THE EXACT RISKS OF SOME PRE-TEST AND STEIN-TYPE REGRESSION ESTIMATORS UNDER BALANCED LOSS

Judith A. Giles[1], David E.A. Giles[1] and Kazuhiro Ohtani[2]

[1]Department of Economics, University of Victoria, P.O. Box 3050, Victoria, B.C., CANADA, V8W 3P5
[2]Faculty of Economics, Kobe University, Rokkodai, Nada-ku, Kobe 657, JAPAN

Key words and phrases: linear regression; shrinkage estimators; estimation precision; goodness of fit; conditional inference.

ABSTRACT

In regression analysis we are often interested in using an estimator which is "precise" and which simultaneously provides a model with "good fit". In this paper we consider the risk properties of several estimators of the regression coefficient vector under "balanced" loss. This loss function (Zellner, 1994) reflects both of the described attributes. Under a particular form of balanced loss, we derive the predictive risk of the pre-test estimator which results after a test for exact linear restrictions on the coefficient vector. The corresponding risks of Stein-rule and positive-part Stein-rule estimators are also established. The risks based on loss functions which allow only for estimation precision, or only for goodness of fit, are special cases of our results, and we draw appropriate comparisons. In particular, we show that some of the well-known results under (quadratic) precision-only loss are *not* robust to our generalization of the loss function.

1. INTRODUCTION

In regression analysis we are often concerned both with the precision with which the coefficient vector is estimated, and with the "goodness of fit" of the overall model. However, typically this dual interest has not been taken into account in the construction of estimators in this context, or in the formulation of the loss function when comparing estimators' performances on the basis of "risk".

This has motivated Zellner (1986a, 1994, 1996) to suggest the (quadratic)

"Balanced" Loss Function (BLF), which explicitly takes account of *both* goodness of fit *and* estimation precision. Zellner (1994) analyses several standard estimation problems, including the linear regression model, and develops estimators which are optimal relative to a BLF. Rodrigues and Zellner (1994) extend this analysis by generalizing the loss function still further, and consider a further application. In this paper we use the quadratic BLF as the basis for examining two somewhat related regression problems.

First, we investigate the consequences of adopting a common "pre-testing" strategy when estimating the regression coefficient vector. Specifically, we consider the case where *either* (unrestricted) Ordinary Least Squares (OLS) *or* Restricted Least Squares (RLS) is used, depending on the outcome of a prior test of the validity of certain exact linear restrictions on the coefficients. The BLF risks of the OLS and RLS estimators are compared with that of the Pre-Test Estimator (PTE), and the results are shown to differ from their well known counterparts based on the standard (quadratic) precision-only loss function (*e.g.,* Judge and Bock, 1978).

Second, we consider Stein-Rule (SR) and Positive-part Stein-Rule (PSR) estimators for the regression coefficient vector (*e.g.,* Stein, 1956; James and Stein, 1961; and Baranchik, 1964; Stein, 1966; respectively), using (quadratic) BLF risk as the basis of comparison with OLS estimation. It is found that the standard risk-dominance results that are associated with the usual quadratic precision-only loss generalise naturally to the BLF case, provided that the shrinkage factor is chosen appropriately.

By considering the properties of these regression estimators under the BLF we add to our understanding of the robustness of certain key results to the choice of performance measure. Other related studies which use the BLF include Ohtani *et al.* (1996), Wan (1994), and Cunneen and Giles (1994). Ohtani *et al.* (1996) examine the exact risk of a pre-test estimator for regression coefficients after a pre-test for homoskedasticity, and Wan (1994) considers the risks of the inequality constrained least squares and pre-test estimators. Cunneen and Giles (1994) simulate the risks of pre-test and other estimators in the context of non-independent regression errors. All of these studies find that some of the results under quadratic loss extend to the case of a BLF. Other recent examples of research which focus on robustness of quadratic loss results to the choice of loss function include studies which examine the use of the asymmetric LINEX loss function (Varian, 1974; Zellner; 1986b), of which quadratic loss is a special case. Recent LINEX risk evaluations in the regression context include those of Giles and Giles (1993a, 1996), Ohtani *et al.* (1993) and the references cited therein.

Section 2 introduces the model and notation more formally, and the "pre-test" estimation problem is discussed and analysed in Section 3. Section 4 deals with the SR and PSR estimators, and some general discussion is given in Section 5. Proofs of the main analytic results appear in Appendices A and B.

2. MODEL AND NOTATION

We are concerned with estimating the coefficients in the linear regression model

$$y = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1.1}$$

where $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \boldsymbol{\sigma}^2 I)$, X is nonstochastic, of full rank, and $(T \times k)$, β is $(k \times 1)$ and y and ε are $(T \times 1)$.

If $\bar{\boldsymbol{\beta}}$ is some estimator of β, and $L(\bar{\boldsymbol{\beta}}, \beta)$ is a loss function which takes zero value if $\bar{\boldsymbol{\beta}} = \beta$ and is otherwise strictly positive, then the risk associated with $\bar{\boldsymbol{\beta}}$ is

$$R(\bar{\boldsymbol{\beta}}) = E[L(\bar{\boldsymbol{\beta}}, \boldsymbol{\beta})] \tag{1.2}$$

where the expectation is taken over the sample space.

For example, if we have quadratic loss, then

$$L_Q(\bar{\boldsymbol{\beta}}, \boldsymbol{\beta}) = (\bar{\boldsymbol{\beta}} - \boldsymbol{\beta})'A(\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}) \tag{1.3}$$

where A is a nonstochastic matrix of weights. Choosing $A = S = X'X$, it is well known that the risk associated with using $X\bar{\beta}$ as a *predictor* of E(y|X) is the same as the risk of using $\bar{\beta}$ as an *estimator* of β if the regressors are orthonormal. We shall exploit this idea to make the results below independent of the X data.

The loss function in (1.3) focuses only on estimator (or predictor) *precision*. Zellner's (1994) BLF is a generalization of (1.3) to

$$L_B(\bar{\beta},\beta) = w(y - X\bar{\beta})'(y - X\bar{\beta}) + (1-w)(\bar{\beta} - \beta)'S(\bar{\beta} - \beta) \qquad (1.4)$$

where $w \in [0,1]$ is nonstochastic, and the first term in (1.4) clearly allows for "goodness of fit".

3. PRE-TEST ESTIMATION

Suppose that in estimating (1.1), we are also interested in the possibility that β should be constrained. In particular, suppose that, potentially, β satisfies $R\beta = r$, where R and r are known and nonstochastic, R is $(m\times k)$ and of rank m, and r is $(m\times 1)$. Such restrictions are encountered frequently in applied regression analysis. In particular, this formulation allows for "zero" restrictions on one or more elements of β. So, the estimation of the regression model after a prior test for the significance of regressors is covered by our analysis here.

If the restrictions in question are valid, there is an incentive to impose them and use the Restricted Least Squares estimator (which is also the restricted maximum likelihood estimator) of β. Ignoring them in this case, and using the Ordinary Least Squares (unrestricted maximum likelihood) estimator will result in inefficient estimation. On the other hand, if the restrictions are invalid, their imposition will result in an RLS estimator which is biased, but it may have smaller MSE than OLS (*e.g.*, see Toro-Vizcarrondo and Wallace, 1968). This same problem is discussed in terms of empirical Bayes and hierarchical Bayes estimation by Ghosh *et al.* (1989, 1990).

Clearly, there is an incentive to (pre-) test the hypothesis, H_0: $R\beta = r$ (*vs.* H_A: $R\beta \neq r$) and choose the RLS or OLS estimator according to the outcome of this test.

In the special case of a single zero restriction on one coefficient when k = 2, this problem was first considered by Bancroft (1944), who derived the bias of the least squares pre-test estimator of the other coefficient. The (quadratic) risk properties of the above general pre-test estimator of β are discussed, for example, by Judge and Bock (1978), the choice of an optimal significance level for the pre-test is considered by Brook (1976), and much of the recent pre-testing literature associated with inference in the context of the regression model is surveyed by Giles and Giles (1993b).

A uniformly most powerful invariant size-α test of H_0 against H_A is based on the statistic $\mathbf{f} = [(\mathbf{r}-\mathbf{R}\mathbf{b}_O)'[\mathbf{R}S^{-1}\mathbf{R}']^{-1}(\mathbf{r}-\mathbf{R}\mathbf{b}_O)]/(\mathbf{m}\tilde{\sigma}^2)$, where $\mathbf{b}_O = S^{-1}X'y$ is the OLS estimator of β in (1.1) and $\tilde{\sigma}^2 = (y-X\mathbf{b}_O)'(y-X\mathbf{b}_O)/(T-k)$. The statistic f is noncentral F-distributed with m and ν (=T-k) degrees of freedom, and noncentrality parameter $\lambda = (\mathbf{r}-\mathbf{R}\boldsymbol{\beta})'[\mathbf{R}S^{-1}\mathbf{R}']^{-1}(\mathbf{r}-\mathbf{R}\boldsymbol{\beta})/(2\sigma^2)$. We reject H_0 if f >c, where $\int_0^c \mathbf{p}(\mathbf{f})\mathbf{df} = (1-\alpha)$ and p(f) is central F with m and ν degrees of freedom. The associated pre-test estimator of β is $\mathbf{b}_P = I_{[0,c]}(\mathbf{f})\mathbf{b}_R + I_{(c,\infty)}(\mathbf{f})\mathbf{b}_O = \mathbf{b}_O + (\mathbf{b}_R-\mathbf{b}_O)I_{[0,c]}(\mathbf{f})$, where $\mathbf{b}_R = \mathbf{b}_O + S^{-1}\mathbf{R}'[\mathbf{R}S^{-1}\mathbf{R}']^{-1}(\mathbf{r}-\mathbf{R}\mathbf{b}_O)$ is the RLS estimator of β, and $I_{[a,b]}(f)$ is an indicator function which is unity if $f \in [a,b]$, and zero otherwise.

The risks of b_0, b_R and b_P are well documented under quadratic (precision-only) loss (and Giles, 1993, considers the corresponding problem under absolute error loss). Ghosh *et al.* (1990) derive the properties of certain Bayes estimators for this problem, under quadratic predictive loss. In particular, they show that every pre-test estimator of $X\beta$ is dominated by a corresponding modified empirical Bayes estimator. Our purpose here is to re-consider such pre-test estimation under the more general BLF in (1.5). Under this loss structure, the risks of these three estimators above are given in the following theorem.

Theorem 3.1

Under the quadratic BLF, (1.5), the risks of b_O, b_R and b_P are given by

$$\mathbf{R}(\mathbf{b}_O) = \sigma^2[w\nu + (1-w)k] \tag{3.1}$$

$$R(\mathbf{b}_R) = R(\mathbf{b}_O) + \sigma^2[m(2w-1)+2\lambda] \tag{3.2}$$

$$R(\mathbf{b}_P) = R(\mathbf{b}_O) + \sigma^2[(m(2w-1)+4\lambda(1-w))P_2 + 2\lambda(2w-1)P_4] \tag{3.3}$$

where $\mathbf{P}_i = \Pr[F'_{(m+i,\nu;\lambda)} < (cm/(m+i))];\ i=0,1,...$ and $F'_{(n_1,n_2;\delta)}$ is noncentral F with n_1 and n_2 degrees of freedom, and noncentrality parameter δ.

Proof

See Appendix A.

Clearly, $R(b_R) = R(b_O)$ when $\lambda=\lambda_O=m(1-2w)/2$, and $\lambda_O=0$ if w=0.5. For w> (≥) 0.5, b_O risk-dominates b_R for all $\lambda \geq (>)\ 0$. For w < 0.5, $R(b_R) < R(b_O)$ for $\lambda < \lambda_O$. Further it is clear that b_O risk-dominates b_P and b_P risk-dominates b_R for all $\lambda \geq (>)$ o if w > (≥) 0.5.

When w < 0.5, the situation is ambiguous, as is typified in[1] Figure 1, which corresponds to the traditional quadratic loss situation.[2] In that particular case (w = 0) it is known that $R(b_O) = R(b_P)$ at $\lambda = \lambda_1$, where $\lambda_1 \in [m/4, m/2]$ (*e.g.*, see Judge and Bock , 1973, p. 73). These bounds on the intersection of $R(b_O)$ and $R(b_P)$ are readily generalised when w ∈ [0, 0.5], and a BLF is used.

Corollary 3.1

When w ≤ 0.5, the intersection of $R(b_O)$ and $R(b_P)$ occurs at $\lambda = \lambda^*$, which is bounded by $m(1-2w)/4 \leq \lambda^* \leq m(1-2w)/2$.

Proof

First,

$$\Delta = R(\mathbf{b}_O) - R(\mathbf{b}_P) = \sigma^2\{[m(1-2w)-2\lambda]P_2 + (P_4-P_2)2\lambda(1-2w)\}$$

and as $P_2 > P_4$, a sufficient condition for $\Delta \leq 0$ is $\lambda \geq m(1-2w)/2$, if w ≤ 0.5.

Second,

$$\Delta = \sigma^2\{[m(1-2w)-4\lambda]P_2 + 4\lambda w(P_2-P_4) + 2\lambda P_4\}$$

and as $P_2 > P_4$, a sufficient condition for $\Delta \le 0$ is $\lambda \ge m(1\text{-}2w)/4$, if $w \le 0.5$.

These results are illustrated further in Figures 2 to 4. The well known result associated with this particular pre-test estimation problem, namely that pre-testing can never be the best of the three strategies considered, and may be the worst (as in Figure 1) is clearly negated if goodness of fit is given sufficient weight in the loss function (*i.e.*, if $w \ge 0.5$). Then, pre-testing is always "second-best" to simply applying OLS with no prior testing of H_0, and naively applying the RLS estimator is the worst of the three strategies in terms of BLF risk. It is also clear from these figures that when $w \ge 0.5$ the "optimal" choice of pre-test critical value is $c = 0$, which corresponds to OLS estimation. Also, when $0 < w < 0.5$ the "optimal" choice of c (under a minimax regret criterion) will be less than that suggested by[3] Brook (1976).

The inadmissibility of pre-test strategies is well known, beyond the evidence provided by Ghosh *et al.* (1990). For example, Cohen (1965) demonstrates this, under quadratic loss, by exploiting the fact that such estimators are discontinuous functions of the data. Judge *et al.* (1983) demonstrate that the traditional regression pre-test estimators are dominated by other pre-test strategies which are considered as mixtures of Stein-rule estimators. The inadmissibility of pre-test strategies in general, under the balanced loss function, is not established formally here, but follows intuitively from the arguments of Cohen (1965).

4. STEIN-RULE ESTIMATION

A Stein-rule estimator of β in (1.1) is

$$\mathbf{b}_{SR} = [1-p(y-X\mathbf{b}_O)'(y-X\mathbf{b}_O)/\mathbf{b}_O'S\mathbf{b}_O]\mathbf{b}_O \tag{4.1}$$

where $p \in [0, 2(k-2)/(\nu+2)]$, for $k \geq 3$. It is well known that $R(b_{SR}) < R(b_O)$ under quadratic loss for $k \geq 3$, and that the "optimum" choice of p is $p^* = (k-2)/(\nu+2)$, in terms of maximising the risk-gain in using b_{SR} over b_O. Further, the positive-part Stein-rule estimator,

$$\mathbf{b}_{PSR} = \max[0,\ 1-p(y-X\mathbf{b}_O)'(y-X\mathbf{b}_O)/\mathbf{b}_O'S\mathbf{b}_O]\mathbf{b}_O \tag{4.2}$$

is known to risk-dominate b_{SR} under the same conditions as in the case of quadratic loss, so the latter is inadmissible, even though it is minimax.

To facilitate our analysis of b_{SR} and b_{PSR} under the BLF, it is convenient to define a different pre-test estimator of β, namely

$$\mathbf{b}_T = I_{(a/c^*,\infty)}(f^*)(1-a/f^*)\mathbf{b}_O \tag{4.3}$$

where I is a conventional indicator function; $f^* = [(\mathbf{b}_O'S\mathbf{b}_O)/(k\tilde{\sigma}^2)]$; $a = p\nu/k$; c^* is the critical value for the pre-test of H_0: $X\beta = 0$ *vs* H_A: $X\beta \neq 0$; and $f^* \sim F'_{(k,\nu;\theta)}$, where $\theta = \beta'S\beta/2\sigma^2$. Then,

(i) $b_T = b_O$; if $p = 0$.

(ii) $b_T \to b_{SR}$; if $c^* \to \infty$.

(iii) $b_T = b_{PSR}$; if $c^* = 1$.

Theorem 4.1

Under the quadratic BLF, (1.5), the risk of b_T is given by

$$\begin{aligned} R(b_T) = \sigma^2\{w(\nu+k) + 2\theta + 2\sum_{d=0}^{\infty}\frac{e^{-\theta}\theta^d}{d!}[(1-2w)g_d(0,1) + p^2g_d(2,1) \\ + (1-w)(2p\theta g_{d+1}(1,0) - 2\theta g_{d+1}(0,0) - 2pg_d(1,1))]\} \end{aligned} \tag{4.4}$$

where

$$g_\tau(i,j) = \Gamma(\frac{k}{2}+\tau+j-i)\Gamma(\frac{\nu}{2}+i)/(\Gamma(\frac{k}{2}+\tau)\Gamma(\frac{\nu}{2}))$$

$$\times(1-I_{c_0}(\frac{k}{2}+\tau+j-i,\frac{\nu}{2}+i)) \quad , \quad i,j=0,1,2,...,$$

$I_{c_0}(.,.)$ is the incomplete beta function and $c_0 = p/(p+c^*)$.

Proof

See Appendix B.

Clearly, $R(b_T) = R(b_O)$ when $p = 0$. It is also obvious that when $c^* \to \infty$, $R(b_T) \to R(b_{SR})$, which is given by,

$$R(b_{SR}) = \sigma^2\{\nu w+k(1-w)+p\nu(\nu+2)[p-2(1-w)(k-2)/(\nu+2)]E(1/\chi^{2'}_{(k;\theta)})\} \quad .(4.5)$$

We can write (4.5) as

$$R(b_{SR}) = R(b_O)+\sigma^2 p\nu(\nu+2)[p-2(1-w)(k-2)/(\nu+2)]E(1/\chi^{2'}_{(k;\theta)}) \qquad (4.6)$$

which implies that b_{SR} risk-dominates b_O for all $p \in [0, 2(1-w)(k-2)/(\nu+2)]$, $k \geq 3$. Also, using (4.6) it is straightforward to show that the maximum risk gain in using b_{SR} over b_O occurs when $p^* = (1-w)(k-2)/(\nu+2)$.

When $w = 0$, (4.5) becomes

$$R(b_{SR}|w=0) = \sigma^2\{k+p\nu(\nu+2)[p-2(k-2)/(\nu+2)]E(1/\chi^{2'}_{(k;\theta)})\}$$

which is the usual expression for $R(b_{SR})$ under quadratic loss. (See, for example, Judge and Bock, 1978, p. 188). Finally, when $w=1$, $p^*=0$ then $b_{SR} = b_{PSR} = b_O$. These results are illustrated in Figures 5 and 6 where $p = p^*$. Clearly, if the value of p which is optimal when $w = 0$ is used instead when $w \neq 0$, then there is no guarantee that b_{SR} and b_{PSR} will risk-dominate b_O. This is illustrated in Figures 7 and 8.

5. DISCUSSION

In this paper we have examined the risk properties of several estimators of the coefficient vector in the standard linear regression model under the so-called "balanced" loss function. This loss structure takes account of goodness of fit, as well as estimation precision.

In looking at this estimation problem after a preliminary test for the validity of a set of exact linear restrictions on the regression coefficients, we find that the well known risk results which hold under conventional (precision-only) quadratic loss also hold under a balanced loss function as long as less than 50% weight is given to goodness of fit. On the other hand, if less than 50% weight is given to estimation precision in the loss function, then the strategies of pre-testing or applying the restricted maximum likelihood estimator are strictly dominated by the strategy of simply applying the unrestricted maximum likelihood estimator. This, of course, reflects the fact that here the maximum likelihood and least squares estimators coincide, and the underlying goodness of fit component in the loss function is quadratic in form.

We have also shown that the unrestricted maximum likelihood estimator is itself inadmissible relative to balanced loss, unless all of the weight is given to goodness of fit. This result is established by considering both Stein-rule and positive-part Stein-rule estimators of the coefficient vector (which coincide with the maximum likelihood estimator when estimation precision gets zero weight in the loss function). Our results show that if at least 50% weight (w) is given to goodness of fit then the optimal estimator among those considered in this paper is the positive-part Stein-rule estimator with a shrinkage factor of $p^* = (1-w)(k-2)/(\nu+2)$, where k is the number of regressors, and ν is the degrees of freedom. It is never preferable to pre-test. This is a natural generalization of the corresponding conventional result under quadratic loss. Moreover, our findings tenatively suggest a case for the positive-part estimator when $w < 0.5$.

Clearly, the particular balanced loss function considered here has many interesting applications. The use of a penalty function which allows for both the goodness of fit of the model, as well as the precision of parameter estimation, has considerable appeal in regression analysis. Work in progress by the authors explores some of the associated pre-test issues in various directions.

APPENDIX A: PROOF OF THEOREM 3.1

$$R(\mathbf{b}_O) = wE[(y-X\mathbf{b}_O)'(y-X\mathbf{b}_O)] + (1-w)E[(\mathbf{b}_O-\beta)'S(\mathbf{b}_O-\beta)]$$

where $S = X'X$, $\mathbf{b}_O = S^{-1}X'y$ and $0 \le w \le 1$. Now, $E[(y-X\mathbf{b}_O)'(y-X\mathbf{b}_0)] = \sigma^2\nu$, $\nu = T-k$, as $(y-X\mathbf{b}_O)'(y-X\mathbf{b}_0) = \mathbf{u}'M\mathbf{u}/\sigma^2 \sim \chi^2_{(\nu)}$ where $M = I - XS^{-1}X'$. Further, $E[(\mathbf{b}_O-\beta)'S(\mathbf{b}_O-\beta)] = \sigma^2 k$ (see, for example, Judge and Bock, 1978; Giles and Giles, 1993b). Equation (3.1) then follows directly.

$$R(\mathbf{b}_R) = wE[(y-X\mathbf{b}_R)'(y-X\mathbf{b}_R)] + (1-w)E[(\mathbf{b}_R-\beta)'S(\mathbf{b}_R-\beta)]$$

where $\mathbf{b}_R = \mathbf{b}_O + S^{-1}R'[RS^{-1}R']^{-1}(\mathbf{r}-R\mathbf{b}_O)$. As $(y-X\mathbf{b}_R)'(y-X\mathbf{b}_R)/\sigma^2 \sim \chi^{2'}_{(\nu+\mathbf{m};\lambda)}$ where $\lambda = (R\beta-\mathbf{r})'[RS^{-1}R']^{-1}(R\beta-\mathbf{r})/(2\sigma^2)$, it follows immediately that $E[(y-X\mathbf{b}_R)'(y-X\mathbf{b}_R)] = \sigma^2(\nu+\mathbf{m}+2\lambda)$. It is also straightforward to show that $E[(\mathbf{b}_R-\beta)'S(\mathbf{b}_R-\beta)] = \sigma^2(k-\mathbf{m}+2\lambda)$ (see, for example, Judge and Bock, 1978; Giles and Giles, 1993b) and so $R(b_R)$ follows.

Finally,

$$R(\mathbf{b}_P) = wE[(y-X\mathbf{b}_P)'(y-X\mathbf{b}_P)] + (1-w)E[(\mathbf{b}_P-\beta)'S(\mathbf{b}_P-\beta)] \quad . \qquad (A.1)$$

Let $u_P = y - Xb_P$ and $u_O = y - Xb_O$, so that

$$\begin{aligned} \mathbf{u}_P'\mathbf{u}_P &= \mathbf{u}_O'\mathbf{u}_O - 2\mathbf{u}_O'X(\mathbf{b}_R-\mathbf{b}_O)I_{[0,c]}(f) + (\mathbf{b}_R-\mathbf{b}_O)'S(\mathbf{b}_R-\mathbf{b}_O)I_{[0,c]}(f) \\ &= \mathbf{u}_O'\mathbf{u}_O + (\mathbf{b}_R-\mathbf{b}_O)'S(\mathbf{b}_R-\mathbf{b}_O)I_{[0,c]}(f) \end{aligned}$$

as $\mathbf{u}_O'X = \mathbf{0}$. Then,

$$E[(y - X\mathbf{b}_P)'(y - X\mathbf{b}_P)] = E(\mathbf{u}_P'\mathbf{u}_P)$$

$$= E(\mathbf{u}_O'\mathbf{u}_O) + E[(\mathbf{b}_R - \mathbf{b}_O)'S(\mathbf{b}_R - \mathbf{b}_O)I_{[0,c]}(f)] \quad . \qquad (A.2)$$

Now,

$$E[(\mathbf{b}_R - \mathbf{b}_O)'S(\mathbf{b}_R - \mathbf{b}_O)I_{[0,c]}(f)] = E[I_{[0,c]}(f)(r - R\mathbf{b}_O)'[RS^{-1}R']^{-1}(r - R\mathbf{b}_O)]$$

$$= \sigma^2(mP_2 + 2\lambda P_4) \qquad (A.3)$$

using Lemma 1 of Clarke *et al.* (1987), with $\mathbf{P_i} = \mathbf{Pr}[\mathbf{F}'_{(m+i,\nu;\lambda)} \leq (\mathbf{cm}/(\mathbf{m}+\mathbf{i}))]$ for i=0,1,2,... . Further, from Judge and Bock (1978), for example,

$$E[(\mathbf{b}_P - \boldsymbol{\beta})'S(\mathbf{b}_P - \boldsymbol{\beta})] = \sigma^2[k + (4\lambda - m)P_2 - 2\lambda P_4]. \qquad (A.4)$$

The desired result follows by substituting (A.3) and $E(\mathbf{u}_O'\mathbf{u}_O) = \sigma^2\nu$ into (A.2) and then the subsequent result along with (A.4) into (A.1).

APPENDIX B: PROOF OF THEOREM 4.1

The SR estimator of β is $\mathbf{b}_{SR} = (1 - p\mathbf{u}_O'\mathbf{u}_O/\mathbf{b}_O'S\mathbf{b}_O)\mathbf{b}_O$ where p is a constant such that o ≤ p ≤ 2(1-w)(k-2)/(ν+2) for k ≥ 3, and the PSR estimator is $\mathbf{b}_{PSR} = \max[0, 1 - p\mathbf{u}_O'\mathbf{u}_O/\mathbf{b}_O'S\mathbf{b}_O]\mathbf{b}_O$. Let a = pν/k and $\mathbf{f}^* = (\mathbf{b}_O'S\mathbf{b}_O/k)/(\mathbf{u}_O'\mathbf{u}_O/\nu)$, so that $\mathbf{b}_{PSR} = I_{(a,\infty)}(f^*)(1 - a/f^*)\mathbf{b}_O$, where $I_{(a,\infty)}(f^*) = 0$ if $f^* \leq a$ and 1 otherwise. To derive the risk of the SR and PSR estimators we define a PTE of β, $\mathbf{b}_T = I_{(a/c^*,\infty)}(f^*)(1 - a/f^*)\mathbf{b}_O$ where c^* is the critical value for the pre-test of H_0: Xβ = 0 *vs* H_A: Xβ ≠ 0 (see, for example, Judge and Bock, 1978, p.184; Ohtani, 1993). Then, $f^* \sim F'_{(k,\nu;\theta)}$ where $\boldsymbol{\theta} = \boldsymbol{\beta}'S\boldsymbol{\beta}/2\sigma^2$ and note that

(i) $\mathbf{b}_T = \mathbf{b}_O$ when p = 0,

(ii) $\mathbf{b}_T \rightarrow \mathbf{b}_{SR}$ when $c^* \rightarrow \infty$,

(iii) $\mathbf{b}_T = \mathbf{b}_{PSR}$ when $c^* = 1$.

So,

$$R(\mathbf{b}_T) = wE[(y-X\mathbf{b}_T)'(y-X\mathbf{b}_T)] + (1-w)E[(\mathbf{b}_T-\boldsymbol{\beta})'S(\mathbf{b}_T-\boldsymbol{\beta})]$$
$$= wT_1 + (1-w)T_2 \quad . \tag{B.1}$$

We consider first T_2:

$$T_2 = E[(\mathbf{b}_T-\boldsymbol{\beta})'S(\mathbf{b}_T-\boldsymbol{\beta})] = E(\mathbf{b}_T'S\mathbf{b}_T) - 2\boldsymbol{\beta}'SE(\mathbf{b}_T) + \boldsymbol{\beta}'S\boldsymbol{\beta} \quad . \tag{B.2}$$

Now,

$$\begin{aligned} E(\mathbf{b}_T'S\mathbf{b}_T) &= E[I_{(a/c^*,\infty)}(f^*)\,\mathbf{b}_O'S\mathbf{b}_O - 2pI_{(a/c^*,\infty)}(f^*)(u_O'u_O) \\ &\qquad + p^2 I_{(a/c^*,\infty)}(f^*)(u_O'u_O)^2(\mathbf{b}_O'S\mathbf{b}_O)^{-1}] \\ &= E[I_{(a/c^*,\infty)}(f^*)\mathbf{b}_O'S\mathbf{b}_O - 2pI_{(a/c^*,\infty)}(f^*)\left(\frac{u_O'u_O}{\mathbf{b}_O'S\mathbf{b}_O}\right)(\mathbf{b}_O'S\mathbf{b}_O) \\ &\qquad + p^2 I_{(a/c^*,\infty)}(f^*)\left(\frac{u_O'u_O}{\mathbf{b}_O'S\mathbf{b}_O}\right)^2(\mathbf{b}_O'S\mathbf{b}_O)] \quad . \end{aligned} \tag{B.3}$$

To evaluate this expression further we require the following Lemma:

Lemma 1

For i,j = 0,1,2,... ; k > 2(i-j),

$$E[I_{(a/c^*,\infty)}(f^*)(u_O'u_O/\mathbf{b}_O'S\mathbf{b}_O)^i(\mathbf{b}_O'S\mathbf{b}_O)^j] = (2\sigma^2)^j\sum_{d=0}^{\infty}\frac{e^{-\theta}\theta^d}{d!}\,g_d(i,j)$$

where

$$g_d(i,j) = \frac{\Gamma(\frac{k}{2}+d+j-i)\Gamma(\frac{\nu}{2}+i)}{\Gamma(\frac{k}{2}+d)\Gamma(\frac{\nu}{2})}\,[1-I_x(\frac{k}{2}+d+j-i,\frac{\nu}{2}+i)] \quad ,$$

$x = p/(p+c^*)$ and I(.,.) is the incomplete beta function.

Proof

Let $\gamma = u_O'u_O/\sigma^2 \sim \chi^2_{(\nu)}$ and $\tau = b_O'Sb_O/\sigma^2 \sim \chi^{2'}_{(k;\theta)}$. Then,

$$
\begin{aligned}
E[I_{(a/c^*,\infty)}(f^*)(u_O'u_O/b_O'Sb_O)^i(b_O'Sb_O)^j] &= E[I_{(a/c^*,\infty)}(f^*)(\gamma/\tau)^i(\tau\sigma^2)^j] \\
&= E[I_{(p/c^*,\infty)}(\tau/\gamma)(\gamma/\tau)^i(\tau\sigma^2)^j] \\
&= \sum_{d=0}^{\infty}\frac{e^{-\theta}\theta^d}{d!}(\sigma^2)^j[2^{(k+\nu)/2+d}\Gamma(\frac{k}{2}+d)\Gamma(\frac{\nu}{2})]^{-1} \\
&\quad\times\int\int_{(\tau/\gamma)>(p/c^*)}\gamma^{\nu/2+i-1}\tau^{j-i+k/2+d-1}\exp[-(\gamma+\tau)/2]d\gamma\, d\tau \quad, \qquad \text{(B.4)}
\end{aligned}
$$

as γ and τ are independently distributed quadratic forms.

Let $A_d = e^{-\theta}\theta^d(\sigma^2)^j[2^{(k+\nu)/2+d}\,\Gamma(\frac{k}{2}+d)\Gamma(\frac{\nu}{2})d!]^{-1}$, $z_1 = (\tau/\gamma)$ and $z_2 = \gamma$, so that (B.4) is

$$
\begin{aligned}
&\sum_{d=0}^{\infty} A_d\int_{p/c^*}^{\infty}\int_0^{\infty} z_2^{\nu/2+i-1}\,(z_1z_2)^{k/2+d-1+j-i}\,\exp[-(z_1z_2+z_2)/2]z_2dz_2dz_1 \\
&= \sum_{d=0}^{\infty} A_d\int_{p/c^*}^{\infty}\int_0^{\infty} z_2^{(k+\nu)/2+j+d-1}z_1^{k/2+d-1+j-i}\,\exp[-(1+z_1)z_2/2]dz_2dz_1 \quad. \qquad \text{(B.5)}
\end{aligned}
$$

Now, let $z_3 = (1+z_1)z_2/2$ and $z_1 = z_1$. Then (B.5) is

$$
\begin{aligned}
&\sum_{d=0}^{\infty} A_d\int_{p/c^*}^{\infty} z_1^{k/2+d-1+j-i}(1+z_1)^{-(k+\nu)/2+j+d}2^{(k+\nu)/2+j+d}dz_1\int_0^{\infty} z_3^{(k+\nu)/2+j+d-1}\exp(-z_3)dz_3 \\
&=\sum_{d=0}^{\infty} A_d\Gamma(\frac{\nu+k}{2}+j+d)\int_{p/c^*}^{\infty} z_1^{k/2+d-1+j-i}\left(\frac{2}{1+z_1}\right)^{(k+\nu)/2+j+d}dz_1 \quad. \qquad \text{(B.6)}
\end{aligned}
$$

Finally, let $z_4 = z_1/(1+z_1)$ and $x = p/(p+c^*)$, so that (B.6) is

$$\sum_{d=0}^{\infty} A_d \Gamma(\frac{\nu+k}{2}+j+d)\, 2^{(k+\nu)/2+j+d} \int_x^1 z_4^{k/2+d+j-i-1}(1-z_4)^{\nu/2+i-1} dz_4$$

$$=\sum_{d=0}^{\infty} A_d \Gamma(\frac{\nu+k}{2}+j+d)\, 2^{(k+\nu)/2+j+d} B\left(\frac{k}{2}+d+j-i, \frac{\nu}{2}+i\right)\left(1-I_x\left(\frac{k}{2}+d+j-i, \frac{\nu}{2}+i\right)\right)$$

where B(.,.) is the beta function. The lemma then follows directly.

Applying Lemma 1 repeatedly to (B.3) we have

$$E(b_T' S b_T) = 2\sigma^2 \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} \left(g_d(0,1) - 2p g_d(1,1) + p^2 g_d(2,1)\right) \quad . \qquad \text{(B.7)}$$

To complete the expression for T_2 we need $S \times E(b_T)$.

$$\begin{aligned} S\times E(b_T) &= S\times E[I_{(a/c^*,\infty)}(f^*) b_O - I_{(a/c^*,\infty)}(f^*)(a/f^*) b_O] \\ &= S \iint_{\bar{R}} b_O p(b_O) p(u_O' u_O) db_O d(u_O' u_O) \\ &\quad - Sp \iint_{\bar{R}} b_O (u_O' u_O / b_O' S b_O) p(b_O) p(u_O' u_O) db_O d(u_O' u_O) \end{aligned} \qquad \text{(B.8)}$$

where $\bar{R}$ is the region defined by $\bar{R} = [(b_O, u_O' u_O) \mid b_O' S b_O / u_O' u_O > p/c^*]$ and $p(b_O)$ and $p(u_O' u_O)$ are the probability density functions of b_O and $u_O' u_O$ respectively. We require the following Lemma to proceed further.

Lemma 2

Let Q(i) be defined by

$$Q(i) = \iint_{\bar{R}} b_O (u_O' u_O / b_O' S b_O)^i p(b_O) p(u_O' u_O) db_O d(u_O' u_O) \quad , \ i = 0,1,2,\ldots \ .$$

Then

$$SQ(i) = S\beta \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_{d+1}(i,0) \ .$$

Proof

Let

$$G(i) = E[I_{(a/c^*,\infty)}(f^*)(u_O'u_O/b_O'Sb_O)^i]$$

$$= \int\int_{\bar{R}} (u_O'u_O/b_O'Sb_O)^i p(b_O)p(u_O'u_O)db_O d(u_O'u_O)$$

$$= \int\int_{\bar{R}} (u_O'u_O/b_O'Sb_O)^i [(2\pi)^{k/2}|\sigma^2 S^{-1}|^{0.5}]^{-1} \exp[-(b_O-\beta)'S(b_O-\beta)/(2\sigma^2)]$$

$$\times p(u_O'u_O)\, d(b_O) d(u_O'u_O)$$

as $b_O \sim N(\beta, \sigma^2 S^{-1})$. Then,

$$\frac{\partial G(i)}{\partial \beta} = \int\int_{\bar{R}} (u_O'u_O/b_O'Sb_O)^i [(2\pi)^{k/2}|\sigma^2 S^{-1}|^{0.5}]^{-1} [(Sb_O/\sigma^2)$$

$$\times \exp\left(-\frac{1}{2\sigma^2}(b_O-\beta)'S(b_O-\beta)\right)$$

$$-(S\beta/\sigma^2)\exp\left(-\frac{1}{2\sigma^2}(b_O-\beta)'S(b_O-\beta)\right)]\, p(u_O'u_O)\, db_O d(u_O'u_O)$$

$$= (S/\sigma^2)\int\int_{\bar{R}} b_O(u_O'u_O/b_O'Sb_O)^i p(b_O)\, p(u_O'u_O)\, db_O\, d(u_O'u_O) - (S\beta/\sigma^2)G(i)$$

$$= (S/\sigma^2)Q(i) - (S\beta/\sigma^2)G(i) \quad . \tag{B.9}$$

Further,

$$G(i) = \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_d(i,0) \tag{B.10}$$

and $\partial\theta/\partial\beta = S\beta/\sigma^2$ so that using (B.10) we have

$$\frac{\partial G(i)}{\partial \beta} = -(S\beta/\sigma^2)\sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_d(i,0) + (S\beta/\sigma^2)\sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_{d+1}(i,0)$$

$$= -(S\beta/\sigma^2)G(i) + (S\beta/\sigma^2)\sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_{d+1}(i,0) \quad . \tag{B.11}$$

Equating (B.9) and (B.11) gives the required result.

Using Lemma 2 with i=0 and i=1 in (B.8) we have

$$S\times E(\mathbf{b}_T) = S\beta \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_{d+1}(0,0) - pS\beta \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_{d+1}(1,0)$$

so that

$$\begin{aligned} T_2 &= 2\sigma^2 \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} [g_d(0,1) - 2pg_d(1,1) + p^2 g_d(2,1)] \\ &\qquad -2\beta' S\beta \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} [g_{d+1}(0,0) - pg_{d+1}(1,0)] + \beta' S\beta \\ &= 2\sigma^2 \ \Big(\theta + \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} [g_d(0,1) - 2pg_d(1,1) + p^2 g_d(2,1) - 2\theta g_{d+1}(0,0)] \\ &\qquad + 2\theta p g_{d+1}(1,0)] \ \Big) \quad . \end{aligned} \tag{B.12}$$

Finally, we require

$$\begin{aligned} T_1 &= E[(y - X\mathbf{b}_T)'(y - X\mathbf{b}_T)] \\ &= E(\mathbf{u}_O'\mathbf{u}_O) + E(\mathbf{b}_O' S\mathbf{b}_O) - E(\mathbf{b}_O' S\mathbf{b}_O I_{(a/c^*,\infty)}(f^*)) \\ &\qquad + E(\mathbf{b}_O' S\mathbf{b}_O I_{(a/c^*,\infty)}(f^*)(a^2/f^{*2})) \quad . \end{aligned} \tag{B.13}$$

As $(\mathbf{u}_O'\mathbf{u}_O/\sigma^2) \sim \chi^2_{(v)}$ and $(\mathbf{b}_O' S\mathbf{b}_O/\sigma^2) \sim \chi^{2'}_{(k;\theta)}$, it follows that $E(\mathbf{u}_O'\mathbf{u}_O) = \sigma^2 v$ and $E(\mathbf{b}_O' S\mathbf{b}_O) = \sigma^2(k+2\theta)$. Further, using Lemma 1

$$E[\mathbf{b}_O' S\mathbf{b}_O I_{(a/c^*,\infty)}(f^*)] = 2\sigma^2 \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_d(0,1)$$

and

$$E[\mathbf{b}_O' S\mathbf{b}_O I_{(a/c^*,\infty)}(f^*)(a^2/f^{*2})] = 2\sigma^2 p^2 \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} g_d(2,1)$$

so that (B.13) becomes

$$T_1 = \sigma^2 \left(\nu + k + 2\theta - 2 \sum_{d=0}^{\infty} \frac{e^{-\theta}\theta^d}{d!} [g_d(0,1) - p^2 g_d(2,1)] \right) . \qquad \text{(B.14)}$$

Substituting (B.12) and (B.14) into (B.1) and rearranging terms gives the required result.

NOTES

1. The evaluations for the figures in this paper were undertaken with FORTRAN code written by the authors, incorporating routines from Press *et al.* (1986), and executed on a VAXstation 4000 under VMS 5.5. In all of the figures $\sigma^2 = 1$, which results in no loss of generality.

2. In Figures 1-4, two choices of c are considered: $c = 4.35$ ($\alpha = 0.05$), and $c = 1.895$ ($\alpha = 0.185$). The latter choice is Brook's (1976) "optimal" c in the case of quadratic loss.

3. That is, regardless of the value of m or ν, the "optimal" value of c will lie between zero and two in value, approximately.

ACKNOWLEDGEMENTS

The authors thank the referee for helpful suggestions. This work, undertaken while the first two authors were members of the Department of Economics, University of Canterbury, was supported in part by University of Canterbury Research Grants #1770901, #1787148, and #1894183. We are grateful to Matthew Cunneen for his assistance.

BIBLIOGRAPHY

Bancroft, T.A. (1944). "On biases in estimation due to the use of preliminary tests of significance," *Annals of Mathematical Statistics*, 15, 190-204.

Baranchik, A.J. (1964). "Multiple regression and estimation of the mean of a multivariate normal distribution," Technical Report No. 51, Department of Statistics, Stanford University.

Brook, R.J. (1976). "On the use of a regret function to set significance points in prior tests of estimation," *Journal of the American Statistical Association*, 71, 126-131.

Clarke, J.A., Giles, D.E.A. and Wallace, T.D. (1987). "Estimating the error variance in regression after a preliminary test of restrictions on the coefficients," *Journal of Econometrics*, 34, 293-304.

Cohen, A. (1965). "Estimates of the linear combinations of parameters in the mean vector of a multivariate distribution," *Annals of Mathematical Statistics*, 36, 299-304.

Cunneen, M.C. and Giles, J.A. (1994). "Autocorrelation pre-testing using a balanced loss function," mimeo., Department of Economics, University of Victoria.

Ghosh, M., Saleh, A.K.Md.E. and Sen, P.K. (1989). "Empirical Bayes subset estimation in regression models," *Statistics and Decisions*, 7, 15-35.

Ghosh, M., Lee, L.C. and Littell, R.C. (1990). "Empirical and hierarchical Bayes estimation of the response function," *Statistics and Decisions*, 8, 299-330.

Giles, D.E.A. (1993). "Pre-test estimation in regression under absolute error loss," *Economics Letters*, 41, 339-343.

Giles, J.A. and Giles, D.E.A. (1993a). "Preliminary-test estimation of the regression scale parameter when the loss function is asymmetric," *Communications in Statistics: Theory and Methods*, A22, 1709-1734.

Giles, J.A. and Giles, D.E.A. (1993b). "Pre-test estimation and testing in econometrics: Recent developments," *Journal of Economic Surveys*, 7, 145-196.

Giles, J.A. and Giles, D.E.A. (1996). "Risk of a homoskedasticity pre-test estimator of the regression scale under LINEX loss," *Journal of Statistical Planning* *and Inference*, 50, 21-35.

James, W. and Stein, C. (1961). "Estimation with quadratic loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability,* Berkeley: University of California Press, 361-379.

Judge, G.G. and Bock, M.E. (1978). *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, Amsterdam: North-Holland.

Judge, G.G., Yancey, T.A. and Bock, M.E. (1983). "Pre-test estimation under squared error loss," *Economics Letters*, 11, 347-352.

Ohtani, K. (1993). "A comparison of the Stein-rule and positive part Stein-rule estimators in a misspecified linear regression model," *Econometric Theory,* 9, 668-679..

Ohtani, K., Giles, D.E.A. and Giles, J.A. (1993). "The risk behaviour of a pre-test estimator in a regression model with possible heteroskedasticity under the LINEX loss function," mimeo., Faculty of Economics, Kobe University.

Ohtani, K., Giles, D.E.A. and Giles, J.A. (1996). "The exact risk performance of a pre-test estimator in a heteroskedastic linear regression model under the balanced loss function," *Econometric Reviews*, forthcoming.

Press, W.H., Flannery, B.P., Teukolsky, S.A. and Vetterling, W.T. (1986). *Numerical Recipes: The Art of Scientific Computing*, Cambridge: Cambridge University Press.

Rodrigues, J. and Zellner, A. (1994). "Weighted balanced loss function and estimation of the mean time to failure," *Communications in Statistics: Theory and Methods*, A23, 3609-3612.

Stein, C. (1956). "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1,* Berkeley: University of California Press, 197-206.

Stein, C. (1966). "An approach to the recovery of inter-block information on balanced incomplete block designs," *Research Papers in Statistics*, (F. N. David, Ed.). New York: Wiley, 351-366.

Toro-Vizcarrondo, C. and Wallace, T.D. (1968). "A test of the mean square error criterion for restrictions in linear regression," *Journal of the American*

Statistical Association, 63, 558-572.

Varian, H. (1974). "A Bayesian approach to real estate assessment," *Studies in Bayesian Econometrics and Statistics in Honor of Leonard J. Savage*, *Vol. 1,* Amsterdam: North-Holland, 195-208.

Wan, A.T.K. (1994). "Risk comparisons of the inequality constrained least squares and other related estimators under balanced loss," *Economics Letters*, 46, 203-210.

Zellner, A. (1986a). "Further results on Bayesian minimum expected loss (MELO) estimates and posterior distributions for structural coefficients," *Advances in Econometrics, Vol. 5,* (D. Slottje, Ed.). Greenwich CT: JAI Press, 171-182.

Zellner, A. (1986b). "Bayesian estimation and prediction using asymmetric loss functions," *Journal of the American Statistical Association*, 81, 446-451.

Zellner, A. (1994). "Bayesian and non-Bayesian estimation using balanced loss functions," *Statistical Decision Theory and Related Topics, V,* (S.S. Gupta and J.O. Berger, Eds.). New York: Springer-Verlag, 377-390.

Zellner, A. (1996). "The finite sample properties of simultaneous equations' estimates and estimators: Bayesian and non-Bayesian approaches," *Journal of Econometrics,* forthcoming.

Offending Command = restore
Error = nametype : invalidrestore

Stack =
--nostringval--
--nostringval--