Sankhyā : The Indian Journal of Statistics
1985, Volume 47, Series B, pt. 3, pp. 401-431

A GENERALIZED MULTICOLLINEARITY INDEX FOR REGRESSION ANALYSIS*

By P. A. V. B. SWAMY
Federal Reserve Board
J. S. MEHTA
Temple University
S. S. THURMAN
Congressional Budget Office
and
N. S. IYENGAR
Indian Statistical Institute

SUMMARY. In this paper we have developed a formula measuring multicollinearity which takes different values for different coefficient estimates and/or for different equations. It takes the correct value of zero for an orthogonal model and a non zero value for a nonorthogonal or multicollinear model. Applying this measure of multicollinearity to a few empirical examples, we have found that if one forces a biased regression estimator to satisfy the minimax conditions, then the other goal of reducing multicollinearity may not be realized. We favor choosing from the near minimax estimators one that minimizes the absolute value of the estimate of multicollinearity.

1. Introduction

The multicollinearity problem may arise when some or all of the explanatory variables in a regression are highly correlated with one another. At least a brief discussion of this problem is, of course, included in every econometrics textbook. In addition, an extensive discussion of the methods of diagnosing multicollinearity (ill conditioning) makes up the empirical core of Belsley, Kuh and Welsch's (1980) book. In this book, Belsley *et al.* propose a method for assessing multicollinearity and compare their method with other methods proposed in the numerical analysis literature; see also Belsley (1984) and comments by Cook, Gunst, Snee and Marquardt, and Wood followed by Belsley's reply. To this list of diagnostics, the method employed by Theil (1971, p. 179) should be added. While all these methods

*Paper presented at the twenty-second annual conference of the Indian Econometric Society, Bangalore, January 2-4, 1984.

*Views expressed in this paper are those of the authors and do not reflect the views of the Organisations to which they belong.

AMS (1980) *subject classification* : 62J05, 62P20.

Key words and phrases : Regression Analysis, Real Multicollinearity, the Rotterdam Model, Ridge Parameters, Constrained Least Squares, Near Minimax Solution.

including Theil's may provide misleading information about multicollinearity, Theil's method is superior to the other methods, as shown by Swamy and Mehta (1985). This paper proposes a modified Theil's method which retains all the advantages of Theil's method, but is devoid of its disadvantages.

Exact multicollinearity is a special case of multicollinearity : the one in which the exact linear dependencies among the explanatory variables exist. In this case, the individual regression coefficients are not identified; see Goldberger (1968, pp. 22–23). The present paper focuses on multicollinearity, to the neglect of exact multicollinearity, since tools for handling the latter are given in Chipman (1964). Another reason for this focus is that exact multicollinearity may be rare, but some degree of multicollinearity is very common in economic data.

In Section 2 we suggest a new measure of multicollinearity suited to biased estimation by extending Theil's measure. Section 3 is devoted to applying this measure to the Rotterdam demand model. Concluding remarks are given in Section 4.

2. A REGRESSION MODEL AND THE MULTICOLLINEARITY EFFECT

2.1 *The model.* The regression model with which we will be concerned in this section is

$$\boldsymbol{y} = \boldsymbol{X\beta}+\boldsymbol{u}, \qquad \ldots \quad (1)$$

where $\boldsymbol{y}$ = $(n\times 1)$ vector of observations on the dependent variable,

$\boldsymbol{X}$ = $(n\times K)$ "fixed" matrix of rank K of observations on the explanatory variables,[1]

$\boldsymbol{\beta}$ = $(K\times 1)$ vector of coefficients, and

$\boldsymbol{u}$ = $(n\times 1)$ vector of disturbances.

We assume further that $E\boldsymbol{u} = 0$ and $E\boldsymbol{uu}' = \sigma^2 I$.

2.2 *The necessity of prior information for multicollinearity.* Associated with model (1), there can be a multicollinearity problem. As indicated in econometrics texts (see, e.g., Goldberger, 1968, p. 80), the formal definition of near exact multicollinearity is :

[1]Since X is known, we can use one of the currently available. numerically stable computer programs to compute the numerical rank of X. If the numerical rank of X is less than K, we may decide not to assume that X has full column rank. We are aware of the fact that the numerical rank of X is only suggestive of its theoretical rank. The former is not even a unique property of X but depends on factors, such as the details of the computational algorithm, the values of tolerance parameters used in the computation, and the effects of machine round-off errors.

Definition 1 (Multicollinearity) : Multicollinearity is the situation which arises when some or all of the explanatory variables are so highly correlated one with another that it becomes very difficult, if not impossible, to disentangle their influences and obtain a reasonably precise estimate of their separate effects. Multicollinearity is a matter of degree rather than of all or nothing.

To clarify the implications of this definition, we briefly review what multicollinearity means for the least squares estimation. When the matrix $\boldsymbol{X}'\boldsymbol{X}$ is nearly singular, some or all of the explanatory variables will be highly correlated one with another and near exact multicollinearity may arise. By definition, the least squares estimating equation $\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b} = \boldsymbol{X}'\boldsymbol{y}$ with the solution vector $\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$ is ill conditioned if small relative changes in the givens (the $\boldsymbol{X}'\boldsymbol{X}$'s and the $\boldsymbol{X}'\boldsymbol{y}$'s) result in large relative changes in the solution $\boldsymbol{b}$. The cause of such sensitivity is the near singularity of $\boldsymbol{X}'\boldsymbol{X}$ or, equivalently, some high auxiliary coefficients of determination resulting from the least squares fit of each explanatory variable on the remaining explanatory variables. Statistically, the near singularity of $\boldsymbol{X}'\boldsymbol{X}$ is manifest in large variances of the elements of $\boldsymbol{b}$ (large diagonal elements of $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$). These variances give confidence intervals for the elements of $\boldsymbol{\beta}$ which are too wide to be of any practical value; we will be very uncertain of the population value of $\boldsymbol{\beta}$. Numerically, the near singularity of $\boldsymbol{X}'\boldsymbol{X}$ means that inconsequential relative shifts in the observed data can produce consequential relative shifts in the coefficient estimates, a situation that is prima facie troublesome. We demonstrate now that the near singularity of $\boldsymbol{X}'\boldsymbol{X}$ is necessary but not sufficient for the presence of real multicollinearity or for the presence of real statistical and numerical problems related to multicollinearity.

The near singularity of $\boldsymbol{X}'\boldsymbol{X}$ means that some or all of the columns of $\boldsymbol{X}$ are nonorthogonal. In practice, it is possible that the regression coefficients corresponding to the nonorthogonal columns of $\boldsymbol{X}$ are indeed zero and the investigator includes these columns in $\boldsymbol{X}$ not knowing a priori that the corresponding coefficients are zero. In this case, the multicollinearity problem and the consequent numerical and statistical problems are not real. To distinguish real multicollinearity from apparent multicollinearity, we need the following definition of an orthogonal model :

Definition 2 (Orthogonal Model) : When $\boldsymbol{X}'\boldsymbol{X}$ is diagonal, model (1) is orthogonal if the prior covariance matrix of $\boldsymbol{\beta}$ is also diagonal. When $\boldsymbol{X}'\boldsymbol{X}$ is nondiagonal and the prior covariance matrix of $\boldsymbol{\beta}$ is diagonal, model (1) can still be orthogonal, provided the elements of $\boldsymbol{\beta}$ corresponding to the nonorthogonal columns of X are all zero.

Model (1) is nonorthogonal if it is not orthogonal. Definition 2 clarifies Definition 1 further. Real multicollinearity cannot arise if model (1) is orthogonal. It cannot also arise in some nonorthogonal models. For example, if the prior covariance matrix of $\boldsymbol{\beta}$ is nondiagonal, then model (1) is nonorthogonal. In this case, according to Definition 1 model (1) does not suffer from multicollinearity if either $X'X$ is diagonal or $X'X$ is nondiagonal and the elements of $\boldsymbol{\beta}$ corresponding to the nonorthogonal columns of X are all zero. Whether or not the prior covariance matrix of $\boldsymbol{\beta}$ is diagonal, real multicollinearity arises if $X'X$ is nondiagonal and the elements of $\boldsymbol{\beta}$ corresponding to the nonorthogonal columns of X are not all zero. Multicollinearity that arises when $X'X$ is nondiagonal and the elements of $\boldsymbol{\beta}$ corresponding to the nonorthogonal columns of X are all zero, is only apparent. This shows that the nondiagonality of $X'X$ is necessary but not sufficient for the presence of real multicollinearity. In other words, without utilizing any prior beliefs about the plausible values of $\boldsymbol{\beta}$, it is not possible to detect the presence of real multicollinearity on the basis of $X'X$ alone.

One should at least be prepared to assume, as some non-Bayesians may be willing to do, that the population values of the elements of $\boldsymbol{\beta}$ corresponding to the nonorthogonal columns of X are not all zero before concluding from an analysis of a nondiagonal $X'X$ matrix that real multicollinearity exists.[2] But then one should also allow for the possibility of this assumption being wrong. Any method which does not use prior information about $\boldsymbol{\beta}$ may not be successful at diagnosing the presence of real multicollinearity. For example, Theil (1971, pages. 169 and 179) has attempted to measure multicollinearity without any reference to the prior information about $\boldsymbol{\beta}$. His measure uses $X'X$ and the sample information about $\boldsymbol{\beta}$ in the form of the t-ratios which are routinely used to test the hypothesis on the elements of $\boldsymbol{\beta}$. Therefore, the ability of Theil's measure to give reliable estimates of the degree of real multicollinearity depends on the power of the tests of hypotheses on $\boldsymbol{\beta}$ based on the Student t-statistics and on the form of $X'X$. If this power is low because of the near singularity of $X'X$, then the absolute value of Theil's measure may either underestimate the degree of real multicollinearity or provide misleading information about multicollinearity, as shown by Swamy and Mehta (1985). These authors also show that while Theil's

[2]At the simplest level of analysis, the practitioner believes (has priors, therefore) that at least *some* of the regressors in an economic relationship may have nonzero explanatory powers for the regressand.

measure gives misleading information about multicollinearity with probability less than 1, any measure which even ignores the sample information about $\boldsymbol{\beta}$ and concentrates only on the form of $X'X$ gives misleading information about multicollinearity with probability 1. In short, not to use the prior information about $\boldsymbol{\beta}$ is bad enough in diagnosing multicollinearity. Ignoring even the sample information about $\boldsymbol{\beta}$ is likely to be disastrous to any attempts to detect and assess real multicollinearity. For this reason, we take the dependence on $X'X$ as well as on the prior and sample information about $\boldsymbol{\beta}$ as a primary characteristic a measure of multicollinearity should possess. In what follows we develop a measure with such a characteristic.

Before we do so, it is important to sound Goldberger's warning. Faced with the problem of fitting model (1) when some columns of X are nearly linearly dependent, one might be tempted to "eliminate" or "reduce" the multicollinearity by a transformation of X. Goldberger (1968, p. 87) warns us to resist this temptation. While the transformed regression could yield a reliable estimate for a coefficient vector, it is not the parameter of interest, i.e., $\boldsymbol{\beta}$. The reduced impression is obtained not by transforming X but by transforming the problem being answered. For this reason, we try to assess multicollinearity without changing model (1). Everytime we change the dependent variable or the set of explanatory variables in (1), we face a new multicollinearity problem.

Prior information about $\boldsymbol{\beta}$ is needed not only to detect multicollinearity but also to solve the multicollinearity problem, as pointed out by Goldberger (1968, p. 82). We, therefore, consider two types of prior information: (i) the non-Bayesian type, and (ii) the Bayesian type. The former is of the form $W\boldsymbol{\beta}$ lying on or within the ellipsoid $(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}})'W'\Delta_1^{-1}W(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}}) = r^2$ with known W, $\bar{\boldsymbol{\beta}}$, Δ_1, and r, and the latter is of the form $W\boldsymbol{\beta}$ distributed a priori with mean $W\bar{\boldsymbol{\beta}}$ and covariance matrix $W\Delta W'$, where W is a known rectangular matrix with full row rank. Here $\bar{\boldsymbol{\beta}}$ and Δ may be unknown.[3]

2.3 *A class of biased estimators.* To incorporate the non-Bayesian type prior information the method of constrained least squares is proposed. Based

[3]It would be more usual to write the more commonly known R here in place of W, but the former is used to denote the multiple correlation coefficient later in this paper. The exact constraints which Goldberger (1968, 82-83) considers are the limiting cases of Bayesian type prior information when $W\Delta W'$ for some W goes to a null matrix. A good example of where we have prior information about $W\boldsymbol{\beta}$, and not about $\boldsymbol{\beta}$, is given by a distributed lag model which employs Shiller's (1973) smoothness priors.

on this method, the constrained least squares estimator of $\boldsymbol{\beta}$ subject to $(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}})'\boldsymbol{W}'\Delta_1^{-1}\boldsymbol{W}(\boldsymbol{\beta}-\bar{\boldsymbol{\beta}}) = r^2$ is

$$\boldsymbol{b}_c = (\boldsymbol{X}'\boldsymbol{X}+\sigma^2\mu\boldsymbol{W}'\Delta_1^{-1}\boldsymbol{W})^{-1}(\boldsymbol{X}'\boldsymbol{y}+\sigma^2\mu\boldsymbol{W}'\Delta_1^{-1}\boldsymbol{W}\bar{\boldsymbol{\beta}}), \quad \ldots \quad (2)$$

where μ is chosen such that $(\boldsymbol{b}_c-\bar{\boldsymbol{\beta}})'\boldsymbol{W}'\Delta_1^{-1}\boldsymbol{W}(\boldsymbol{b}_c-\bar{\boldsymbol{\beta}}) = r^2$; see Swamy and Mehta (1983, p. 367). For the practical situation where r^2 is unknown, Thurman, Swamy and Mehta (1984) develop an approximation to (2), in which a value determined by a statistic replaces μ. The estimator (2) with this value of μ is nearly minimax.

In the case where the Bayesian prior distribution of $\boldsymbol{W}\boldsymbol{\beta}$ is available, Thurman, Swamy and Mehta (1984) and Kashyap, Swamy, Mehta and Porter (1984) prove that the prior covariance matrix of $\boldsymbol{\beta}$ implied by the prior covariance matrix of $\boldsymbol{W}\boldsymbol{\beta}$ need not be nonsingular. If it is singular, then the posterior probability density function (pdf) for $\boldsymbol{\beta}$ exists only on a subspace. Any analysis of this posterior p.d.f should take into account the information about the subspace.[4] However, Chipman's (1976, 603–617) or Rao's (1973, 305–306) method of deriving the posterior mean works whether or not the prior covariance matrix of $\boldsymbol{\beta}$ is singular. Using this method, we can show that the posterior mean of $\boldsymbol{\beta}$ is

$$\boldsymbol{b}_G = E\boldsymbol{\beta}+\text{cov}(\boldsymbol{\beta})[\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}+\text{cov}(\boldsymbol{\beta})]^{-1}(\boldsymbol{b}-E\boldsymbol{\beta}), \quad \ldots \quad (3)$$

where $E\boldsymbol{\beta}$ and $\text{cov}(\boldsymbol{\beta})$ are respectively the prior mean and the prior covariance matrix of $\boldsymbol{\beta}$; for a derivation of the estimator (3), see Thurman, Swamy and Mehta (1984) and Kashyap *et al.* (1984). After showing that the estimator (3) satisfies all the exact restrictions on $\boldsymbol{\beta}$ implied by the singular matrix, $\text{cov}(\boldsymbol{\beta})$, Kashyap *et al.* (1984) discuss an operational version of (3) which is nearly minimax.

The estimators (2) and (3) are not algebraically equivalent if $\text{cov}(\boldsymbol{\beta})$ is singular. In any event, it should be remembered that both of these estimators use some type of prior information about $\boldsymbol{\beta}$ and both are biased. Hence they are appropriate for developing a multicollinearity measure with the desired characteristic.

2.4 *The modified coefficient of determination.* We begin with the following specific member of the class of biased estimators introduced in Section 2.3 :

$$\boldsymbol{b}^* = (\boldsymbol{X}'\boldsymbol{X}+s^2\hat{\mu}\boldsymbol{W}'\Delta_1^{-1}\boldsymbol{W})^{-1}\boldsymbol{X}'\boldsymbol{y}, \quad \ldots \quad (4)$$

where $s^2 = (\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b})'(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b})/(n-K)$ and $\hat{\mu}$ is the value of μ selected according to the Thurman, Swamy and Mehta (1984) empirical rule : $\hat{\mu}$ is equal to

[4]Shiller's (1973) analysis, for one, overlooks this information.

zero if the value of the quadratic form $b'W'[2\hat{\mu}^{-1}\Delta_1+s^2W(X'X)^{-1}W']^{-1}Wb$ at $\hat{\mu}=0$ is greater than 1, $\hat{\mu}$ is equal to the numerical solution to the equation $b'W'[2\hat{\mu}^{-1}\Delta_1+s^2W(X'X)^{-1}W']^{-1}Wb = 1$ if the quadratic form $b'W'[2\hat{\mu}^{-1}\Delta_1+s^2W(X'X)^{-1}W']^{-1}Wb$ increases monotonically from 0 to $(b'W'[W(X'X)^{-1}W']^{-1}Wb/s^2) > 1$ as $\hat{\mu}$ increases from 0 to ∞, and $\hat{\mu}^{-1}$ is equal to 0 if the quadratic form $b'W'[2\hat{\mu}^{-1}\Delta_1+s^2W(X'X)^{-1}W']^{-1}Wb$ increases monotonically from 0 to $(b'W'[W(X'X)^{-1}W']^{-1}Wb/s^2) \leqslant 1$ as $\hat{\mu}$ increases from 0 to ∞. Usually, $\hat{\mu}$ determined by this rule will be a finite, positive quantity.

The discrepancy vector implied by the estimator (4) is

$$\hat{u}^* = y - Xb^*. \quad \ldots \quad (5)$$

Premultiplying each side of the equation $y = Xb^*+\hat{u}^*$ by its own transpose gives

$$y'y = b^{*\prime}X'Xb^*+2b^{*\prime}X'\hat{u}^*+\hat{u}^{*\prime}\hat{u}^*. \quad \ldots \quad (6)$$

We define the modified coefficient of multiple determination as

$$R^{*2} = \frac{b^{*\prime}X'Xb^*+2b^{*\prime}X'\hat{u}^*}{y'y}. \quad \ldots \quad (7)$$

Clearly,

$$1-R^{*2} = \frac{\hat{u}^{*\prime}\hat{u}^*}{y'y}. \quad \ldots \quad (8)$$

The derivations of R^{*2} using the members of the biased class other than b^* are straightforward and hence they are ignored. A simple geometric illustration of the modified coefficient of multiple determination given by (7) for the case $n = 3$ and $K = 2$ is shown in Figure 1. The mental extension of Figure 1 to higher dimensions is straightforward.

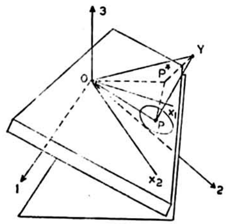

Figure 1. Geometric representation of modified coefficient of multiple determination ($n = 3$, $K = 2$).

Figure 1 shows the sample space when $n = 3$, the coordinate axes being labelled 1, 2 and 3 to correspond to the three observations on the dependent and two independent variables; see Draper and Smith (1981, p. 493). The two columns of $\boldsymbol{X}$ define two points $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and the vectors OX_1 and OX_2 define the two-dimensional plane OX_1X_2. The vectors $\boldsymbol{y}$, $\boldsymbol{Xb}$, $\boldsymbol{y}-\boldsymbol{X}\boldsymbol{b}$ and $\boldsymbol{Xb}^*$ define the lines OY, OP, YP and OP^* respectively. YP is perpendicular to the plane OX_1X_2 representing the orthogonal projection of Y onto OX_1X_2. The nonorthogonal projection involved in biased estimation is evidenced by the nonperpendicular line YP^*. Further, both P and P^* are the points in the same OX_1X_2-plane. For these reasons, YP is perpendicular to PP^*. These results show that $1-R^{*2}$ defined in (8) is simply the ratio of the squared distance from Y to P^* to the squared distance from O to Y. That is,

$$1-R^{*2} = \frac{(YP^*)^2}{(OY)^2}$$

$$= \frac{(YP)^2+(PP^*)^2}{(OY)^2} \quad \text{(by the Pythagorean theorem)}$$

or

$$R^{*2} = \frac{(OP)^2-(PP^*)^2}{(OY)^2}$$

$$= R^2-\frac{(PP^*)^2}{(OY)^2}, \qquad \dots \quad (9)$$

where R^2 is the square of the multiple correlation coefficient (or the coefficient of multiple determination) defined in Theil (1971, 164–165). It can be seen from (9) that $R^{*2} \leqslant R^2$. By introducing bias into the coefficient estimates we do not increase R^2. Further, unlike R^2, R^{*2} may be negative; in that case the square root of R^{*2} is not computed.

2.5 *The modified incremental contributions of explanatory variables.* Following Theil (1971, 168-169) we measure the contribution of each explanatory variable to the explanation of the variation in the dependent variable by measuring the increase in the square of the multiple correlation coefficient due to adding the h-th explanatory variable to the model when $K-1$ other explanatory variables are already included. To measure this increase in the case of biased estimation, we consider the truncated model $\boldsymbol{y} = \boldsymbol{X}_{-h}\boldsymbol{\beta}_{-h}+\boldsymbol{u}_h$ in which the h-th independent variable is not included but the $\boldsymbol{X}-1$ other variables are included. Let the biased estimates of the coefficients of the truncated model be

$$\boldsymbol{b}^*_{-h} = [\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}+s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}]^{-1}\boldsymbol{X}'_{-h}\boldsymbol{y}, \qquad \dots \quad (10)$$

where X_{-h} is obtained by deleting the h-th column from X, $(W'\Delta_1^{-1}W)_{-h,-h}$ is obtained by deleting the h-th row and h-th column from $W'\Delta_1^{-1}W$. The corresponding modified coefficient of determination is determined from

$$(1-R^{*2}_{-h})y'y = \hat{u}_h^{*\prime}\hat{u}_h^{*}, \qquad \dots \quad (11)$$

where $\hat{u}_h^{*} = y - X_{-h}b^{*}_{-h}$. The modified incremental contribution of the h-th variable is then

$$R^{*2} - R^{*2}_{-h}. \qquad \dots \quad (12)$$

This contribution cannot be negative as long as the same $\hat{\mu}$ and s^2 are employed in both (4) and (10), and $(W'\Delta_1^{-1}W)_{-h,-h}$ instead of some other matrix is employed in (10). This is because the biased estimation procedure minimizes $u'u$ or maximizes $-u'u$ subject to the restriction that $\beta'W'\Delta_1^{-1}W\beta = r^2$. Similarly, the estimate b^{*}_{-h} is determined by maximizing $-u'u$ subject to the same constraint that $\beta'W'\Delta_1^{-1}W\beta = r^2$ and the additional constraint that the h-th element of β is zero. Constrained maximization with one additional constraint cannot yield a higher modified coefficient of determination than the same constrained maximization without the additional constraint.

It is shown in the appendix to the paper that the sum of the modified incremental contributions of the K explanatory variables is exactly equal to their total contribution R^{*2} if and only if (iff) both $X'X$ and $W'\Delta_1^{-1}W$ are diagonal. The values of the modified incremental contributions should supplement the partial-regression leverage plots described in Belsley, Kuh and Welsch (1980, p. 30).

2.6 *A multicollinearity measure.* The measure of multicollinearity we propose here is

$$\tilde{m} = R^{*2} - \sum_{h=1}^{K}(R^{*2} - R^{*2}_{-h}). \qquad \dots \quad (13)$$

The derivation as well as the properties of the multicollinearity measure using other biased estimators of β introduced in Section 2.3 are exactly similar to those of (13) and thus will not be presented here. Now it follows from the definitions of R^{*2} and R^{*2}_{-h} that

$$\tilde{m}y'y = y'(A_x - B_x)y, \qquad \dots \quad (14)$$

where

$$A_x = (K-1)X(X'X+s^2\hat{\mu}W'\Delta_1^{-1}W)^{-1}X'X(X'X+s^2\hat{\mu}W'\Delta_1^{-1})W^{-1}X'$$
$$-\,2(K-1)X(X'X+s^2\hat{\mu}W'\Delta_1^{-1}W)^{-1}X'$$

and

$$\boldsymbol{B}_x = \sum_{h=1}^{K} \boldsymbol{X}_{-h}(\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}+s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h})^{-1}\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}$$

$$\times (\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}+s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h})^{-1}\boldsymbol{X}'_{-h}$$

$$-2\sum_{h=1}^{K} \boldsymbol{X}_{-h}(\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}+s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h})^{-1}\boldsymbol{X}'_{-h}.$$

From (14) we see that

$$\Pr(-\infty < \tilde{m}_L \leqslant \tilde{m} \leqslant \tilde{m}_U < \infty) = 1, \qquad \ldots \quad (15)$$

where $\tilde{m}_L$ and $\tilde{m}_u$ are the smallest and the largest eigenvalues of $A_x - B_x$ respectively. The derivation of the distribution function of $\tilde{m}$ is facilitated by the fact that

$$\Pr(\tilde{m} \leqslant m) = \Pr[\boldsymbol{y}'(\boldsymbol{A}_x - \boldsymbol{B}_x - m\boldsymbol{I})\boldsymbol{y} \leqslant 0] \qquad \ldots \quad (16)$$

which depends (as it should) on the conditional distribution of $\boldsymbol{y}$ given $\boldsymbol{X}$, $\boldsymbol{\beta}$, and σ^2, and the prior moments of $\boldsymbol{\beta}$.

We show in the Appendix to the paper that $\tilde{m}$ is degenerate at 0 iff both $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}$ are diagonal and is nondegenerate taking any particular value with probability 0 otherwise. This means that the case where model (1) is orthogonal is readily discovered by the measure (13). Since the measure (13) depends on both $\boldsymbol{X}'\boldsymbol{X}$ and the prior information about $\boldsymbol{\beta}$, a nondegenerate distribution of $\tilde{m}$ may accurately assign probabilities to different intervals of values the degree of real multicollinearity can take when $\boldsymbol{X}'\boldsymbol{X}$ is non-diagonal. However, different biased estimators of $\boldsymbol{\beta}$ utilizing different prior information yield different nondegenerate distributions of $\tilde{m}$ for the same model with nonorthogonal explanatory variables. This is reasonable because the degree of real multicollinearity depends not only on the form of $\boldsymbol{X}'\boldsymbol{X}$ but also on the plausible values of $\boldsymbol{\beta}$. We do not look at the value of $\tilde{m}$ if $\boldsymbol{X}'\boldsymbol{X}$ is diagonal and $\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}$ is nondiagonal, since there is no multicollinearity in this case. Prior information can reduce the multicollinearity in the data —for example, $\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}$ is diagonal and dominates the near-singular $\boldsymbol{X}'\boldsymbol{X}$. Indeed, the measure (13) shows whether the procedure of combining a prior information with the data via a biased estimator reduces the multicollinearity in the data. However, it should be noted that just as the R^2's of equations with different dependent variables are not directly comparable (Goldberger, 1968, p. 130), so the values of $\tilde{m}$ for different equations are not comparable either. We only compare the values of $\tilde{m}$ given by different coefficient estimators for the same equation.

In Swamy and Mehta (1985), Theil's (1971, p. 179) measure of multicollinearity is shown to be superior to the multicollinearity measures considered in the numerical analysis literature. The measure (13) is an improvement over Theil's measure because the former is based on both the sample and prior information and the latter is based on only the sample information. If the sample information is weak, as it is whenever $X'X$ is near singular, a nondegenerate distribution of Theil's measure may not accurately assign probabilities to different intervals of values the degree of real multicollinearity can take.

The range $(\tilde{m}_L, \tilde{m}_u)$ of $\tilde{m}$ may cover both negative and positive values. Since the degenerate distribution of $\tilde{m}$ at 0 represents the absence of multicollinearity, we measure the degree of multicollinearity present in (1) by the absolute value of $\tilde{m}$. In other words, the distance of a value of $\tilde{m}$ from zero measures the departure of model (1) from orthogonality when $\boldsymbol{X}'\boldsymbol{X}$ is nondiagonal and model (1) is defined broadly to include the data as well as the prior information about $\boldsymbol{\beta}$. We may avoid the difficult problem of computing the probabilities in (16) by adopting the rule : In model (1), multicollinearity is serious if the value of (13) is closer to the value of $\tilde{m}_L$ or $\tilde{m}_U$ than to zero and $X'X$ is nondiagonal. The computationally simpler bounds for $\tilde{m}$ are given by

$$m_L^* = R^{*2} - K[\max_{1 \leq h \leq K} (R^{*2} - R^{*2}_{-h})]$$

and

$$m_U^* = R^{*2} - K[\min_{1 \leq h \leq K} (R^{*2} - R^{*2}_{-h})] \leqslant R^{*2} \qquad \ldots \quad (17)$$

2.7 *Further analysis.* Further insights can be gained by looking into the relationships among the partial correlation coefficients and the measures in (12) and (13). These relationships can be derived geometrically as follows. Consider equation (1) with $K = 2$. The points Y, X_1 and X_2 shown in Figure 2 are defined by the vector $\boldsymbol{y}$ and the two columns $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ of X respectively, see, e.g., Wonnacott and Wonnacott (1970, p. 306).

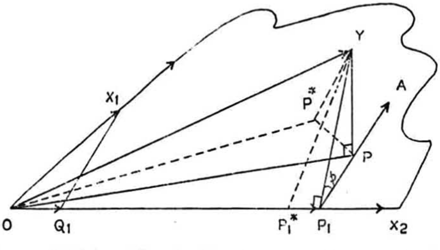

Figure 2. Geometrical interpretation of modified partial and multiple correlations ($K = 2$)

In Figure 2, the line YP_1 is the perpendicular from Y onto OX_2 representing the orthogonal projection of Y onto OX_2 and the point P_1 represents the fitted value when $\boldsymbol{y}$ is regressed on $\boldsymbol{x}_2$. Therefore YP_1 is the least squares residual vector obtained from a least squares fit of $\boldsymbol{y}$ on $\boldsymbol{x}_2$. The lines OX_1 and OX_2 define a plane. Suppose X_1Q_1 is the perpendicular from X_1 onto OX_2. That is, X_1Q_1 is obtained from the orthogonal projection of $\boldsymbol{x}_1$ onto OX_2. Then Q_1 represents the fitted value when $\boldsymbol{x}_1$ is regressed on $\boldsymbol{x}_2$ and X_1Q_1 is the least squares residual vector obtained from a least squares fit of $\boldsymbol{x}_1$ on $\boldsymbol{x}_2$. Since P_1A is parallel to X_1Q_1, ϕ is the angle between the lines YP_1 and X_1Q_1. The square of the cosine of the angle ϕ, denoted r_1^2, is[5]

$$r_1^2 = (\cos\phi)^2 = \frac{(PP_1)^2}{(YP_1)^2} = 1 - \frac{(YP)^2}{(YP_1)^2} \qquad \dots \quad (18)$$

where use is made of the fact that YP is perpendicular to the plane OX_1X_2. Now let YP^* represent a nonorthogonal projection of Y onto OX_1X_2—plane and let YP_1^* represent a nonorthogonal projection of Y onto OX_2. Then, from (18) we have

$$\begin{aligned} 1-r_1^2 &= \frac{(YP)^2/(OY)^2}{(YP_1)^2/(OY)^2} \\ &= \frac{[(YP^*)^2-(P^*P)^2]/(OY)^2}{[(YP_1^*)^2-(P_1^*P_1)^2]/(OY)^2} \\ &= \frac{1-R^{*2}-[(P^*P)^2/(OY)^2]}{1-R_{-1}^{*2}-[(P_1^*P_1)^2/(OY)^2]} \,. \end{aligned} \qquad \dots \quad (19)$$

It is clear from Theil (1971, p. 174) that

$$1-r_1^2 = \left(1+\frac{t_1^2}{n-K}\right)^{-1} \qquad \dots \quad (20)$$

where t_1 is the usual t-ratio or the ratio of the first element of $\boldsymbol{b}$ to the square root of the first diagonal element of $s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Equating the righthand sides of equations (19) and (20) gives

$$1-R_{-1}^{*2}-[(P_1^*P_1)^2/(OY)^2] = \left(1+\frac{t_1^2}{n-K}\right)[1-R^{*2}-((P^*P)^2/(OY)^2)]$$

from which it follows that

$$R^{*2}-R_{-1}^{*2} = \frac{t_1^2}{n-K}\,[1-R^{*2}-((P^*P)^2/(OY)^2)] + \frac{[(P_1^*P_1)^2-(P^*P)^2]}{(OY)^2} .$$

---

[5]It may be noted that the $\cos\phi$ is equal to a partial correlation coefficient if $\boldsymbol{x}_2$ is a vector of unit elements or when $\boldsymbol{x}_2$ is not a vector of unit elements if $\boldsymbol{y}$, $\boldsymbol{x}_1$, and $\boldsymbol{x}_2$ are expressed as deviations from their respective means. This point is not clear from Theil (1971, 172-173).

More generally, for $h = 1, 2, ..., K > 2$

$$R^{*2}-R^{*2}_{-h}$$

$$= \frac{t_h^2}{(n-1)}(1-R^2)+\frac{[(b_{-h}-b^*_{-h})'X'_{-h}X_{-h}(b_{-h}-b^*_{-h})-(b-b^*)'X'X(b-b^*)]}{y'y}$$

$$= \frac{t_h^2}{(n-K)}(1-R^2)+[(R^2_{-h}-R^{*2}_{-h})-(R^2-R^{*2})] \qquad \dots \quad (21)$$

where $b_{-h} = (X'_{-h}X_{-h})^{-1}X'_{-h}y$ is the least squares estimate of coefficients in the truncated model $y = X_{-h}\beta_{-h}+u_h$, b^*_{-h} is as in (10) and R^2 and R^2_{-h} are the same as the square of the multiple correlation coefficient and R_h^2 defined in Theil (1971, pp. 164 and 169) respectively. The result in (9) implies that $(R^2_{-h}-R^{*2}_{-h}) \geqslant 0$ and $(R^2-R^{*2}) \geqslant 0$.

It has been shown by Theil (1971, p. 175, Problem 3.3) that the first term on the right-hand side of equation (21) is equal to the incremental contribution of the h-th independent variable implied by the unbiased least squares estimator b. Therefore, it follows from (21) that when we arrange the independent variables in descending order of their incremental contributions $(R^{*2}-R^{*2}_{-h})$ or the corresponding squared t-ratios (t_h^2), we do not get the same arrangement unless the second term on the right-hand side of (21) is zero. This second term is zero in the case of unbiased estimation.

The difference between the incremental contribution (12) and the corresponding incremental contribution implied by b (or the first term on the right-hand side of (21)) is equal to the second term on the right-hand side of (21). A change of procedure from that dictated by the minimum variance linear unbiasedness criterion can increase or decrease the incremental contribution of the h-th variable according as this second term is positive or negative. Since the second term cannot always be positive for every h, we cannot say that a biased estimation will always push the absolute value of the multicollinearity measure (13) toward zero regardless of the values of $\hat{\mu}$, W and Δ_1 even though R^2 in Theil's (1971, pp. 175 and 179) measure, $R^2-\sum_{h=1}^{K}(R^2R^2_{-h})$, of multicollinearity based on the estimator b is always bigger than R^{*2}, as shown in (9).

It is found by Theil's (1971, p. 175, (3.12)) result that equation (21) can be written alternatively as

$$R^{*2}-R^{*2}_{-h} = r_h^2(1-R^2_{-h})+[(R^2_{-h}-R^{*2}_{-h})-(R^2-R^{*2})]. \qquad \dots \quad (22)$$

The first term on the right-hand side of this equation shows that when an unbiased method of estimation is employed, the incremental contribution of the h-th explanatory variable is equal to r_h^2 times a factor which is at most equal to 1. This factor states that in the case of unbised estimation, given r_h^2, the larger the proportion of the variation in the dependent variable accounted for by the $K-1$ other explanatory variables, the smaller the incremental contribution of the h-th variable. In this instance, the h-th independent variable may be a minor variable of little consequence. If we change to a biased estimation method this minor variable may become a variable of major contributing significance depending on the sign of the second term on the right-hand side of (22). In view of the importance of this second term, let us study it further.

Subtracting (10) from $\boldsymbol{b}_{-h}$ gives

$$\begin{aligned}\boldsymbol{b}_{-h}-\boldsymbol{b}^*_{-h} &= \{(\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h})^{-1}-[\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}+s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}]^{-1}\}\boldsymbol{X}'_{-h}\boldsymbol{y}\\ &= (\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h})^{-1}\{I-[\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}+s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}\\ &\quad -s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}][\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}+s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}]^{-1}\}\boldsymbol{X}'_{-h}\boldsymbol{y}\\ &= s^2\hat{\mu}(\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h})^{-1}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}\boldsymbol{b}^*_{-h}.\end{aligned}$$

Therefore,

$$\begin{aligned}&(\boldsymbol{b}_{-h}-\boldsymbol{b}^*_{-h})'\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h}(\boldsymbol{b}_{-h}-\boldsymbol{b}^*_{-h})\\ &= s^4\hat{\mu}^2\boldsymbol{b}^{*\prime}_{-h}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}(\boldsymbol{X}'_{-h}\boldsymbol{X}_{-h})^{-1}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h}\boldsymbol{b}^*_{-h}. \qquad \ldots (23)\end{aligned}$$

Similar algebraic operations give

$$(\boldsymbol{b}-\boldsymbol{b}^*)'\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{b}-\boldsymbol{b}^*) = s^4\hat{\mu}^2\boldsymbol{b}^{*\prime}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}\boldsymbol{b}^*. \qquad \ldots (24)$$

Notice that (23) minus (24) is not always positive for every h. The second term on the right hand-side of (21) cannot take substantial values if $\hat{\mu}$ is very small. In other words the incremental contribution of the h-th variable implied by the biased estimator (4) cannot be very different from that implied by the unbiased estimator $\boldsymbol{b}$ if $\hat{\mu}$ takes a tiny value.

It should be noted that when we use the estimator (2) or (3) in place of the estimator (4) used in (12), the result (22) shows that the use of certain values of $\bar{\boldsymbol{\beta}}$ (or $E\boldsymbol{\beta}$) and $\hat{\mu}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}$ (or cov($\boldsymbol{\beta}$)) may result in implausible values for the incremental contributions. From this we infer that sufficient caution should be exercised in choosing the values of $E\boldsymbol{\beta}$ and cov($\boldsymbol{\beta}$). We should choose among these values by reflecting on their operating characteristics. The operating characteristic of an estimator of $\boldsymbol{\beta}$ can be taken to consist of the absolute value of the estimate of multicollinearity based on that estimator.

In many complex settings such as those of econometric estimation, any operating characteristic that is sensible for the treatment of optimality or admissibility requires considerations additional to the multicollinearity effect and incremental contributions, e.g., reflecting the mean square error of an estimator. We suggest choosing from the estimators, (2), (3) and their operational versions like (4), one that has desirable operating characteristics. The relative risk properties of these estimators are discussed in Thurman, Swamy and Mehta (1984).

However, any attempt to find an estimator with desirable operating characteristics cannot succeed if it failed to take note of the conflict between minimaxity and the conditioning problem pointed out by Casella (1980). That is, there is an incompatibility between the criterion of reducing the risk of the estimator (2) or (3) below that of $\boldsymbol{b}$ for every possible value of $\boldsymbol{\beta}$ and the criterion of lowering the condition number of the basic data matrix X (a measure equal to the square root of the ratio of the largest to the smallest eigenvalues) by augumenting the rows of X by some other rows. As emphasized by Casella (1980, p. 1052), some compromise between the two criteria is needed, and the empirical examples of Section 3 are aimed at illustrating such a compromise. In view of this conflict, not every kind of prior information about $\boldsymbol{\beta}$ which solves the numerical problem will also solve the statistical problem. For example, the prior values W, $\bar{\boldsymbol{\beta}}$ and $\mu^{-1}\Delta_1$ may be such that the condition number of the matrix $(X'X+\sigma^2\mu W'\Delta_1^{-1}W)$ inverted in (2) is much smaller than the condition number of $X'X$ but the risk of the estimator (2) based on these values is larger than that of $\boldsymbol{b}$. Thus, caution should be exercised in formulating and using prior information.

3. Application of measures to the Rotterdam model

How the compromise advanced by Casella (1980, p. 1052) works in practice can be conveyed by presentation of some illustrative examples with the various estimates of incremental contributions and multicollinearity provided. In this section we work with the Rotterdam model in absolute prices which for fourteen commodity groups takes the following form :

$$\bar{w}_{it}Dq_{it} = \gamma_i DQ_t + \sum_{j=1}^{13} \pi_{ij}(Dp_{jt} - Dp_{14t}) + u_{it}$$

$$(i = 1, 2, \ldots, 13;\ t = 1, 2, \ldots, 31) \quad \ldots \quad (25)$$

where the dependent variables are the (annual) quantity log-changes of the individual commodity groups multiplied by the corresponding value share averages of the years $t-1$ and t, γ_i is the marginal value share of the i-th commodity group, DQ_t is the sum of the dependent variables of all 14 demand equations, the π's are the price coefficients, Dp_{jt} is the (annual) price log-change of the j-th commodity group and u_{it} is a disturbance term; see Theil (1975, pp. 24, 39–42, 48–49). We estimate the model (25) using the Dutch data for the years from 1922-23 through 1938-39 and 1949-50 through 1962-63 given in Theil (1975, 264–265). As in Theil (1975, 41–42), we assume that

$$Eu_{it} = 0 \text{ for every } i \text{ and } t \text{ and } Eu_{is}u_{jt} = \begin{cases} \omega_{ij} & \text{if } s = t; \\ 0 & \text{if } s \neq t. \end{cases} \quad \dots \quad (26)$$

Since we have illustrative purposes in mind, we estimate each equation in (25) separately ignoring the symmetry constraints $\pi_{ij} = \pi_{ji}$ for $i \neq j = 1, 2, \dots, 13$. [A ridge procedure which utilizes such symmetry constraints is developed in Swamy and Mehta, 1983]. We can combine the 31 observations for each i in the form

$$\boldsymbol{y}_i = \boldsymbol{X\beta}_i + \boldsymbol{u}_i \ (i = 1, 2, \dots, 13) \quad \dots \quad (27)$$

where $\boldsymbol{y}_i$ and $\boldsymbol{u}_i$ are 31-element column vectors having $\bar{w}_{it}Dq_{it}$ and u_{it} as their respective t-th elements, $\boldsymbol{\beta}_i = (\gamma_i, \pi_{i1}, \dots, \pi_{i13})'$ and $\boldsymbol{X}$ is the 31×14 matrix of observations on the independent variables in (25). Now we apply the formulas in Section 2 to each equation in (27). Since (25) does not necessarily represent a distributed lag model, we set $\boldsymbol{W} = \boldsymbol{I}$. [For an application of the estimator (4) with $\boldsymbol{W} \neq \boldsymbol{I}$ to a distributed lag model, see Thurman, *et al.* (1984) and Kashyap *et al.* (1984). The following coefficient estimators are applied to each equation in (27) to compute the incrememtal contribution and multicollinearity.

1. *The least squares* (*LS*) *estimator* : $\boldsymbol{b}$.

2. *The SMR estimator* : The Swamy, Mehta and Rappoport (1978) estimator (4) with W = any orthogonal matrix and $\boldsymbol{\Delta}_1 = \boldsymbol{I}$.

3. *Casella's* (1980) *estimator* : The estimator (2) wth $\boldsymbol{W}\bar{\boldsymbol{\beta}} = 0$, $\boldsymbol{W}$ = an orthogonal matrix which diagonalizes $\boldsymbol{X}'\boldsymbol{X}$, $\sigma^2 = s^2$ and $\mu\boldsymbol{\Delta}_1^{-1} = \boldsymbol{D}$, a diagonal matrix whose i-th diagonal elememt is

$$\frac{2(K-2)\lambda_i^2\lambda_{\max}^{-1}(n-K)}{(n-K+2)(\boldsymbol{b}'\boldsymbol{X}'\boldsymbol{X}\boldsymbol{b}+\bar{g}s^2+c)} \quad \dots \quad (28)$$

where $K = 14$, $n = 31$, λ_i is the i-th eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$, $\lambda_{\max}$ is the largest eigenvalue of $\boldsymbol{X}'\boldsymbol{X}$, $\bar{g} = 2(n-K)(K-2)(n-K+2)^{-1}$ or 0 and $c = 0$.

4. *The HKB estimator*: The Hoerl, Kennard and Baldwin (1975) estimator which is the same as the estimator (2) with $\boldsymbol{W}\bar{\boldsymbol{\beta}} = 0$, $\boldsymbol{W} =$ any orthogonal matrix, $\boldsymbol{\Delta}_1 = \boldsymbol{I}$, $\sigma^2 = s^2$ and

$$\mu = \frac{K}{\boldsymbol{b}'\boldsymbol{D}_x\boldsymbol{b}}, \qquad \ldots \quad (29)$$

where $\boldsymbol{D}_x$ is the diagonal matrix having the same diagonal elements as the matrix $\boldsymbol{X}'[\boldsymbol{I}-\boldsymbol{l}(\boldsymbol{l}'\boldsymbol{l})^{-1}\boldsymbol{l}']\boldsymbol{X}$ with $\boldsymbol{l} = (1, 1, \ldots, 1)'$, a $n \times 1$ vector of 1's.

As we have argued in Section 2.2, real multicollinearity cannot arise unless the coefficients corresponding to the nonorthogonal columns of $\boldsymbol{X}$ are nonzero. The Swamy and Mehta (1983) ridge estimates show that of 182 estimates of the coefficients in (25) only 37 are significantly different from zero. Therefore, the measure (13) can precisely measure real multicollinearity if it is based on the prior information which precisely represents the population values of the coefficients in (25).

The values of s^2 and the measures (7), (12) and (13) when the LS estimator $\boldsymbol{b}$ is used in place of $\boldsymbol{b}^*$ used in defining these measures, are given in Table 1 for all the 13 equations in (27). The figures in parentheses below each estimated value of the multicollinearity are the values of (m_L^*, m_u^*) defined in (17). As expected, our measure (13) indicates different degrees of multicollinearity for different equations in (27). The value of the Thisted (1980) measure for estimation is 1·745 which gives the impression that there is no high degree of collinearity in Theil's (1975) data on the variables in (25). This impression has been contradicted by the value of the measure (13) for the 10th equation in (27). For this equation the $\tilde{m}$ value of (13) is 0.6 which is closer to the value of 0.84 of m_u^* than to zero.

The rankings of independent variables by the values of the mincremetal contributions are also given in Table 1. The figures in parentheses next to each estimated value of the incremental contribution are the ranks of independent variables when they are ordered according to decreasing values of the incremental contributions. It can be seen from these ranks that the contribution to explanation provided by either the income or the relative own price term is not the largest in 5 of the 13 equations. In 8 equations the incremental contribution of the independent variable with the first rank is even smaller than the fraction of the variation in the dependent variable not accounted for by all the included explanatory variables. The 6-th equation in (27) does not fit the data well. Much of the variation in the dependent variable of the 6-th equation remains unaccounted for.

Table 2 reports the estimates of (7), (12) and (13) for the SMR estimator. The values of s^2 times $\hat{\mu}$ are also reported in this table. The estimated incremental contributions, multicollinearity and R^{*2} are of the same order of magnitude in Tables 1 and 2, the margins of difference being generally negligible. This is because the values of $s^2\hat{\mu}$ employed in (4) are very small. It is clear from the simulation study of Hoerl, Kennard and Baldwin (1975) that the optimal values of the ridge parameter will be close to zero when the error variance is small. Since the error variance estimates given in Table 1 are tiny, the values of $\hat{\mu}$ given in Table 2 may be optimal. Still, the LS estimator and the SMR estimator yield different rankings for the independent variables of the 6-th equation in (27). A comparison of the values in the rows labelled "Multicollinearity" in Tables 1 and 2 shows that a small increase in the value of $s^2\hat{\mu}$ in (4) from zero can resut in a detectable increase or decrease in the absolute value of the multicollinearity measure (13). Specifically for the 6-th equation in (25), the multicollinearity estimate decreased from 0·097 to 0·080 as we increased the value of $s^2\hat{\mu}$ from 0 to 0·0355 whereas for the 3-rd equation the multicollinearity estimate increased from 0·221 to 0·230 as we increased the value of $s^2\hat{\mu}$ from 0 to 0·0015. More importantly, for the 10-th equation where there is a high degree of multicollinearity the biased SMR estimator leaves us with about the same measure of multicollinearity as the unbiased LS estimator. The change in the multicollinearity estimates using the SMR estimator is generally "minuscule" for this example.

Tables 3 and 4 set out the estimates of (7), (12) and (13) implied by Casella's estimates of coefficients. The differences between the estimates given in these two tables are entirely attributable to the differences in the value of $\bar{g}$ used. Table 3 is based on the value $\bar{g} = 2(K-2)(n-K)(n-K+2)^{-1}$ and Table 4 is based on the value $\bar{g} = 0$. The estimates of (7), (12) and (13) implied by the HKB estimates of coefficients are given in Table 5. The range of the diagonal elements of s^2D in (28) and the value of s^2 times (29) are alse given in these tables. The figures in parentheses next to each estimate of the incremental contribution and also below each estimate of the multicollinearity have the same interpretation as those in Tables 1 and 2. Estimates of incremental contributions, multicollinearity and R^{*2}, implied by the LS, SMR, and Casella coefficient estimates, may be compared by referring to Tables 1, 2, 3 and 4. There are cases in which the choice of coefficient estimator or $\bar{g}$-value makes a detectable difference in the estimates of (7), (12) and (13). The operating characteristics of Casella's estimator depend on the value of $\bar{g}$ used. Our interpretation of Tables 1, 2, 3 and 4, then is that in terms of multicollinearity there is no demonstrable payoff to the use of

TABLE 1. INCREMENTAL CONTRIBUTIONS, MULTICOLLINEARITY EFFECT, MULTIPLE CORRELATION, AND ERROR VARIANCE ESTIMATE: LEAST SQUARES RESULTS

	Regression Equations in (25)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
I. Incremental contributions of income	.048 (7)	.079 (2)	.083 (2)	.018 (7)	.084 (2)	.000(13)	.102 (1)	.046 (4)	.043 (4)	.150 (1)	.101 (2)	.247 (1)	.151 (1)
relative prices of													
(1) bread	.009 (8)	.001(12)	.023 (7)	.097 (2)	.078 (3)	.007 (5)	.009 (7)	.136 (1)	.043 (3)	.000(14)	.002(11)	.022 (5)	.019 (7)
(2) groceries	.006 (9)	.003(10)	.037 (5)	.012 (8)	.018 (8)	.018 (4)	.003(10)	.021(10)	.014 (8)	.003 (8)	.004(10)	.000(13)	.029 (5)
(3) dairy products	.063 (5)	.002(11)	.206 (1)	.009(11)	.018 (7)	.020 (2)	.001(13)	.023 (9)	.007 (9)	.013 (4)	.012 (8)	.000(12)	.020 (6)
(4) vegetables & fruit	.062 (6)	.000(14)	.000(14)	.261 (1)	.157 (1)	.004 (7)	.016 (6)	.001(13)	.000(13)	.001(10)	.027 (5)	.000(14)	.000(13)
(5) meat	.226 (2)	.012 (7)	.074 (3)	.010 (9)	.000(12)	.053 (1)	.024 (3)	.085 (2)	.020 (6)	.000(11)	.004 (9)	.021 (6)	.014 (8)
(6) fish	.137 (4)	.017 (6)	.040 (4)	.000(14)	.036 (4)	.001(11)	.000(14)	.031 (5)	.016 (7)	.005 (7)	.000(13)	.001 (9)	.004(12)
(7) beverages	.307 (1)	.027 (4)	.000(13)	.022 (6)	.022 (6)	.002(10)	.020 (5)	.012(11)	.064 (2)	.007 (5)	.000(14)	.000 (7)	.006(11)
(8) tobacco	.005(10)	.020 (5)	.007(10)	.042 (3)	.009(11)	.002 (9)	.030 (2)	.078 (3)	.000(12)	.001 (0)	.021 (7)	.004 (8)	.000(14)
(9) pastry, chocolate, ice cream	.002(11)	.005 (9)	.011 (9)	.000(13)	.001(14)	.006 (6)	.001(12)	.026 (7)	.243 (1)	.006 (6)	.024 (6)	.001(10)	.039 (3)
(10) clothing and other textiles	.000(13)	.057 (3)	.015 (8)	.037 (4)	.002(13)	.000(14)	.006 (8)	.011(12)	.000(14)	.027 (3)	.030 (4)	.034 (3)	.013 (9
(11) footwear	.002(12)	.001(13)	.003(11)	.001(12)	.026 (5)	.000(12)	.020 (4)	.025 (8)	.002(11)	.000(12)	.200 (1)	.000(11)	.038 (4)
(12) other durables	.000(14)	.124 (1)	.024 (6)	.009(10)	.010(10)	.020 (3)	.005 (9)	.000(14)	.002(10)	.030 (2)	.063 (3)	.158 (2)	.009(10)
(13) water, light, & heat	.162 (3)	.007 (8)	.001(12)	.026 (5)	.019 (9)	.002 (8)	.002(11)	.020 (6)	.021 (5)	.000(13)	.001(12)	.032 (4)	.118 (2)
II. Multicollinearity **(13) (m^*_L, m^*_U)	−.333 (−3.51, .69)	.498 (−.89, .85)	.221 (−2.15, .75)	.223 (−2.89, .77)	.144 (−1.58, .61)	.007 (−5, .23)	.463 (−.73, .7)	.310 (−1.07, .83)	.371 (−2.55, .85)	.600 (−1.25, .84)	.354 (−2.07, .86)	.420 (−2.51, .95)	.211 (−1.7, .69)
III. Total contribution of explanatory variables (R^2)	.696	.854	.754	.707	.624	.233	.703	.834	.847	.845	.861	.950	.604
IV. Error variance estimate (s^2)	.771×10	.264×10	.373×10	.374×10	.121×10	.822×10	.201×10	.116×10	.823×10	.301×10	.403×10	.645×10	.434×10

Thisted's measure of multicollinearity effect = 1.745.

TABLE 2. INCREMENTAL CONTRIBUTIONS, MULTICOLLINEARITY EFFECT, MULTIPLE CORRELATION, AND RIDGE PARAMETER: THE SMR ESTIMATES

	Regression Equations in (25)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
I. Incremental contributions of income	.047 (7)	.079 (2)	.079 (2)	.018 (7)	.083 (2)	.003 (8)	.098 (1)	.046 (4)	.043 (4)	.149 (1)	.101 (2)	.247 (1)	.165 (1)
relative prices of (1) bread	.009 (8)	.001(12)	.021 (7)	.098 (2)	.080 (3)	.005 (7)	.006 (7)	.136 (1)	.043 (3)	.000(14)	.002(11)	.022 (5)	.018 (7)
(2) groceries	.005 (9)	.002(10)	.035 (5)	.011 (8)	.016 (8)	.012 (3)	.003(10)	.021(10)	.014 (8)	.003 (8)	.004(10)	.000(13)	.027 (5)
(3) dairy products	.063 (5)	.002(11)	.203 (1)	.008(11)	.017 (7)	.020 (2)	.000(13)	.023 (9)	.007 (9)	.013 (4)	.012 (8)	.000(12)	.018 (6)
(4) vegetables & fruit	.062 (6)	.000(14)	.000(14)	.262 (1)	.167 (1)	.003 (9)	.017 (6)	.001(13)	.000(13)	.001 (9)	.027 (5)	.000(14)	.000(13)
(5) meat	.226 (2)	.012 (7)	.073 (3)	.000 (9)	.006(12)	.050 (1)	.023 (3)	.085 (2)	.030 (6)	.000(11)	.004 (9)	.021 (6)	.014 (8)
(6) fish	.136 (4)	.017 (6)	.049 (4)	.000(14)	.035 (4)	.001(13)	.000(14)	.031 (5)	.016 (7)	.005 (6)	.000(13)	.001 (9)	.004(12)
(7) beverages	.307 (1)	.027 (4)	.000(13)	.021 (6)	.022 (6)	.008 (5)	.019 (5)	.012(11)	.064 (2)	.007 (5)	.000(14)	.009 (7)	.006(11)
(8) tobacco	.005(10)	.020 (5)	.006(10)	.042 (3)	.010(11)	.002(11)	.027 (2)	.078 (3)	.000(12)	.001(10)	.021 (7)	.004 (8)	.000(14)
(9) pastry, chocolate, ice cream	.002(11)	.006 (9)	.010 (9)	.000(13)	.001(14)	.005 (6)	.001(12)	.025 (7)	.243 (1)	.005 (7)	.024 (6)	.001(10)	.041 (3)
(10) clothing and other textiles	.000(13)	.057 (3)	.015 (8)	.037 (4)	.022(13)	.000(14)	.005 (8)	.011(12)	.000(14)	.027 (3)	.039 (4)	.034 (3)	.013 (9)
(11) footwear	.002(12)	.001(13)	.004(11)	.001(12)	.028 (5)	.001(12)	.022 (4)	.025 (8)	.002(11)	.000(12)	.209 (1)	.000(11)	.037 (4)
(12) other durables	.000(14)	.124 (1)	.023 (6)	.009(10)	.011(10)	.011 (4)	.005 (9)	.000(14)	.002(10)	.029 (2)	.063 (3)	.158 (2)	.008(10)
(13) water, ligh & heat	.161(3)	.008 (8)	.001(12)	.026 (5)	.011 (9)	.002(10)	.003(11)	.029 (6)	.021 (5)	.000(13)	.001(12)	.032 (4)	.118 (2)
II. Multicollinearity $(\overset{\bullet}{m}_L, \overset{\bullet}{m}_U)$ (13)	−.332 (−3.60, .69)	.499 (−.88, .85)	.230 (−2.10, .75)	.222 (−2.9, .76)	.138 (−1.58, .61)	.080 (−.50, .20)	.467 (−.67, .69)	.309 (−1.07, .83)	.370 (−2.56, .85)	.601 (−1.24, .84)	.353 (−2.06, .86)	.421 (−2.51, .95)	.217 (−1.65, .69)
III. Total contribution of explanatory variables (R^{*2})	.693	.853	.750	.765	.619	.205	.698	.833	.846	.843	.861	.950	.690
IV. Ridge parameter ($s^2\mu$)	.0012	.0011	.0015	.0019	.0027	.355	.0028	.0009	.0009	.0011	.0009	.0002	.0014

TABLE 3. INCREMENTAL CONTRIBUTIONS, MULTICOLLINEARITY EFFECT, MULTIPLE CORRELATION, AND RIDGE PARAMETERS: CASELLA'S ESTIMATES

$$\bar{g} = 2(K-2)(n-K)\ (n-K+2)^{-1}$$

	Regression Equations in (25)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
I. Incremental contributions of income	.048 (7)	.070 (2)	.083 (2)	.017 (7)	.084 (2)	.000(12)	.102 (1)	.045 (4)	.042 (4)	.149 (1)	.101 (2)	.247 (1)	.170 (1)
relative prices of (1) bread	.009 (8)	.001(12)	.021 (7)	.097 (2)	.070 (3)	.00, (5)	.008 (7)	.136 (1)	.043 (3)	.000)11)	.002(11)	.002 (5)	.010 (7)
(2) groceries	.006 (9)	.002(10)	.036 (5)	.012 (8)	.017 (7)	.019 (3)	.002(10)	.021(10)	.014 (8)	.003 (8)	.003(10)	.000(11)	.028 (5)
(3) dairy products	.063 (5)	.002(11)	.202 (1)	.009 (9)	.017 (8)	.018 (4)	.000(13)	.023 (9)	.007 (9)	.013 (4)	.012 (8)	.000(12)	.020 (6)
(4) vegetables & fruit	.002 (6)	.000(13)	.000(13)	.259 (1)	.158 (1)	.004 (7)	.014 (9)	.001(13)	.000(12)	.001 (9)	.027 (5)	.000(13)	.001(13)
(5) meat	.226 (2)	.012 (7)	.067 (3)	.009(11)	.004(12)	.047(1)1	.024 (3)	.085 (2)	.020 (6)	.000(12)	.004 (9)	.021 (6)	.014 (8)
(6) fish	.136 (4)	.017 (6)	.048 (4)	.000(13)	.030 (4)	.001(10)	.000(14)	.031 (5)	.016 (7)	.005 (6)	.001(12)	.001(10)	.004(12)
(7) beverages	.307 (1)	.027 (4)	.000(14)	.022 (6)	.022 (6)	.001(11)	.019 (4)	.012(11)	.064 (2)	.007 (5)	.000(14)	.009 (7)	.000(11)
(8) tobacco	.005(10)	.020 (5)	.007(10)	.042 (3)	.000(11)	.002 (9)	.030 (2)	.078 (3)	.000(13)	.001(10)	.021 (7)	.005 (8)	.000(14)
(9) pastry, chocolate, ice cream	.002(11)	.005 (9)	.010 (9)	.000(14)	.000(14)	.006 (6)	.001(12)	.026 (7)	.242 (1)	.005 (7)	.023 (6)	.001 (9)	.038 (3)
(10) clothing and other textiles	.000(13)	.056 (3)	.013 (8)	.037 (4)	.002(13)	.000(13)	.005 (8)	.011(12)	.000(14)	.026 (3)	.038 (4)	.034 (3)	.013(09)
(11) footwear	.002(12)	.000(14)	.001(11)	.001(12)	.027 (5)	.000(14)	.017 (5)	.024 (8)	.002(10)	.000(13)	.207 (1)	.000(14)	.073 (4
(12) other textiles	.000(14)	.125 (1)	.024 (6)	.000(10)	.010 (9)	.020 (2)	.004 (9)	.000(14)	.002(11)	.030 (2)	.063 (3)	.158 (2)	.009(10)
(13) water, light, & heat	.162 (3)	.008 (8)	.001(12)	.026 (5)	.010(10)	.002 (8)	.002(11)	.020 (6)	.021 (5)	.000(14)	.001(13)	.032 (4)	.117 (2)
II. Multicollinearity $(\mathring{m}_L, \mathring{m}_U)$ (13)	−.334 (−3.30, .70)	.500 (−1.77, .85)	.226 (−1.89, .74)	.225 (−2.61, .76)	.146 (−1.44, .62)	.008 (−.39, .23)	.466 (−.63, .70)	.310 (−.94, .83)	.372 (−2.3, .85)	.600 (−1.09, .84)	.355 (−1.83, .86)	.421 (−2.26, .95)	.213 (1−.52, .69)
III. Total contribution of explanatory variables (R^{*2})	.695	.852	.740	.761	.618	.225	.695	.833	.846	.848	.858	.950	.690
Range of $s^2 \times$ values in (28)*	(0, .195)	(0, .097)	(0, .160)	(0, .152)	(0, .237)	(0, .442)	(0. .190)c	(0, .110)	(0, .102)	(0, .103)	(0, .093)	(0, .034)	(0, .196)

*The minimum of (28) is not exactly equal to 0 but is of the order of 10^{-4} or smaller for all equations.

TABLE 4. INCREMENTAL CONTRIBUTIONS, MULTICOLLINEARITY EFFECT, MULTIPLE CORRELATION, AND RIDGE PARAMETERS: CASELLA'S ESTIMATES $\bar{g} = 0$

	Regression Equations in (25)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
I. Incremental contributions of income	.048 (7)	.070 (2)	.082 (2)	.017 (7)	.083 (2)	.000(12)	.000 (1)	.045 (4)	.042 (4)	.149 (1)	.101 (2)	.247 (1)	.169 (1)
relative prices of													
(1) bread	.009 (8)	.001(12)	.021 (7)	.006 (2)	.074 (3)	.008 (5)	.008 (7)	.136 (1)	.043 (3)	.000(11)	.002(11)	.022 (5)	.019 (7)
(2) groceries	.005 (9)	.002(10)	.036 (5)	.012 (8)	.017 (7)	.017 (3)	.002(10)	.021(10)	.014 (8)	.003 (8)	.003(10)	.000(11)	.027 (5)
(3) dairy products	.063 (5)	.002(11)	.149 (1)	.008 (9)	.015 (8)	.012 (4)	.003(13)	.023 (9)	.007 (9)	.013 (4)	.012 (8)	.000(12)	.020 (6)
(4) vegetables & fruit	.062 (6)	.000(13)	.000(12)	.254 (1)	.157 (1)	.004 (7)	.012 (6)	.001(13)	.000(12)	.001 (9)	.027 (5)	.000(13)	.000(13)
(5) meat	.226 (2)	.012 (7)	.063 (3)	.000(11)	.001(13)	.025 (1)	.024 (3)	−.085 (2)	.020 (6)	.000(12)	.004 (9)	.021 (6)	.014 (8)
(6) fish	.135 (4)	.017 (6)	.048 (4)	.000(13)	.036 (4)	.001 (8)	.000(14)	.031 (5)	.016 (7)	.005 (6)	.001 (2)	.001 (9)	.004(12)
(7) beverages	.307 (1)	.027 (4)	.000(13)	.022 (6)	.022 (6)	.000(13)	.018 (4)	.012(11)	.064 (2)	.007 (5)	.000(14)	.000 (7)	.006(11)
(8) tobacco	.005(10)	.019 (5)	.007(10)	.042 (3)	.009(11)	.001 (9)	.030 (2)	.078 (3)	.000(13)	.001(10)	.020 (7)	.005 (8)	.000(14)
(9) pastry, chocolate, ice cream	.002(11)	.005 (9)	.010 (9)	.000(14)	.000(14)	.006 (6)	.001(12)	.026 (7)	.242 (1)	.005 (7)	.023 (6)	.001(10)	.038 (3)
(10) clothing and other textiles	.000(13)	.056 (3)	.013 (8)	.037 (4)	.002(12)	.000(14)	.005 (8)	.010(12)	.000(14)	.026 (3)	.038 (4)	.034 (3)	.013 (9)
(11) footwear	.002(12)	.000(14)	.000(14)	.001(12)	.027 (5)	.001(10)	.015 (5)	.024 (8)	.001(11)	.000(13)	.207 (1)	.000(14)	.037 (4)
(12) other textiles	.000(14)	.125 (1)	.023 (6)	.000(10)	.010(10)	.020 (2)	.004 (9)	.000(14)	.002(10)	.030 (2)	.063 (3)	.158 (2)	.009(10)
(13) water, light, & heat	.162 (3)	.008 (8)	.001(11)	.026 (5)	.011 (9)	.001(11)	.002(11)	.029 (6)	.021 (5)	.000(14)	.001(13)	.032 (4)	.116 (2)
II. Multicollinearity $(\overset{*}{m}_L, \overset{*}{m}_U)$ (13)	−.334 (−3.30, .69)	.500 (−.77, .85)	.228 (−1.21, .73)	.225 (−2.55, .76)	.146 (−1.44, .61)	.006 (−.14, .18)	.467 (−.64, .69)	.310 (−.94, .83)	.372 (−2.30, .85)	.601 (−1.10, .84)	.355 (−1.83, .86)	.421 (−2.26, .95)	.214 (−1.53, .69)
III. Total contribution of explanatory variables (R^{*2})	.693	.851	.730	.757	.610	.183	.686	.832	.845	.842	.857	.950	.685
Range of $s^2 \times$ values in (28)*	(0, .302)	(0, .118)	(0. .226)	(0, .210)	(0, .417)	(0, .277)	(0, .202)	(0, .137)	(0, .125)	(0, 127)	(0, .111)	(0, .030)	(0, .305)

*The minimum of (28) is not exactly equal to 0 but is of the order of 10^{-4} or smaller for all equations.

25b

TABLE 5. INCREMENTAL CONTRIBUTIONS, MULTICOLLINEARITY EFFECT, MULTIPLE CORRELATION, AND RIDGE PARAMETER: THE HKB ESTIMATES

I.	Regression Equations in (25)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Incremental contributions of income	.003 (9)	.068 (1)	.005 (5)	.020 (3)	.022 (4)	.001 (4)	.032 (2)	.054 (2)	.041 (2)	.089 (1)	.074 (2)	.176 (1)	.050 (2)
relative prices of													
(1) bread	.000(14)	.003 (9)	.000(13)	.058 (2)	.042 (2)	.000(14)	.006 (9)	.050 (3)	.026 (3)	.005(11)	.000(13)	.000(13)	.000(13)
(2) groceries	.006 (7)	.035 (4)	.000(14)	.002 (7)	.001(12)	.001 (5)	.023 (5)	.000(14)	.013 (7)	.057 (2)	.021 (6)	.005(10)	.006 (8)
(3) dairy products	.019 (6)	.000(11)	.121 (1)	.001 (9)	.018 (5)	.005 (2)	.005(11)	.007(10)	.016 (6)	.021 (5)	.040 (3)	.029 (3)	.001 (9)
(4) vegetables & fruit	.041 (4)	.000(13)	.000(11)	.251 (1)	.044 (1)	.000(12)	.030 (3)	.007(11)	.003(10)	.000(14)	.008 (9)	.000(14)	.001(10)
(5) meat	.062 (3)	.000(10)	.077 (2)	.000(13)	.024 (3)	.015 (1)	.003(12)	.083 (1)	.012 (8)	.009 (6)	.005(12)	.021 (4)	.000(11)
(6) fish	.087 (2)	.024 (7)	.065 (3)	.000(14)	.005 (9)	.000(13)	.011 (6)	.034 (7)	.002(12)	.005(10)	.009 (8)	.003(12)	.011 (5)
(7) beverages	.095 (1)	.041 (3)	.000(12)	.001(10)	.000(14)	.002 (3)	.027 (4)	.000(13)	.011 (9)	.039 (4)	.007(10)	.012 (7)	.011 (4)
(8) tobacco	.001(11)	.025 (6)	.000(10)	.011 (5)	.002(11)	.000 (7)	.001(13)	.040 (6)	.001(13)	.005 (9)	.033 (4)	.007 (9)	.007 (6)
(9) pastry, chocolate, ice cream	.006 (8)	.016 (8)	.001 (9)	.000(12)	.008 (7)	.000 (9)	.006 (8)	.026 (8)	.162 (1)	.006 (8)	.026 (5)	.003(11)	.046 (3)
(10) clothing and other textiles	.000(13)	.065 (2)	.012 (4)	.005 (6)	.000(13)	.000(10)	.001(14)	.012 (9)	.017 (5)	.047 (3)	.000(14)	.008 (8)	.000(14)
(11) footwear	.002(10)	.000(14)	.002 (6)	.000(11)	.013 (6)	.000 (8)	.055 (1)	.041 (4)	.000(14)	.000(13)	.150 (1)	.015 (6)	.006 (7)
(12) other textiles	.000(12)	.027 (5)	.001 (7)	.002 (8)	.007 (8)	.000(11)	.005(10)	.002(12)	.002(11)	.006 (7)	.016 (7)	.172 (2)	.000(12)
(13) water, light, & heat	.025 (5)	.000(12)	.001 (8)	.015 (4)	.002(10)	.001 (6)	.011 (7)	.039 (6)	.021 (4)	.002(12)	.007(11)	.015 (5)	.052 (1)
II. Multicollinearity $(m^{\bullet}_L, m^{\bullet}_U)$ •(13)	−.070 (-1.05, .276)	.199 (-.44, .50)	.236 (-1.17, .519)	.064 (-3.08, .432)	.028 (-.40, .217)	.002 (-.19, .029)	.057 (-.49, .257)	.082 (-.69, .476)	.078 (-1.86, .404)	.118 (-.34, .41)	.168 (-1.54m .533)	.274 (-1.72, .741)	.068 (-46.46, .258)
III. Total contribution of explanatory variables ($R^{\bullet 2}$)	.276	.504	.519	.432	.217	.029	.272	.474	.404	.410	.563	.741	.258
IV. Ridge parameter ($s^2 \times (29)$)	.113	.161	.146	.170	.297	1.777	.458	.112	.142	.223	.100	.036	.236

Casella's estimator. For 7 (or 8) of the 13 equations considered, the SMR estimator rates higher in terms of low multicollinearity than the LS (or Casella) estimator. This shows that the measure (13) gives results which are consistent with Casella's (1980) analytical result that there is a conflict between minimaxity and the multicollinearity problem. The additional information utilized by minimax (or near minimax) estimators can increase (or decrease) the multicollinearity in the data.

This conclusion is further strengthened by the values in Table 5. The reductions in R^{*2} and multicollinearity, and the changes in incremental contributions through the use of the HKB estimator are clearly and dramatically reflected in the values of this table. These reductions and changes are the consequences of the fact that the value of (29) is much larger than $\hat{\mu}$. That is, by using a sufficiently large value for μ in the HKB estimator we may be able to push the estimate of multicollinearity toward zero. Unfortunately, the HKB estimator with a large value of μ may be inferior to the LS estimator by the criterion of smaller mean square error. This follows from the fact that s^2 times (29) far exceeds the smallest diagonal element of s^2D in (28) given in Tables 3 and 4 for all the 13 equations and the HKB estimator is not (near) minimax with respect to the quadratic loss functions. These results point out to econometricians that there exist possible dangers involved in applying the standard approach to the problem of multicollinearity. Use of additional information as an aid in estimation to reduce multicollinearity may result in inefficient estimates. This should dissuade us from relying on the criterion of reducing multicollinearity exclusively.

On the other hand, the SMR and Casella estimators are nearly minimax and minimax respectively, with respect to the quadratic loss functions but do not lead to impressive reductions in the multicollinearity. The HKB estimator rates higher in terms of low multicollinearity than the other estimators considered, but does so with the sacrifice of a minimax condition. The biased estimators which are not minimax or nearly minimax appear to cope easily with multicollinearity that an unbiased and (near) minimax biased estimators find difficult. Therefore, we may not be able to reduce the multicollinearity in the data by using a minimax estimator of $\boldsymbol{\beta}$.

The problem of multicollinearity is difficult to solve if minimaxity is an absolutely essential property for an estimator. Berger (1982, p. 81) points out that any minimax estimator having uniformly smaller risk than the LS estimator ($\boldsymbol{b}$) will have substantially smaller risk only in a fairly small region of the parameter space. Ideally, we hope to find a minimax estimator that has substantially smaller risk than $\boldsymbol{b}$ in a small region which covers the "true" value of $\boldsymbol{\beta}$. Since such a region is not known, we, like Berger (1982, p. 83), feel that minimaxity is not an absolutely essential property for an estimator, particularly when minimaxity is dependent, as it usually is, on the mathematically convenient loss structure assumed. For this reason, we solve the multicollinearity problem by choosing a "near' minimax estimator which leads to a smaller measure of multicollinearity (13) than the LS estimator. Thus our formula (13) can play an important role in the choice of estimators.

We consider it less important in every way to determine the value of μ for which the multicollinearity estimate is smallest in absolute value rather than the value of μ which yields a near minimax biased estimator of $\boldsymbol{\beta}$. The value of $\hat{\mu}$ in (4) leads to such an estimator in most circumstances. By using the SMR estimator, we may both improve upon the squared error loss of the LS estimator (Swamy, Mehta and Rappoport, 1978) and not worsen the multicollinearity problem. The suggestion advanced is that every problem should be studied and if it is not known which loss function is acceptable or which estimator is best, the SMR estimator which has bounded second-order moments even in a highly collinear situation should be used.

4. Conclusions

The question of whether any of the biased estimation methods suggested in the statistics literature can cope with the multicollinearity problem better than the unconstrained least squares method cannot be answered without knowing the estimates of multicollinearity implied by both types of estimators. In order to answer this question we have developed in this paper a formula for measuring multicollinearity which takes different values for different coefficient estimators and/or for different equations. It takes the correct value of zero for an orthogonal model and a nonzero value for a multicollinear or nonorthogonal model. Applying this measure of multicollinearity to a few empirical examples, we have found that if one forces a biased regression estimator to satisfy the minimax conditions, then the other goal of reducing multicollinearity will not be realized. The principle for estimation emphatically rejected by us is that of minimizing the single measure, viz., the absolute value of the estimate of multicollinearity. We favor instead choosing from the near minimax estimators one that minimizes the absolute value of the estimate of multicollinearity. The estimator which is satisfactory from this point of view is the SMR estimator (4) which nearly satisfies Casella's

(1980, p. 1049) minimax conditions and does not worsen the multicollinearity problem. Casella's (1980, p. 1049) and HKB's (1975) estimators are not satisfactory because the former does not solve the multicollinearity problem and the latter may lead to inefficient estimates.

Appendix

Proposition : The measure (13) is degenerate at 0 iff $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}$ are *both* diagonal.

Proof : Recall (14) which is

$$\boldsymbol{y}'\boldsymbol{y}\widetilde{m} = \boldsymbol{y}'(\boldsymbol{A}_x-\boldsymbol{B}_x)\boldsymbol{y}. \qquad \ldots \quad (A.1)$$

It is now convenient to distinguish between the cases where $\boldsymbol{\Delta}_1^{-1} = 0$ and $\boldsymbol{\Delta}_1^{-1} \neq 0$.

Case 1. $\boldsymbol{\Delta}_1^{-1} = 0$. The quadratic form (A.1) is a continuous random variable with its matrix equal to $\boldsymbol{A}_x-\boldsymbol{B}_x$ which is fixed. Therefore (A.1) = 0 with probability 0 if $\boldsymbol{A}_x-\boldsymbol{B}_x$ is not null and = 0 with probability 1 if $\boldsymbol{A}_x-\boldsymbol{B}_x$ is null. From (A.1) it follows that $\boldsymbol{A}_x-\boldsymbol{B}_x = 0$ iff

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \sum_{h=1}^{K} [\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \boldsymbol{X}_{-h}(\boldsymbol{X}_{-h}'\boldsymbol{X}_{-h})^{-1}\boldsymbol{X}_{-h}'] = 0 \qquad \ldots \quad (A.2)$$

or, equivalently, iff

$$\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \sum_{h=1}^{K} \boldsymbol{M}_{-h}\boldsymbol{x}_h(\boldsymbol{x}_h'\boldsymbol{M}_{-h}\boldsymbol{x}_h')^{-1}\boldsymbol{x}_h'\boldsymbol{M}_{-h} = 0 \qquad \ldots \quad (A.3)$$

where $\boldsymbol{M}_{-h} = \boldsymbol{I} - \boldsymbol{X}_{-h}(\boldsymbol{X}_{-h}'\boldsymbol{X}_{-h})^{-1}\boldsymbol{X}_{-h}'$ and $\boldsymbol{x}_h$ is the h-th column of $\boldsymbol{X}$ and use is made of the matrix identity given in Theil (1971, p. 682, (B.23)).

Recognizing that both matrices in (A.3) reduce to the same matrix, $\sum_{h=1}^{K} \boldsymbol{x}_h(\boldsymbol{x}_h'\boldsymbol{x}_h)^{-1}\boldsymbol{x}_h'$, when $\boldsymbol{X}'\boldsymbol{X}$ is diagonal, we see that condition (A.3) is true if $\boldsymbol{X}'\boldsymbol{X}$ is diagonal; $\widetilde{m} = 0$ with probability 1 in this case.

To prove the necessity, suppose that condition (A.3) is true, so that $\widetilde{m}$ takes the value 0 with probability 1. Then the matrix

$$[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \sum_{h=1}^{K} \boldsymbol{M}_{-h}\boldsymbol{x}_h(\boldsymbol{x}_h'\boldsymbol{M}_{-h}\boldsymbol{x}_h)^{-1}\boldsymbol{x}_h'\boldsymbol{M}_{-h}]\boldsymbol{X}$$

$$\times \boldsymbol{X}'[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' - \sum_{h=1}^{K} \boldsymbol{M}_{-h}\boldsymbol{x}_h(\boldsymbol{x}_h'\boldsymbol{M}_{-h}\boldsymbol{x}_h)^{-1}\boldsymbol{x}_h'\boldsymbol{M}_{-h}] \qquad \ldots \quad (A.4)$$

must be null. This matrix simplifies to

$$\sum_{h=1}^{K} \hat{\boldsymbol{x}}_h \hat{\boldsymbol{x}}_h', \qquad \dots \quad (A.5)$$

where $\hat{\boldsymbol{x}}_h = \boldsymbol{X}_{-h}(\boldsymbol{X}_{-h}'\boldsymbol{X}_{-h})^{-1}\boldsymbol{X}_{-h}'\boldsymbol{x}_h$, which involves only sums of squares and products. The fact that (A.5) is a null matrix implies that $\hat{\boldsymbol{x}}_h = 0$ for $h = 1, 2, \dots, K$ which, in turn, implies that $\boldsymbol{X}'\boldsymbol{X}$ is diagonal. Hence $\widetilde{m}$ cannot take the value zero with positive probability if $\boldsymbol{X}'\boldsymbol{X}$ is not diagonal. However, $\widetilde{m}$ can take the value 0 for a specific value of $\boldsymbol{y}$ when $\boldsymbol{A}_x - \boldsymbol{B}_x$ is not null. In the cases in which $\boldsymbol{X}'\boldsymbol{X}$ is not diagonal, $\widetilde{m} \neq 0$ for all values of $\boldsymbol{y}$, except possibly in certain points, of which any finite interval contains at most a finite number.

Case 2. $\boldsymbol{\Delta}_1^{-1} \neq 0$. First, the value of $\boldsymbol{y} = 0$ (a vector of zeros) occurs with probability zero. In that event, $s^2\hat{\mu} = 0$, $\boldsymbol{A}_x - \boldsymbol{B}_x$ is fixed and null or nonnull as in Case 1 and the variable (A.1) takes the value zero with probability zero. Next, $\boldsymbol{y}$ can take an interval of values with positive probability. If this interval does not contain zero, then the matrix $\boldsymbol{A}_x - \boldsymbol{B}_x$ is random and has a constant rank over the entire interval. This rank depends only on the ranks of $\boldsymbol{X}'\boldsymbol{X}$ and $\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W}$. Consequently $\boldsymbol{y}$ cannot be an eigenvector of $\boldsymbol{A}_x - \boldsymbol{B}_x$ corresponding to its zero eigenvalue. (If it were, then (A.1) would be equal to zero with probability 1.) The variable (A.1) takes the value zero with positive probability iff the matrix $\boldsymbol{A}_x - \boldsymbol{B}_x$ is null for an interval of values of $\boldsymbol{y}$ not containing 0. We see from (14) that for the values of $\hat{\mu}s^2$ corresponding to these values of $\boldsymbol{y}$, $\boldsymbol{A}_x - \boldsymbol{B}_x = 0$ iff

$$2\boldsymbol{A}_{1x} - \boldsymbol{A}_{2x} = 0 \qquad \dots \quad (A.6)$$

where

$$\boldsymbol{A}_{1x} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + s^2\hat{\mu}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})^{-1}\boldsymbol{X}' - \sum_{h=1}^{K} [\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + s^2\hat{\mu}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})^{-1}\boldsymbol{X}' - \boldsymbol{X}_{-h}(\boldsymbol{X}_{-h}'\boldsymbol{X}_{-h} + s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h})^{-1}\boldsymbol{X}_{-h}'] \qquad \dots \quad (A.7)$$

and

$$\begin{aligned}\boldsymbol{A}_{2x} = {} & \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + s^2\hat{\mu}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + s^2\hat{\mu}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})^{-1}\boldsymbol{X}' \\ & - \sum_{h=1}^{K} \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + s^2\hat{\mu}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X} + s^2\hat{\mu}\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})^{-1}\boldsymbol{X}' \\ & + \sum_{h=1}^{K} \boldsymbol{X}_{-h}(\boldsymbol{X}_{-h}'\boldsymbol{X}_{-h} + s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h})^{-1}\boldsymbol{X}_{-h}'\boldsymbol{X}_{-h}(\boldsymbol{X}_{-h}'\boldsymbol{X}_{-h} \\ & + s^2\hat{\mu}(\boldsymbol{W}'\boldsymbol{\Delta}_1^{-1}\boldsymbol{W})_{-h,-h})^{-1}\boldsymbol{X}_{-h}'. \end{aligned} \qquad \dots \quad (A.8)$$

Since the matrices in (A.8) are obtained by squaring each matrix in (A.7), any matrix in (A.7) cannot cancel with any matrix in (A.8) and $\boldsymbol{A}_x - \boldsymbol{B}_x$ is null iff $\boldsymbol{A}_{1x} = \boldsymbol{0}$ and $\boldsymbol{A}_{2x} = \boldsymbol{0}$.

Defining

$$\boldsymbol{X} = [\boldsymbol{I}, \boldsymbol{0}] \begin{bmatrix} \boldsymbol{X} \\ s\sqrt{\hat{\mu}}\Delta_1^{-1/2}\boldsymbol{W} \end{bmatrix} = \boldsymbol{J}\boldsymbol{X}_e \qquad \ldots \quad (A.9)$$

we can rewrite (A.7) as

$$\boldsymbol{A}_{1x} = \boldsymbol{J}\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{J}' - \sum_{h=1}^{K} [\boldsymbol{J}\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{J}' - \boldsymbol{J}\boldsymbol{X}_{e,-h}(\boldsymbol{X}_{e,-h}'\boldsymbol{X}_{e,-h})^{-1}\boldsymbol{X}_{e,-h}'\boldsymbol{J}'] \qquad \ldots \quad (A.10)$$

where $\boldsymbol{X}_{e,-h}$ is formed from $\boldsymbol{X}_e$ by removing its h-th column, $\boldsymbol{x}_{e,h}$, and similarly we can rewrite (A.8) as

$$\boldsymbol{A}_{2x} = \boldsymbol{J}\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{J}'\boldsymbol{J}\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{J}' - \sum_{h=1}^{K} [\boldsymbol{J}\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{J}'\boldsymbol{J}\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{J}' - \boldsymbol{J}\boldsymbol{X}_{e,-h}(\boldsymbol{X}_{e,-h}'\boldsymbol{X}_{e,-h})^{-1}\boldsymbol{X}_{e,-h}'\boldsymbol{J}'\boldsymbol{J}\boldsymbol{X}_{e,-h}(\boldsymbol{X}_{e,-h}'\boldsymbol{X}_{e,-h})^{-1}\boldsymbol{X}_{e,-h}'\boldsymbol{J}']. \qquad \ldots \quad (A.11)$$

Again using the matrix identity in Theil (1971, p. 682, (B.23)), we can show that (A.10) is equivalent to.

$$\boldsymbol{A}_{1x} = \boldsymbol{J}[\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e' - \sum_{h=1}^{K} \boldsymbol{M}_{e,-h}\boldsymbol{x}_{e,h}(\boldsymbol{x}_{e,h}'\boldsymbol{M}_{e,-h}\boldsymbol{x}_{e,h})^{-1}\boldsymbol{x}_{e,-h}'\boldsymbol{M}_{e,-h}]\boldsymbol{J}' \quad \ldots \quad (A.12)$$

where $\boldsymbol{M}_{e,-h} = \boldsymbol{I} - \boldsymbol{X}_{e,-h}(\boldsymbol{X}_{e,-h}'\boldsymbol{X}_{e,-h})^{-1}\boldsymbol{X}_{e,-h}'$. By an argument parallel to that underlying (A.3)-(A.5), we see that the matrix

$$\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e' - \sum_{h=1}^{K} \boldsymbol{M}_{e,-h}\boldsymbol{x}_{e,h}(\boldsymbol{x}_{e,h}'\boldsymbol{M}_{e,-h}\boldsymbol{x}_{e,h})^{-1}\boldsymbol{x}_{e,h}'\boldsymbol{M}_{e,-h} \qquad \ldots \quad (A.13)$$

is null iff $\boldsymbol{X}_e\boldsymbol{X}_e'$ is diagonal. The matrix (A.12) is null iff (A.13) is null. Hence $\boldsymbol{A}_{1x}$ is null iff $\boldsymbol{X}_e'\boldsymbol{X}_e$ is diagonal.

First adding and subtracting the matrix

$$\sum_{h=1}^{K} \boldsymbol{J}\boldsymbol{X}_{e,-h}(\boldsymbol{X}_{e,-h}'\boldsymbol{X}_{e,-h})^{-1}\boldsymbol{X}_{e,-h}'\boldsymbol{J}'\boldsymbol{J}\boldsymbol{X}_e(\boldsymbol{X}_e'\boldsymbol{X}_e)^{-1}\boldsymbol{X}_e'\boldsymbol{J}' \qquad \ldots \quad (A.14)$$

to (A.11) and then applying the matrix identity in Theil (1971, p. 682, (B.23)), we have

$$A_{2x} = J[X_e(X_e'X_e)^{-1}X_e' - \sum_{h=1}^{K} M_{e,-h}x_{e,h} \cdot (x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}]J'JX_e(X_e'X_e)^{-1}X_e'H' - \sum_{h=1}^{K} JX_{e,-h}(X_{e,-h}'X_{e,-h})^{-1}X_{e,-h}'J'JM_{e,-h}x_{e,h} \cdot (x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}J'. \quad \ldots \text{(A.15)}$$

To rewrite this matrix in a convenient form, it is useful to add and subtract the matrix

$$\sum_{h=1}^{K} JX_e(X_e'X_e)^{-1}X_e'J'JM_{e,-h}x_{e,h} \cdot (x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}J'. \quad \ldots \text{(A.16)}$$

Doing so, we find that (A.15) changes to

$$A_{2x} = J[X_e(X_e'X_e)^{-1}X_e' - \sum_{h=1}^{K} M_{e,-h}x_{e,h} \cdot (x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}]J'J[X_e(X_e'X_e)^{-1}X_e' - \sum_{h=1}^{K} M_{e,-h}x_{e,h}(x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}]J' - \sum_{h=1}^{K} JM_{e,-h}x_{e,h}(x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'{}_eM_{,-h}J' \cdot \sum_{h \neq h'=1}^{K} JM_{e,-h}x_{e,h'}(x_{e,h'}'M_{e,-h}x_{e,h'})^{-1}x_{e,h'}'M_{e,-h'}J'. \quad \ldots \text{(A.17)}$$

The right-hand side matrices in this equation do not cancel one with another when either $X'X$ or $W'\Delta_1^{-1}W$ is nondiagonal.

Finally, we verify from (A.6), (A.12) and (A.17) that

$$A_x - B_x = 2J[X_e(X_e'X_e)^{-1}X_e' - \sum_{h=1}^{K} M_{e,-h}x_{e,h}(x_{e,h}'M_{e,-h}x_{e,h})^{-1}M_{e,-h}]J' - J[X_e(X_e'X_e)^{-1}X_e' - \sum_{h=1}^{K} M_{e,-h}x_{c,h}(x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}]J' \times J[X_e(X_e'X_e)^{-1}X_e' - \sum_{h=1}^{K} M_{e,-h}x_{e,h}(x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}]J' + \sum JM_{e,-h}x_{e,h}(x_{e,h}'M_{e,-h}x_{e,h})^{-1}x_{e,h}'M_{e,-h}J' \sum_{h \neq h=1}^{K} JM_{e,h'}x_{e,h'} \times (x_{e,h}'M_{e,h'}x_{e,h'})^{-1}x_{e,h'}'M_{e,-h'}J'. \quad \ldots \text{(A.18)}$$

In conjunction with the results in (A.13) and (A.17), this equation suffices to show that $\boldsymbol{A}_x - \boldsymbol{B}_x$ is null iff $\boldsymbol{X}_e'\boldsymbol{X}_e$ is diagonal.

Proposition 1 also implies that the sum of the modified incremental contributions of the K explanatory variables is exactly equal to their total contribution R^{*2} iff $\boldsymbol{X}_e'\boldsymbol{X}_e$ is diagonal.

References

Belsley, D. A. (1984): Demeaning conditioning diagnostics through centering (with discussion), *The American Statistician*, **38**, 73-93.

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

Berger, J. (1982): Selecting a minimax estimator of a multivariate normal mean, *Annals Statist*, **10**, 81-92.

Casella, G. (1980): Minimax ridge regression estimation, *Ann. Statist*, **8**, 1036-1056.

Chipman, J. S. (1964): On least squares with insufficient observations. *Jour. Amer. Statist. Assoc.*, **59**, 1078-1111.

——— (1976): Estimation and aggregation in econometrics: An application of the theory of generalized inverses. *Generalized Inverses and Applications*, ed. M. Zuhair Nashed, Academic Press, New York, 549-769.

Draper, N. R. and Smith, H. (1981): *Applied Regression Analysis*, 2nd ed., Wiley, New York.

Goldberger, A. S. (1968): *Topics in Regression Analysis*, The Macmillan Company, London.

Hoerl, A. D., Kennard, R. W., and Baldwin, K. F. (1975): Ridge regression: some simulations. *Communications in Statistics*, **4**, 105 -123.

Kashyap, A. K., Swamy, P. A. V. B., Mehta, J. S., and Porter, R. D. (1984): Estimating distributed lag relationships using near-minimax procedures, *Special Studies Paper*, Federal Reserve Board, Washington, D.C.

Poirier, D. J. (1976): *The Econometrics of Structural Change*, North-Holland Publishing Company, Amsterdam.

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

Shiller, R. J. (1973): A distributed lag estimator derived from smoothness priors, *Econometrica*, **41**, 775-788.

Swamy, P. A. V. B. and Mehta, J. S. (1983): Ridge regression estimation of the Rotterdam model. *Journal of Econometrics*, **22**, 365-390.

——— (1985): On a neglected measure of multicollinearity. *Special Studies Paper*, Federal Reserve Board, Washington, DC.

Swamy, P. A. V. B., Mehta, J. S., and Rappoport, R. N. (1978): Two methods of evaluating Hoerl and Kennard's ridge regression. *Commiunications in Statistics*, **7**, 1133-1165.

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

——— (1975): *Theory and Measurement of Consumer Demand*, 1, North-Holland Publishing Company.

THISTED, R. A. (1980): Comment on "A critique of some ridge regression methods, by Gary Smith and Frank Campbell", *Jour. Amer. Statist. Assoc.*, 75, 81-86.

THURMAN, S. S., SWAMY, P. A. V. B., and MEHTA, J. S. (1984): An examination of distributed lag model coefficients estimated with smoothness priors. *Special Studies Paper*, Federal Reserve Board, Washington, DC.

WONNACOTT, R. J. and WONNACOTT, T. H. (1970): *Econometrics*, Wiley, New York.

Paper received : July, 1984.

Revised : March, 1985.