1972

# Minimum Second Moment Estimation with Simultaneous Equation Systems

A. L. Nagar

R. A. L. Carter

RESEARCH REPORT 7209

# MINIMUM SECOND MOMENT ESTIMATION WITH SIMULTANEOUS EQUATION SYSTEMS

by

A. L. Nagar and R. A. L. Carter

# MINIMUM SECOND MOMENT ESTIMATION WITH SIMULTANEOUS EQUATION SYSTEMS*

by

A. L. Nagar and R. A. L. Carter

## 1. Introduction

The minimum second moment (MSM) estimator was defined by Nagar [1] to be the k class estimator which minimizes the determinant of the matrix of second moments about the parameters. This is achieved by the appropriate selection of a parameter $\varkappa$, where $k = 1 + \frac{\varkappa}{T}$ and T is the sample size. However, since $\varkappa$ contains population information it is necessary, in practise, to estimate it using two stage least squares (2SLS). This note discusses the distribution of the resulting $\hat{\varkappa}$ for the case where there is only one endogenous variable on the right-hand side of the equation to be estimated. We use Sawa's [3] results to establish the range of $\varkappa$ for which the moments of the k class estimators exist. If these moments do not exist MSM is undefined and cannot be employed. We give a straightforward way of obtaining a non-stochastic approximation to the optimum $\varkappa$. Finally we give some Monte Carlo results on the small sample distribution of $\hat{\varkappa}$ and on the performance of MSM using the approximate optimum $\varkappa$.

## 2. Notation and Assumptions

We write the linear simultaneous equation system as

(1) $$Y_w \Gamma + X\beta = U$$

where $Y_w$ is a T x M matrix of T observations on M endogenous variables,

---

*We would like to thank T. H. Wonnacott for helpful discussion and the Canada Council for financial assistance. Mrs. Johnson assisted with the computations reported on in section 6. Residual errors are ours exclusively.

X is a T x K matrix of T observations on K exogenous variables, U is a T x M matrix of unobservable random disturbances, $\Gamma$ and $\beta$ are, respectively, M x M and K x M matrices of unknown parameters.

We make the following assumptions:

(A.1) $\Gamma$ is non-singular. Therefore we can write the reduced form of (1) as

(2) $Y_w = X\Pi_w + \bar{V}_w$, where $\Pi_w = -\beta\Gamma^{-1}$ and $\bar{V}_w = U\Gamma^{-1}$

(A.2) The T rows of U are independent random drawings from an M variate normal distribution with means zero and a positive definite covariance matrix $\Sigma$. That is: $\frac{1}{T} EU'U = \Sigma$ and $U \sim N(0,\Sigma)$.

(A.3) X is a non-stochastic matrix of rank $K \leq T$.

Assumption (A.2) implies

(3) $\bar{V}_w \sim N(0,\Omega)$ where $\Omega = \Gamma'^{-1}\Sigma\Gamma^{-1}$

We write the equation whose coefficients are to be estimated as

(4) $y = \gamma\, y_1 + X_1\beta + u$

where y and $y_1$ are endogenous vectors, $X_1$ is a T x K submatrix of X, u is a column of random normal disturbances, $\gamma$ is a scalar parameter and $\beta$ is a $K_1$ x 1 vector of parameters. The entities in (4) are derived from those in (1) after the imposition of identifying restrictions and normalization. There are $K_2$ exogenous variables excluded from equation (4) so $K = K_1 + K_2$.

We will write the reduced form equations of the two endogenous variables in (4) as

(5) $y = X\pi + \bar{v}$ and

(6) $y_1 = X\pi_1 + \bar{v}_1$.

(A.4) Equation (4) is over-identified so that $\bar{v} - \bar{v}_1\gamma = u$.

## 3. The Optimum Value of $\kappa$

The k class estimates c and b of $\gamma$ and $\beta$ are given by

$$(7) \qquad \begin{bmatrix} c \\ b \end{bmatrix} = \begin{bmatrix} y_1'y_1 - kv_1'v_1 & y_1'X_1 \\ X_1'y_1 & X_1'X_1 \end{bmatrix}^{-1} \begin{bmatrix} y_1' - kv_1' \\ X_1' \end{bmatrix} y$$

where $v_1 = y_1 - X(X'X)^{-1}X'y_1$; the least squares estimate of $\bar{v}_1$ and

$$(8) \qquad k = 1 + \frac{\kappa}{T}$$

The MSM value of $\kappa$ is ([1], p. 580)

$$(9) \qquad \kappa = K - 2(m + K_1) - 3 - \frac{tr(Q\ C_1)}{tr(Q\ C_2)}$$

where m = the number of right hand endogenous variables (1 in our case)

$$Q = \begin{bmatrix} \bar{y}_1'\bar{y}_1 & \bar{y}_1'X_1 \\ X_1'\bar{y}_1 & X_1'X_1 \end{bmatrix}^{-1} \text{ with } \bar{y}_1 = X\,\pi_1$$

$$C_1 = \frac{1}{\sigma^2}\, qq' \text{ with } q = \frac{1}{T}\begin{bmatrix} E\ \bar{v}_1'u \\ 0 \end{bmatrix} \text{ and } \sigma^2 = \frac{1}{T}\, E\ u'u$$

$$C = \begin{bmatrix} \frac{1}{T}\, E\ \bar{v}_1'\bar{v}_1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$C_2 = C - C_1$$

We can assume, without loss of generality, that $\Omega = \frac{1}{T}E\, \bar{V}_w'\bar{V}_w = I.$[1]

Then, for m = 1 we have:

$$C = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

Using the identifying restriction (A.4) we have:

$$q = \frac{1}{T}\begin{bmatrix} E\, \bar{v}_1'(\bar{v} - \bar{v}_1\gamma) \\ 0 \end{bmatrix} = \begin{bmatrix} -\gamma \\ 0 \end{bmatrix}.$$

Also,

$$\sigma^2 = \frac{1}{T} E\, u'u = \frac{1}{T} E(\bar{v} - \bar{v}_1\gamma)'(\bar{v} - \bar{v}_1\gamma)$$

$$= 1 + \gamma^2 ,$$

so that

$$C_1 = \frac{1}{1+\gamma^2}\begin{bmatrix} \gamma^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore,

$$C_2 = \frac{1}{1+\gamma^2}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}.$$

---

[1] We can always transform the model to make $\Omega = I$. Since $\Sigma$ is positive definite, we can find a non-singular matrix $\Psi$ such that $\Psi\Psi' = \Sigma^{-1}$. Then define $P = \Gamma\Psi'$, $Y_w^* = Y_w P$ and $\Gamma = P^{-1}\Gamma^*$. The transformed model can then be written as $Y_w^*\Gamma^* + X\beta = u$ with $\Omega^* = \Gamma^{*\prime -1}\Sigma\, \Gamma^{*-1} = I$. Note that the transformation leaves $X_1\beta$, u and the identifying restrictions unaffected

Now if we write

$$Q = \begin{bmatrix} w_{11} & w_{22} \\ w_{21} & w_{22} \end{bmatrix}$$

then

$$\text{tr } QC_2 = w_{11}\left(\frac{1}{1+\gamma^2}\right),$$

$$\text{tr } QC_1 = w_{11}\left(\frac{\gamma^2}{1+\gamma^2}\right)$$

and

(10) $\kappa = a - \frac{1}{\gamma^2}$ where $a = K - 2K_1 - 5$ which is constant for a given model.

## 4. The Distribution of $\hat{\kappa}$

In practice $\gamma$ must be estimated, say by 2SLS, in order to evaluate (10) which becomes $\hat{\kappa} = a - \frac{1}{\hat{\gamma}^2}$. Asymptotically $\hat{\gamma}$ is normally distributed with mean $\gamma$ and variance $\sigma^2(\hat{\gamma})$. The density function for $\hat{\gamma}$ is

$$f(\hat{\gamma}) = \frac{1}{\sigma(\hat{\gamma})\sqrt{2\pi}}\, e^{\left[-\frac{1}{2}(\hat{\gamma} - \gamma)^2/\sigma^2(\hat{\gamma})\right]}$$

Then the density function for $\hat{\kappa}$ is

(11) $$f(\hat{\kappa}) = \frac{c}{2\sigma(\hat{\gamma})\sqrt{2\pi}}\,(a-\hat{\kappa})^{-\frac{3}{2}}\, e^{\left\{-\frac{1}{2}\left[(a-\hat{\kappa})^{-\frac{1}{2}} - \gamma\right]^2/\sigma^2(\hat{\gamma})\right\}}$$

where c is a scalar whose purpose is to ensure

$$\int_{-\infty}^{a} f(\hat{\kappa})d\,\hat{\kappa} = 1 .$$

The range of $\hat{\gamma}$ is $-\infty \leq \hat{\gamma} \leq \infty$ so that the range of $\hat{\kappa}$ is $-\infty \leq \hat{\kappa} \leq a$.

The density function (1) is somewhat cumbersome. It may be more useful, therefore, to consider the asymptotic confidence limits for $\hat{\kappa}$. We will show the 95% confidence limits but other limits are easily found. Asymptotically, we know that

$$\Pr(\gamma_1 \leq \hat{\gamma} \leq \gamma_2) = .95 \quad \text{when} \quad \begin{aligned} \gamma_1 &= \gamma - 1.96\,\sigma(\hat{\gamma}) \\ \gamma_2 &= \gamma + 1.96\,\sigma(\hat{\gamma}). \end{aligned}$$

Therefore,

$$\Pr(\hat{\gamma}^2 \leq \gamma_2^2) = \Pr\left(\frac{1}{a - \hat{\kappa}} \leq \gamma_2^2\right)$$

$$= \Pr\left(\hat{\kappa} \leq a - \frac{1}{\gamma_2^2}\right)$$

and

$$\Pr\left(\hat{\kappa} \leq a - \frac{1}{[\gamma + 1.96\,\sigma(\hat{\gamma})]^2}\right) = .95 \tag{12}$$

Sawa [3] has shown that the moments of the k class estimator do not exist for values of k greater than one.[2] Under these circumstances, of course, the MSM estimator does not exist. In order to ensure that $k \leq 1$ we must have

---

[2] Sawa's finding also has implications for the unbiased k class (UBK) estimator which is obtained by setting $\kappa = K_2 - m - 1$ ([3], p. 579). Since $K_2$ and m are both positive integers, and since we must have $\kappa \leq 0$ for UBK to exist, this estimator is restricted to the cases where:

i) $K_2 = m$ and equation in question is just identified;

ii) $K_2 - m = 1$ so that $\kappa = 0$ and UBK is 2SLS.

In case (i) above UBK is a good substitute for 2SLS which does not have finite moments in the just identified case.

$\hat{\varkappa} \le 0$. Therefore, we are interested in knowing under what circumstances the right side of the inequality in (12) is, or is not, zero (or negative). We can identify two cases.

i) $a \le 0$ or $K_2 - K_1 \le 5$. In this case we can see by (10) that $\varkappa$ is certainly negative and we can use the MSM estimator with confidence.

ii) $a > 0$. Since a is an integer number this is equivalent to $a \ge 1$ or $K_2 - K_1 \ge 6$. In this case, which is more likely with a large model, we must rely on the term $[\gamma + 1.96\ \sigma(\hat{\gamma})]$ to produce a zero (or negative) limit in (12). If this term is small enough, i.e., somewhat less than 1, we would have $\frac{1}{[\gamma + 1.96\ \sigma(\hat{\gamma})]^2} > a$. The term in question will be small if; both $\gamma$ and $\sigma(\hat{\gamma})$ are small, or if $\gamma$ is negative and about the same size as $\sigma(\hat{\gamma})$. Clearly there is a large set of possibilities outside these two conditions so that the use of MSM with a positive a is not recommended.

## 5. A Non-Stochastic Approximation to $\varkappa$

The MSM estimator was derived under the assumption that $\varkappa$ is non stochastic. The previous section shows that, under some conditions we could obtain estimates of $\varkappa$ which would lead to MSM being undefined. In addition, the approximate small sample standard errors of the MSM estimates [1] also depend upon $\varkappa$ being non stochastic. Preliminary Monte Carlo evidence suggests that the use of a stochastic $\hat{\varkappa}$ can lead to occasional computed values for the small sample MSM standard errors that are much worse estimates of the standard deviation of the sampling distribution than are the computed asymptotic standard errors. For these reasons we were induced to seek a non-stochastic approximation to $\kappa$.

From (10) we see that, since a is always known once the model has been specified, the stochastic nature of $\hat{\kappa}$ arises from the necessity to estimate $\gamma^2$. However, if $\gamma$ were large enough the term $\frac{1}{\gamma^2}$ would be small and could safely be ignored. The size of $\gamma$ depends upon the units in which $y_1$ is measured. Therefore, for $a \leq 0$, we obtain a non-stochastic approximation by scaling the variable $y_1$ by $10^{-n}$ and letting $\kappa \doteq a$. The choice of the scale factor n will be influenced by the sample size (after scaling $\frac{1}{\gamma^2}$ should be about the same size as $\frac{1}{T}$) and the researcher's prior notions about the size of $\gamma$.

## 6. Monte Carlo Results

The population used in the Monte Carlo experiments was:

$$y_1 = .7\, y_2 - .8\, x_1 - .7\, x_2 + u_1$$

$$y_2 = - 2.5\, y_1 - 1.5\, x_3 + 1.0\, x_4 + u_2$$

$$\Sigma = \frac{1}{T} E\, U'U = \begin{bmatrix} 1.49 & 1.80 \\ 1.80 & 7.25 \end{bmatrix}$$

$$\Omega = \frac{1}{T} E\, V'V = \begin{bmatrix} 1.0 & 0 \\ 0 & 1.0 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 2000 & 0 & 0 & 0 \\ 0 & 2000 & 0 & 0 \\ 0 & 0 & 2000 & 0 \\ 0 & 0 & 0 & 2000 \end{bmatrix}$$

One hundred samples of size 20 were generated from the reduced form using sets of disturbances V which were random drawings from a bivariate normal distribution with zero means and covariance $\Omega$. The 20 x 4 matrix X of exogenous data is the first four columns of a 20 x 20 orthogonalized matrix of random drawings from a uniform distribution. X is constant over the 100 samples. Each $x_i$ has zero mean.

Each sample was used to compute three kinds of MSM estimates using three different versions of $\varkappa$; population $\varkappa = a - \frac{1}{\gamma^2}$, a 2SLS estimate of $\varkappa = \hat{\varkappa} = a - \frac{1}{\hat{\gamma}^2}$ and a non-stochastic approximation to $\varkappa = \tilde{\varkappa} = a$ with the right-hand endogenous variable multiplied by $10^{-1}$. The standard errors of each set of coefficient estimates were computed by both the asymptotic formula and by Nagar's finite sample formula [1]. This procedure produces frequency distributions of coefficient estimates, and standard error estimates, by each of the three MSM estimators and a frequency distribution of $\hat{\varkappa}$.

The frequency distribution of $\hat{\varkappa}$ may give some hint as to the form of the finite sample distribution of this statistic. Table 1 gives summary statistics of the empirical distribution

Table 1

Frequency Distribution of $\hat{\varkappa}$: Summary Statistics

| | Equation 1 | Equation 2 |
|---|---|---|
| Population Value | -7.04 | -5.16 |
| Minimum | -7.84 | -5.27 |
| Maximum | -6.36 | -5.10 |
| Mean | -7.06 | -5.16 |
| Standard Deviation | .266 | .0338 |
| Skewness | - .140 | - .624 |
| Kurtosis | 3.16 | 3.40 |

The quality of $\hat{\kappa}$ from equation 2 is somewhat better than that from equation 1. For the second equation the mean of the frequency distribution is equal to the population value and the dispersion is small. Of course $\gamma$ for equation 2 is -2.5 compared to a $\gamma$ of .7 for equation 1. Therefore, the portion of $\hat{\kappa}$ which is stochastic, $\frac{1}{\hat{\gamma}^2}$ is much smaller for equation 2 than for equation 1. For both equations the range of $\hat{\kappa}$ is entirely negative so the MSM estimator exists for both cases.

The square of the skewness coefficient and the kurtosis coefficient together define which one of Pearson's curves best describes our empirical distribution. Reference to a chart by Pearson and Hartley ([2], p.234) shows that $\hat{\kappa}$ for equation 2 follows a type I curve. Type I curves have finite upper and lower bounds. $\hat{\kappa}$ has a finite upper bound but an infinite lower bound. Type IV curves have infinite upper and lower bounds. The result for equation 2 is, therefore, more credible than that for equation 1.

Pearson and Hartley also provide a table giving upper 5% points under Pearson curves ([2], p.230) for various skewness and kurtosis values. Inspection of this table shows that for equation 1 $\Pr\left(\frac{\hat{\kappa} - \mu_{\hat{\kappa}}}{\sigma_{\hat{\kappa}}} < 1.64\right) = .95$ and for equation 2 $\Pr\left(\frac{\hat{\kappa} - \mu_{\hat{\kappa}}}{\sigma_{\hat{\kappa}}} < 1.82\right) = .95$. If we substitute the population values of $\kappa$ for $\mu_{\hat{\kappa}}$ and the standard deviation of the frequency distribution for $\sigma_{\hat{\kappa}}$ we obtain for equation 1 $\Pr(\hat{\kappa} < - 6.61) \doteq .95$ and $\Pr(\hat{\kappa} < -5.10) \doteq .95$ for equation 2. We, of course, would expect the empirical confidence limit to be negative because a = -5 for both equations of this model.

In Table 2 we present the empirical bias and empirical root mean square error for MSM using the three alternate methods of deriving $\hat{\kappa}$,

Table 2

Performance of MSM with Alternative $\varkappa$ Values

| Coefficient | Population $\varkappa$ Bias | Population $\varkappa$ RMSE | $\hat{\varkappa}$ by 2SLS Bias | $\hat{\varkappa}$ by 2SLS RMSE | $\tilde{\varkappa} = K_1 - 2K - 5$ Bias | $\tilde{\varkappa} = K_1 - 2K - 5$ RMSE |
|---|---|---|---|---|---|---|
| $\gamma_2 = .7$ | -.00201 | .0452 | -.00449 | .0461 | -.00122 | .0454 |
| $\beta_1 = -.8$ | -.000214 | .0443 | .00160 | .0448 | -.000784 | .0444 |
| $\beta_2 = -.7$ | .00424 | .0351 | .00580 | .0354 | .00374 | .0352 |
| $\gamma_1 = -2.5$ | .0677 | .247 | .0678 | .247 | .0654 | .247 |
| $\beta_3 = -1.5$ | .0306 | .145 | .0306 | .145 | .0297 | .145 |
| $\beta_4 = 1.0$ | -.0165 | .114 | -.0165 | .114 | -.0159 | .114 |

Of course the three versions of MSM we deal with are the same asymptotically. The figures in Table 2 suggest that for finite samples the use of our non-stochastic approximation to $\varkappa$ (with appropriate scaling of the right hand endogenous variable) gives estimates that are better than those obtained when $\varkappa$ is estimated by 2SLS; every bias is smaller and 3 out of 6 RMSE values are smaller for $\tilde{\varkappa}$ than for $\hat{\varkappa}$. Indeed, in terms of bias $\tilde{\varkappa}$ outperforms even population $\varkappa$ for every parameter. For the second equation all methods were judged equally good on the basis of RMSE.[3] In general we would expect that where there are significant differences in the performances of the three MSM versions $\tilde{\varkappa}$ will give better estimates than $\hat{\varkappa}$ but where there are only small differences in performance the ranking is not clear.

When $\hat{\varkappa}$ (by 2SLS) was used to compute the small sample MSM standard errors the results were observed to become unstable for cases where $\hat{\varkappa}$ had a fairly

---

[3] We do not consider it useful to report results to more than 3 significant digits.

large variance. For instance, for equation 1, whose $\hat{\kappa}$ (from Table 1) was widely dispersed, we obtained finite MSM standard errors some 50 times as large as the average in 1 case out of 100 for each coefficient. This did not occur with equation 2, for which $\hat{\kappa}$ had a very small variance, nor did it occur when a non-stochastic $\kappa$ is used. That is, this problem was solved by the employment of our approximation $\tilde{\kappa}$

## 7. Conclusion

We conclude that the optimum value of k, when there is one endogenous variable on the right hand side, is given by

$$(13) \quad k = \begin{cases} 1 + \dfrac{K_2 - K_1 - 5}{T} ; \text{ if } K_2 - K_1 \le 5 \text{ . } y_1 \text{ to be scaled.} \\ 1; \text{ if } K_2 > 5. \end{cases}$$

## References

[1] Nagar, A. L., "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations," _Econometrica_, vol. 17, no. 4 (Oct. 1949).

[2] Pearson, E. S. and H. O. Hartley, _Tables for Statisticians and Econometricians_, vol. I, Cambridge, Cambridge University Press, 1966.

[3] Sawa, J., _Finite Sample Properties of the k-Class Estimators_, mimeographed.