Computer Science in Economics and Management 1 (1988) 21–30.

Two- or Three-Stage Least Squares?

DAVID A. BELSLEY*
Department of Economics, Boston College, Chestnut Hill, MA 02167, U.S.A.

(Received: December 1987)

Abstract. Two elements enter the choice between 2 and 3SLS for full-system estimation: statistical efficiency and computational cost. 2SLS always has the computational edge, but 3SLS can be more efficient, a relative advantage that increases with the strength of the interrelations among the error terms. A measure of these interrelations is thus helpful in making the choice, and, when there are only two equations, this has suggested using a high pairwise error correlation as an indicator of when to use 3SLS. In larger systems of equations, however, these pairwise correlations can remain small even though more general interrelations give 3SLS the relative advantage. More general indicators are therefore needed, and this paper suggests three such and demonstrates their efficacy.

Key words. Condition number, determinant, eigenvalues, multiple correlations, relative efficiency, simultaneous-equations estimation, singular values.

1. Introduction

In estimating a system of simultaneous equations, the question often arises whether to use two or three stages of least squares. Clearly 2SLS is computationally cheaper, and, whereas 3SLS is known asymptotically to be more efficient, this need not be so for small samples. 3SLS, then, becomes the estimator of choice only when (1) the researcher considers a gain in efficiency to be important relative to computational cost and (2) when the potential for such a gain is high. This paper suggests measures for this potential that are at once both simple and more general than the previously suggested indicator, high pairwise correlations.

That 3SLS need not possess greater efficiency than 2SLS for small samples is readily motivated as follows: it is well known that 2SLS and 3SLS are equivalent when there is no cross-equation covariation (Theil, 1971). 3SLS's asymptotic efficiency arises, then, from exploiting nonzero cross-equation covariation. In practice, of course, samples are finite and this cross-equation covariation must be estimated. Thus, when the true (but unknown) cross-equation covariation is small, it can be more efficient to impose the restriction that it is zero, which is what 2SLS does, than to use an estimate of it, as does 3SLS. As a result, one would expect 2SLS to be more efficient when the cross-equation covariation is small and for 3SLS to become more worthwhile as this covariation becomes larger. Indeed, for a two-equation simultaneous system,

* Professor of Economics, Boston College, and Principal Research Associate, Center for Computational Research in Economics and Management Science, MIT. All computation was done on the TROLL system at MIT. My thanks go to Josh Charap for his able research assistance. This research was sponsored in part by the National Science Foundation under grant #IST-8420614.

Mikhail(1975) has demonstrated with Monte Carlo studies that 2SLS had the smaller mean square error when the between-equation correlation was $\varrho = .18$, but 3SLS became the winner when ϱ was .76.

It would seem, then, that estimates of the cross-equation correlation coefficients should indicate when 3SLS was likely to be worthwhile. And indeed high cross-equation correlations are sufficient indicators of this condition, but they are not necessary. For, as we shall see, in larger systems of equations, it is quite possible for all cross-equation correlations to be small even though the equations' error terms are tightly linked through more general multiple correlations. 2SLS, of course, continues to ignore these more general relations, while 3SLS does not. Hence, 3SLS can have greater small-sample efficiency than 2SLS even when pairwise correlations are small. A more general measure than pairwise correlations is therefore needed if one is to be apprised of when this is so, and this paper suggests several such generalizations.

The next section demonstrates the inadequacy of pairwise correlations for indicating the presence of more general correlations among the error terms. The subsequent section presents three measures that are more appropriate indicators of the presence and strength of these multiple relationships. Next, Monte Carlo results are described which demonstrate the efficacy of these suggested measures. A concluding section provides speculative considerations for further research.

2. The Inadequacy of Pairwise Correlations

Consider the general linear model of G simultaneous equations

$$\mathbf{y}^T(t)\mathbf{\Gamma} + \mathbf{x}^T(t)\mathbf{B} = \mathbf{u}^T(t), \tag{1}$$

where $\mathbf{y}$ is a G-vector of endogenous variables, $\mathbf{x}$ is a K-vector of exogenous variables, and $\mathbf{u}$ is a G-vector of error terms distributed with mean $\mathbf{0}$ and variance-covariance matrix $\mathbf{\Sigma}$. We denote by $\mathbf{R}$ the correlation matrix corresponding to $\mathbf{\Sigma}$ (i.e. $\mathbf{R} \equiv \mathbf{D}^T\mathbf{\Sigma}\mathbf{D}$, where $\mathbf{D}^{-1} \equiv \operatorname{diag} \mathbf{\Sigma}^{1/2}$).

Our purpose in this section is simply to show that it is possible for the off-diagonal elements of $\mathbf{R}$ (the pairwise correlations between the elements of $\mathbf{u}$) to be small even when there is more generally a strong multiple correlation among the elements of $\mathbf{u}$ (for example, some element of $\mathbf{u}$ is highly correlated with some linear combination of the other elements of $\mathbf{u}$). This demonstration is important, of course, to show that the off-diagonal elements of $\mathbf{R}$ are a sufficient, but not necessary, indicator of the potential for 3SLS.

Assume, then, a perfect relation among the elements of $\mathbf{u}$, that is, some $\mathbf{c} \neq \mathbf{0}$ such that $\operatorname{var}(\mathbf{c}^T\mathbf{u}) = 0$. A little reflection will convince the reader that, under these conditions, we can minimize the largest absolute pairwise correlation among the elements of $\mathbf{u}$ by making them all equal. This is seen for $G = 3$ in Figure 1 where we have plotted the variates $\mathbf{u} \equiv (u_1, u_2, u_3)$. The perfect linear relation among the u's is depicted by their lying in a two-dimensional space, and the pairwise correlations between the elements of $\mathbf{u}$ are depicted by the angles between them. (A little elementary

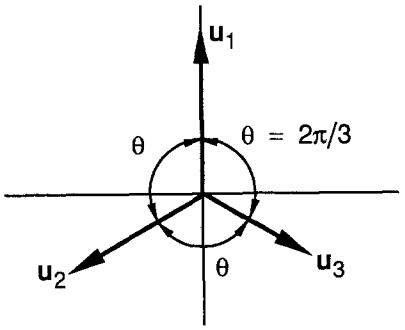

Fig. 1.

trigonometry will show that, if ζ_1 and ζ_2 are vectors representing two random variables with mean zero, then $\text{cor}(\zeta_1, \zeta_2) = \cos\theta$, where θ is the angle between ζ_1 and ζ_2.) It is clear that, starting from the equal-angle situation depicted, no one correlation may be made smaller without making another one larger. Thus, the largest absolute pairwise correlation here may be as small as $.5 = |\cos(2\pi/3)|$.

Indeed, more generally, we find for the G-vector $\mathbf{u}$ that it is possible for there to be a perfect linear relation among its elements while the largest absolute pairwise correlation can be as small as $1/(G-1)$. Thus, if there are 20 equations in the system, there could be perfect cross-equation covariation even though no two equation's error terms were absolutely correlated by more than $1/19 = \approx .05$.

This result is readily seen by noting that the equi-correlation matrix

$$\mathbf{R}^*(\varrho) \equiv \begin{bmatrix} 1 & \varrho & \varrho & \cdot & \cdot & \cdot & \varrho \\ \varrho & 1 & \varrho & \cdot & \cdot & \cdot & \varrho \\ \varrho & \varrho & 1 & \cdot & \cdot & \cdot & \varrho \\ \cdot & \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot & \cdot \\ \varrho & \varrho & \varrho & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

is (a) singular for $\varrho = -1/(G-1)$, and (b) positive definite for $-1/(G-1) < \varrho < 1$. Both of these properties follow directly from the fact that the G eigenvalues of $\mathbf{R}^*(\varrho)$ are $1-\varrho$ (with multiplicity $G-1$) and $1+(G-1)\varrho$, all of which are positive for $-1/(G-1) < \varrho < 1$ and some of which are zero for $\varrho = 1$ and $\varrho = -1/(G-1)$ (Rao, 1973).

3. Three Indicators for 3SLS.

From the preceding, we know that small absolute pairwise correlations between the error terms in a simultaneous system of equations need not indicate the adequacy of 2SLS; more general relations may still exist among the elements of $\mathbf{u}$ that could give

3SLS the edge. To detect these more general dependencies, a measure more general than pairwise correlations is needed. Three such measures are suggested here, all based on the error correlation matrix $\mathbf{R}$: its determinant $\det(\mathbf{R})$, its smallest eigenvalue $\lambda_{\min}$, and its condition number $\kappa(\mathbf{R})$.

$\det(\mathbf{R})$. Clearly, $\det(\mathbf{R}) = 0$ if and only if there is some $\mathbf{c} \neq \mathbf{0}$ such that $\operatorname{var}(\mathbf{c}^T\mathbf{u}) = 0$. This situation defines quite generally what we have loosely been calling a perfect linear relation among the elements of $\mathbf{u}$. Also we note that $\det(\mathbf{R}) = 1$ when there is no correlation of any sort among the elements of $\mathbf{u}$, for here $\mathbf{R} = \mathbf{I}$. In general, $0 \leqslant \det(\mathbf{R}) \leqslant 1$. The first inequality follows simply from $\mathbf{R}$'s being positive definite. The second follows since $\det(\mathbf{R}) = \Pi_i \lambda_i$ and $\operatorname{tr}(\mathbf{R}) = \Sigma_i \lambda_i = G$, where the λ_i are the eigenvalues of $\mathbf{R}$. It is straightforward to show that the maximum of a product of nonnegative magnitudes whose sum is constant occurs when their values are equal, and, in this instance, this means each $\lambda_i = 1$. $\det(\mathbf{R})$, then, is an index of general multiple correlation. The closer is $\det(\mathbf{R})$ to zero, the tighter the existence of some general correlation among the elements of $\mathbf{u}$. The closer is $\det(\mathbf{R})$ to unity, the smaller any such general level of correlation.

$\lambda_{\min}$. A second indicator arises from the recognition that a tight correlation among the elements of $\mathbf{u}$ occurs if there exists a unit-vector $\mathbf{c}$ such that $\operatorname{var}(\mathbf{c}^T\mathbf{u})$ is small. Without normalization, however, this notion has a major drawback: for a given $\mathbf{c}$, without changing any of the angles (pairwise correlations) among the elements of $\mathbf{u}$, we can make $\operatorname{var}(\mathbf{c}^T\mathbf{u})$ as small or as large as we wish merely by rescaling these u_i's. To normalize the problem, we consider $\operatorname{var}(\mathbf{c}^T\boldsymbol{\eta})$ of the standardized variables

$$\boldsymbol{\eta} = (\eta_1, \eta_2, \eta_3) \equiv \{u_1/\sqrt{\operatorname{var}(u_1)}, u_2/\sqrt{\operatorname{var}(u_2)}, u_3/\sqrt{\operatorname{var}(u_3)}\}.$$

Our interest is in examining when the minimum of

$$\operatorname{var}(\mathbf{c}^T\boldsymbol{\eta}) = \mathbf{c}^T\mathbf{R}\mathbf{c} \tag{2}$$

is small subject to $\mathbf{c}^T\mathbf{c} = 1$. The solution to this problem is a familiar result for eigenvalues of real symmetric matrices, namely that the minimum of (2) subject to $\mathbf{c}^T\mathbf{c} = 1$ is $\lambda_{\min}$, the minimum eigenvalue of $\mathbf{R}$ (Rao, 1973). Thus, a tight correlation exists among the elements of $\mathbf{u}$ when $\lambda_{\min}$ is close to zero. Furthermore, it is readily shown that $\lambda_{\min} \leqslant 1$ and that it assumes this upper bound of 1 if and only if $\mathbf{R} = \mathbf{I}$, that is, if and only if there is an absence of any correlation. This first result follows from the facts that $\operatorname{tr}(\mathbf{R}) = \Sigma_i \lambda_i = G$ and that $\lambda_i > 0$ for all i; the second reflects the necessary existence of an orthogonal matrix $\mathbf{C}$ giving $\mathbf{C}^T\mathbf{R}\mathbf{C} = \boldsymbol{\Lambda} \equiv \operatorname{diag}(\lambda_1, \ldots, \lambda_G)$. Thus, $\lambda_{\min}$ also behaves like a measure of general correlation, ranging between zero and unity, with the extremes determining perfect correlation and noncorrelation, respectively.

$\kappa(\mathbf{R})$. A third measure of general correlation, related to $\lambda_{\min}$, is the condition number (Belsley, Kuh, and Welsch, 1980) of $\mathbf{R}$

$$\kappa(\mathbf{R}) \equiv \frac{\lambda_{\max}}{\lambda_{\min}} \geqslant 1. \tag{3}$$

The motivation of $\kappa(\mathbf{R})$ is similar to that given $\lambda_{\min}$ above except here we accept the presence of a general multiple correlation among the elements of $\mathbf{u}$ (equivalently $\boldsymbol{\eta}$) if

there are two unit-vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ such that $\text{var}(\mathbf{c}_1^T\boldsymbol{\eta}) \ll \text{var}(\mathbf{c}_2^T\boldsymbol{\eta})$. We already know that (2) is minimized at $\lambda_{\min}$, and it is similarly shown that (2) is maximized at $\lambda_{\max}$. Hence, a strong multiple correlation among the elements of $\mathbf{u}$ occurs when $\kappa(\mathbf{R})$ is large.* Further, it is clear that $\kappa(\mathbf{R})$ assumes its minimum value 1 when $\mathbf{R} = \mathbf{I}$. (No simple transformation of $\kappa(\mathbf{R})$ provides a meaningful correlation-like behavior. One might try to base such a transform on a single-parameter matrix like $\mathbf{R}^*(\varrho)$ defined in the previous section. Recalling that the roots of $\mathbf{R}^*(\varrho)$ are $1 - \varrho$ with a multiplicity $G - 1$ and $1 + (G - 1)\varrho$, for $\varrho > 0$ we have $\kappa(\mathbf{R}^*(\varrho)) = (1 + (G - 1)\varrho)/(1 - \varrho)$, while for $\varrho < 0$, $\kappa(\mathbf{R}^*(\varrho)) = (1 - \varrho)/(1 + (G - 1)\varrho)$. Only for $G = 2$ is $\mathbf{R}^*(\varrho)$ always the actual correlation matrix $\mathbf{R}$ and does a symmetric relation exist between k($\mathbf{R}$) and ϱ, namely $\kappa(\mathbf{R}) = (1 + |\varrho|)/(1 - |\varrho|)$ or $\varrho = (\kappa(\mathbf{R}) - 1)/(\kappa(\mathbf{R}) + 1)$.)

The detailed behavior of these three indicators in differing situations is a matter of empirical research that is beyond the scope of this paper. Some specifics are, however, clear. $\lambda_{\min}$ necessarily provides the least information since it is based only on one of $\mathbf{R}$'s eigenvalues. det($\mathbf{R}$), by contrast, is the product of all $\mathbf{R}$'s eigenvalues, and this measure can become very small through the joint presence of several modest eigenvalues, no one of which need be very small. Likewise, $\kappa(\mathbf{R})$ must necessarily tell different stories about situations that would be treated the same by $\lambda_{\min}$. Using the same information that showed $\lambda_{\min} \leqslant 1$, we can show that $\lambda_{\max} \leqslant G$. For a given $\lambda_{\min}$, then, $\kappa(\mathbf{R})$ must be greater the fewer are the number of other small λ's, that is, the smaller the number of different multiple correlations that coexist among the elements of $\mathbf{u}$.

In any event, it is reasonable to suppose that these several measures, based as they are on the correlation matrix $\mathbf{R}$ of the error structure of the system (1), are more general indicators of the potential for gains in efficiency through using 3SLS. Presumably, the closer is any one of them toward indicating a strong general correlation among the u's, the greater is this potential, and this regardless of the absolute magnitudes of the pairwise correlations.

4. A Monte Carlo Experiment

In this section a Monte Carlo experiment is conducted to demonstrate the previously described phenomena, namely, (1) that 3SLS can have greater small-sample efficiency than 2SLS despite low cross-equation correlations so long as a strong, more general correlation exists, and (2) that this situation can be effectively assessed by the indicators suggested above. No attempt at completeness is made here. It is the purpose of this study only to demonstrate these phenomena and to suggest tentative conclusions.

* Readers familiar with Belsley, Kuh, and Welsch (1980) (BKW) will recognize the parallel between the development here and that in chapter 3 of BKW. The standardization to the η_i's employed above is exactly analogous to the column equilibration used there. Indeed, the argument of Appendix 3C of BKW can be used directly to show that, among all the possible scalings of the u_i's, that which produces the correlation matrix $\mathbf{R}$ is the one that is guaranteed to produce a condition number that is nearly minimum, and therefore most meaningful for our current needs. More generally, the dualism that exists between a variable space like that used here and an observation space like that used in BKW is examined in Dempster (1969).

A final section for speculative considerations, however, suggests those areas where further research will be most fruitful.

THE MODEL

The following six-equation model is employed in the Monte Carlo experiment:

$$\begin{aligned}
y_1 &= 5 + 0.3y_2 + 0.5y_6 - 0.6x_1 + 1.0x_2 + u_1, \\
y_2 &= 10 + 0.5y_1 - 0.3x_3 + 0.7x_4 + u_2, \\
y_3 &= -6 - 0.4y_1 + 0.2y_4 + 1.0x_5 + 0.5x_6 + u_3, \\
y_4 &= 40 + 0.5y_3 - 2.0x_2 + 0.6x_3 + u_4, \\
y_5 &= -9 + 0.4y_3 - 0.3y_6 + 0.6x_4 + 0.3x_5 + u_5, \\
y_6 &= 15 + 0.2y_5 + 0.7x_1 + 0.2x_6 + u_6,
\end{aligned} \tag{4}$$

with $\mathbf{u} \sim N_6(0, \Sigma)$.

This model was chosen to be large enough ($G = 6$) to be interesting, to insure each equation to be overidentified, and to be close to a model that could occur in common econometric practice.

THE EXPERIMENT

The equi-correlation matrix introduced above,

$$\mathbf{R}^*(\varrho) = (1 - \varrho)\mathbf{I} + \varrho\boldsymbol{\iota}\boldsymbol{\iota}^T, \tag{5}$$

affords an excellent test environment for this study, for we can move smoothly from a situation where the u's are completely independent when $\varrho = 0$ to one where they possess a single perfect multiple correlation as ϱ goes to $-1/(G - 1) = -.2$, recalling here that $G = 6$. Thus, the largest absolute pairwise correlation never exceeds .2 while all the other indicators attain their extreme values: $\lambda_{\min} \to 0$, $\det(\mathbf{R}^*(\varrho)) \to 0$, and $\kappa(\mathbf{R}^*(\varrho)) \to \infty$. In the experiment to follow, then, nine values of ϱ are chosen to span this range: $\varrho = 0, -.02, -.05, -.07, -.10, -.125, -.15, -.19$ and $-.199$.

For each value of ϱ, a run of 100 replications is made. Each replication consists of 40 observations whose y's are generated subject to (4) for a fixed set of x's and whose u's are generated according to $N_6(\mathbf{0}, \boldsymbol{\Sigma})$. $\boldsymbol{\Sigma}$ is determined to have a corresponding correlation matrix (5) and to have variances that produce a signal-to-noise of 20 for the structural equations (4). This value is chosen to mimic good-quality economic data providing R^2's in the neighborhood of .9 for the estimated structural equations. Following Mikhail (1975), the x's are chosen randomly from uniform distributions of differing ranges. In particular, the x_i's are chosen, respectively, from uniforms 5–10, 0.5–5.0, 15–30, 2–8, 8–22, and 10–14.

The data for each replication are used to estimate (4) by both 2SLS and 3SLS, providing for each ϱ 100 estimates of each of the 27 parameters by each estimator.

THE RESULTS

Figures 2–4 summarize the results. For each ϱ, a root-mean-square-error (RMSE) about the true parameter value is calculated for each parameter and each estimator. A relative RMSE is then calculated as RRMSE $\equiv$ $\text{RMSE}_{2SLS}/\text{RMSE}_{3SLS}$. Clearly, a RRMSE $>$ 1 favors 3SLS, while RRMSE $<$ 1 favors 2SLS.

Figure 2 shows for each ϱ the number of coefficients out of the 27 for which 3SLS proved superior (RRMSE $>$ 1). A line between 13 and 14 is drawn to denote the half-way point. Figure 3 shows for each ϱ the mean relative RMSE (across the 27 parameters) as well as the largest and smallest relative RMSE. A line is drawn at 1.0; mean relative RMSE's above this line denote when, on the average, 3SLS outperforms 2SLS and inversely.

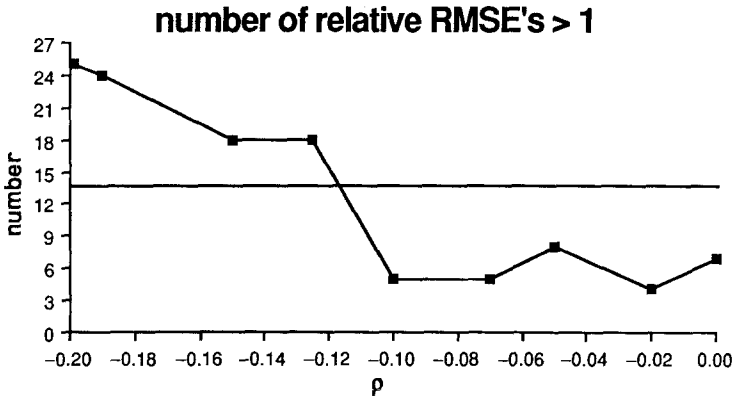

Fig. 2.

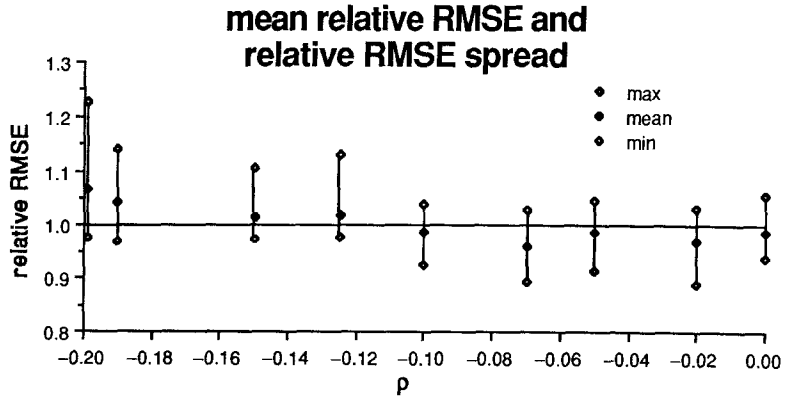

Fig. 3.

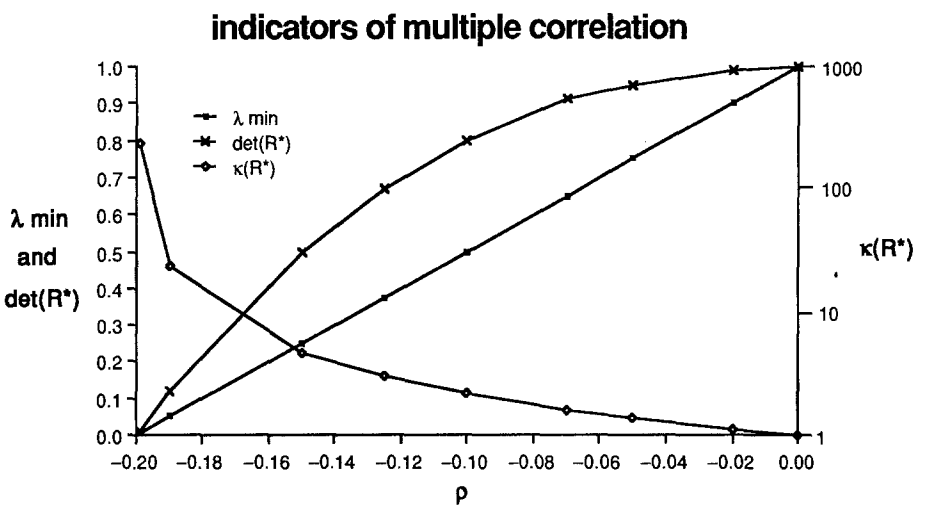

Fig. 4.

Both of these summary figures show that 3SLS breaks even with 2SLS for ϱ somewhere between $-.1$ and $-.125$ and is wholly in the lead by the time ϱ approaches $-.2$. It is clear that this situation could not be assessed solely from examination of the low magnitude of ϱ. Examination of Figure 4, however, shows that all three of the suggested general indicators point to this outcome. Taking very conservative values, for example, we note that 3SLS is certainly dominant by the time $\lambda_{\min}$ has dropped to .05, det(**R***) has dropped to .12, and κ(**R***) has exceeded 23. These values are all extremes in their respective ranges, while they correspond to an absolute ϱ of only .19, a value most would consider small indeed.

5. Conclusions and Speculative Considerations

Consonant with expectations, an indicator more general than pairwise correlations among the errors of a system of equations is necessary to determine when 3SLS is likely to posses greater small-sample efficiency than 2SLS and, thus, when the added computational cost of 3SLS is likely to be worthwhile. For larger systems of equations, the presence of strong multiple correlations among these error terms can give the edge to 3SLS even when the largest absolute pairwise correlation is small.

Any of the three suggested general indicators, $\lambda_{\min}$, det(**R**), or κ(**R**), would seem to provide this more appropriate information. While it is beyond the present study to provide a detailed comparison of these three measures or interpretation of their values, it would seem safe to say that 3SLS would possess good small-sample relative efficiency for values of $\lambda_{\min}$ and det(**R**) in the neighborhood of .1 and for values of κ(**R**) above 20–30. This latter magnitude is wholly consistent with values found in BKW (1980) to denote the presence of strong collinear dependencies among columns of a data matrix in linear regression.

It should be emphasized that the experimental situation employed in this study is narrow by design, serving only to demonstrate that pairwise correlations cannot adequately indicate when the added computation of 3SLS becomes worthwhile whereas the three suggested general indicators can. In this it succeeds, but it is not intended, and does not pretend, to provide definitive information about comparing and interpreting these three indicators in more general contexts. Such a necessarily substantial study would have to address the following issues:

(1) What are the general rules for how small $\lambda_{\min}$ and $\det(\mathbf{R})$ must be and how large $\kappa(\mathbf{R})$ must be before 3SLS has greater small-sample efficiency than 2SLS?
(2) How do these rules depend on
   (a) the number of equations, G?
   (b) the number of coexisting multiple correlations within $\mathbf{u}$? and additionally, on
   (c) sample size? and
   (d) degree of overidentification?
(3) Does the information from one of these indicators dominate that of another? or are there differing situations favoring specific indicators?
(4) Are there wide classes of econometric situations in which simple rules of interpretation can be effective?

Some speculation on these questions is in order. Issues 2c and 2d arise since it is well known that the performance of 3SLS always bests that of 2SLS as the sample size increases indefinitely, but that the two are equivalent as the degree of overidentification goes to zero. One could conjecture, then, that a stronger degree of intercorrelation among the errors would increase the attractiveness of 3SLS the smaller the sample size and the greater the degree of overidentification.

Issues 2a and 2b are of central interest. We can see from the eigenvalues of $\mathbf{R}^*(\varrho)$ that all three indicators depend on G. Different rules or thresholds may be needed, therefore, for different system sizes. It would obviously be nicer if single thresholds would apply (at least over a wide range of practical values for G).

In the example presented above, there was only one multiple correlation among the elements of $\mathbf{u}$ as $\varrho \to -.2$ (since only one of the G eigenvalues of $\mathbf{R}^*(\varrho)$ goes to zero). In typical econometric situations there could be coexisting multiple correlations. It is to be supposed that the more such relations, the greater the relative advantage for 3SLS. The three indicators, however, differ in their ability to assess such distinctions. Thus, $\lambda_{\min}$ can easily remain unchanged as the number of such multiple correlations increases while both $\det(\mathbf{R})$ and $\kappa(\mathbf{R})$ will respond to such alterations. For this reason, it is to be expected that these later two indicators will be more generally useful. Indeed, one can further bolster the value of $\kappa(\mathbf{R})$ in this regard by considering a full set of condition indexes for $\mathbf{R}$ à la BKW (1980) rather than the condition number $\kappa(\mathbf{R})$ alone.

References

Belsley, D. A., Kuh, Ed. and Welsch, R. E. (1980) *Regression Diagnostics: Identifying Influential Observations and Sources of Collinearity*, John Wiley, New York.

Dempster, A. P. (1969) *Elements of Continuous Multivariate Analysis*, Addison–Wesley, Reading, Mass.

Mikhail, W. M. (1975) A comparative Monte Carlo study of the properties of econometric estimators, *J. Am. Statistical Assoc.* **70**, 94–104.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, 2nd Edn, John Wiley, New York.

Theil, H. (1971) *Principles of Econometrics*, John Wiley, New York.