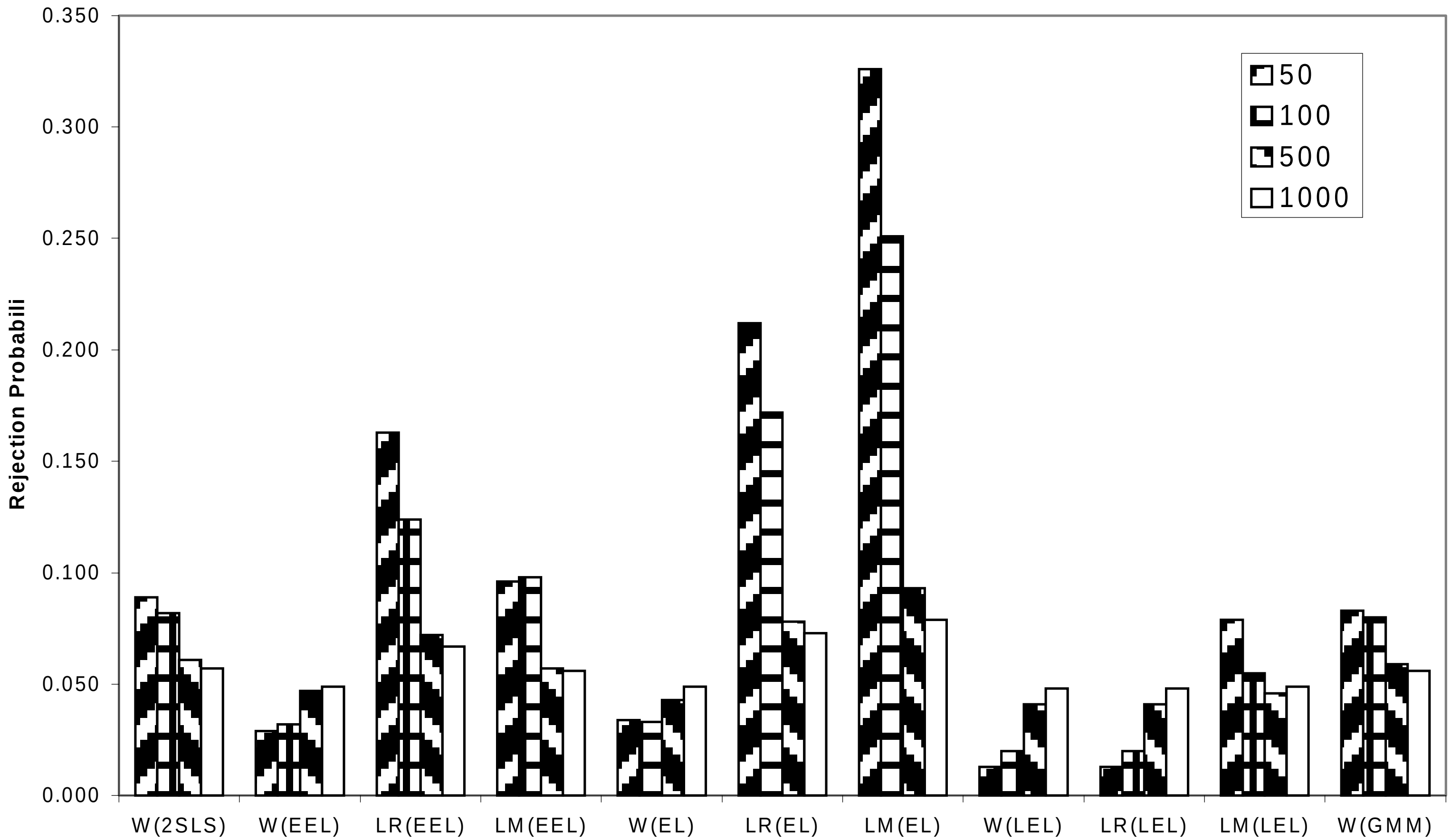

Figure 3. Testing the validity of moment conditions. Case 3: sqrt(variance) is proportional to the square of the first exogenous variable for each observation, or $\sigma_i = \left(.5z_{i1}^2\right)\sigma$.

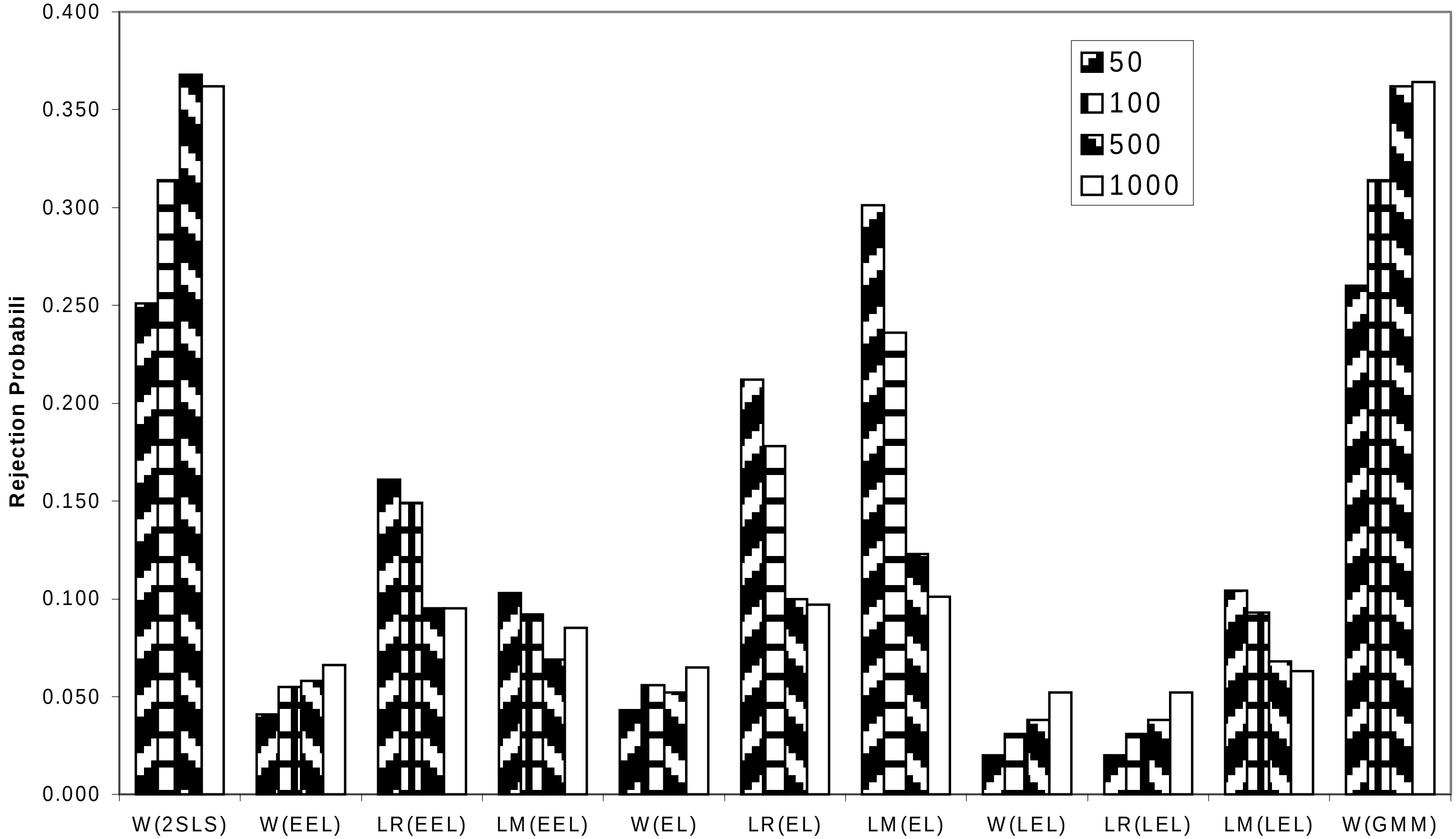

Figure 4. Testing the validity of moment conditions. Case 4: multiplicative heteroskedasticity, or $\sigma_i = \exp(.5 z_{1i})\sigma$.