Faculty Working Papers

MULTICOLLINEARITY AND THE CHOICE OF ESTIMATOR UNDER SQUARED ERROR LOSS

G. G. Judge and M. E. Bock

#122

College of Commerce and Business Administration
University of Illinois at Urbana-Champaign

FACULTY WORKING PAPERS

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

August 13, 1973

MULTICOLLINEARITY AND THE CHOICE OF ESTIMATOR UNDER SQUARED ERROR LOSS

G. G. Judge and M. E. Bock

#122

5 July, 1973

MULTICOLLINEARITY AND THE CHOICE OF ESTIMATOR UNDER SQUARED ERROR LOSS

G. G. Judge and M. E. Bock

University of Illinois

Urbana-Champaign

Using the traditional linear statistical model the impact of multicollinearity on the choice among conventional, pre test and variants of the Stein-James estimators is analytically evaluated under a squared error loss measure of goodness for both the conditional mean forecasting and parameter estimation problems.

5 July, 1973

MULTICOLLINEARITY AND THE CHOICE OF ESTIMATOR UNDER SQUARED ERROR LOSS

G. G. Judge and M. E. Bock
University of Illinois

1. Introduction

Most econometric ventures are experiments in non-experimental model building, and most economic observations are generated by a system where everything depends on everything else, and variables tend to move together over time. Attempts to capture the parameters of economic relations from these passively generated data means that in many cases the investigator must work with highly but not perfectly correlated sample information for the explanatory variables in conventional linear statistical models. The multicollinear characteristics of some non-experimental economic data imply that society's experimental design is such that the sample data in many cases are not rich enough to support a statistical search of the parameter space and permit the unknown parameters to be accurately estimated. Given this situation emphasis in the literature has centered on (i) detecting the existence, measuring the extent, and identifying the cause of multicollinearity [9, 17], and (ii) mitigating its impact through for example, the use of principal components [14] and factor analysis [16], or through procedures for enriching the sample information by say combining sample and external information, in the form of linear restrictions [20,21].

Against this background of work, we are interested in the impact of the incidence of near collinearity on the performance of and the choice among estimators for the general (non-orthogonal) regression model. As a basis for analyzing this choice problem, we broaden the family of estimators considered beyond the class of linear unbiased estimators, use squared error loss as the measure of goodness, and review and present new analytical results relative to the performance of conventional, restricted [22], preliminary test [1,2,5], and, Stein-rule estimators [11]. In particular, we generalize the conditional omitted variable specifications of Feldstein [10] and Toro-Vizcarrondo and Wallace [20], Ashar [1], and Wallace [21], as they relate to the multicollinearity problem, develop the analytical risk function for a variety of new Stein-James estimators [15], and explore how the incidence and extent of substantial collinearity conditions the performance of these estimators.

The plan of the paper is as follows. In section 2 the statistical models, estimators, and measures of goodness are specified. In section 3 the performance of the alternative estimators under the prediction goal is reviewed. In section 4 the risk functions for the estimators under the estimation goal are developed and a basis for how near collinearity conditions estimator choice is specified. Last, section 5 contains a summary of results and concluding remarks.

2. The Statistical Models, Estimators and Measures of Goodness

Assume the linear statistical model

(2.1) $$\underset{\sim}{y} = X\underset{\sim}{\beta} + \underset{\sim}{e},$$

where $\underset{\sim}{y}$ is a (T x 1) vector of observations, X is a known (T x K) matrix of non-stochastic variables of rank K, $\underset{\sim}{\beta}$ is a (K x 1) vector of unknown parameters and $\underset{\sim}{e}$ is a (T x 1) vector of unobservable normal random variables with mean $\underset{\sim}{0}$ and covariance $\sigma^2 I_K$:

The conventional estimators

The unrestricted least squares estimator based on the sample information contained in (2.1) is

(2.2) $\underline{b} = (X'X)^{-1}X'\underline{y} = S^{-1}X'\underline{y}$,

where $\underline{b}$ is distributed normally with mean $\underline{\beta}$ and covariance $\sigma^2 S^{-1}$. As is well known for the model (2.1), $\underline{b}$ is the maximum likelihood estimator, is unbiased, and under a quadratic loss measure is minimax.

When there is almost a perfect linear relation between two or more of the regressors in this linear statistical model, the S matrix is almost singular, and the smallest characteristic root approaches zero. One way traditionally used to cope with near collinearity is to enrich the sample data by taking into account other information. If we do this, following Toro-Vizcarrondo and Wallace [20], Chipman and Rao [7], Wallace [21], and Feldstein [10], in the form of J linear restrictions or general linear hypotheses about the unknown parameters in $\underline{\beta}$, we may specify this information as

(2.3a) $\underline{\delta} = \underline{0}$, with $R\underline{\beta} - \underline{r} = \underline{\delta}$,

where $\underline{r}$ is a (J x 1) vector of known elements, R is a (J x K) known matrix with rank J, and $\underline{0}$ is a (J x 1) null vector, $\underline{\delta}$ is a (J x 1) vector representing specification errors in the external information and $J \leq K$. Since our focus is on multicollinearity we will for expository purposes assume the conventional null hypothesis case $R=I_K$, and J=K, and $\underline{r}=\underline{0}$. Carrying through the more general case, as in [4,22], presents no complications but increases the algebra and decreases the expository sharpness of the results. Under this specification we can now write (2.3a) as

(2.3b) $\underline{\delta} = \underline{\beta}$.

In each case in terms of the estimators, test statistics, etc., we will give both the general and special case.

The restricted least squares or general linear hypothesis estimator, which makes use of both the sample and exact prior information or linear hypotheses, (2.1) and (2.3), is

(2.4) $$\hat{\beta} = b - S^{-1}R'(RS^{-1}R')^{-1}(Rb - r) = 0.$$

where $\hat{\beta}$ has mean square error $E(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = \delta\delta'$.

In applied work there is generally uncertainty concerning the statistical model that generated the data, and from the standpoint of multicollinearity there is a question concerning the linear restrictions or hypotheses to use. In cases such as this the investigator usually proceeds by statistical testing of the hypotheses followed by estimation. Thus, it is conventional when deciding questions concerning the truth or falsity of the general linear hypotheses or restrictions (2.3) to use likelihood ratio procedures of the traditional or Toro-Vizcarrondo and Wallace [20] variety and test the hypothesis $H: R\beta = r$ or $\beta = 0$, against not H, by using the test statistic

(2.5) $$u = (Rb-r)'(RS^{-1}R')^{-1}(Rb-r)/K\hat{\sigma}^2 = b'Sb/K\hat{\sigma}^2 .$$

H is rejected if u is greater than some critical value c, where u is distributed as the noncentral F distribution with J and T-K degrees of freedom and noncentrality parameter

(2.6) $$(R\beta-r)'(RS^{-1}R')^{-1}(R\beta-r)2\sigma^2 = \beta'S\beta'/2\sigma^2 .$$

The value of c is determined, for a given level of the test, α, by

(2.7) $$\int_c^{\infty} dF(u) = \alpha,$$

where F is a central F distribution with K and T-K degrees of freedom. By accepting H, we take $\hat{\beta}$ as our estimate of β, and by rejecting H, we use b the unrestricted least squares estimate.

In this conventional two stage testing or conditional omitted variables procedure, estimation is dependent on a test of significance, which implies the use of the preliminary test estimator [2,5],

$$\hat{\hat{\beta}} = I_{(0,c)}(u)\hat{\beta} + I_{[c,\infty)}(u)b = I_{[c,\infty)}(u)b, \tag{2.8}$$

where $I_{(0,c)}(u)$ and $I_{[c,\infty)}(u)$ are indicator functions which are one if u falls in the interval subscripted and zero otherwise. Thus, when one makes use of preliminary tests of significance in post data model evaluation in the case of a highly correlated set of regressors, this is the actual estimator used by researchers. The mean and covariance for this estimator has been derived by Bock, et.al. [5].

The Stein-rule estimators

As an alternative to the above conventional estimators, one extension of Stein-James estimator[11], for the K means of a multi normal distribution with identity covariance, to the regression model, results in the following Stein rule for combining the least squares and restricted least squares estimators[12]

$$\beta^* = (1-c^*/u)(b-\hat{\beta}) + \hat{\beta} = (1-c^*/u)b, \tag{2.9}$$

where c^* is a positive number.

Baranchik [3] and Stein [18] have modified the Stein-James estimator into a positive part version which may be extended for the statistical model (2.2) and (2.3) in the following way,

$$\beta^+ = (1-c^*/u)^+(b-\hat{\beta}) + \hat{\beta} = I_{[c^*,\infty)}(u)(1-c^*/u)b, \tag{2.10}$$

which has the form $\beta^+ = \hat{\beta} = 0$ if $u < c^*$,

and $\beta^+ = (1-c^*/u)b$ if $u \geq c^*$.

Building on this work, Sclove, et.al. [15] have developed a modified version of the Stein-James estimator which may be extended to our specification in the form

$$(2.11) \qquad \underset{\sim}{\beta}^{**} = I_{[c,\infty)}(u)(1-c^*/u)^+(\underline{b}-\underline{\hat{\beta}}) + \underset{\sim}{\hat{\beta}} = I_{[c,\infty)}(u)(1-c^*/u)\underline{b}$$

which has the form $\underline{\beta}^{**} = \underline{\hat{\beta}} = \underline{0}$ if $u \leq c$,

and $\underline{\beta}^{**} = (1-c^*/u)\underline{b}$, if $u > c$.

When $c = c^*$, the modified version (2.11) is the same as the extension of the positive part estimator (2.10).

The measure of goodness

Given this set of estimators we will evaluate the impact of multicollinearity on the choice of estimator by making use of the quadratic loss function

$$(2.12) \qquad L(\underline{\tilde{\beta}},\underline{\beta}) = (\underline{\tilde{\beta}}-\underline{\beta})'(\underline{\tilde{\beta}}-\underline{\beta}) = ||\underline{\tilde{\beta}}-\underline{\beta}||^2,$$

where $\underline{\tilde{\beta}}$ is any particular estimator with risk

$$(2.13) \qquad \rho(\underline{\tilde{\beta}},\underline{\beta}) = E[L(\underline{\tilde{\beta}},\underline{\beta})] = E(\underline{\tilde{\beta}}-\underline{\beta})'(\underline{\tilde{\beta}}-\underline{\beta}).$$

In comparing the risk function of two estimators we will say that the estimator $\underline{\tilde{\beta}}$ is superior to $\underline{\bar{\beta}}$ if

$$(2.14) \qquad E(\underline{\tilde{\beta}}-\underline{\beta})'(\underline{\tilde{\beta}}-\underline{\beta}) - E(\underline{\bar{\beta}}-\underline{\beta})'(\underline{\bar{\beta}}-\underline{\beta}) < 0, \text{ for all } \underline{\beta},$$

i.e. if the risk of the estimator $\underline{\tilde{\beta}}$ is less than $\underline{\bar{\beta}}$ over the region of the parameter space considered. In general, risk functions for alternative estimators cross. In other words, the difference in the risk of the estimators change sign for different regions of the parameter space. When this happens for the estimators considered in this paper we will identify the point(s) in the parameter space where the risk functions cross.

The reparametrized model

Although we are concerned with a situation where a near linear relation(s) exists between the explanatory variables, we have assumed that the X'X matrix is of full rank. Since it simplifies the algebra, we perform a canonical reduction on the statistical model (2.1) and the restrictions (2.3) and work with the following reparametrized model:

(2.15) $$\underline{y} = XS^{-\frac{1}{2}}S^{\frac{1}{2}}\underline{\beta} + \underline{e} = Z\underline{\theta} + e$$

and

(2.16) $$\underline{\delta}_o = \underline{\theta},$$

where $S^{\frac{1}{2}}$ is a positive definite symmetric matrix with $S^{\frac{1}{2}}S^{\frac{1}{2}} = S$, $Z'Z= I_K$, $\underline{\theta} = S^{\frac{1}{2}}\underline{\beta}$ and $Z'Z = S^{-\frac{1}{2}}(X'X)S^{-\frac{1}{2}} = I_K$. An estimator $\tilde{\underline{\theta}}$ for $\underline{\theta}$ yields an estimator $S^{-\frac{1}{2}}\,\tilde{\underline{\theta}} = \tilde{\underline{\beta}}$ for $\underline{\beta}$. This equivalent model leads to the least squares or maximum likelihood estimator for $\underline{\theta}$,

(2.17) $$\underline{\omega} = (Z'Z)^{-1}Z'\underline{y} = Z'\underline{y},$$

and the restricted least squares estimator

(2.18) $$\hat{\underline{\theta}} = \underline{0}.$$

The likelihood ratio test statistic becomes

(2.19) $$u = \underline{\omega}'\underline{\omega}\ (T-K)/K(\underline{y}-Z\underline{\omega})'(\underline{y}-Z\underline{\omega}) = \underline{\omega}'\underline{\omega}/K\hat{\sigma}^2$$

which has a non-central $F(\lambda, K, T-K)$ distribution with K and T-K degrees of freedom with

(2.20) $$\lambda = \underline{\theta}'\underline{\theta}/2\sigma^2,$$

and its use implies the preliminary test estimator for $\underline{\theta}$

(2.21) $$\hat{\hat{\underline{\theta}}} = \underline{\omega} - I_{(o,c)}(u)\underline{\omega}.$$

In terms of the reparametrized model the extension of the Stein-James estimator (2.10) becomes

(2.22) $\underline{\theta}^* = (1-c/u)\underline{\omega}$,

with the corresponding changes made for the extension of the positive part and the modified Sclove estimators (2.10) and (2.11).

3. Choice of Estimator under the Prediction Goal

In this section we consider the impact of multicollinearity on estimator choice when the objective is one of conditional mean forecasting. Our interest centers on comparing the risk functions

(3.1) $E(X\bar{\underline{\beta}} - X\underline{\beta})'(X\bar{\underline{\beta}} - X\underline{\beta}) = E(\bar{\underline{\beta}} - \underline{\beta})'X'X(\bar{\underline{\beta}} - \underline{\beta})$,

which weights the quadratic form $E(\bar{\underline{\beta}}-\underline{\beta})'(\bar{\underline{\beta}}-\underline{\beta})$ with the cross product matrix $S = X'X$, where $\bar{\underline{\beta}}$ is any of the six estimators for $\underline{\beta}$ that were developed in the previous section.

Since for the reparametrized model

(3.2) $$\rho(\bar{\underline{\theta}},\underline{\theta}) = E(\bar{\underline{\theta}}-\underline{\theta})'(\bar{\underline{\theta}}-\underline{\theta}) = E(\bar{\underline{\beta}}-\underline{\beta})'S^{\frac{1}{2}}S^{\frac{1}{2}}(\bar{\underline{\beta}}-\bar{\underline{\beta}})$$
$$= E[(X\bar{\underline{\beta}}-X\underline{\beta})'(X\bar{\underline{\beta}}-X\underline{\beta})],$$

we can make our comparisons with an unweighted risk function in terms of $\underline{\theta}$. Since the explanatory variables, Z, are orthogonal, multicollinearity is not a problem, and therefore, in the conditional mean forecasting case, conventional results regarding the risk of the alternative estimators hold and may, from the work of [11, 15, 18, 21, 22, 24], be summarized as follows:

i) In terms of (3.2) the risk of the unrestricted least squares estimator $\underline{\omega}$ is $\rho(\underline{\omega},\underline{\theta}) = \sigma^2 K$, and the risk of the restricted least squares estimator $\hat{\underline{\theta}}$ is $\rho(\hat{\underline{\theta}},\underline{\theta}) = \underline{\theta}'\underline{\theta}$. If the restrictions are correct, $\underline{\theta} = \underline{0}$, $\rho(\hat{\underline{\theta}},\underline{\theta}) < \rho(\underline{\omega},\underline{\theta})$. If $\underline{\theta} \neq \underline{0}$, in order that $\rho(\hat{\underline{\theta}},\underline{\theta}) - \rho(\underline{\omega},\underline{\theta}) \leq 0$, then $\sigma^2 K - \underline{\theta}'\underline{\theta}$ must be non-negative and this implies the condition $\underline{\theta}'\underline{\theta}/\sigma^2 \leq K$ or in terms of the noncentrality parameter for the test statistic, $\lambda \leq K/2$, must be satisfied. The risk for the restricted estimator, in terms of $\underline{\theta}$ is unbounded, and we have the typical situation when the risk functions for the two estimators cross.

ii) Under (3.2), with a weighted squared error loss measure of goodness in terms of $\underline{\beta}$, from the work of Cohen [8], Sclove, et.al. [15], and Bock, et.al.[5] on the preliminary test estimator $\hat{\hat{\underline{\theta}}}$, if the restrictions are correct, $\underline{\theta} = \underline{0}$, and $\rho(\hat{\hat{\underline{\theta}}},\underline{\theta}) < \rho(\underline{\omega},\underline{\theta})$. If $\underline{\theta} \neq 0$, then it is necessary that $\underline{\theta}'\underline{\theta}/\sigma^2 \geq K/2$ or $\lambda \geq K/4$ in order for the risk of the preliminary test estimator to be smaller than that of the least square estimator. Alternatively for $\rho(\hat{\hat{\underline{\theta}}},\underline{\theta}) > \rho(\underline{\omega},\underline{\theta})$ then, $\underline{\theta}'\underline{\theta}/\sigma^2 > K$ or $\lambda > K/2$. Although there are conditions under which the pre test estimator has a smaller risk than the conventional least squares estimator, the pre test estimator is inferior to the least squares estimator over an infinite interval of the parameter space of $\lambda = \underline{\theta}'\underline{\theta}/2\sigma^2$.

iii) In terms of the measure of goodness reflected by (3.2) and from the work of James and Stein[11], if $K > 2$ and $0 < c < 2(K-2)/(T-K+2)$, then the Stein-James estimator $\underline{\theta}^*$ is uniformly superior to the least squares estimator $\underline{\omega}$. The optimal choice of c is $c_o = (K-2)/(T-K+2)$.

iv) Under (3.2) the Stein-James positive part estimator

(3.3) $\underline{\theta}^+ = (1/c^*/u)^+\underline{\omega} = I_{[c^*,\infty)}(u)(1-c^*/u)\underline{\omega}$, where $o < c^* \leq 2(K-2)/(T-K+2)$ or $0 < c^* \leq 2c_o$, is uniformly superior to the Stein-James estimator (2.24) and thus, demonstrates its inadmissibility under squared error loss [11,18]. In addition as Bock has shown [4], if $c \leq c^*$ and $K \geq 3$, for comparable values of c, the positive part estimator dominates the preliminary test estimator.

v) The positive part version of the Sclove, et.al. [15] modified Stein-James estimator is

(3.4) $\underset{-}{\theta}^{**} = I_{[c,\infty)}(u)(1-c^*/u)^+ \underset{-}{\omega}$, where $o < c^* < 2(K-2)/(T-K+2)$.

If $c \leq 2(K-2)/(T-K+2)$ let $c^* = c$, then $\underset{-}{\theta}^{**} = \underset{-}{\theta}^+ = \underset{-}{\hat{\hat{\theta}}} - c/u\ I_{[c,\infty)}(u)\underset{-}{\omega}$.

Alternatively, if $c > 2(K-2)/(T-K+2)$ let $c^* \leq 2(K-2)/(T-K+2)$.

Then $\underset{-}{\theta}^{**} = \underset{-}{\hat{\theta}} - c^*/u\ I_{[c,\infty)}(u)\underset{-}{\omega}$.

If the value of c is equal to or less than c*, then (3.4) is the conventional positive part estimator, and this estimator is uniformly superior over the range of the parameter space to the least squares (2.18), pre test (2.23) and Stein-James (2.24) estimators. If c > c* the modified Sclove positive part estimator (3.4) is uniformly superior to the conventional preliminary test estimator (2.23). When $c \leq 2c_o$ the modified Stein-James estimator (3.4) provides a minimax substitute for the conventional preliminary test estimator (2.23). The estimator given in (3.4) is in reality a preliminary test estimator with the outcome of the of the preliminary test, at a level of significance dictated by the value of c, resulting in either a selection of the Stein-James positive part estimator or the restricted least squares estimator.

In summary, in the prediction case, although the X's may be "almost collinear", as long as X is of rank K the conventional results for estimator choice under quadratic loss hold. The Stein-James positive part estimator (3.3) dominates and thus provides a minimax counterpart for the conventional (2.18), Stein-James (2.24), and pre test (2.22) estimators, when $K \geq 3$ and $0 \leq c \leq 2c_o$, and the Sclove estimator (3.4) dominates the conventional preliminary test estimator over the whole range of c. It would appear that with or without multicollinearity, if we are willing to leave the class of unbiased estimators, for prediction purposes a version of the Stein-James preliminary test estimator should be our choice. It should be remarked here that although the pre test Stein-James estimator is minimax, when the conditions $K > 2$ and $0 < 2c_o \leq (K-2)/(T-K+2)$ are fulfilled, this estimator along with others using

Stein rules, are not admissible. Strawderman [19] has developed an estimator of this general form, that is both minimax and admissible, when certain conditions are fulfilled (one being that $K \geq 5$).

4. The Choice of Estimator Under the Estimation Goal

If we are interested in a measure of goodness involving an unweighted risk function in terms of the originial parameters β, then under squared error loss and the reparametrized model of the last section

$$(4.1) \qquad \rho(\tilde{\beta},\beta) = E[(\tilde{\beta}-\beta)'(\tilde{\beta}-\beta)] = E[(\tilde{\theta}-\theta)'D(\tilde{\theta}-\theta)] = \rho_1(\tilde{\theta},\theta)$$

where $D = S^{-1}$ and $\tilde{\theta} = S^{\frac{1}{2}}\tilde{\beta}$. An unweighted risk function for β implies a weighted risk function for θ in the reparametrized model and indicates why the $X'X = I$ case traditionally analyzed in the statistical literature is not sufficient for gauging estimator performance for the general (usual) case when the emphasis is on estimation and $X'X$ is some positive definite symmetric matrix.

From the standpoint of multicollinearity this focus on parameter estimation is relevant since we are concerned with the implications or incidence of multicollinearity on the comparative sampling performance of the alternative estimators of β.

Risk of traditional estimators

Using the measure of performance reflected by (4.1) the risk for the conventional least squares estimator is

$$(4.2) \qquad \rho(b,\beta) = E[(b-\beta)'(b-\beta)] = E[(\omega-\theta)'D'(\omega-\theta)] = E[(\omega-\theta)'S^{-1}(\omega-\theta)]$$
$$= \rho_1(\omega,\theta) = \sigma^2 \operatorname{tr} S^{-1} .$$

In contrast the risk for the restricted least squares estimator, in the context of Wallace [21], and Yancey, et.al. [22] is

$$\rho(\hat{\beta},\beta) = E[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)] = E[(\hat{\theta}-\theta)'D(\hat{\theta}-\theta)] = \rho_1(\hat{\theta},\theta) \tag{4.3}$$
$$= \rho_1(\omega,\theta) + \theta'S^{-1}\theta - \sigma^2 trS^{-1} = \theta'S^{-1}\theta.$$

The difference in the risk of the least squares estimator (4.2) and the risk of the restricted estimator (4.3) is

$$\rho(b,\beta) - \rho(\hat{\beta},\beta) = \sigma^2 trS^{-1} - \theta'S^{-1}\theta \tag{4.4}$$

which is non negative if

$$\sigma^2 trS^{-1} \geq \theta'S^{-1}\theta. \tag{4.5}$$

From the work of Wallace [21], and Yancey, et.al. [22] is

$$\sigma^2 trS^{-1} \geq d_L\, \theta'\theta = d_L 2\lambda\sigma^2 \tag{4.6}$$

where the d_i are the roots of S^{-1}, with d_L being the largest, and $t_i = d_i/\sum_{i=1}^{K} d_i$.

From (4.6) the difference in the risks (4.4) will be non-negative if

$$\lambda \leq 1/2t_L = (1/2)\sum_{i=1}^{K} d_i/d_L = (1/2)trS^{-1}/d_L . \tag{4.7}$$

Alternatively equation (4.6) will be non positive and the risk of the conventional estimator less than the restricted estimator if

$$\lambda \geq 1/2t_S = (1/2)\sum_{i=1}^{K} d_i/d_S = (1/2)trS^{-1}/d_S = \lambda_1 \tag{4.8}$$

where d_S is the smallest root of S^{-1}. Therefore, the risk functions of the unrestricted least squares and restricted least squares estimators cross for some value of λ in $[\lambda_o,\lambda_1]$ where

$$1/2t_L \leq \lambda_o \leq \lambda_1 \leq 1/2t_S. \tag{4.9}$$

Since the incidence of collinearity between the explanatory variables means that the smallest root of $X'X = S$ approaches zero, the largest root of S^{-1} approaches infinity. As the degree of collinearity increases the largest root d_L of S^{-1} increases and the range of the parameter space, in terms of λ, over which the risk of the restricted least squares estimator is less than that of the least squares estimator, approaches one half. The interval of uncertainty in terms of the equality of the risk functions depends of course on the relative sizes of the roots, d_L and d_S. Therefore, the degree of collinearity in the explanatory variables affects, for a given λ, the location in the parameter space where the risk for one estimator is equal to less than or greater than that of another estimator, and thus, the choice of the estimator.

Risk of the pre test estimator

Alternatively following Wallace [21], Feldstein [10], and others in using either a new or conventional test statistic along with a preliminary test of significance rule, we now compare analytically using (4.1), the risks for the resulting pre test estimator with that of the least squares estimator and analyze the impact of collinearity on estimator choice.

The risk for the pre test estimator from the work of Bock, et.al. [6] is

$$\text{(4.10)} \qquad \rho_1(\hat{\underline{\beta}},\underline{\beta}) = E[(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta})] = E[(\hat{\underline{\theta}}-\underline{\theta})'D(\hat{\underline{\theta}}-\underline{\theta})]$$
$$= \rho_1(\underline{\omega},\underline{\theta}) - \sigma^2 p_1 \mathrm{tr}S^{-1} + (2p_1 - p_2)\,\underline{\theta}'S^{-1}\underline{\theta}$$

where p_j is the probability of a random variable with a non-central f distribution being smaller than a constant, i.e., $p_j = \Pr[X_{2(\lambda,K+2j)}/X_{2(T-K)} \leq cK/T-K]$, and all other symbols were previously defined in connection with (3.6) and (4.3).

Since the risk depends on both λ and $\underline{\theta}'S^{-1}\underline{\theta}$ the risk for the pre test estimator may be bounded by

$$(4.11) \qquad \rho_1(\hat{\hat{\beta}},\beta)/\sigma^2 \leq (1-p_1)\mathrm{tr}S^{-1} - 2t_L\lambda \mathrm{tr}S^{-1}(p_2-2p_1)$$

and

$$(4.12) \qquad \rho_1(\hat{\hat{\beta}},\beta)/\sigma^2 \geq (1-p_1)\mathrm{tr}S^{-1} - 2t_S\lambda \mathrm{tr}S^{-1}(p_2-2p_1),$$

where t_S and t_L are defined as in (4.6), (4.7), and (4.8).
The difference in the risk of the pre test and conventional least squares estimators (4.2) and (4.10) is

$$(4.13) \qquad \rho(b,\beta) - \rho(\hat{\hat{\beta}},\beta) = \sigma^2\mathrm{tr}S^{-1}[p_1 + 2(p_2 - 2p_1)]\theta'S^{-1}\theta/2\sigma^2\mathrm{tr}S^{-1}.$$

This difference in the risk functions (4.13) will be non-negative if $\lambda \leq 4t_L$ and non-positive if $\lambda \geq 1/2t_S$. Therefore the risk functions for the preliminary test and least squares risk estimators cross for some values of λ in the interval

$$(4.14) \qquad 1/4t_L \leq \lambda = \theta'\theta/2\sigma^2 \leq 1/2t_S.$$

It is not possible to be more precise about the relation between the risk functions unless more information exists about δ_o or θ.

This outcome reflected by (4.14) contrasts to the range of uncertainty for λ in the pre test prediction case of $J/4 \leq \lambda \leq J/2$. Thus again, the degree of collinearity, as reflected by the size of the roots of the S^{-1} matrix, conditions the range of λ over which the risk of the pre test estimator is less than that of the least squares estimator. In the face of multicollinearity this range of uncertainty may be very large since the smallest root of S approaches zero and the largest root of S^{-1} approaches infinity. Thus, the interval of the parameter space where there is a gain from testing, may be very small

indeed, and the losses for the pre test estimator relative to the least squares estimator may be positive and large over a significant interval of the parameter space. The Monte Carlo results of Feldstein [10] appear consistent with the analytical results presented above.

Risk of the Stein Rule estimators

Let us now consider the impact of collinearity on the choice between the conventional least squares and the extensions of the Stein-James estimators. The risk of an extension of the Stein-James estimator, $\underline{\beta}^* = (1-c/u)\underline{b} = (1-c/u)S^{-\frac{1}{2}}\underline{\omega} = S^{-\frac{1}{2}}\theta^*$, is

$$\rho_1(\underline{\beta}^*,\underline{\beta}) = E[(\underline{\theta}^*-\underline{\theta})'D(\underline{\theta}^*-\underline{\theta})] = \sigma^2 trS^{-1} + \sigma^2 trS^{-1} c^*(T-K) \tag{4.15}$$

$$E\{(1/(K+2H)(K-2+2H))[c^*(T-K+2)-2(K-2)+(\underline{\theta}'S^{-1}\underline{\theta}/\underline{\theta}'\underline{\theta}(trS^{-1}))2H$$

$$(c^*(T-K+2)-2(\underline{\theta}'\underline{\theta}(trS^{-1}/\underline{\theta}'S^{-1}\underline{\theta} -2))]\}$$

where H is a Poisson random variable with parameter $(\lambda/2)$.

In order for the expression between the brackets to be zero or negative and thus the risk of the Stein-James estimator to be equal to or less than the risk of the least squares estimator over the range of the parameter space (i.e. be uniformly superior) then

$$\sum_{i=1}^{K} d_i \,/\, d_L = trS^{-1} / d_L \geq 2, \tag{4.16a}$$

and

$$c \leq 2d_L^{-1}\,(trS^{-1} - 2d_L)/(T-K+2). \tag{4.16b}$$

Thus the uniform superiority of this extension of the Stein-James estimator to the conventional estimators for the general regression model depends not only on the number of explanatory variables or hypotheses, as was the requirement for the orthonormal or prediction case, but also on whether or not trS^{-1} divided by the largest root of S^{-1} is equal to or larger than 2. If $trS^{-1}/d_L \leq 2$, then for no value of $c > 0$ does this extension of the Stein-James dominate the least squares estimator. Since the degree of collinearity is related to the magnitude of the roots of S^{-1} it therefore affects whether or not the risk functions cross at some point in the parameter space and thus has a direct impact on the choice of estimator. In addition, if multicollinearity exists, then at least one of the roots of X'X will approach zero, and thus d_L, the largest root of S^{-1}, will approach infinity and trS^{-1}/d_L may well be less than 2. That this situation may often occur in econometric work can be seen from the following first order correlation matrix for four explanatory variables of sample size 10, which was initially generated to reflect the characteristics of economic time series data and used in a Monte Carlo study [23]:

$$(4.17) \qquad \begin{bmatrix} 1.00 & & & \\ .58 & 1.00 & & \\ .76 & .28 & 1.00 & \\ .44 & .29 & .87 & 1.00 \end{bmatrix}$$

This correlation matrix, which is certainly not atypical of that reflected by much passively generated economic sample data, has one root for the X'X matrix which is .000006 and small relative to the other roots. Thus $trS^{-1}/d_L < 2$ and in fact is very close to one.

These same conditions or requirements ($trS^{-1}/d_L \geq 2$) also hold in order for this particular extension of the Stein-James positive part estimator, $\underline{\beta}^{+}$, and the extension of the Sclove-Stein-James preliminary test estimator, $\underline{\beta}^{**}$, to dominate the least squares and conventional preliminary test estimators respectively. This means that when we are concerned with the risk for the estimation case the appearance of 3 or more regressors and a suitable small c do not insure, as they did in the prediction case, that the risk of these various extensions of Stein rule estimators, will be less over the entire parameter space than conventional and pre test estimators. As a consequence we have a new rule (4.16a) for determining the degree of collinearity that is permissible to permit these extensions of the Stein rule estimators to dominate the other sampling theory estimators.

5. Concluding Remarks

In terms of impact on and choice among estimators, for the orthonormal and general regression models under a squared error loss measure of goodness, multicollinearity appears to have the following affects when all or a subset of the regressors are almost perfectly collinear, but when the X regression matrix has full column rank:

i) If the objective is conditional mean forecasting, conventional results for the choice between estimators hold; i.e. (a) under conditions normally found in practice the Stein or modified Stein rule dominate the conventional least squares and preliminary test estimators and (b) although the risk of the pre test estimator is smaller than that of the least squares estimators over a part of the parameter space there is an interval of infinite length of the space where this superiority does not hold.

ii) If the objective or emphasis is on estimation, the incidence of multicollinearity (a) conditions the interval of the parameter space when the risk of the conventional pre test estimator can be said to be less than or exceed that of the least squares estimator, (b) means that in the case when the X regressor matrix is "borderline" full rank, and thus at least one root of the relevant weight matrix approaches infinity, the range of uncertainty as to estimator choice between the pre test and least squares estimator goes over almost the entire range of the parameter space, (c) means the appearance of 3 or more regressors and a small c does not insure that these extensions of the Stein-James estimators will dominate the least squares and conventional preliminary test estimators, (d) means that in order for these extensions of the Stein-James estimators to be uniformly superior to conventional estimators, the ratio of the sum of the characteristic roots of the S^{-1} matrix (4.5), to the largest root of this matrix must be equal to or greater than 2, rather than the traditional condition $K > 2$, (e) means that if this ratio is not greater than two then some members of the family of potential risk functions for the extensions of the Stein-James estimators in a given problem cross the risk function for the least squares estimator, and (f) means that if the conditions (4.16a and b) are fulfilled, the extension of the Stein rule pre test estimator (2.11), for the general model, is uniformly superior to the conventional pre test estimator over the parameter space, but like the conventional pre test estimator its risk function crosses that of the least squares estimator for large values of the critical value c or small values of α, the level of the test; (2.11) is only a minimax estimator for smaller values of c or larger values of α than are ordinarily used.

These analytical results suggest that when multicollinearity is present to the extent that the $\mathrm{tr}S^{-1} < 2d_L$, under a squared error loss measure of goodness the restricted, pre test and the family of Stein rule estimators are superior (smaller risk) to the least squares estimator only over a very small

interval of the parameter space and are inferior (larger risk) over a large and in some cases infinite range of the parameter space. Unless the researcher has great confidence that his linear hypotheses $R\beta - r = \delta = 0$ are true, under the risk measure we have employed when multicollinearity is present, he has much to lose and very little to gain by broadening the class of estimators and using the two stage procedure.

6. References

[1] Ashar, V.G. (1970): "On the Use of Preliminary Tests in Regression," Unpublished Thesis, North Carolina State University, Raleigh.

[2] Bancroft, T.A. (1944): "On Biases in Estimation Due to the Use of Preliminary Tests of Significance," Annals of Mathematical Statistics, Vol. 15, pp. 190-204.

[3] Baranchik, A.J. (1964): "Multiple Regression and the Estimation of the Mean of a Multivariate Normal Distribution," Stanford University Technical Report, no. 51.

[4] Bock, M.E. (1973): "Minimax Estimators of the Mean of a Multivariate Distribution," Research Paper, University of Illinois.

[5] Bock, M.E. (1972): "A Comparison of the Risk Functions for Preliminary Test and Positive Part Estimators," Research Paper, University of Illinois.

[6] Bock, M.E., T.A. Yancey, and G.G. Judge (1972): "The Statistical Consequences of Preliminary Test Estimators in Regression," Journal of American Statistical Association, forthcoming.

[7] Chipman, J.S. and M.M. Rao (1964): "The Treatment of Linear Restrictions in Regression Analysis," Econometrica, Vol. 32, pp. 198-209.

[8] Cohen, A. (1965): "Estimates of the Linear Combination of Parameters in the Mean Vector of a Multivariate Distribution," Annals of Mathematical Statistics, Vol. 46, pp. 78-87.

[9] Farrar, D.E. and R.R. Glauben (1967): "Multicollinearity in Regression Analysis: The Problem Revisited," Review of Economics and Statistics, Vol. XLIX, pp. 92-107.

[10] Feldstein, M.S. (1971): "Multicollinearity and the Mean Square Error of Alternative Estimators," to appear in Econometrica.

[11] James, W. and C. Stein (1961): "Estimation with Quadratic Loss," Proceedings of Fourth Berkeley Symposium on Mathematical Statistic Problems, University of California Press, Berkeley, pp. 361-379.

[12] Judge, G.G., M.E. Bock, and T.A. Yancey (1973): "On Post Data Model Evaluation," to appear in Review of Economics and Statistics.

[13] Larson, H.J. and T.A. Bancroft (1963a): "Sequential Model Building for Prediction in Regression Analysis," Annals of Mathematical Statistics, Vol. 34, pp.462-79.

[14] Massey, W.F. (1965): "Principal Components Regression in Exploratory Statistical Research," Journal of the American Statistical Association, Vol. 60, pp. 234-256.

[15] Sclove, S.L., C. Morris, and R. Radhakrishnan (1972): "Non Optimality of Preliminary-Test Estimators for the Multinormal Mean," Annals of Mathematical Statistics, Vol. 43, no. 5, pp. 1481-90.

[16] Scott, J.T. (1966): "Factor Analysis and Regression," Econometrica Vol. 34, pp. 552-62.

[17] Silvey, S.D. (1969): "Multicollinearity and Imprecise Estimation," Journal of the Royal Statistical Society, Vol. 31 (Series B), pp. 539-52.

[18] Stein, C. (1966): "An Approach to the Recovery of Interblock Information in Balanced Incomplete Block Designs," Research Papers in Statistics: Festschrift for J. Neyman, F.N. David (ed.), John Wiley and Sons, Inc., New York, pp. 351-66.

[19] Strawderman, W.E. and A. Cohen (1971): "Admissibility of Estimators of the Mean Vector of a Multivariate Normal Distribution with Quadratic Loss," Annals of Mathematical Statistics, Vol. 42, pp. 270-96.

[20] Toro-Vizcarrondo, C. and T.D. Wallace (1968): "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," Journal of the American Statistical Association, Vol. 63, pp. 558-72.

[21] Wallace, T.D. (1971): "Weaker Criteria and Tests for Linear Restrictions in Regression," to appear in Econometrica.

[22] Yancey, T.A., G.G. Judge, and M.E. Bock (1973): "Wallace's Mean Square Error Criterion for Testing Linear Restrictions in Regression: A Tighter Bound," Forthcoming in Econometrica.

[23] Yancey, T.A., M.E. Bock, and G.G. Judge (1972): "Some Finite Sample Results for Theil's Mixed Regression Estimator," Journal of American Statistical Association, Vol. 67, pp. 176-79.

[24] Zellner, A. and W. Vandaele (1972): "Bayes-Stein Estimators for k Means, Regression, and Simultaneous Equation Models," Unpublished Paper, University of Chicago.