Faculty Working Papers

PROPERTIES OF SOME PRELIMINARY TEST ESTIMATORS IN REGRESSION USING A QUADRATIC LOSS CRITERION

M. E. Bock, T. A. Yancey, G. G. Judge

#62

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

September 6, 1972

PROPERTIES OF SOME PRELIMINARY TEST ESTIMATORS IN REGRESSION USING A QUADRATIC LOSS CRITERION

M. E. Bock, T. A. Yancey, G. G. Judge

#62

PROPERTIES OF SOME PRELIMINARY TEST ESTIMATORS IN REGRESSION USING A QUADRATIC LOSS CRITERION

M.E. Bock, T.A. Yancey, G.G. Judge*
University of Illinois at Urbana-Champaign

This study is concerned with deriving the properties of the preliminary test estimator for the general linear normal regression model and determining the conditions necessary for the risk of this estimator to exceed or be less than the conventional one under a quadratic loss criterion. A test procedure and the problem of choosing an optimal level of significance for the test are discussed.

1. Introduction

In much of the work concerned with estimating the parameters of behavioral and technical relations, there is uncertainty as to the appropriate model to be used. As a consequence, investigators begin with an initial set of specifications and then modify their models by testing the statistical significance of some or all of a class of hypotheses. This process makes the model and thus the estimation procedure dependent on the outcome of the tests of hypotheses and leads to, what has been termed in the literature, preliminary test or sequential estimators. Fortunately, this class of statistical procedures has been studied, starting with Bancroft in the early 1940's, by Mosteller (1948), Kitagawa (1963), Huntsberger (1965), Larson and Bancroft (1963a,1963b), Bancroft (1964), to determine the properties of the resulting statistics in terms of their means and mean square errors. Cohen (1965) showed that under certain assumptions for estimation with quadratic loss,

*The authors have benefited from papers by, and comments from, T.D. Wallace, S.L. Sclove and T.A. Bancroft.

the preliminary test estimator is inadmissible. Unfortunately, he did not suggest a superior estimator. Toro-Vizcarrondo (1968) and Wallace (1971) suggest a practical procedure for determining the estimator to use based on a test of compatibility of sample and exact prior information in a regression model and in so doing implied a preliminary test estimator. Ashar (1970) studied the conditional omitted variable (preliminary test) estimator for the regression model. In an unpublished paper, Sclove et al. (1970) show when certain conditions are fulfilled that the preliminary test estimator is dominated by the positive part version of the James-Stein (1961) estimator. Unfortunately, the conclusions flowing from this result are of limited significance for practitioners since (i) only the orthonormal regressor case is considered and extension to the non-orthonormal or general case is not direct since in reparametrizing the model the measure of goodness is changed; (ii) the number of regressors must be strictly greater than 2; (iii) the critical value of the test statistic is constrained to lie within a range that implies, for the usual sample sizes and numbers of regressors, a risk function very close to that of the conventional estimator; (iv) the risk for the positive part estimator is, over the range of critical test values that are appropriate, approximately equal to the preliminary test estimator[1/] and (v) the risk of the positive part and preliminary test estimators are only analyzed for comparable values of the level of the test. In addition, Strawderman and Cohen (1971, pp. 284-285) have shown, following the results of Sacks (1963), that the James-Stein (1961) estimator fails to satisfy the conditions necessary for a generalized Bayes estimator and thus this estimator is inadmissible. For the same reason, the Stein-James (1966) positive part estimator is also inadmissible.

[1/] See Sclove, et al. (1970, p. 9).

In reviewing the literature, it would appear that although many investigators have not understood the properties of the preliminary test estimator or the possible distortion of subsequent inferences from the use of a preliminary test of significance based on the data of the investigation, this estimator is widely used in practice. Given this state of affairs, a study of the properties of the estimator and the characteristics of its risk function under a squared error loss criterion, are of interest and value. Within this context, the purpose of this paper, which is to a large degree expository in nature, is to analyze, for the general linear normal regression model, (i) the properties of the preliminary test estimator implied by a two-stage testing estimation procedure; (ii) the characteristics of the risk function for the preliminary test and restricted estimators; (iii) the conditions under which the risk of the preliminary test estimator is greater than, less than or equal to the conventional and restricted estimators; (iv) the decision problem of choosing an optimal level of the test and (v) the implications of the results for model specification, conditional mean forecasting and aggregation over micro relations or pooling data.

The statistical models, estimators and tests are given in Section 2. The risk function for the preliminary test estimator is derived and compared with other estimators in Sections 3 and 4. The optimal choice of the level of the test, the sampling properties of the sequential estimator and the risk for the conditional mean forecasting case is given in Sections 5, 6 and 7. Some theorems and lemmas necessary for the results given in the test are given in the Appendices.

2. The Statistical Models and Estimators

Assume the linear hypothesis model

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

where $\underset{\sim}{y}$ is a $(T \times 1)$ vector of observations, X is a $(T \times K)$ matrix of non-stochastic variables of rank K, $\underset{\sim}{\beta}$ is a $(K \times 1)$ vector of unknown parameters and $\underset{\sim}{e}$ is a $(T \times 1)$ vector of unobservable normal random variables with

$$(2.2)\qquad E(\underset{\sim}{e}) = \underset{\sim}{0} \quad \text{and} \quad E(\underset{\sim}{e}\underset{\sim}{e}') = \sigma^2 I,$$

where I is an identity matrix of order T.

Using the sample information, specifications (2.1) and (2.2), and defining $S = X'X$, the unrestricted least squares estimator is

$$(2.3)\qquad \underset{\sim}{b} = S^{-1}X'\underset{\sim}{y},$$

where $\underset{\sim}{b}$ is distributed normally with

$$(2.4a)\qquad E(\underset{\sim}{b}) = \underset{\sim}{\beta},$$

$$(2.4b)\qquad E(\underset{\sim}{b}-\underset{\sim}{\beta})(\underset{\sim}{b}-\underset{\sim}{\beta})' = \sigma^2 S^{-1},$$

and an unbiased estimate of σ^2 is given by

$$(2.4c)\qquad \hat{\sigma}^2 = \frac{(\underset{\sim}{y}-X\underset{\sim}{b})'(\underset{\sim}{y}-X\underset{\sim}{b})}{T-K}.$$

As is well known for the model (2.1) and (2.2), $\underset{\sim}{b}$ is the maximum likelihood estimator, and is unbiased.

In addition to the sample information (2.1), suppose additional information which consists of J linear restrictions is perceived as

$$(2.5a)\qquad R\underset{\sim}{\beta} - \underset{\sim}{r} = \underset{\sim}{0},$$

where $\underset{\sim}{r}$ is a $(J \times 1)$ vector of known elements, R is a $(J \times K)$ known matrix

with rank J, and $\underline{0}$ is a $(J \times 1)$ null vector. The true relationship among parameters is assumed to be

$$(2.5b) \qquad R\underline{\beta} - \underline{r} = \underline{\delta},$$

where $\underline{\delta}$ is a $(J \times 1)$ vector representing specification errors in the perceived information, which are zero if that information is correct.

The restricted least squares estimator, which makes use of both the sample and exact prior information or linear hypotheses, (2.1) and (2.5), is

$$(2.6) \qquad \hat{\underline{\beta}} = \underline{b} - S^{-1}R'(RS^{-1}R')^{-1}(R\underline{b}-\underline{r}),$$

where $\hat{\underline{\beta}}$ is normally distributed with mean

$$(2.7a) \qquad E(\hat{\underline{\beta}}) = \underline{\beta} - S^{-1}R'(RS^{-1}R')^{-1}\underline{\delta},$$

variance

$$(2.7b) \qquad E(\hat{\underline{\beta}}-E\hat{\underline{\beta}})(\hat{\underline{\beta}}-E\hat{\underline{\beta}})' = \sigma^2[S^{-1} - S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}]$$

and mean square error

$$(2.7c) \qquad E(\hat{\underline{\beta}}-\underline{\beta})(\hat{\underline{\beta}}-\underline{\beta})' = \sigma^2S^{-1} - \sigma^2S^{-1}R'(RS^{-1}R')^{-1}RS^{-1} + S^{-1}R'(RS^{-1}R')^{-1}\underline{\delta}\,\underline{\delta}'(RS^{-1}R')^{-1}RS^{-1}.$$

If the restriction hypotheses are correct, $\underline{\delta} = \underline{0}$, the restricted least squares estimators are unbiased and have smaller variances (mean square errors) than do the unrestricted least squares estimators. If the prior restrictions are incorrect, $\underline{\delta} \neq \underline{0}$, use of a quadratic loss function involves a trade-off between variance and bias and results in the following risk function for $\hat{\underline{\beta}}$:

$$(2.7d) \qquad E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) = \sigma^2\mathrm{tr}S^{-1} - \sigma^2\mathrm{tr}S^{-1}R'(RS^{-1}R')^{-1}RS^{-1} + \mathrm{tr}S^{-1}R'(RS^{-1}R')^{-1}\underline{\delta}\text{-}\underline{\delta}'(RS^{-1}R')^{-1}RS^{-1}.$$

Using this criterion to appraise performance, the equality restricted estimator is defined to be better than the unrestricted estimator if (2.7d) is

smaller than the trace of (2.4b).

In order to test the compatibility of the sample information (2.1) and the linear hypotheses (2.5a), it is conventional to use the test statistic:

$$(2.8) \qquad u = (R\underline{b}-\underline{r})'(RS^{-1}R')^{-1}(R\underline{b}-\underline{r})/J\hat{\sigma}^2.$$

If the restrictions (2.5a) are correct, u has a central F distribution with J and T-K degrees of freedom and conventional two stage test procedures such as those found in Chipman and Rao (1964) and Rao (1945) may be used. If the linear restrictions are incorrect, u is distributed as a non-central F distribution with J and T-K degrees of freedom and non-centrality parameter

$$(2.9) \qquad \lambda = \frac{\underline{\delta}'(RS^{-1}R')^{-1}\underline{\delta}}{2\sigma^2}.$$

Wallace (1971) suggests that instead of using the traditional test and assuming the linear restrictions are correct, we determine values of λ for which the risk of the restricted estimator (2.7d) is less than that of the unrestricted estimator. Given these critical values of λ, a parameter in the distribution of u, Wallace tests whether or not λ is small enough to insure that the risk for $\hat{\underline{\beta}}$ is as small or smaller than that of $\underline{b}$. Thus, the hypothesis, H_o, that λ is less than or equal to a critical value, is tested against not H_o, by using u and rejecting H_o if $u \geq F^{\alpha}_{(J,T-K,\lambda_o)} = c$. The value of c is determined, for a given level of the test, α, by

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

and λ_o is the value of λ for which the risk of the restricted estimator is less than the unrestricted estimator. By accepting H_o, we take $\hat{\underline{\beta}}$ as our estimate of $\underline{\beta}$, and by rejecting H_o, we use the unrestricted least squares estimate.

In either the conventional or Wallace two stage testing procedures, estimation is dependent on a preliminary test of significance, which implies the use of the preliminary test estimator,

$$(2.10)\qquad \hat{\hat{\beta}} = I_{(0,c)}(u)\hat{\beta} + I_{[c,\infty)}(u)b,$$

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.

It is useful, for the development of the risk function of $\hat{\hat{\beta}}$, to write $\hat{\hat{\beta}}$ as

$$(2.11)\qquad \hat{\hat{\beta}} = b - I_{(0,c)}(u)S^{-1}R'(RS^{-1}R')^{-1}R[b-S^{-1}R'(RS^{-1}R')^{-1}r].$$

If, as is the case in much of applied work, we follow the decision rule suggested by conventional testing procedures or by Wallace, the preliminary test estimator $\hat{\hat{\beta}}$ results, and it becomes important to know the sampling properties of this estimator and its performance relative to the conventional estimator (2.4) and other estimators such as (2.6).

3. The Risk Function of the Preliminary Test Estimator

In deriving the properties of the preliminary test estimator, $\hat{\hat{\beta}}$, use is made of the following quadratic loss function:

$$(3.1)\qquad L(\hat{\hat{\beta}},\beta,) = ||\hat{\hat{\beta}}-\beta||^2 = (\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta),$$

where the estimator $\hat{\hat{\beta}}$ is defined by (2.10), and its risk is

$$(3.2a)\qquad R(\hat{\hat{\beta}},\beta,) = E[L(\hat{\hat{\beta}},\beta,\sigma^2)] = E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta).$$

In order to compare the risk functions of different estimators, by using methods based on the work of Stein (1966) and Sclove et al. (1970), it is convenient to transform the random variables appearing in (3.2a) and in the

argument of the test statistic (2.8). It is to this sequence of transformation that we now turn.

Using $\hat{\hat{\underline{\beta}}}$ derived in (2.11), the risk function of $\hat{\hat{\underline{\beta}}}$ becomes

$$
\begin{aligned}
(3.3)\qquad E(\hat{\hat{\underline{\beta}}}-\underline{\beta})'(\hat{\hat{\underline{\beta}}}-\underline{\beta}) \;=\; & E(\underline{b}-\underline{\beta})'(\underline{b}-\underline{\beta}) - 2E(\underline{b}-\underline{\beta})'[I_{(0,c)}(u)S^{-1}R'(RS^{-1}R')^{-1}R \\
& \cdot[\underline{b}-S^{-1}R'(RS^{-1}R')^{-1}\underline{r}]] \\
& + EI_{(0,c)}(u)[S^{-1}R'(RS^{-1}R')^{-1}R \\
& \cdot[\underline{b}-S^{-1}R'(RS^{-1}R')^{-1}\underline{r}]]'[S^{-1}R'(RS^{-1}R')^{-1} \\
& R[\underline{b}-S^{-1}R'(RS^{-1}R')^{-1}\underline{r}]].
\end{aligned}
$$

The first term on the right side of the equality is $\sigma^2 trS^{-1}$ and S^{-1} may be written using $P^{-1}(P^{-1})' = S^{-1}$. In addition, an orthogonal transformation Q is chosen to diagonalize the idempotent matrix $(P^{-1})'R'(RS^{-1}R')^{-1}RP^{-1}$, which is of rank J, giving J characteristic roots of one and K-J zero roots. Consequently, (3.3) may be written as

$$
\begin{aligned}
(3.4)\qquad E(\hat{\hat{\underline{\beta}}}-\underline{\beta})'(\hat{\hat{\underline{\beta}}}-\underline{\beta}) \;=\; & \sigma^2 trS^{-1} - 2EI_{(0,c)}(u)(QP\underline{b}-QP\underline{\beta})'Q(P^{-1})'P^{-1}Q'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} \\
& \cdot[QP\underline{b}-Q(P^{-1})'R'(RS^{-1}R')^{-1}\underline{r}] \\
& + EI_{(0,c)}(u)[QP\underline{b}-Q(P^{-1})'R'(RS^{-1}R')^{-1}\underline{r}]'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} \\
& \cdot Q(P^{-1})'P^{-1}Q'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}[QP\underline{b}-Q(P^{-1})'R'(RS^{-1}R')^{-1}\underline{r}].
\end{aligned}
$$

where I_J is an identity matrix of order J and $\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}$ is of order K. Defining $\underline{w} = [QP\underline{b}-Q(P^{-1})'R'(RS^{-1}R')^{-1}\underline{r}]$, where $\underline{w}$ is a normally distributed random vector with mean $\underline{\eta} = QP\underline{\beta}-Q(P^{-1})'R'(RS^{-1}R')^{-1}\underline{r}$ and covariance matrix $\sigma^2 I$, the risk function of $\hat{\hat{\underline{\beta}}}$ is

$$(3.5)\qquad E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta) = \sigma^2 \mathrm{tr} S^{-1} - 2EI_{(0,c)}(u)(QP\underline{b}-QP\underline{\beta})'\begin{pmatrix} A_1 & 0 \\ A_3' & 0 \end{pmatrix}\underline{w} + EI_{(0,c)}(u)\underline{w}'\begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}\underline{w},$$

where

$$(3.6)\qquad Q(P^{-1})'P^{-1}Q' = A = \begin{pmatrix} A_1 & A_3 \\ A_3' & A_2 \end{pmatrix},$$

and A_1 and A_2 are of order J and K-J, respectively.

The second term of equation (3.5) may be written as

$$(3.7)\qquad -2E[I_{(0,c)}(u)(\underline{w}'-\underline{\eta}')\begin{pmatrix} A_1 & 0 \\ A_3' & 0 \end{pmatrix}\underline{w}.$$

The elements of $\underline{w}$ are independent. Partitioning the $(K \times 1)$ vectors $\underline{w}'$ and $\underline{\eta}'$ as vectors $(\underline{w}_1' \ \underline{w}_2')$ and $(\underline{\eta}_1' \ \underline{\eta}_2')$, each with J and K-J, respectively, the risk function of $\hat{\hat{\underline{\beta}}}$ becomes

$$(3.8)\qquad E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta) = \sigma^2 \mathrm{tr} S^{-1} - EI_{(0,c)}(u)\underline{w}_1'A_1\underline{w}_1 - 2EI_{(0,c)}(u)\underline{w}_2'A_3'\underline{w}_1 + 2\underline{\eta}_1'A_1 EI_{(0,c)}(u)\underline{w}_1 + 2\underline{\eta}_2'A_3'EI_{(0,c)}(u)\underline{w}_1.$$

The evaluation of the risk function of $\hat{\hat{\underline{\beta}}}$ now requires transforming the test statistic u to a function of $\underline{w}_1'\underline{w}_1$.

a. <u>A Reformulation of the Test Statistic, u</u>

Using the operations and notation defined above, the test statistic u may be written as

(3.9)

$$u = \frac{[QP\underline{b}-Q(P^{-1})'R'(RS^{-1}R')^{-1}\underline{r}]'Q(P^{-1})'R'(RS^{-1}R')^{-1}RP^{-1}Q'[QP\underline{b}-Q(P^{-1})'R'(RS^{-1}R')^{-1}\underline{r}]}{J\hat{\sigma}^2}$$

or

$$(3.10) \qquad u = \frac{\underset{\sim}{w}_1' \underset{\sim}{w}_1}{J\hat{\sigma}^2} .$$

If a (J × J) orthogonal matrix C_1 is chosen so that $C_1' A_1 C_1$ becomes a diagonal matrix D_1, the test statistic u may be expressed as

$$(3.11) \qquad u = \frac{\underset{\sim}{u}^{*\prime} \underset{\sim}{u}^*}{J\hat{\sigma}^2} ,$$

where $\underset{\sim}{u}^* = C_1 \underset{\sim}{w}_1$ has mean $\underset{\sim}{\xi}^* = C_1 \underset{\sim}{\eta}_1$ and variance $\sigma^2 I_J$.

b. A Reformulation of the Risk Function for $\hat{\hat{\underset{\sim}{\beta}}}$

The test statistic presented in (3.10) may now be used as the argument for the indicator function in (3.8) giving

$$(3.12) \qquad E(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta})'(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta}) = \sigma^2 trS^{-1} - \sigma^2 EI_{(0,c^*)}\left(\frac{\underset{\sim}{w}_1'\underset{\sim}{w}_1}{\sigma^2}\right)\frac{\underset{\sim}{w}_1'}{\sigma} A_1 \frac{\underset{\sim}{w}_1}{\sigma} + 2\sigma\underset{\sim}{\eta}_1' A_1 E \cdot\left[I_{(0,c^*)}\left(\frac{\underset{\sim}{w}_1'\underset{\sim}{w}_1}{\sigma^2}\right)\frac{\underset{\sim}{w}_1}{\sigma}\right]$$

where $\frac{J\hat{\sigma}^2 c}{\sigma^2} = c^*$. In order to continue the evaluation of (3.12), the two following theorems, proven in Appendix A, are required:

Theorem 1: If the (J × 1) vector $\frac{\underset{\sim}{w}_1}{\sigma}$ is distributed $N(\frac{\underset{\sim}{\eta}_1}{\sigma}, I)$, then

$$E\left[I_{(0,c^*)}\left(\frac{\underset{\sim}{w}_1'\underset{\sim}{w}_1}{\sigma^2}\right)\frac{\underset{\sim}{w}_1}{\sigma}\Big|\hat{\sigma}^2\right] = \frac{\underset{\sim}{\eta}_1}{\sigma} E\left[I_{(0,c^*)}(\chi^2_{(\lambda,J+2)})\,|\hat{\sigma}^2\right].$$

Theorem 2: If the (J × 1) vector $\frac{\underset{\sim}{w}_1}{\sigma}$ is distributed $N(\frac{\underset{\sim}{\eta}_1}{\sigma}, I)$ and A_1 is a positive definite symmetric matrix, then

$$E\left[I_{(0,c^*)}\left(\frac{\underset{\sim}{w}_1'\underset{\sim}{w}_1}{\sigma^2}\right)\frac{\underset{\sim}{w}_1'}{\sigma} A_1 \frac{\underset{\sim}{w}_1}{\sigma}\Big|\hat{\sigma}^2\right] = E\left[I_{(0,c^*)}(\chi^2_{(\lambda,J+2)})\,|\hat{\sigma}^2\right] trA_1 + E\left[I_{(0,c^*)}(\chi^2_{(\lambda,J+4)})\,|\hat{\sigma}^2\right]\frac{\underset{\sim}{\eta}_1'}{\sigma} A_1 \frac{\underset{\sim}{\eta}_1}{\sigma} .$$

Utilizing Theorems 1 and 2, (3.12) can be written as

$$(3.13)\qquad E(\hat{\hat{\underline{\beta}}}-\underline{\beta})'(\hat{\hat{\underline{\beta}}}-\underline{\beta}) = \sigma^2 \mathrm{tr} S^{-1} - \sigma^2 E[E[I_{(0,c^*)}(\chi^2_{(\lambda,J+2)})|\hat{\sigma}^2]]\mathrm{tr}A_1 - \underline{\eta}_1' A_1 \underline{\eta}_1 E[E[I_{(0,c^*)}(\chi^2_{(\lambda,J+4)})|\hat{\sigma}^2]] + 2\underline{\eta}_1' A_1 \underline{\eta}_1 E[E[I_{(0,c^*)}(\chi^2_{(\lambda,J+2)})|\hat{\sigma}^2]].$$

Recognizing that

$$E[E[I_{(0,c^*)}(\chi^2_{(\lambda,J+\ell)})|\hat{\sigma}^2]] = EI_{(0,\frac{cJ}{T-K})}\left(\frac{\chi^2_{(\lambda,J+\ell)}}{\chi^2_{(T-K)}}\right) = \Pr\left(\frac{\chi^2_{(\lambda,J+\ell)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right),$$

and using the orthogonal transformation C_1 again,

$$(3.14)\qquad E(\hat{\hat{\underline{\beta}}}-\underline{\beta})'(\hat{\hat{\underline{\beta}}}-\underline{\beta}) = \sigma^2 \mathrm{tr} S^{-1} - \sigma^2 \Pr\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right)\sum_{i=1}^{J} d_i^{(1)} - \sum_{i=1}^{J} d_i^{(1)}\xi_i^{*2}\Pr\left(\frac{\chi^2_{(\lambda,J+4)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right) + 2\sum_{i=1}^{J} d_i^{(1)}\xi_i^{*2}\Pr\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right),$$

where $d_i^{(1)}$ and ξ_i^* are the characteristic roots of A_1 and the elements of $\underline{\xi}^*$, respectively.

Furthermore, the non-centrality parameter for the distribution of u, given in (2.9), can be written as:

$$(2.9a)\qquad \lambda = \frac{\sum_{i=1}^{J}\xi_i^{*2}}{2\sigma^2}.$$

c. <u>Characteristics of the Risk Function</u>

If in line with the specifications in the previous section, and by defining $h_\lambda(\ell) = \Pr(\chi^2_{(\lambda,J+\ell)}/\chi^2_{(T-K)} < cJ/(T-K))$ and $t_i = d_i^{(1)}/\sum_{j=1}^{J} d_j^{(1)}$ which implies

$t_i > 0$ and $\sum_{i=1}^{J} t_i = 1$, (3.14) may be expressed as

$$(3.15) \qquad E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta) = \sigma^2 \mathrm{tr} S^{-1} - \sigma^2 (\sum_{i=1}^{J} d_i^{(1)})[h_\lambda(2)+2(h_\lambda(4)-2h_\lambda(2)) \sum_{i=1}^{J} t_i \left(\frac{\xi_i^{*2}}{2\sigma^2}\right)].$$

Written in this form, the risk of the preliminary test estimator (3.16) is seen to be a function of both λ and $\sum_{i=1}^{J} t_i \xi_i^{*2}/2\sigma^2$, where λ appears through the functions $h_\lambda(\ell)$. Thus, a given value of λ does not completely determine the risk functions of $\hat{\hat{\beta}}$, and one must know in addition the values of the ξ_i^{*2} which appear in $\sum_{i=1}^{J} t_i \xi_i^{*2}/2\sigma^2$.

In order to determine the largest and smallest risk values that $\hat{\hat{\beta}}$ may take for a given $\lambda = \sum_{i=1}^{J} \xi_i^{*2}/2\sigma^2$, we may choose t_L and t_S as the t_i with the largest and smallest values, respectively, and vary the ξ_i^{*2}'s. The value of the risk function (3.15) is largest, for a given λ, when only the ξ_L^{*2} associated with t_L, is non-zero which means that $\xi_L^{*2} = 2\sigma^2\lambda$. Alternatively, the value of the risk function (3.15) is smallest when the ξ_i^{*2} are varied so that only the ξ_S^{*2} associated with t_S is non-zero, and hence $\xi_S^{*2} = 2\sigma^2\lambda$.[2/] Thus, given λ,

$$(3.16a) \qquad E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta) \leq \sigma^2 \mathrm{tr} S^{-1} - \sum_{i=1}^{J} d_i^{(1)} \sigma^2 [h_\lambda(2)+2(h_\lambda(4)-2h_\lambda(2))\lambda t_L],$$

and

$$(3.16b) \qquad E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta) \geq \sigma^2 \mathrm{tr} S^{-1} - \sum_{i=1}^{J} d_i^{(1)} \sigma^2 [h_\lambda(2)+2(h_\lambda(4)-2h_\lambda(2))\lambda t_S].$$

[2/] Note that $1 > h_\lambda(2) = \Pr\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right) > \Pr\left(\frac{\chi^2_{(\lambda,J+4)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right) = h_\lambda(4) > 0$, since $\chi^2_{(\lambda,J+2)}$ is stochastically larger than $\chi^2_{(\lambda,J+4)}$.

Furthermore, by varying the ξ_i^{*2} under the restriction $\lambda = \sum_{i=1}^{J} \xi_i^{*2}/2\sigma^2$, the rest of the preliminary test estimator, $E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta)$, can assume any value from (3.16b) to (3.16a). There is only one point for each equation for which the value of the right side is $\sigma^2 trS^{-1}$, the risk of the conventional estimator. (3.16a) and (3.16b) are equal when $\lambda = 0$.

The characteristics of the risk functions (3.16a) and (3.16b) are reflected in Figure 1, for the situation where $\alpha = .05$, $\sigma^2 = 1$, $J = 2$, $T-K = 10$, $\sum_{i=1}^{J} d_i^{(1)} = 1$, $t_L = .9$, $t_S = .1$ and $E(b-\beta)'(b-\beta) = 2$.

4. Comparison of the Risk Functions

a. Conventional and Preliminary Test Estimators

We now wish to determine conditions under which the risk function of the preliminary test estimator, $\hat{\hat{\beta}}$, is less than or greater than that of the unrestricted least squares estimator, b, in terms of λ. Subtracting (3.15) from the risk function for b, (2.4), we have

$$(4.1) \qquad E(b-\beta)'(b-\beta)-E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta) = (\sum_{i=1}^{J} d_i^{(1)})\sigma^2[h_\lambda(2)+2(h_\lambda(4)-2h_\lambda(2))\sum_{i=1}^{J} t_i\left(\frac{\xi_i^{*2}}{2\sigma^2}\right)].$$

For a given value of λ, the risk function least favorable to the preliminary test estimator is equation (3.16a), where $\xi_L^{*2} = 2\sigma^2\lambda$. Therefore, for a fixed λ, the smallest possible value of (4.1) is

$$(4.2a) \qquad (\sum_{i=1}^{J} d_i^{(1)})\sigma^2[h_\lambda(2)+2(h_\lambda(4)-2h_\lambda(2))t_L\lambda],$$

and the risk of the conventional estimator b is at least as large as that for the preliminary test estimator $\hat{\hat{\beta}}$ if

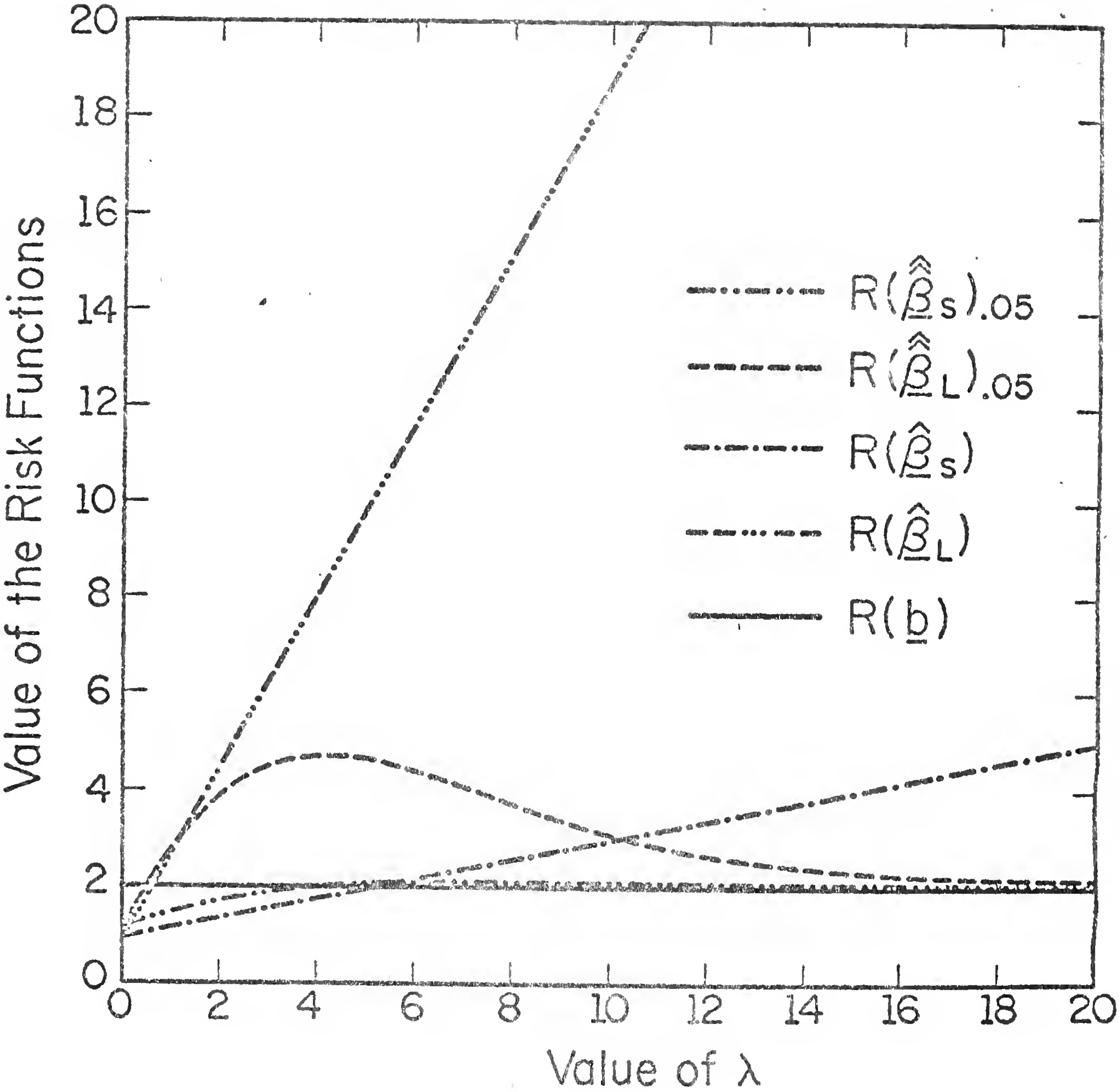

Figure 1.

Risk Functions for the Conventional Restricted and Preliminary Test Estimators.

(4.3a) $$\lambda \leq \frac{1}{2t_L[2 - \frac{h_\lambda(4)}{h_\lambda(2)}]} .$$

Alternatively, the risk function <u>most favorable</u> to the preliminary test estimator is equation (3.16b), where $\xi_S^{*2} = 2\sigma^2\lambda$.

For a given value of λ, the largest possible value of (4.1) is

(4.2b) $$(\sum_{i=1}^{J} d_i^{(1)})\sigma^2[h_\lambda(2)+2(h_\lambda(4)-2h_\lambda(2))t_S\lambda],$$

which lies above (4.2a) for every λ. Making use of (4.2b), the condition for the risk of $\underline{b}$ to be less than or equal to that of $\hat{\hat{\underline{\beta}}}$ is

(4.3b) $$\lambda \geq \frac{1}{2t_S[2 - \frac{h_\lambda(4)}{h_\lambda(2)}]} .$$

Since $h_\lambda(4)$ and $h_\lambda(2)$ are complicated functions of λ, it is difficult, in general, to solve for the equality of the risk functions involving t_L, i.e., for λ_o such that $f(\lambda_o) = 0$, where $f(\lambda) = \lambda - \frac{1}{2t_L[2 - \frac{h_\lambda(4)}{h_\lambda(2)}]}$. We do, however, know that $\lambda_o \geq 1/4t_L$, since $\frac{h_\lambda(4)}{h_\lambda(2)} \geq 0$, and $\lambda_o \geq \frac{1}{2t_L(2-\omega_o)}$ if $T-K \geq 2$, since $\frac{h_\lambda(4)}{h_\lambda(2)} \geq \omega_o$, where $\omega_o = \left(\frac{c}{\frac{T-K}{J} + c}\right)$. [3/]

Correspondingly, the same type of reasoning applies to finding the equality of the risk functions involving t_S, i.e., for λ_1 such that $g(\lambda_1) = 0$, where $g(\lambda) = \lambda - \frac{1}{2t_S[2 - \frac{h_\lambda(4)}{h_\lambda(2)}]}$, with $g(\lambda) < 0$ if $\lambda < \lambda_1$. Using t_S,

---

[3/] See Theorem 1 in Appendix B.

since $\frac{h_\lambda(4)}{h_\lambda(2)} \leq 1$, the risk for the preliminary test estimator is less than that of the conventional estimator if $\lambda_1 \leq \frac{1}{2t_S}$.

Therefore, the equality of the risk functions of the preliminary test and unrestricted least squares estimators have the following bounds:

(4.3c) $$\frac{1}{4t_L} \leq \lambda_o \leq \lambda_1 \leq \frac{1}{2t_S},$$

with the lower bound replaced by $\frac{1}{2t_L(2-\omega_o)}$ if $T-K \geq 2$.

In order to depict this situation graphically, the case that formed the basis for Figure 1 is used and the neighborhood of the origin is enlarged in Figure 2 in which the actual values of λ_o and λ_1 are identified.

As a special case if the characteristic roots, $d_i^{(1)}$, are the same (for example, X'X is a scalar matrix) and thus, $t_1 = t_2 = \ldots = t_J$, then t_L and t_S equal 1/J, and (3.16a) is equal to (3.16b). Under this situation, the conditions for the preliminary test estimator to be less than or equal to the conventional estimator are

(4.3d) $$\frac{J}{4} \leq \lambda_o = \lambda_1 \leq \frac{J}{2},$$

where the lower bound is replaced by $\frac{J}{2(2-\omega_o)}$ if $T-K \geq 2$. This result is consistent with that derived by Sclove, et al. (1970) for the orthonormal regressor case.

We may summarize the conclusions to this point as follows:

If $\lambda \notin (\lambda_o,\lambda_1)$ or $\lambda \notin (\frac{1}{4t_L}, \frac{1}{2t_S})$ and λ is known, we can decide if (4.1) is positive or negative. Furthermore, even if λ is known and $\lambda \in (\lambda_o,\lambda_1)$ with $\lambda_o \neq \lambda_1$, one cannot determine whether or not the risk function of $\underset{\sim}{b}$ exceeds that of $\hat{\hat{\underset{\sim}{\beta}}}$ without knowing the value of $\sum_{i=1}^{J} \frac{t_i \xi_i^{*2}}{2\sigma^2}$.

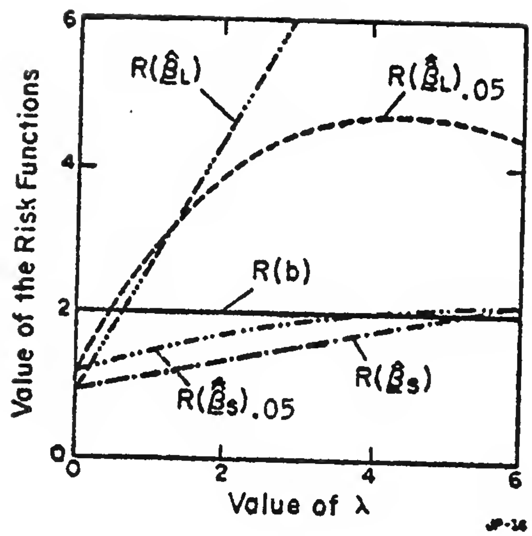

Figure 2.

Risk Functions for the Conventional Restricted and Preliminary Test Estimators in the Neighborhood of $\lambda = 0$.

The risk of $\hat{\hat{\beta}}$ at the origin where $\lambda = 0$, a consequence of $\delta = 0$, is

$$\sigma^2 \mathrm{tr}S^{-1} - \Pr\left(\frac{\chi^2_{(J+2)}}{\chi^2_{(T-K)}} \le \frac{cJ}{T-K}\right)\sigma^2 \sum_{i=1}^{J} d_i^{(1)},$$

which is smaller than $\sigma^2 \mathrm{tr}S^{-1}$, the risk of b. This can be seen from (3.15) since $\lambda = 0$ implies $\frac{\xi_i^{*2}}{2\sigma^2} = 0$, $i = 1,\ldots,J$. At $\lambda = 0$, the risk function (3.16a) is equal to (3.16b).

Alternatively, since $\lambda h_\lambda(\ell)$ and $h_\lambda(\ell)$ approach 0 as λ approaches ∞, (3.15), (3.16a) and (3.16b) approach $\sigma^2 \mathrm{tr}S^{-1}$ as $\lambda \to \infty$. This implies that the risk of $\hat{\hat{\beta}}$ approaches that of b from above as $\lambda \to \infty$ (see Figure 1).

Finally, it should be noted that the terms $h_\lambda(2)$ and $h_\lambda(4)$ are the probabilities of ratios of random variables, being less than a constant and they depend on the critical value c or the level of the test α. Therefore, as $\alpha \to 0$, $h_\lambda(\ell) \to 1$ and the risk of the preliminary test estimator, $\hat{\hat{\beta}}$, approaches that of the restricted least squares estimator, $\hat{\beta}$, since in a repeated sampling context, $\hat{\beta}$ is used more frequently as an estimator of β for all λ. Alternatively, as $\alpha \to 1$ and $h_\lambda(\ell) \to 0$, the risk function for $\hat{\hat{\beta}}$ approaches that for the conventional estimator, b.

b. <u>Conventional and Restricted Estimators</u>

To facilitate a comparison of the risk of the equality restricted least squares estimator, $\hat{\beta}$, with the conventional estimator, b, we note from the derivation given in Appendix C that

(4.4) $$E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) = \sigma^2 \mathrm{tr}S^{-1} - \sigma^2 \sum_{i=1}^{J} d_i^{(1)} + \sum_{i=1}^{J} d_i^{(1)} \xi_i^{*2}.$$

$$= \sigma^2 \mathrm{tr}S^{-1} - \sigma^2 \sum_{i=1}^{J} d_i^{(1)} + \sigma^2\left(\sum_{j=1}^{J} d_j^{(1)}\right)[2 \sum_{i=1}^{J} t_i \left(\frac{\xi_i^{*2}}{2\sigma^2}\right)].$$

Written in this form, it is clear that, in general, (4.4) is not just a function of λ, but one must know the ξ_i^{*2} which appear in $\sum_{i=1}^{J} t_i\left(\frac{\xi_i^{*2}}{2\sigma^2}\right)$. As a counterpart for (3.17a) and (3.17b), for a fixed value of λ,

$$\text{(4.5a)} \qquad E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) \leq \sigma^2 \mathrm{tr}S^{-1} - \sigma^2 \sum_{i=1}^{J} d_i^{(1)} + \sigma^2\left(\sum_{j=1}^{J} d_j^{(1)}\right)2t_S\lambda,$$

and

$$\text{(4.5b)} \qquad E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) \geq \sigma^2 \mathrm{tr}S^{-1} - \sigma^2 \sum_{i=1}^{J} d_i^{(1)} + \sigma^2\left(\sum_{j=1}^{J} d_j^{(1)}\right)2t_L\lambda.$$

Thus, the risk function (4.4) can assume any value from (4.5b) to (4.5a), by varying the ξ_i^{*2}'s, under the restriction that $\sum_{i=1}^{J} \frac{\xi_i^{*2}}{2\sigma^2} = \lambda$. Making use of (4.4) or (4.5a) and (4.5b), when $\lambda = 0$, the risk of $\hat{\underline{\beta}} = \sigma^2 \mathrm{tr}S^{-1} - \sigma^2 \sum_{i=1}^{J} d_i^{(1)} > 0$. Also, as λ goes to infinity, the expressions on the right side of (4.5a) and (4.5b), and thus (4.4), go to infinity. The characteristics of the risk function for t_S and t_L in (4.5a) and (4.5b), respectively, for the example given in Section 3c, are given in Figures 1 and 2.

Making use of (3.2b), and comparing the risk of $\hat{\underline{\beta}}$ and $\underline{b}$, we have

$$\text{(4.6)} \qquad E(\underline{b}-\underline{\beta})'(\underline{b}-\underline{\beta})-E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) = \sigma^2 \sum_{i=1}^{J} d_i^{(1)}\left[1-2\sum_{i=1}^{J} t_i\left(\frac{\xi_i^{*2}}{2\sigma^2}\right)\right].$$

By varying ξ_i^{*2} where $\lambda = \sum_{i=1}^{J} \frac{\xi_i^{*2}}{2\sigma^2}$, (4.6) may assume any value from

$$\text{(4.7a)} \qquad \sigma^2 \sum_{i=1}^{J} d_i^{(1)}[1 - 2t_L\lambda]$$

to

$$\text{(4.7b)} \qquad \sigma^2 \sum_{i=1}^{J} d_i^{(1)}[1 - 2t_S\lambda], \quad \text{for fixed } \lambda.$$

Therefore, (4.6) will be non-negative if

(4.8a) $$\lambda \leq \frac{1}{2t_L}$$

and (4.6) will be non-positive if

(4.8b) $$\lambda \geq \frac{1}{2t_S}.$$

If $t_S \neq t_L$ and $\lambda \in (\frac{1}{2t_L}, \frac{1}{2t_S})$, one cannot determine the sign of (4.6) even if λ is known precisely. Thus, as in the case of the preliminary test estimator, it is necessary to know the value of $\sum_{i=1}^{J} t_i \left(\frac{\xi_i^{*2}}{2\sigma^2}\right)$.

c. <u>Preliminary Test and Restricted Estimators</u>

By making use of (3.16) and (4.4), the difference between the risk functions for $\hat{\hat{\beta}}$ and $\hat{\beta}$ may be expressed as

(4.9) $$E(\hat{\hat{\beta}}-\beta)'(\hat{\hat{\beta}}-\beta)-E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) =$$

$$= \sigma^2(\sum_{i=1}^{J} d_i^{(1)})[1-h_\lambda(2)-2(1+h_\lambda(4)-2h_\lambda(2))(\sum_{i=1}^{J} t_i \left(\frac{\xi_i^{*2}}{2\sigma^2}\right))].$$

Proceeding as before, for a given value of λ, (4.9) may assume any value between

(4.10a) $$\sigma^2(\sum_{i=1}^{J} d_i^{(1)})[1-h_\lambda(2)-2(1+h_\lambda(4)-2h_\lambda(2))(\lambda t_L)]$$

and

(4.10b) $$\sigma^2(\sum_{i=1}^{J} d_i^{(1)})[1-h_\lambda(2)-2(1+h_\lambda(4)-2h_\lambda(2))(\lambda t_S)],$$

if we let the values of the ξ_i^{*2} vary under the restriction that $\sum_{i=1}^{J} \frac{\xi_i^{*2}}{2\sigma^2} = \lambda$. Now $1-h_\lambda(4) \geq 1-h_\lambda(2)$ implies $1+h_\lambda(4)-2h_\lambda(2) \geq 0$ and the difference in the risk functions given by (4.10a) (and thus (4.9)) will be non-negative if

(4.11) $$\lambda \leq \frac{1}{2t_L[2 - \frac{(1-h_\lambda(4))}{(1-h_\lambda(2))}]}.$$

Since $\frac{1-h_\lambda(4)}{1-h_\lambda(2)} \geq 1$, this means that (4.11) is satisfied if

(4.12) $$\lambda_2 \leq \frac{1}{2t_L} .$$

Also, (4.10b) (and thus (4.9)) will be non-positive if

(4.13) $$\lambda \geq \frac{1}{2t_S[2 - \frac{(1-h_\lambda(4))}{(1-h_\lambda(2))}]} .$$

If $w_o \leq \frac{J+2}{J+T-K}$, then $[2 - \frac{(1-h_\lambda(4))}{(1-h_\lambda(2))}] \geq Pr[\frac{\chi^2_{(J)}}{\chi^2_{(T-K)}} \geq \frac{cJ}{T-K}]$, and thus (4.13) is satisfied if

(4.14) $$\lambda_3 \geq \frac{1}{2t_S Pr[\frac{\chi^2_{(J)}}{\chi^2_{(T-K)}} \geq \frac{cJ}{T-K}]} .$$

The inequalities (4.12) and (4.14) are helpful because it is difficult to solve for λ_2 such that (4.11) holds for $\lambda \leq \lambda_2$ and λ_3 such that (4.13) holds for $\lambda \geq \lambda_3$. Therefore, the equality of the risk functions of the preliminary test and restricted least squares estimators have the following bounds:

(4.15) $$\frac{1}{2t_L} < \lambda_2 \leq \lambda_3 < \frac{1}{2t_S Pr[\frac{\chi^2_{(J)}}{\chi^2_{(T-K)}} \geq \frac{cJ}{T-K}]} .$$

As before, if $t_1 = t_2 = \ldots = t_J$, then (4.9) = (4.10a) = (4.10b) and $\lambda_2 = \lambda_3$. This implies $t_i = \frac{1}{J}$, for $i = 1,\ldots,J$, and thus

(4.16) $$\frac{J}{2} < \lambda_2 = \lambda_3 < \frac{J}{2Pr[\frac{\chi^2_{(J)}}{\chi^2_{(T-K)}} \geq \frac{cJ}{T-K}]} .$$

If $\lambda \notin (\lambda_2, \lambda_3)$ or $\lambda \notin \left(\frac{1}{2t_L}, \frac{1}{2t_S \Pr\left[\frac{\chi^2_{(J)}}{\chi^2_{(T-K)}} \geq \frac{cJ}{T-K}\right]}\right)$ and λ is known, we can decide if (4.9) is positive or negative. However, even if λ is known and $\lambda \in (\lambda_2, \lambda_3)$ with $\lambda_2 \neq \lambda_3$ (which occurs if $t_S \neq t_L$), one cannot determine whether or not the risk function of $\hat{\underline{\beta}}$ exceeds that of $\hat{\hat{\underline{\beta}}}$ without knowing the value of

$$\sum_{i=1}^{J} \frac{t_i \xi_i^{*2}}{2\sigma^2} .$$

For $\lambda = 0$, the difference in the risk functions of $\hat{\hat{\underline{\beta}}}$ and $\hat{\underline{\beta}}$ (4.9) is $\sigma^2 (\sum_{i=1}^{J} d_i^{(1)}) \Pr\left[\frac{\chi^2_{(J+2)}}{\chi^2_{(T-K)}} \geq \frac{cJ}{T-K}\right] > 0$. Furthermore, as λ goes to infinity, the right side of (4.9) goes to minus infinity because $\lambda h_\lambda(i)$ and $h_\lambda(i)$ go to zero so (4.10a) and (4.10b) go to minus infinity (see Figure 1).

5. Optimal Choice of α

For λ such that $0 \leq \lambda \leq 1/4t_L$, the risk for the preliminary test estimator is smaller than the conventional estimator regardless of the choice of the level of statistical significance, α, or the critical value, c. However, the choice of α or c does affect the magnitude of the difference between the risk functions that result for each λ.

When α approaches zero, the critical value c approaches infinity, and the risk function for the preliminary test estimator approaches that of the restricted least squares estimator $\hat{\underline{\beta}}$. Alternatively, when α approaches one the risk function for the preliminary test estimator approaches that of the conventional estimator, and the difference in the risk functions tend to zero. A graph of the risk functions for two levels of α is given in Figure 3.

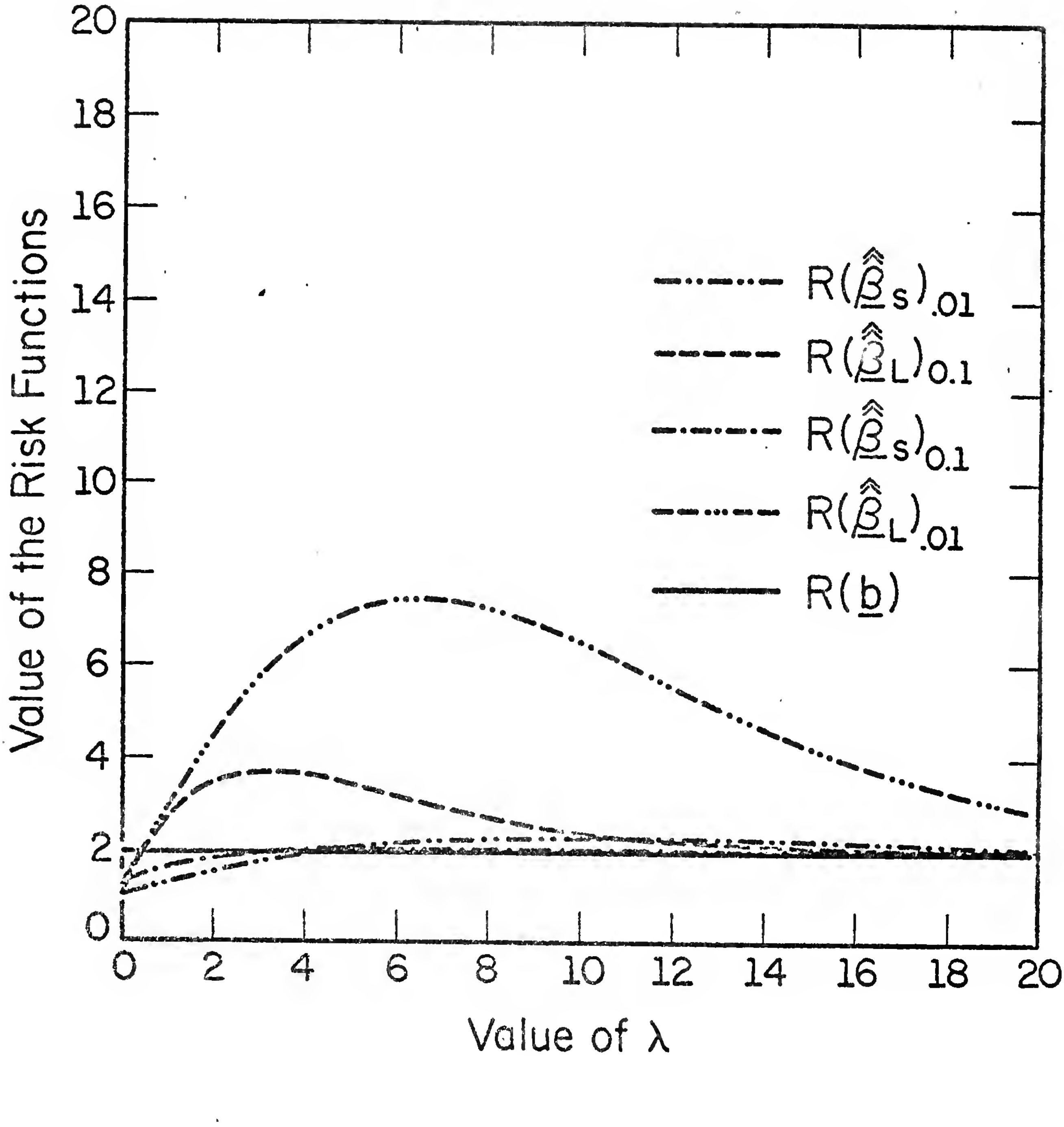

Figure 3.

Risk Functions for the Preliminary Test Estimator When α = .01 and .1.

Since we have expressed the conditions for the risk of the conventional estimator to exceed that of the preliminary test estimator in terms of the non-centrality parameter, λ, of the non-central F distribution, we could follow Toro-Vizcarrondo and Wallace (1968) or Wallace (1970) and propose a test, for example, for the <u>orthonormal</u> <u>regressor</u> <u>case</u> for the hypothesis $\lambda \leq \frac{J}{4}$ against the alternative $\lambda \geq \frac{J}{4}$. For this test, the investigator computes the value for the test statistic and rejects the hypothesis if this value exceeds the critical value based on an α and the non-central F with J and T-K degrees of freedom and a non-centrality parameter of $\frac{J}{4}$. However, it does not matter whether one uses the test statistic with the central F ($F_{T-K,J,\lambda=0}$), the one originally proposed by Toro-Vizcarrondo and Wallace ($F_{T-K,J,\lambda=\frac{J}{2}}$), or the one suggested above, since the critical points for these tests can always be matched up by varying α for one test versus another.

Since in reality λ is unknown and the gain or loss for the preliminary test estimator varies with the choice of c, we are faced with a decision problem, where the optimum critical value, c, or the level of statistical significance, α, depends on the optimality criterion used. If we are interested in a choice of c or α which would minimize the maximum risk, a minimax solution is $c = 0$. Given this trivial result, one alternative is to use the minimax regret in the class of preliminary test estimators. Some results have been obtained in this area for the risk function criterion, by Sawa and Hiromatsu (1970), for the special case when $J = 1$.

Alternatively, if one could specify a tractable prior probability density function for λ, say, for example, a uniform or a chi-square distribution, then a Bayesian extension of these results is possible. Relative to (4.1) for a uniform density for λ on $[0,\infty)$, the integral exists and is finite. When $\alpha = 1$, i.e., $\omega_0 = 0$, the integral is 0. Of course, when $\alpha = 0$, i.e., $\omega_0 = 1$, then the integral is infinite because estimator $\hat{\underline{\beta}}$ is always chosen.

6. The Bias and Covariance Matrix of $\hat{\hat{\beta}}$

It is well known that unrestricted estimator b is unbiased with a covariance matrix $\sigma^2 S^{-1}$. The restricted estimator $\hat{\beta}$ has mean

$$(6.1) \qquad E\hat{\beta} = \beta - S^{-1}R'(RS^{-1}R')^{-1}\delta,$$

and if $\delta \neq 0$, $\hat{\beta}$ is biased with bias $S^{-1}R'(RS^{-1}R')\delta$. Furthermore, $\hat{\beta}$ has a covariance matrix,

$$(6.2) \qquad E(\hat{\beta}-E\hat{\beta})(\hat{\beta}-E\hat{\beta})' = \sigma^2[S^{-1}-S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}].$$

The mean of the preliminary test estimator,

$$(6.3) \qquad E\hat{\hat{\beta}} = EI_{(0,c)}(u)\hat{\beta} + EI_{[c,\infty)}(u)b,$$

is evaluated in Appendix D and gives, from (D.4),

$$(6.4) \qquad E\hat{\hat{\beta}} = \beta + h_\lambda(2)S^{-1}R'(RS^{-1}R')^{-1}\delta.$$

If $\delta = 0$, the preliminary test estimator is unbiased. Otherwise, its bias depends on (i) the probability of a random variable with a non-central F distribution being smaller than a constant determined by the level of the test and the number of restrictions, J, as well as the incorrectness of the restriction through the non-centrality parameter, λ, (ii) the incorrectness of the prior information through δ, and (iii) the matrix $S^{-1}R'(RS^{-1}R')^{-1}$. Thus, the bias is always as small as that of $\hat{\beta}$ given in (6.1).

The covariance matrix of $\hat{\hat{\beta}}$ is derived in Appendix C and is given in (D.12) as

$$(6.5) \qquad E(\hat{\hat{\beta}}-E\hat{\hat{\beta}})(\hat{\hat{\beta}}-E\hat{\hat{\beta}})' = \sigma^2 S^{-1} - P^{-1}Q'\begin{pmatrix} h_\lambda(2)\sigma^2 I_J + [2h_\lambda(2)-h_\lambda(4)+h_\lambda^2(2)]\eta_1\eta_1' & 0 \\ 0' & 0_{K-J} \end{pmatrix} Q(P^{-1})',$$

where P, Q, $h_\lambda(2)$, $h_\lambda(4)$ and η_1 were defined in Section 3.

By using

$$\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} = QPS^{-1}R'(RS^{-1}R')^{-1}RS^{-1}P'Q'$$

and

$$\underline{\eta} = QP(\underline{\beta}-S^{-1}R'(RS^{-1}R')^{-1}\underline{r}),$$

so

$$P^{-1}Q'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})' = S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$$

and

$$\begin{pmatrix} \underline{\eta}_1\underline{\eta}_1' & 0' \\ 0 & 0 \end{pmatrix} = QPS^{-1}R'(RS^{-1}R')^{-1}RS^{-1}P'Q'QP(\underline{\beta}-S^{-1}R'(RS^{-1}R')^{-1}\underline{r}) \cdot(\underline{r}'(RS^{-1}R')^{-1}RS^{-1}-\underline{\beta})P'Q'QPS^{-1}R'(RS^{-1}R')^{-1}RS^{-1}PQ,$$

(6.5) can be expressed as

$$(6.6) \qquad E(\hat{\hat{\underline{\beta}}}-E\hat{\hat{\underline{\beta}}})(\hat{\hat{\underline{\beta}}}-E\hat{\hat{\underline{\beta}}})' = \sigma^2S^{-1} - \sigma^2h_\lambda(2)S^{-1} - [2h_\lambda(2)-h_\lambda(4)+h_\lambda^2(2)]S^{-1}R' \cdot(RS^{-1}R')^{-1}\underline{\delta}\underline{\delta}'(RS^{-1}R')^{-1}RS^{-1}.$$

Hence, the variance of $\hat{\hat{\underline{\beta}}}$ depends on the variance of $\underline{b}$, the error in the restriction, $\underline{\delta}$, and probabilities which are associated with the chance of accepting the hypothesis that $\lambda \leq \lambda_o$ and using the restriction matrix, R.

7. The Risk for the Preliminary Test Estimator $\hat{\hat{\underline{y}}}$

In addition to the risk function for the preliminary test estimator $\hat{\hat{\underline{\beta}}}$, one might be interested in the quadratic loss for conditional mean forecasting and thus the risk function for the estimator $\hat{\hat{\underline{y}}} = X\hat{\hat{\underline{\beta}}}$. For our case, this implies a risk function

$$(7.1) \qquad E(X\hat{\hat{\underline{\beta}}}-X\underline{\beta})'(X\hat{\hat{\underline{\beta}}}-X\underline{\beta}) = E(\hat{\hat{\underline{\beta}}}-\underline{\beta})'X'X(\hat{\hat{\underline{\beta}}}-\underline{\beta}),$$

which weights the elements in the quadratic form, $E(\hat{\hat{\underline{\beta}}}-\underline{\beta})'(\hat{\hat{\underline{\beta}}}-\underline{\beta})$ with elements

from the cross products matrix, X'X. Wallace (1971) considered a risk function of the form (7.1) for the restricted estimator $\hat{\underset{\sim}{\beta}}$ largely because of the simplification it produced in the tables required for testing hypotheses about the non-centrality parameter, λ. These same simplifications occur for the preliminary test estimator.

The value of risk function (7.1) can be developed using the methodology of Section 3 with only minor changes. Equation (3.7) from Section 3 becomes

$$(7.2) \qquad Q(P^{-1})'SP^{-1}Q' = A^* = \begin{pmatrix} A_1^* & A_3^* \\ A_3^{*\prime} & A_2^* \end{pmatrix} = I,$$

since $(P^{-1})'SP^{-1} = I$ and Q is an orthogonal transformation.

As a consequence of the weighting pattern in risk function (7.1), the criterion for the risk function for $\underset{\sim}{b}$ to exceed that of $\hat{\hat{\underset{\sim}{\beta}}}$ is

$$(7.3) \qquad \lambda \leq \frac{1}{4} \sum_{i=1}^{J} \frac{d_i^{(1)}}{d_L} = \frac{J}{4},$$

and for the risk function of $\hat{\hat{\underset{\sim}{\beta}}}$ to exceed that of $\underset{\sim}{b}$ is

$$(7.4) \qquad \lambda \geq \frac{1}{2} \sum_{i=1}^{J} \frac{d_i^{(1)}}{d_S} = \frac{J}{2},$$

since $d_1, \ldots, d_J$ are all ones. Thus, in this special case, which is in line with the results of Section 3b, the minimum value for λ, which is small enough to insure that the risk function for the unrestricted estimator is less than that of the unrestricted estimator, reduces to a single value, $\frac{J}{4}$. It should be noted that if one assumes the orthonormality of regressors, this is equivalent to taking (7.1) as the risk function. The extension of the orthonormal regressor results to the general case is not a direct one as the measure of goodness is changed.

8. Concluding Remarks

Using a quadratic risk function to measure estimator performance, a test procedure proposed by Wallace, and the methods for evaluating the preliminary test statistic developed by Stein and Sclove, we have investigated the properties of the preliminary test estimator for the standard regression problem in which both sample and exact linear prior information or hypotheses about the parameters is utilized. In particular, a preliminary test estimator is evaluated which permits the investigator to utilize both sample and exact, though slightly incorrect, information to improve the estimates when judged by a quadratic risk function criterion over certain regions of the linear hypothesis space. The mean and variance of the preliminary test estimator is specified and the condition for which this estimator is better than the conventional estimator, in a quadratic risk context, is derived in terms of the non-centrality parameter, λ, of a non-central F distribution. It is shown that in order for the preliminary test estimator to be superior to the conventional estimator that $\lambda \leq \frac{1}{4} \sum_{i=1}^{J} \frac{d_i^{(1)}}{d_L} \leq \frac{J}{4}$, which contrasts with the result found by Wallace when comparing the risk functions of the restricted and unrestricted estimators. When the risk functions for the orthonormal regressors and conditional mean forecast cases are compared, the condition for the preliminary test estimator to be superior is that $\lambda \leq \frac{J}{4}$.

The choice of the level of the test, and hence the critical value, conditions the relative gain or loss accruing to the preliminary test estimators for various values of λ and the d_i. At this stage, the choice of an optimal critical point, c, that would satisfy some criterion or lead to an optimum decision rule remains to be resolved. Inasmuch as the advantage of the preliminary test estimator over the usual unrestricted estimator occurs when λ

is confined to an interval $(0, \frac{1}{4} \sum_{i=1}^{J} \frac{d_i^{(1)}}{d_L})$, a Bayesian analysis in which a prior distribution is placed on λ seems to be one natural extension of this work. As noted in the introduction, Sclove et al., in an unpublished paper, have studied estimation preceded by testing for the orthonormal regressors case and reach conclusions compatible with those we have derived for the general model. In addition, it has been shown that for comparable values of c and for K or J greater than 2 and for $0 < c < 2(K-2)|(T+K)$, the Stein-James (1961) positive part estimator strictly dominates the preliminary test estimator. These results should carry over for the general case and what remains to be done is to contrast the risk functions for the two estimators for non-comparable critical values c.

Another line of inquiry would be to alter the exact constraint into a stochastic constraint and following the work of Theil and Goldberger (1961) and Theil (1963), develop a preliminary test estimator for that model. A third line of inquiry is to consider as a criterion a matrix of risk functions rather than a risk function and thus extend the work of Toro-Vizcarrondo and Wallace (1968). The authors are developing these lines of inquiry in other papers at the present time.

Appendix A
Some Theorems and Lemmas

We now turn to Lemmas 1 and 2 to be used in Theorems 1 and 2.

Lemma 1. If the random variable u is $N(\theta,1)$, then

$$EI_{(0,c)}(u^2)u^2 = Pr(\chi^2_{(\frac{\theta^2}{2},3)} < c) + \theta^2 Pr(\chi^2_{(\frac{\theta^2}{2},5)} < c).$$

Proof: If u is $N(\theta,1)$, then u^2 is distributed as non-central $\chi^2_{(\frac{\theta^2}{2},1)}$. Thus, u^2 is distributed as central $\chi^2_{(1+2H)}$ where H is a random variable with a Poisson distribution with parameter $\frac{\theta^2}{2}$.

The expectation,

$$EI_{(0,c)}(u^2)u^2 = E[E[I_{(0,c)}(t_h)t_h|H=h]] = \sum_{h=0}^{\infty} E[I_{(0,c)}(t_h)t_h]\frac{e^{\frac{-\theta^2}{2}}\left(\frac{\theta^2}{2}\right)^h}{h!},$$

where t_h is distributed as $\chi^2_{(1+2h)}$,

$$EI_{(0,c)}(u^2)u^2 = \sum_{h=0}^{\infty}(1+2h)\left\{\frac{\int_0^c t^{\frac{3+2h}{2}-1}e^{\frac{-t}{2}}dt}{2^{\frac{3+2h}{2}}\Gamma\left(\frac{3+2h}{2}\right)}\right\}\frac{e^{\frac{-\theta^2}{2}}\left(\frac{\theta^2}{2}\right)^h}{h!}$$

$$= \sum_{h=0}^{\infty}\frac{\int_0^c t^{\frac{3+2h}{2}-1}e^{\frac{-t}{2}}dt}{2^{\frac{3+2h}{2}}\Gamma\left(\frac{3+2h}{2}\right)}\frac{e^{\frac{-\theta^2}{2}}\left(\frac{\theta^2}{2}\right)^h}{h!} + 2\sum_{h=1}^{\infty}\frac{\int_0^c t^{\frac{3+2h}{2}-1}e^{\frac{-t}{2}}dt}{2^{\frac{3+2h}{2}}\Gamma\left(\frac{3+2h}{2}\right)}\frac{e^{\frac{-\theta^2}{2}}\left(\frac{\theta^2}{2}\right)^h}{(h-1)!}$$

$$= Pr(\chi^2_{(\frac{\theta^2}{2},3)} < c) + \theta^2 Pr(\chi^2_{(\frac{\theta^2}{2},5)} < c)$$

$$= EI_{(0,c)}(\chi^2_{(\frac{\theta^2}{2},3)}) + \theta^2 EI_{(0,c)}(\chi^2_{(\frac{\theta^2}{2},5)})$$

which was to be shown.

The following proof is a variant of one used by Stein (1966) to obtain a similar result.

Lemma 2. If u is $N(\theta,1)$, then

$$EI_{(0,c)}(u^2)u = \theta EI_{(0,c)}(\chi^2_{(\frac{\theta^2}{2},3)}) = \theta Pr(\chi^2_{(\frac{\theta^2}{2},3)} < c).$$

Proof: If u is $N(\theta,1)$, u^2 can be expressed as a central $\chi^2_{(1+2H)}$ where H is distributed as Poisson $\left(\frac{\theta^2}{2}\right)$. The expectation

$$EI_{(0,c)}(u^2)u = \frac{e^{\frac{-\theta^2}{2}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} I_{(0,c)}(u^2)u\, e^{(\frac{-u^2}{2}+\theta u)} du$$

$$= \frac{e^{\frac{-\theta^2}{2}}}{\sqrt{2\pi}} \frac{\partial}{\partial\theta} \{ \int_{-\infty}^{\infty} I_{(0,c)}(u^2) e^{(\frac{-u^2}{2}+\theta u)} du\}$$

$$= e^{\frac{-\theta^2}{2}} \frac{\partial}{\partial\theta} \{e^{\frac{\theta^2}{2}} E(I_{(0,c)}(u^2))\}$$

$$= e^{\frac{-\theta^2}{2}} \frac{\partial}{\partial\theta} \{ \sum_{h=0}^{\infty} \frac{\left(\frac{\theta^2}{2}\right)^h}{h!} EI_{(0,c)}(\chi^2_{(1+2h)})\}$$

$$= \theta e^{\frac{-\theta^2}{2}} \{ \sum_{h=1}^{\infty} \frac{\left(\frac{\theta^2}{2}\right)^{h-1}}{(h-1)!} EI_{(0,c)}[\chi^2_{(3+2(h-1))}]\}$$

$$= \theta Pr(\chi^2_{(\frac{\theta^2}{2},3)} < c).$$

Q.E.D.

Using Lemma 1, we have:

Theorem 1: If the $(J \times 1)$ vector $\underset{\sim}{u}$ is distributed as $N(\underset{\sim}{\theta},I)$, then

$$EI_{(0,c)}(\underset{\sim}{u}'\underset{\sim}{u})\underset{\sim}{u} = \underset{\sim}{\theta} EI_{(0,c)}(\chi^2_{(\frac{\underset{\sim}{\theta}'\underset{\sim}{\theta}}{2},J+2)}) = \underset{\sim}{\theta} Pr(\chi^2_{(\frac{\underset{\sim}{\theta}'\underset{\sim}{\theta}}{2},J+2)} < c).$$

Proof: Let $\underline{u} = (u_1,\ldots,u_J)'$ so $\underline{u}'\underline{u} = \sum_{j=1}^{J} u_j^2$. Conditioning on the u_j's,

$$E[I_{(0,c)}(\underline{u}'\underline{u})\underline{u}] = \{E[E[I_{(0,c-\sum_{j\neq 1} u_j^2)}(u_1^2)u_1 | u_j, j\neq 1, \ldots,$$

$$E[E[I_{(0,c-\sum_{j=1}^{J-1} u_j^2)}(u_J^2)u_J | u_j, j\neq J]]\}',$$

which by Lemma 2 gives

$$E[I_{(0,c)}(\underline{u}'\underline{u})\underline{u}] = \{\theta_1 E[I_{(0,c-\sum_{j=2}^{J} u_j^2)}(\chi^2_{(\frac{\theta_1^2}{2},3)})], \ldots, \theta_J E[I_{(0,c-\sum_{j=1}^{J-1} u_j^2)}(\chi^2_{(\frac{\theta_J^2}{2},3)})]\}'$$

$$= \{\theta_1 E[I_{(0,c)}(\chi^2_{(\frac{\theta_1^2}{2},3)} + \sum_{j=2}^{J} u_j^2)], \ldots, \theta_J E[I_{(0,c)}(\chi^2_{(\frac{\theta_J^2}{2},3)} + \sum_{j=1}^{J-1} u_j^2)]\}'$$

Now, since the sum of independent variables with non-central chi square distributions has a chi square distribution with a non-centrality parameter which is the sum of the non-centrality parameters for the variables summed and degrees of freedom equal to the sum of the degrees of freedom of the individual variables,

$$E[I_{(0,c)}(\underline{u}'\underline{u})\underline{u}] = (\theta_1,\ldots,\theta_J)EI_{(0,c)}(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+2)}) = \underline{\theta}\Pr(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+2)} < c),$$

Q.E.D.

Using Lemma 1, we have:

Theorem 2: If $\underline{u}$ is a $(J \times 1)$ vector distributed as $N(\underline{\theta}, I)$ and A is any positive definite symmetric matrix, then

$$EI_{(0,c)}((\underline{u}'\underline{u})\underline{u}'A\underline{u}) = E[I_{(0,c)}(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+2)})]\mathrm{tr}A + E[I_{(0,c)}(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+4)})]\underline{\theta}'A\underline{\theta}$$

$$= \Pr(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+2)} < c)\mathrm{tr}A + \Pr(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+4)} < c)\underline{\theta}'A\underline{\theta}.$$

Proof: Let P be an orthogonal matrix such that $PAP' = D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ \vdots & \ddots & & \vdots \\ 0 & & & d_J \end{pmatrix}$ where the $d_i > 0$ are the characteristic roots of A. Define the $(J \times 1)$ vector $\underset{\sim}{w} = P\underset{\sim}{u}$. So $\underset{\sim}{w}$ is distributed $N(P\underset{\sim}{\theta}, I)$. We then have

$$E[I_{(0,c)}(\underset{\sim}{u}'\underset{\sim}{u})\underset{\sim}{u}'A\underset{\sim}{u}] = E[I_{(0,c)}(\underset{\sim}{w}'\underset{\sim}{w})\underset{\sim}{w}'D\underset{\sim}{w}]$$

$$= \sum_{i=1}^{J} d_i E[E[I_{(0,c-\sum_{j\neq i} w_j)}(w_i^2)w_i^2 | w_j, j\neq i]],$$

which by Lemma 1 can be expressed as

$$\sum_{i=1}^{J} \{d_i E[I_{(0,c-\sum_{j\neq i} w_j^2)}(\chi^2_{(\frac{(\underset{\sim}{p}_i'\underset{\sim}{\theta})^2}{2},3)})] + (\underset{\sim}{p}_i'\underset{\sim}{\theta})^2 E[I_{(0,c-\sum_{j\neq i} w_j^2)}(\chi^2_{(\frac{(\underset{\sim}{p}_i'\underset{\sim}{\theta})^2}{2},5)})]\},$$

where $\underset{\sim}{p}_i'$ is the i^{th} row of P. Therefore,

$$E[I_{(0,c)}(\underset{\sim}{u}'\underset{\sim}{u})\underset{\sim}{u}'A\underset{\sim}{u}] = \sum_{i=1}^{J} \{d_i E[I_{(0,c)}(\chi^2_{(\frac{\underset{\sim}{\theta}'\underset{\sim}{\theta}}{2},J+2)})] + E[I_{(0,c)}(\chi^2_{(\frac{\underset{\sim}{\theta}'\underset{\sim}{\theta}}{2},J+4)})]\}$$

$$= Pr(\chi^2_{(\frac{\underset{\sim}{\theta}'\underset{\sim}{\theta}}{2},J+2)} < c)trA + Pr(\chi^2_{(\frac{\underset{\sim}{\theta}'\underset{\sim}{\theta}}{2},J+4)} < c)\underset{\sim}{\theta}'A\underset{\sim}{\theta}$$

Q.E.D.

A theorem useful in evaluating the covariance matrix of $\hat{\hat{\underset{\sim}{\beta}}}$ is

<u>Theorem 3</u>: If the $(J \times 1)$ vector $\underset{\sim}{u}$ is distributed normally with mean vector $\underset{\sim}{\theta}$ and covariance matrix I of order J, then

$$EI_{(0,c)}(\underset{\sim}{u}'\underset{\sim}{u})\underset{\sim}{u}\underset{\sim}{u}' = E[I_{(0,c)}(\chi^2_{(\lambda,J+2)})]I_J + E[I_{(0,c)}(\chi^2_{(\lambda,J+4)})]\underset{\sim}{\theta}\underset{\sim}{\theta}',$$

where $\lambda = \underset{\sim}{\theta}'\underset{\sim}{\theta}/2$.

Proof: Let $\underset{\sim}{u} = (u_1,\ldots,u_J)'$, and determine the diagonal and off-diagonal elements of $EI_{(0,c)}(\underset{\sim}{u}'\underset{\sim}{u})\underset{\sim}{u}\underset{\sim}{u}'$. The diagonal elements are of the form

$$E[I_{(0,c)}(\sum_{i=1}^{J} u_i^2)u_i^2] = E[E[I_{(0,c)}(u_i^2)u_i^2|u_j^2,j\neq i]$$

$$= E[E[I_{(0,c^*)}(\chi^2_{(\frac{\theta_i^2}{2},3)})|u_j^2,j\neq i]$$

$$+ \theta_i^2 E[I_{(0,c^*)}(\chi^2_{(\frac{\theta_i^2}{2},5)})|u_j^2,j\neq i]]$$

(by Lemma 1 and letting $c^* = c-\sum_{j\neq i} u_j^2$)

$$E[E[I_{(0,c^*)}(\chi^2_{(\frac{\theta_i^2}{2},3)})|u_j,j\neq i] + \theta_i^2 E[I_{(0,c^*)}(\chi^2_{(\frac{\theta_i^2}{2},5)})|u_j.j\neq i]]$$

$$= E[I_{(0,c)}(\chi^2_{(\frac{\theta_i^2}{2},3)} + \sum_{j\neq i} u_j^2)] + \theta_i^2 E[I_{(0,c)}(\chi^2_{(\frac{\theta_i^2}{2},5)} + \sum_{j\neq i} u_j)]$$

$$= E[I_{(0,c)}(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+2)})] + \theta_i^2 E[I_{(0,c)}(\chi^2_{(\frac{\underline{\theta}'\underline{\theta}}{2},J+4)})].$$

The off-diagonal elements, for $i \neq j$, have the form

$$E[I_{(0,c)}(\sum_{k=1}^{J} u_k^2)u_iu_j] = E[u_jE[I_{(0,c^*)}(u_i^2)u_i|u_k,k\neq i]]$$

$$= E[u_j\theta_iE[I_{(0,c^*)}(\chi^2_{(\frac{\theta_i^2}{2},3)})|u_k,k\neq i]]$$

by Lemma 2 and where $c^* = c-\sum_{j\neq i} u_j^2$. Furthermore,

$$Eu_j\theta_iE[I_{(0,c^*)}(\chi^2_{(\frac{\theta_i^2}{2},3)})|u_k,k\neq i]] = E[u_j\theta_iEI_{(0,c-u_j^2)}(\chi^2_{(\frac{\theta_i^2}{2},3)} + \sum_{\substack{k\neq i\\k\neq j}} u_k^2|u_j)]$$

$$= E[u_j\theta_iE[I_{(0,c-u_j^2)}(\chi^2_{(\sum_{k\neq j}\frac{\theta_i^2}{2},3+J-2)})|u_j]].$$

Now interchanging u_j^2 and $\chi^2_{(\frac{\Sigma\theta_i^2}{2},J+1)}$, we have

$$\theta_i E[E[I_{(0,c-\chi^2_{(\sum_{k\neq j}\frac{\theta_i^2}{2},J+1)})}(u_j^2)u_j \,|\, \chi^2_{(\sum_{k\neq j}\frac{\theta_i^2}{2},J+1)}]]$$

$$= \theta_i\theta_j E[E[I_{(0,c-\chi^2_{(\sum_{i\neq j}\frac{\theta_i^2}{2},J+1)})}(\chi^2_{(\frac{\theta_i^2}{2},3)}) \,|\, \chi^2_{\sum_{k\neq j}\frac{\theta_i^2}{2}}]]$$

by Lemma 2. The unconditional expectations of the off-diagonal elements are

$$\theta_i\theta_j E I_{(0,c)}(\chi^2_{(\sum_{k\neq j}\frac{\theta_i^2}{2},J+1)} + \chi^2_{(\frac{\theta_i^2}{2},3)}) = \theta_i\theta_j E[I_{(0,c)}(\chi^2_{(\lambda,J+4)})],$$

where $\lambda = \sum_{i=1}^{J}\theta_i^2/2$.

Combining the diagonal and off-diagonal components, the matrix may be written as

$$E[I_{(0,c)}(\chi^2_{(\lambda,J+2)})]I_J + E[I_{(0,c)}(\chi^2_{(\lambda,J+4)})]\underline{\theta}\underline{\theta}'.$$

Q.E.D.

Appendix B. Properties of Functions of the Non-Central F Distribution

In this Appendix, a theorem is developed which permits the evaluation of regions where the non-centrality parameter, λ, of the non-central F distribution is either small enough to insure the risk function for $\hat{\hat{\beta}}$ is less than that of b or large enough to insure that the risk function of $\hat{\hat{\beta}}$ is larger than that of b.

Theorem 1: Let

$$h_\lambda(\ell) \equiv \Pr(\chi^2_{(\lambda,J+\ell)}/\chi^2_{(T-K)} \leq cJ/T-K)$$

and

$$w_o = \frac{c}{\frac{T-K}{J} + c} .$$

If $T-K \geq 2$, then $h_\lambda(4)/h_\lambda(2) \geq w_o$.

Proof:

$$\text{(B.1)} \qquad \Pr(\chi^2_{(\lambda,J+\ell)}/\chi^2_{(T-K)} \leq \frac{cJ}{T-K}) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \Pr\left(\frac{\chi^2_{(J+\ell+2k)}}{\chi^2_{(T-K)}} \leq \frac{cJ}{T-K}\right),$$

where

$$\Pr\left(\frac{\chi^2_{(J+\ell+2k)}}{\chi^2_{(T-K)}} \leq \frac{cJ}{T-K}\right) = \int_0^{\frac{cJ}{T-K}} \frac{w^{\frac{J+\ell+2k-1}{2}} dw}{\beta(\frac{J+\ell+2k}{2}, \frac{T-K}{2})(1+w)^{\frac{J+\ell+2k+T-K}{2}}} . \;\; 1/$$

Integrating the density by parts and defining $c_o = \frac{cJ}{T-K}$ yields

---

1/ Lindgren (1968), p. 380.

(B.2) $$\frac{2}{J+\ell+2k}\int_0^{c_o}\frac{(\frac{J+\ell+2k}{2})w^{\frac{J+\ell+2k-1}{2}}\,dw}{\beta(\frac{J+\ell+2k}{2},\ \frac{T-K}{2})(1+w)^{\frac{J+\ell+2k+T-K}{2}}}$$

$$= \frac{2}{J+\ell+2k}\left[\frac{w^{\frac{J+\ell+2k}{2}}}{\beta(\frac{J+\ell+2k}{2},\ \frac{T-K}{2})(1+w)^{\frac{J+\ell+2k+T-K}{2}}}\right]_0^{c_o}$$

$$+ \int_0^{c_o}\frac{2(\frac{J+\ell+2k+T-K}{2})w^{\frac{J+\ell+2k}{2}}\,dw}{(\frac{J+\ell+2k}{2})\beta(\frac{J+\ell+2k}{2},\ \frac{T-K}{2})(1+w)^{\frac{J+\ell+2k+T-K}{2}}},$$

which can be written as

(B.3) $$\Pr\left(\frac{\chi^2_{(\lambda,J+\ell)}}{\chi^2_{(T-K)}} \le c_o\right) \equiv t_{k+\frac{\ell}{2}} = t_{(k+1)+\frac{\ell}{2}} + \frac{c_o^{\frac{J+\ell+2k}{2}}(\frac{2}{J+\ell+2k})}{\beta(\frac{J+\ell+2k}{2},\frac{T-K}{2})(1+c_o)^{\frac{J+\ell+2k+T-K}{2}}}.$$

Using (B.3) recursively,

$$t_{k+\frac{\ell}{2}} = \sum_{j=k}^{\infty}\frac{c_o^{\frac{J+\ell+2j}{2}}(\frac{2}{J+\ell+2j})}{\beta(\frac{J+\ell+2j}{2},\frac{T-K}{2})(1+c_o)^{\frac{J+\ell+T-K+2j}{2}}}$$

since $\lim_{k'\to\infty} t_{k'} = 0$.

Letting the previous definitions of c_o and w_o mean that $c_o = \frac{w_o}{1-w_o}$ and $\frac{1}{1+c_o} = 1-w_o$,

(B.4) $$t_{k+\frac{\ell}{2}} = \sum_{j=k}^{\infty}\frac{\frac{2}{J+\ell+2j}\,w_o^{\frac{J+\ell+2j}{2}}(1-w_o)^{\frac{T-K}{2}}}{\beta(\frac{J+\ell+2j}{2},\frac{T-K}{2})}.$$

It follows from using (B.4) in (B.1) that

$$h_\lambda(\ell) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \sum_{j=k}^{\infty} \frac{\Gamma(\frac{J+\ell+T-K+2j}{2}) w_o^{\frac{J+\ell+2j}{2}} (1-w_o)^{\frac{J+\ell+T-K+2j}{2}}}{\Gamma(\frac{T-K}{2})\Gamma(\frac{J+\ell+2j+2}{2})(\frac{J+\ell+2j}{2})}$$

Hence,

$$\frac{h_\lambda(4)}{h_\lambda(2)} = \frac{e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\sum_{j=k}^{\infty}\frac{w_o^{\frac{J+4+2j}{2}}(1-w_o)^{\frac{T-K}{2}}\Gamma(\frac{J+4+T-K+2j}{2})}{\Gamma(\frac{J+4+2j+2}{2})\Gamma(\frac{T-K}{2})}}{e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\sum_{j=k}^{\infty}\frac{w_o^{\frac{J+2+2j}{2}}(1-w_o)^{\frac{T-K}{2}}\Gamma(\frac{J+2+T-K+2j}{2})}{\Gamma(\frac{J+2+2j+2}{2})\Gamma(\frac{T-K}{2})}}$$

$$= w_o \frac{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\sum_{j=k}^{\infty}\frac{w_o^{\frac{J+2j+2}{2}}\Gamma(\frac{J+2+T-K+2j}{2})(\frac{J+2+T-K+2j}{2})}{\Gamma(\frac{J+2+2j+2}{2})(\frac{J+2+2j+2}{2})}}{\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}\sum_{j=k}^{\infty}\frac{w_o^{\frac{J+2j+2}{2}}\Gamma(\frac{J+2+T-K+2j}{2})}{\Gamma(\frac{J+2+2j+2}{2})}} \geq w_o ,$$

since $\frac{J+2+T-K+2j}{J+2+2j+2} \geq 1$ whenever $T-K \geq 2$, which proves the theorem. Furthermore,

$$\frac{J+2+T-K+2j}{J+2+2j+2} = \frac{(J+2+T-K)+2j}{(J+4)+2j} \leq \frac{J+2+T-K}{J+4} ,$$

for $j = 0,1,2,\ldots$; hence,

$$\frac{h_\lambda(4)}{h_\lambda(2)} \leq w_o\left(\frac{J+2+T-K}{J+4}\right) = w_o\left(1 + \frac{T-K-2}{J+4}\right), \quad \text{when } T-K \geq 2.$$

Appendix C — The Risk Function for $\hat{\underline{\beta}}$

The evaluation of $E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta})$ begins with text equation (2.7d)

$$\text{(C.0)} \qquad E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) = \text{tr covar}(\hat{\underline{\beta}}) + \text{tr (bias } \hat{\underline{\beta}})(\text{bias } \hat{\underline{\beta}})'$$

and the covariance and bias terms are transformed using the transformations described in Section 3a. By (2.7b), the covariance term can be expressed as

$$\text{(C.1)} \qquad = \sigma^2[\text{tr}A - \text{tr}Q(P^{-1})'Q'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})'Q'],$$

$$= \sigma^2[\text{tr}A - \text{tr}\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}A]$$

Equation (C.1) can be expressed as

$$\text{(C.2)} \qquad = \sigma^2 \text{tr}A_2$$

$$= \sigma^2 \sum_{j=1}^{K-J} d_j^{(2)},$$

where $d_j^{(2)}$ are the characteristic roots of A_2 as defined subsequent to definition (3.7).

The bias term in (C.0) is

$$\text{(C.3)} \qquad \text{tr } S^{-1}R'(RS^{-1}R')^{-1}\underline{\delta}\underline{\delta}'(RS^{-1}R')^{-1}RS^{-1}$$

$$= \text{tr}[QPS^{-1}R'(RS^{-1}R')^{-1}(RP^{-1}Q'QP\underline{\beta}-\underline{r})]'Q(P^{-1})'P^{-1}Q'$$

$$\cdot[QPS^{-1}R'(RS^{-1}R')^{-1}(RP^{-1}Q'QP\underline{\beta}-\underline{r})]$$

$$= [\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}QP\underline{\beta}-QPS^{-1}R'(RS^{-1}R')^{-1}\underline{r}]'A[\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}QP\underline{\beta}-QPS^{-1}R'(RS^{-1}R')^{-1}\underline{r}].$$

We note that

(C.4) $$QPS^{-1}R'(RS^{-1}R')^{-1} = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} QPS^{-1}R'(RS^{-1}R')^{-1}.$$

Substituting (C.4) into (C.3), we have

(C.5) $$\begin{aligned} \text{tr}\; S^{-1}R'(RS^{-1}R')^{-1}\underline{\delta}\underline{\delta}'(RS^{-1}R')^{-1}RS^{-1} &= [QP(\underline{\beta}-S^{-1}R'(RS^{-1}R')^{-1}\underline{r})]' \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} A_1 & A_3 \\ A_3' & A_2 \end{pmatrix} \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} \\ &\quad \cdot [QP(\underline{\beta}-S^{-1}R'(RS^{-1}R')^{-1}\underline{r})] \\ &= \sum_{i=1}^{J} d_i^{(1)}(\xi_i^*)^2, \end{aligned}$$

where ξ_i^* and $d_i^{(1)}$ are defined in Sections 3a and 3b.

Consequently, using (C.2) and (C.5),

(C.6) $$E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) = \sigma^2 \sum_{j=1}^{K-J} d_j^{(2)} + \sum_{i=1}^{J} d_i^{(1)}(\xi_i^*)^2.$$

Appendix D — Bias and Covariance for $\hat{\hat{\underset{\sim}{\beta}}}$

In this Appendix, the bias and covariance matrix for $\hat{\hat{\underset{\sim}{\beta}}}$ are determined. Returning to the definition of $\hat{\hat{\underset{\sim}{\beta}}}$ and its derived value in (2.11), and the transformations of Section 3,

$$\text{(D.1)} \qquad \hat{\hat{\underset{\sim}{\beta}}} = \underset{\sim}{b} - I_{(0,c)}(u)P^{-1}Q'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}\underset{\sim}{w} = \underset{\sim}{b} - P^{-1}Q'I_{(0,c)}(u)\begin{pmatrix} \underset{\sim}{w}_1 \\ 0 \end{pmatrix}.$$

Therefore, the expected value of $\hat{\hat{\underset{\sim}{\beta}}}$ is

$$\text{(D.2)} \qquad E\hat{\hat{\underset{\sim}{\beta}}} = \underset{\sim}{\beta} - \sigma P^{-1}Q'\begin{pmatrix} I \\ 0 \end{pmatrix}E[I_{(0,c^*)}\left(\frac{\underset{\sim}{w}_1'\underset{\sim}{w}_1}{\sigma^2}\right)\frac{\underset{\sim}{w}_1}{\sigma}],$$

since $c > u = \dfrac{\underset{\sim}{w}_1'\underset{\sim}{w}_1}{J\hat{\sigma}^2}$ and $\dfrac{\underset{\sim}{w}_1'\underset{\sim}{w}_1}{\sigma^2} < \dfrac{Jc\hat{\sigma}^2}{\sigma^2} = c^*$ as shown in Section 3b.

Using Theorem 1 to evaluate the expectation in (D.2),

$$\text{(D.3)} \qquad E\hat{\hat{\underset{\sim}{\beta}}} = \underset{\sim}{\beta} - P^{-1}Q'\begin{pmatrix} I \\ 0 \end{pmatrix}\underset{\sim}{\eta}_1 P\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right).$$

Noticing that

$$P^{-1}Q'\begin{pmatrix} I \\ 0 \end{pmatrix}\underset{\sim}{\eta}_1 = P^{-1}Q'\begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}[QP\underset{\sim}{\beta}-QPS^{-1}R'(RS^{-1}R')^{-1}(R\underset{\sim}{\beta}-\underset{\sim}{\delta})],$$

as well as

$$\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}QPS^{-1}R'(RS^{-1}R')^{-1}\underset{\sim}{\delta} = QP^{-1}R'(RS^{-1}R')^{-1}\underset{\sim}{\delta},$$

$$\text{(D.4)} \qquad E\hat{\hat{\underset{\sim}{\beta}}} = \underset{\sim}{\beta} + S^{-1}R'(RS^{-1}R')^{-1}\underset{\sim}{\delta}\ Pr\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right),$$

and the amount of bias depends on $\underset{\sim}{\delta}$ and the level of the test, α, through c. A small α implies a larger c and hence increases the probability that given $\underset{\sim}{\delta}$ will be included in $E\hat{\hat{\underset{\sim}{\beta}}}$.

Turning now to the covariance matrix of $\hat{\hat{\underset{\sim}{\beta}}}$, it can be written as

$$\text{(D.5)} \quad \text{Var}(\hat{\hat{\underset{\sim}{\beta}}}) = E(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta})(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta})' - E(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta})E(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta})'$$

$$= \text{MSE}\hat{\hat{\underset{\sim}{\beta}}} - [\Pr\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}} < \frac{cJ}{T-K}\right)]^2[S^{-1}R'(RS^{-1}R')^{-1}\underset{\sim}{\delta}][\underset{\sim}{\delta}'(RS^{-1}R')^{-1}RS^{-1}],$$

but

$$QPS^{-1}R'(RS^{-1}R')^{-1}\underset{\sim}{\delta} = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}\underset{\sim}{\eta},$$

so

$$\text{(D.6)} \quad \text{Var}\ \hat{\hat{\underset{\sim}{\beta}}} = \text{MSE}\hat{\hat{\underset{\sim}{\beta}}} - [\Pr\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}} < \frac{Jc}{T-K}\right)]^2 P^{-1}Q'\begin{pmatrix} \underset{\sim}{\eta}_1\underset{\sim}{\eta}_1' & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})'.$$

From the definition of $\underset{\sim}{w} = QP[\underset{\sim}{b}-S^{-1}R'(RS^{-1}R')^{-1}\underset{\sim}{r}]$, given after (3.4), the $\text{MSE}\hat{\hat{\underset{\sim}{\beta}}}$ is

$$\text{(D.7)} \quad E(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta})(\hat{\hat{\underset{\sim}{\beta}}}-\underset{\sim}{\beta})' = E(\underset{\sim}{b}-\underset{\sim}{\beta})(\underset{\sim}{b}-\underset{\sim}{\beta})' - E[(\underset{\sim}{b}-\underset{\sim}{\beta})I_{(0,c)}(u)\underset{\sim}{w}'Q(P^{-1})'R'(RS^{-1}R')^{-1}RS^{-1}]$$
$$- E[I_{(0,c)}(u)(S^{-1}R'(RS^{-1}R')^{-1}RP^{-1}Q'\underset{\sim}{w})(\underset{\sim}{b}-\underset{\sim}{\beta})']$$
$$+ (S^{-1}R'(RS^{-1}R')^{-1}RP^{-1}Q'E[I_{(0,c)}(u)\underset{\sim}{w}\underset{\sim}{w}']Q(P^{-1})'R'(RS^{-1}R')^{-1}RS^{-1})$$

$$\text{(D.7')} \quad = \sigma^2S^{-1} - P^{-1}Q'EI_{(0,c)}(u)(\underset{\sim}{w}-\underset{\sim}{\eta})\underset{\sim}{w}'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})'$$
$$- P^{-1}Q'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}EI_{(0,c)}(u)\underset{\sim}{w}(\underset{\sim}{w}-\underset{\sim}{\eta})'Q(P^{-1})'$$
$$+ P^{-1}Q'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}EI_{(0,c)}(u)\underset{\sim}{w}\underset{\sim}{w}'\begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})'$$

since

$$Q(P^{-1})'R'(RS^{-1}R')^{-1}RP^{-1}Q' = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad S^{-1} = P^{-1}(P^{-1})' = P^{-1}Q'Q(P^{-1})'.$$

Making use of the results of Section 3b, $u = \frac{\underline{w}_1'\underline{w}_1}{J\hat{\sigma}^2} < c$ so $\frac{\underline{w}_1'\underline{w}_1}{\sigma^2} < \frac{cJ\hat{\sigma}^2}{\sigma^2} = c^*$,

given $\hat{\sigma}^2$, we have for the second term of (D.7')

$$\text{(D.8)} \qquad = -P^{-1}Q'\begin{pmatrix} EI_{(0,c^*)}\left(\frac{\underline{w}_1'\underline{w}_1}{\sigma^2}\right)(\underline{w}_1-\underline{\eta}_1)\underline{w}_1' & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})'.$$

A similar argument holds for the third and fourth terms of (D.7') so

$$\text{(D.9)} \qquad E(\hat{\hat{\underline{\beta}}}-\underline{\beta})(\hat{\hat{\underline{\beta}}}-\underline{\beta})' = \sigma^2S^{-1}-\sigma^2P^{-1}Q'\begin{pmatrix} EI_{(0,c^*)}\left(\frac{\underline{w}_1'\underline{w}_1}{\sigma^2}\right)\frac{\underline{w}_1}{\sigma}\frac{\underline{w}_1'}{\sigma} & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})'$$

$$+P^{-1}Q'\begin{pmatrix} \underline{\eta}_1EI_{(0,c^*)}\left(\frac{\underline{w}_1'\underline{w}_1}{\sigma^2}\right)\underline{w}_1' & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})'$$

$$+P^{-1}Q'\begin{pmatrix} EI_{(0,c^*)}\left(\frac{\underline{w}_1'\underline{w}_1}{\sigma^2}\right)\underline{w}_1\underline{\eta}_1' & 0 \\ .0 & 0 \end{pmatrix}Q(P^{-1})'.$$

Using Theorems 1 and 3 of Appendix A, the risk matrix of $\hat{\hat{\underline{\beta}}}$ is

$$\text{(D.10)} \qquad E(\hat{\hat{\underline{\beta}}}-\underline{\beta})(\hat{\hat{\underline{\beta}}}-\underline{\beta})' = \sigma^2S^{-1}$$

$$-P^{-1}Q'\begin{pmatrix} E[E[I_{(0,c^*)}(\chi^2_{(\lambda,J+2)})|\hat{\sigma}^2]][\sigma^2I_J]-\underline{\eta}_1\underline{\eta}_1'E[E[I_{(0,c^*)}(\chi^2_{(\lambda,J+4)})|\hat{\sigma}^2]] & \\ \quad +2E[EI_{(0,c^*)}(\chi^2_{(\lambda,J+2)})|\hat{\sigma}^2]]\underline{\eta}_1\underline{\eta}_1' & 0 \\ 0 & 0 \end{pmatrix}Q(P^{-1})$$

or after taking the expectation over $\hat{\sigma}^2$,

$$
\text{(D.11)} \quad E(\hat{\hat{\underline{\beta}}}-\underline{\beta})(\hat{\hat{\underline{\beta}}}-\underline{\beta})' = \sigma^2 S^{-1}
$$

$$
-P^{-1}Q'\begin{pmatrix} \begin{pmatrix} EI_{(0,\frac{cJ}{T-K})}\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}}\right)[\sigma^2 I_J]-\underline{\eta}_1\underline{\eta}_1' E[I_{(0,\frac{cJ}{T-K})}\left(\frac{\chi^2_{(\lambda,J+4)}}{\chi^2_{(T-K)}}\right)] \\ +2\underline{\eta}_1\underline{\eta}_1' E[I_{(0,\frac{cJ}{T-K})}\left(\frac{\chi^2_{(\lambda,J+2)}}{\chi^2_{(T-K)}}\right)] \end{pmatrix} & 0 \\ 0 & 0 \end{pmatrix} Q(P^{-1})',
$$

Combining equations (D.6) and (D.11), the variance of $\hat{\hat{\underline{\beta}}}$ is

$$
\text{(D.12)} \quad E(\hat{\hat{\underline{\beta}}}-E\hat{\hat{\underline{\beta}}})(\hat{\hat{\underline{\beta}}}-E\hat{\hat{\underline{\beta}}})' = \sigma^2 S^{-1}-P^{-1}Q'\begin{pmatrix} h(2)\sigma^2 I_J+h(4)\underline{\eta}_1\underline{\eta}_1'-2h(2)\underline{\eta}_1\underline{\eta}_1' & 0 \\ 0 & 0_{K-J} \end{pmatrix} Q(P^{-1})'
$$

$$
-P^{-1}Q'\begin{pmatrix} h^2(2)\underline{\eta}_1\underline{\eta}_1' & 0 \\ 0 & 0 \end{pmatrix} Q(P^{-1})',
$$

where $\underline{\eta}_1\underline{\eta}_1'$ and I_J are $(J \times J)$ matrices, 0_{K-J} is a $(K-J) \times (K-J)$ null matrix and the remaining null matrices are of the proper order for conformability in multiplication.

References

Ashar, V.G. (1970): "On the Use of Preliminary Tests in Regression," unpublished thesis, North Carolina State University, Raleigh.

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.

Bancroft, T.A. (1964): "Analysis and Inference for Incompletely Specified Models Involving the Use of Preliminary Test(s) of Significance," Biometrics, Vol. 20, pp. 427-442.

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

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.

Huntsberger, D.V. (1955): "A Generalization of a Preliminary Testing Procedure for Pooling Data," Annals of Mathematical Statistics, Vol. 26, pp. 734-743.

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.

Kitagawa, T. (1963): "Estimation After Preliminary Tests of Significance," University of California Publications in Statistics, Vol. 3, pp. 147-186.

Larson, H.J. and T.A. Bancroft (1963a): "Sequential Model Building for Prediction in Regression Analysis," Annals of Mathematical Statistics, Vol. 34, pp. 462-479.

Larson, H.J. and T.A. Bancroft (1963b): "Biases in Prediction by Regression for Certain Incompletely Specified Models," Biometrika, Vol. 50, pp. 391-402.

Lindgren, B.W. (1968): Statistical Theory, The MacMillan Co., London, p. 380.

Mosteller, F. (1945): "On Pooling Data," Journal of the American Statistical Association, Vol. 43, pp. 231-242.

Rao, C. (1945): "Markoff's Theorem with Linear Restrictions on Parameters," Sankhya, Vol. 7, pp. 16-19.

Sacks, J. (1963): "Generalized Bayes Solution in Estimation Problems," Annals of Mathematical Statistics, Vol. 34, pp. 751-768.

Sawa, T. and T. Hiromatsu (1971): "Minimax Regret Significance Points for a Preliminary Test in Regression Analysis," Technical Report 39, Institute for Mathematical Studies in the Social Sciences, Stanford University.

Sclove, S.L., C. Morris and R. Radhakrishnan (1970): "Non Optimality of Preliminary-Test Estimators for the Multinormal Mean," to appear in Annals of Mathematical Statistics.

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. 351-366.

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. 270-296.

Theil, H. (1963): "On the Use of Incomplete Prior Information in Regression Analysis," Journal of the American Statistical Association, Vol. 58, pp. 401-414.

Theil, H. and A.S. Goldberger (1961): "On Pure and Mixed Regression Estimation in Economics," International Economic Review, Vol. 2, pp. 65-78.

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. 558-572.

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