Discussion Paper No. 7921 November 1979

ON THE EFFECTS OF MULTICOLLINEARITY UPON THE PROPERTIES OF STRUCTURAL COEFFICIENT ESTIMATORS

Roberto S. Mariano
James B. McDonald
Asher Tishler

ABSTRACT

This paper considers an analytical investigation of multicollinearity in a simultaneous-equations model and focuses on coefficients of endogenous variables. Previous Monte Carlo studies tend to support the notion that higher multicollinearity among exogenous variables causes estimator precision to deteriorate. It is shown in this paper, however, that in the case of simple, partial and multiple correlations, higher multicollinearity can increase or decrease the mean squared error of estimators, depending upon the true model parameter values and the observations on the exogenous variables. Some special cases are identified where a higher degree of multicollinearity brings about less precise estimators.

The analysis leading to this indeterminacy of multicollinearity effects starts from the result that multicollinearity among the exogenous variables will affect the probability distributions of the LIML and k-class estimators (k nonstochastic and $0 \leq k \leq 1$) only through the so-called concentration parameter. Through numerical calculations of concentration parameter values in two simulation studies, we reconcile the apparent conflict between the conclusion from Monte Carlo experiments and the analytical result presented here.

The paper also contains some comments on an approach to making a choice among competing data sets for the exogenous variables. It also suggests a way of choosing additional observation vectors to increase estimator precision in simultaneous systems.

ON THE EFFECT OF MULTICOLLINEARITY UPON THE PROPERTIES OF STRUCTURAL COEFFICIENT ESTIMATORS*

*Roberto S. Mariano***
James B. McDonald
Ashler Tishler

I. INTRODUCTION

The effects of multicollinearity on the estimates of regression coefficients have been analyzed quite extensively in single-equation models. For examples, see Farrar and Glauber [1967], Feldstein [1973] , Klein and Nakamura [1962], Silvey [1969], and Smith [1974]. In this case, it is generally accepted that the higher the degree of multicollinearity, the less precise are the coefficient estimates.

In the context of simultaneous equation models, most of the research is restricted to the empirical side.[1/] Monte Carlo studies, such as those by Summers [1965] and Atkinson [1978] tend to suggest that the inverse relationship between the degree of multicollinearity and precision of structural coefficient estimates also applies in this case. However, the results of such studies are not necessarily applicable extensively over the entire parameter space and may be very specific to the set or sets of parameter values assumed in the experiment. Furthermore, there is the additional complication that conclu-

---

*For presentation at the annual meeting of the Econometric Society to be held in Atlanta, Georgia on December 28-30, 1979.

**The first author, Associate Professor of Economics at the University of Pennsylvania, acknowledges support from NSF Grant SOC79-07964 and the hospitality of the School of Economics, University of the Philippines, where he is a visitor for November-December, 1979.

2. MULTICOLLINEARITY AND THE CONCENTRATION PARAMETER

Consider the structural equation

$$y_1 = y_2\beta + Z_1\gamma + e \tag{2.1}$$

from a complete system of linear stochastic equations,[3] where y_1 and y_2 are $N \times 1$ vectors of endogenous variables, Z_1 is the $N \times K_1$ matrix of included endogenous variables, β (a scalar) and γ $(K_1 \times 1)$ are unknown coefficients and e is an $N \times 1$ vector of disturbance terms. Let $Z = (Z_1 \; Z_2)$ be the $N \times (K_1 + K_2)$ matrix of exogenous variables in the complete system.

The reduced form equations for y_1 and y_2 are

$$\begin{aligned} y_1 &= Z_1\pi_{11} + Z_2\pi_{12} + v_1 \\ y_2 &= Z_1\pi_{21} + Z_2\pi_{22} + v_2 \end{aligned} \tag{2.2}$$

where the error vector $(v_1' \; v_2')$ has a multivariate normal distribution with zero mean and covariance matrix $\Sigma \otimes I_n$. Σ is a 2×2 covariance matrix whose (i, j)th element is denoted by σ_{ij}.

Observations on exogenous variables appear in the probability distributions of the OLS, 2SLS, LIML (as well as all the k-class estimators with k non-stochastic, $0 \leq k \leq 1$) estimators of β only in the concentration parameter,[4] defined as:

$$\mu^2 = (\pi_{22}' \, Z_2' \, \bar{P}_{Z_1} \, Z_2 \, \pi_{22})/\sigma_{22} \tag{2.3}$$

R_{1j} = multiple correlation between Z_1 and the jth excluded exogenous variable,

$r_{ij.1}$ = partial correlation between the ith and jth excluded exogenous variables given Z_1,

λ_j = jth characteristic root of $Z_2'\overline{P}_{Z_1}Z_2$,

V = orthogonal matrix whose jth column is a characteristic vector of $Z_2'\overline{P}_{Z_1}Z_2$ corresponding to λ_j,

α = $V'\pi$.

The following give alternative expressions for the concentration parameter in terms of simple, multiple, and partial correlations as well as characteristic roots.

$$\tau^2 = \pi'Z_2'Z_2 \quad (1-R^2) \tag{2.5a}$$

$$\tau^2 = \pi'(C-A'B^{-1}A)\pi \tag{2.5b}$$

$$\tau^2 = \sum_{i=1}^{K_2} \pi_i^2(1-R_{1i}^2) + 2\sum_{i=1}^{K_2}\sum_{j=i+1}^{K_2} \pi_i\pi_j r_{ij.1}[(1-R_{1i}^2)(1-R_{1j}^2)]^{1/2} \tag{2.5c}$$

$$\tau^2 = \sum_{i=1}^{K_2} \alpha_i^2\lambda_i \tag{2.5d}$$

It follows from (2.5) that

$$\frac{\partial\tau^2}{\partial(Z'Z)} = \begin{pmatrix} B^{-1}A\pi\pi'A'B^{-1} & -2B^{-1}A\pi\pi' \\ -2\pi\pi'A'B^{-1} & 2\pi\pi' \end{pmatrix} \tag{2.6}$$

fashion in (2.6) and (2.7); in (2.8) and the lower right corner in (2.6) the signs of a pair of reduced form coefficients suffice to determine the direction of change in τ^2.

The lower right corner of (2.6) shows the influence of a change in the simple correlation between two excluded exogenous variables on the concentration parameter. A change in collinearity between a specific pair of excluded exogenous variables affects the size of π^2 in the same magnitude, regardless of the initial degree of collinearity between them. As noted in the preceding paragraph, the change itself can bring about an increase or a decrease in τ^2. This parallels the well-known analogous result in multiple regression.

The upper left corner of (2.6) shows the effects on τ^2 of changes in the correlation between two included exogenous variables. Such a change will affect τ^2 only if Z_1 and Z_2 are not orthogonal. In such cases that $Z_1'Z_2 \neq 0$, the effect on τ^2 depends on π and on the initial collinearity within Z_1 as well as on the collinearity between Z_1 and Z_2.

Another commonly-used indicator of multicollinearity (e.g., Summers [1965] and Atkinson [1979]) is the determinant of $Z'Z$. Note that

$$|Z'Z| = |Z_1'Z_1| \quad |Z_2'\overline{P}_{z_1} Z_2|$$

and that given the expression for μ^2 in (2.3), it would be more appropriate to consider the determinant of $Z_2'\bar{P}_{Z_1}Z_2$. But even here, it is only in special cases that an increase in the determinant of $Z_2'\bar{P}_{Z_1}Z_2$ would lead to an increase in μ^2 (regardless of the values of π). A necessary and sufficient condition for this to hold is that such an increase in $|Z_2'\bar{P}_{Z_1}Z_2|$ should not cause a decrease in any of the characteristic roots of $Z_2'\bar{P}_{Z_1}Z_2$. If at least one of the characteristic roots decreases, then the effect on μ^2 of such an increase in $|Z_2'\bar{P}_{Z_1}Z_2|$ may be positive or negative depending on values of α defined in (2.4). More specifically,

$$\frac{\partial\tau^2}{\partial|Z_2'\bar{P}_{Z_1}Z_2|} = \sum_{i=1}^{K_2} \frac{\partial\tau^2}{\partial\lambda_i} \cdot \frac{\partial\lambda_i}{\partial|Z_2'\bar{P}_{Z_1}Z_2|}$$

$$= \sum_{i=1}^{K_2} \alpha_i^2 \frac{\partial\lambda_i}{\partial|Z_2'\bar{P}_{Z_1}Z_2|} . \qquad (2.10)$$

3. CONCENTRATION PARAMETER VALUES IN SOME MONTE CARLO STUDIES

The indeterminate effect of multicollinearity, concluded in section 2, seems to clash directly with results in Summers and Atkinson which indicate that higher multicollinearity causes estimator precision to deteriorate. To reconcile these studies with the results of the preceding section, we present the calculated values of the concentration parameter (and other relevant parameters and parameter functions germane to the distributions of estimators).

The values of the concentration parameters reported for the Summers study correspond to his correctly specified models associated with a sample of size twenty and were obtained from the reported correlation matrices and calculated variances. His correctly specified model associated with a sample of size forty would have concentration parameters equal to twice the values reported in Table 1.[5/] The difference between Summers' A and B experiments is the "degree of multicollinearity." The determinant of the correlation matrix associated with the exogenous variables in experiment A is .76 and .0056 for experiment B. A similar measure of multicollinearity is used in Atkinson [1978]. Numerical values of this measure are included in Table 1.

The tabulation indicates that, with the exception of two cases, the choice of parameter values and exogenous observations is such that higher levels of multicollinearity are associated with lower values of the concentration parameter. There is little wonder then that, in the light of the proposition that MSE decreases with an increase in the concentration parameter, these studies would tend to support the inverse relationship between multicollinearity and estimator precision.

Table 1

CONCENTRATION PARAMETER VALUES AND MULTICOLLINEARITY IN SOME MONTE CARLO STUDIES

SUMMMERS (EXPERIMENTS 3 A,B)

	Equation 1		Equation 2	
Data	1	2	1	2
Concentration Parameter	146.3	7.11	90.24	48.12
Determinant of Correlation Matrix	.76	.0056	.76	.0056

ATKINSON

Concentration Parameters

	Equation 1				Equation 2		
Para-meters	Date Set 1	2	3	4	Data Set 1	2	3
1	90.56	60.43	42.39	11.81	35.49	32.41	15.16
2	40.58	16.87	3.36	1.68	143.09	130.71	61.13
3	31.11	29.13	17.84	3.55	12.91	7.29	7.40
4	2392.71	1104.97	746.54	276.10	132.33	105.59	68.56
5	2521.04	945.03	526.77	241.08	1139.38	531.10	653.88
6	32.85	18.51	13.08	4.14	15.38	12.24	3.82
Determinant of Correla-tion Matrix	.697	.110	.028	.003	.697	.110	.028

In the Atkinson study, there are two cases where an increase in multicollinearity level caused the concentration parameter to increase (data sets 2 and 3, parameter sets 3 and 5 in equation 2). These exceptions, however, are far too few to have any significant bearing on the aggregate conclusion obtained. But they serve to illustrate the indeterminacy discussed in the preceding section in the effects of multicollinearity. If more such cases had occurred, the evidence pointing to the relationship between multicollinearity and estimator precision would have been less conclusive.

4. PRESCRIPTIONS AND CONCLUSIONS

The study of multicollinearity presented here is by no means complete. However, some lessons and prescriptions can be gleaned from it. The analysis of multicollinearity effects through the concentration parameter led to the conclusion that higher multicollinearity may or may not be favorable to estimator precision. That Monte Carlo studies tend to show lower precision with higher multicollinearity is explained by the relative values of the concentration parameter implied by the choice of data sets reflecting varying degrees of multicollinearity.

The main interest really lies in the relation between precision of estimators on one hand and exogenous observations on the other. Instead of taking a detour through multicollinearity then, we might

FOOTNOTES

[1] The paper by Klein and Nakamura [1962] contains a different discussion of multicollinearity in simultaneous systems. They consider the occurrence of extreme or tail values in the probability distributions of ordinary least squares (OLS), two-stage least-squares (2SLS) and limited-information maximum likelihood (LIML) estimators. They reasoned that based upon computational characteristics it is likely for LIML to be more sensitive to the presence of multicollinearity than 2SLS, and 2SLS more sensitive than OLS.

[2] As far as we know, this conclusion has not been mentioned in the literature although it readily follows from available results, such as Richardson [1968], Richardson and Wu [1971], Sawa [1969, 1972], Mariano [1972, 1975], Mariano and Sawa [1972], Anderson and Sawa [1973] and Basman [1974].

[3] One can view multicollinearity as a specification problem and use this approach to resolve difficulties arising out of it. This paper does not go into issues involved in this approach. We assume that we have a properly specified simultaneous-equations model albeit containing exogenous variables which may be observed to be highly collinear.

[4] In some recent studies the concentration paramer is defined as

$$(\sigma^{11}\beta^2 + 2\beta\sigma^{12} + \sigma^{22}) \quad \pi_{22}'Z_2'\bar{P}_{z_1} Z_2\pi_{22}$$

where $(\sigma^{ij}) = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}^{-1}$ and the associated expressions are modifed accodingly, Anderson [1977]. Basman [1958] and McDonald [1970] consider a generalization of the concentration parameter fo structural equations containing more than two endogenous variable

It should be mentioned that in a given application μ^2 is unknown. The structural econometric estimation program ECOMP III (see Richardson and Rohr [1975]) calculates estimates of μ^2, $\hat{\mu}^2$, for each structural equation under consideration. These estimates are obtained from unrestricted maximum likelihood estimates obtai from the reduced form and are distributed as a non-central F stati with K_2 and N-K degrees of freedom with μ^2 as the associated non-centrality parameter, Kshirsagar [1972].

[5] If the actual exogenous data series is used and more than two significant digits are retained in the calculations, the value of the concentration parameters associated with the B experiment are $\mu_1^2 = 7.8$, and $\mu_2^2 = 48.6$. Comparable calculations could not be carried out for the A experiments due to the unavailability of the associated exogenous data series. Perhaps it should be mentio that the time path for Z_{t2} depicted in Figure 1 in Summers [1965 incorrect. This was verified with Summers by letter.

REFERENCES

Anderson, T.W. "Asymptotic Expansions of Estimates in Simultaneous Equations for Alternative Parameter Sequences," Econometrica, 45(1977), pp. 509-518.

Anderson, T.W. and T. Sawa. "Distributions of Estimates of Coefficients of a Single Equation in a Simultaneous System and Their Asymptotic Expansions," Econometrica, 41(1973), pp. 683-714.

Atkinson, E.S. "Small-Sample Properties of Simultaneous Equation Estimators with Multicollinearity," Journal of the American Statistical Association, 73(1978), pp. 719-723.

Basmann, R.L. "An Experimental Investigation of Some Small Sample Properties of (GCL) Estimators of Structural Equations: Some Preliminary Results," Richland Washington: General Electric Company, Hanford Laboratories, 1958.

____________. "Remarks Concerning the Application of Exact Finite Sample Distribution Functions of GCL Estimators in Econometric Statistical Inference," Journal of the American Statistical Association, 58(1963), pp. 943-976.

____________. "Exact Finite Sample Distributions for Some Econometric Estimators and Test Statistics: A Survey and Appraisal," Frontiers of Quantitative Economics, II(1974), ed. M.D. Intriligator and D.A. Kendrick, Amsterdam: North Holland Publishing Company.

Farrar, E.D. and R.R. Glauber. "Multicollinearity in Regression Analysis: The Problem Re-visited," Review of Economics and Statistics, 49(1967), pp. 92-107.

Feldstein, M.S. "Multicollinearity and the Mean Square Error of Alternative Estimators," Econometrica, 41(1973), pp. 337-345.

Klein, L.R. and M. Nakamura. "Singularity in the Equation System of Econometrics: Some Aspects of the Problem of Multicollinearity," International Economic Review, 3(1962), pp. 274-299.

Kshirsagar, A.M. Multivariate Analysis, New York: Marcel Deker Inc., 1972.

Mariano, R. "The Existence of Moments of the Ordinary Least Squares and Two-stage Least Squares Estimators," Econometrica, 40(1972) pp. 643-652.

Mariano, R.S. "Some Large-Concentration Parameter Asymptotics for the K-Class Estimators," Journal of Econometrics, 3(1975), pp. 171-178.

Mariano, R. and T. Sawa. "The Exact Finite-Sample Distribution of the Limited-Information Maximum Likelihood Estimator in the Case of Two Included Endogenous Variables," Journal of the American Statistical Association, 67(1972), pp. 159-163.

McDonald, J.B. "On the Distribution Functions of Some GCL Statistics Associated with the Structural Representative of an Econometric Model," Ph.D. Thesis, Purdue University, 1970.

Mikhail, W.M. "A Comparative Monte Carlo Study of the Properties of Econometric Estimators," Journal of the American Statistical Association, 70(1975), pp. 94-104.

Neelman, D. Multicollinearity in Linear Economic System, Netherlands: Tilburg University Press, 1973.

Neiswanger, W.A. and J.A. Yancey. "Parameter Estimates and Autonomous Growth," Journal of the American Statistical Association, 54(1959), pp. 389-402.

Richardson, D.H. "The Exact Distribution of a Structural Coefficient Estimator," Journal of the American Statistical Association, 63(1968), pp. 1214-1226.

Richardson, D.H. and R.J. Rohr. "A Program for Estimating Econometric Models--ECOMP 3," Econometrica, 43(1975), pp. 177-178.

Richardson, D.H. and De-Min Wu. "Least Squares and Grouping Method Estimators in the Errors in Variables Model," Journal of the American Statistical Association, 65(1970), pp. 724-748.

__________. "A Note on the Comparison of Ordinary and Two-Stage Least Squares Estimators," Econometrica, 39(1971), pp. 973-981.

Sawa, T. "The Exact Sampling Distribution of Ordinary Least Squares and Two-Stage Least Squares Estimators," Journal of the American Statistical Association, 64(1969), pp. 923-937.

__________. "Finite Sample Properties of K-Class Estimators," Econometrica, 40(1972), pp. 653-680.

Silvey, D.S. "Multicollinearity and Imprecise Estimator," Journal of the Royal Statistical Society Series B, 31(1969), pp. 539-552.

Smith, G. "Multicollinearity and Forecasting," Cowles Foundation Discussion Paper No. 383, December, 1974.

Summers, R. "A Capital Intensive Approach to the Small Sample Properties of Various Simultaneous Equation Estimators," Econometrica, 33(1965), pp. 1-41.