

## OPTIMAL CRITICAL VALUES FOR PRE-TESTING IN REGRESSION

BY T. TOYODA AND T. D. WALLACE

In this paper we derive and present optimal critical points for pre-tests in regression using a minimum average relative risk criterion. We use the same type risk functions as Sawa and Hiromatsu [8] who, in a recent paper in this journal, derived pre-test critical values using a minimax regret criterion. Since James-Stein type estimators can be shown to dominate any pre-test estimator for the risk functions used here and in [8], no normative claims are made for the critical values we give. However, the use of pre-testing procedures continues in practice and the results given here, contrasted with other results, add to information about the character of costs and returns to such practices.

### 1. INTRODUCTION

RECENTLY, JUDGE and his associates have derived risk functions and matrices for pre-test estimators in different contexts, and have called attention to the dominance of James-Stein type estimators where continuous weight functions of the tests statistic are used to "average" the restricted and unrestricted estimators (see [1, 2, and 7]). Sawa and Hiromatsu [8] obtained minimax regret critical values for pre-test estimators where they define regret as the difference between pre-test risk and the minimum of risk for unrestricted and restricted least squares over the range of the noncentrality parameter in the probability density function on the test statistic. They tabulated their minimax regret critical value for the case of a single restriction. Brook [3], with somewhat different scalar risk functions than Sawa and Hiromatsu, derived minimax regret critical values for multiple restrictions, and examined the effect of non-orthogonal  $X$  matrices for risk functions that concentrate on the parameter space of the regression coefficients themselves rather than on conditional forecasting of the dependent variable.

In contrast to Sawa and Hiromatsu and Brook, the pre-test critical values that we derive are more favorable to unrestricted least squares. For example, using the same risk functions as Sawa and Hiromatsu or Brook, minimum average risk of the pre-test estimator leads to opting for ordinary, unrestricted least squares unless the number of restrictions is five or greater, no matter the (finite) number of regressors and observations. That is, our "optimal" critical value for a pre-test is zero unless the restrictions imply a reduction of five or more in the parameter space. In contrast, even with only one potential restriction, Sawa and Hiromatsu find that the minimax critical value for a pre-test is about 1.8, depending upon degrees of freedom. However, when the number of restrictions is very large, our results are very close to the findings by Brook. That is, optimal critical values are about 2.0 for a large number of potential restrictions for either the minimax regret or the relative minimum average risk criteria.

2. RISK FUNCTIONS FOR PRE-TEST ESTIMATORS

Consider a general linear model

$$(1) \quad Y = X\beta + \varepsilon; \quad \varepsilon \sim N(0, \sigma^2 I),$$

where  $Y$  is, say,  $T \times 1$ ,  $X$  is  $T \times k$ ,  $\beta$  is  $k \times 1$ , and  $\varepsilon$  is  $T \times 1$ . Usual assumptions hold about observability of  $Y$ ,  $X$ , non-randomness of  $X$ , etc., and we assume that the rank of  $X$  is  $k$ .

In choosing an estimator for  $\beta$ , the investigator may have some set of restrictions in mind, such as potential equality of a subset of the coefficients across two data regimes.

Write the potential restrictions as

$$(2) \quad H'\beta = h$$

where  $H'$  is  $m \times k$  with rank  $m$  and  $h$  is  $m \times 1$ . Both  $H'$  and  $h$  are explicit.

Consider a pre-test estimator for  $E(Y|X)$ :

$$(3) \quad X\beta^* = \begin{cases} Xb & \text{if } u \geq \lambda, \\ X\hat{\beta} & \text{if } u < \lambda, \end{cases}$$

where  $b$  is the unrestricted least squares estimator for  $\beta$ ,  $\hat{\beta}$  is the restricted least squares estimator,  $\lambda$  is a critical value for accepting or rejecting some hypothesis related to the restriction in equation (2) and  $u$  is the calculated value of a test statistic. That is,

$$(4) \quad u = \frac{(H'b - h)(H'S^{-1}H)^{-1}(H'b - h)}{m\hat{\sigma}^2}$$

where  $S = X'X$  and  $\hat{\sigma}^2$  is an unbiased estimator of the regression variance in the unrestricted case, and  $u$  is distributed as the noncentral  $F$  with  $m$  and  $T - k$  degrees of freedom and noncentrality  $\theta$ , where

$$(5) \quad \theta = \frac{(H'\beta - h)(H'S^{-1}H)^{-1}(H'\beta - h)}{2\sigma^2}.$$

Under a null hypothesis that  $\theta = 0$ ,  $\lambda$  can be thought of as a critical value for the central  $F$ .<sup>1</sup>

A risk function for the pre-test estimator is defined as

$$(6) \quad R(X\beta^*; m, T - k, \lambda : \theta) = E(X\beta^* - X\beta)'(X\beta^* - X\beta) \\ = E(\beta^* - \beta)'X'X(\beta^* - \beta).$$

By concentrating on the conditional forecast of  $Y$  as the parameter space rather than  $\beta$  itself, risk defined by equation (6) becomes a function only of degrees of freedom, noncentrality, and the choice of pre-test critical value.<sup>2</sup>

<sup>1</sup> Other null hypotheses would yield other definitions of  $\lambda$ . For this paper, the null is immaterial and only has residual implication for type one errors (see Table I).

<sup>2</sup> See Brook [3, pp. 15-20]. Perhaps an intuitively more appropriate concern would be with  $E(\beta^* - \beta)'(\beta^* - \beta)$ , which, however, is not scale invariant to measuring variables in  $X$ —that is, risk is then a function of  $X'X$ .

First, note that for  $\lambda = 0$  one would always choose unrestricted least squares since  $u$  is nonnegative and for  $\lambda = +\infty$ ,  $X\beta^*$  become  $X\hat{\beta}$  with probability one. Therefore, from Wallace [13],

$$(7) \quad R(X\beta^*; m, T - k, 0 : \theta) = R(Xb) = \sigma^2 k$$

and

$$(8) \quad R(X\beta^*; m, T - k, +\infty : \theta) = R(X\hat{\beta}) = \sigma^2 [k - m + 2\theta].$$

For all  $\lambda$ , Brook [3] derived the risk function and found

$$(9) \quad R(X\beta^*; m, T - k, \lambda : \theta) = \sigma^2 \{k - m + mr(\theta, \lambda) + 2\theta[1 - 2r(\theta, \lambda) + s(\theta, \lambda)]\},$$

where

$$(10) \quad r(\theta, \lambda) = P \left\{ F'(m + 2, T - k; \theta) \geq \frac{m\lambda}{m + 2} \right\}$$

and

$$(11) \quad s(\theta, \lambda) = P \left\{ F'(m + 4, T - k; \theta) \geq \frac{m\lambda}{m + 4} \right\}.$$

The letter,  $P$ , denotes probability, and  $F'$  has the non-central  $F$  density. Thus,  $r$  and  $s$  are both integrals, bounded by zero and one. Note that both  $r$  and  $s$  go to zero as  $\lambda$  increases (for fixed, finite degrees of freedom and  $\theta$ ) and to one as  $\lambda$  goes to zero, confirming equations (7) and (8).

Figure 1 depicts three risk functions; two for the extreme choice of  $\lambda$  and one for  $\lambda \in (0, +\infty)$ .  $R(X\hat{\beta})$  and  $R(Xb)$  are, of course, independent of  $\lambda$  since  $\lambda$  is pre-selected at  $+\infty$  and zero, respectively, to yield those functions.  $R(X\beta^*)$  for  $\lambda$  in the open half-line graphs differently as different  $\lambda$  are chosen.

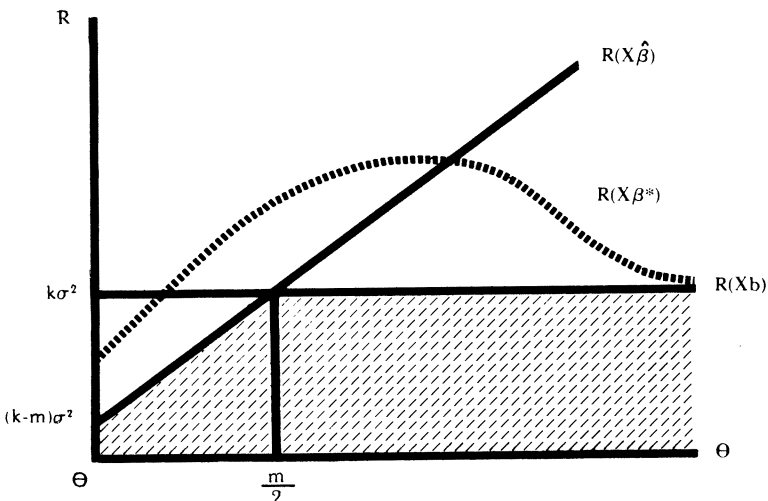


FIGURE 1—Risk functions for pre-test estimator.

Brook and, in a somewhat more limited context, Sawa and Hiromatsu sought to choose values of  $\lambda$  that minimized maximum regret given no particular prior knowledge of  $\theta$ . This involved choosing a pre-test critical value to balance maximum regret for  $\theta \in [0, m/2]$  for  $R(X\beta^*)$  versus  $R(X\hat{\beta})$  with maximum regret for

$$\theta \in [m/2, +\infty)$$

as between  $R(X\beta^*)$  and  $R(Xb)$ .

The best that one can do by least squares is indicated by the upper boundary of the shaded area in Figure 1.<sup>3</sup> Optimal choice of  $\lambda$  for  $\theta$  not known consists of choosing a value of  $\lambda$  that pulls  $R(X\beta^*)$  as close to the bound as possible. Thus, an alternative to the Sawa and Hiromatsu criterion with a diffuse prior on  $\theta$  is to choose  $\lambda$  to minimize the area bounded above by  $R(X\beta^*)$  and below by  $R(X\hat{\beta})$  for  $\theta \in [0, m/2)$  and  $R(Xb)$  for  $\theta \in [m/2, +\infty]$ .<sup>4</sup>

### 3. MINIMUM AVERAGE RISK CRITERION AND OPTIMAL SOLUTION

We define the relative risk of  $X\beta^*$  with respect to the always unrestricted and the always restricted estimators as

$$(12) \quad \{R(X\beta^*) - \min [R(Xb), R(X\hat{\beta})]\} / \sigma^2.$$

There may be several alternative ways of choosing optimal  $\lambda$  which minimizes the risk (9) in some sense, including a minimax, a minimax regret, and a Bayesian minimum expected loss approaches. The decision criterion adopted here is a simple but reasonable one under innocence of any a priori knowledge of skewness of the distribution of  $\theta$ . That is, we want to minimize the average relative risk over the whole range of the nuisance parameter  $\theta$ .<sup>5</sup> This means that we want to minimize

$$(13) \quad G(\lambda) = \int_0^\infty \{R(X\beta^*) - \min [R(Xb), R(X\hat{\beta})]\} d\theta / \sigma^2 \\ = -m \int_0^\infty \{1 - r(\theta, \lambda)\} d\theta + 2 \int_0^\infty \theta \{2(1 - r(\theta, \lambda)) \\ - (1 - s(\theta, \lambda))\} d\theta + m \int_0^{m/2} d\theta - 2 \int_0^{m/2} \theta d\theta.$$

The convergence of the improper integral  $G(\lambda)$  for the range  $\lambda \in [0, +\infty)$  is proved in Appendix 1. From Figure 1 it is clear that  $G(+\infty)$  does not converge; i.e., that

<sup>3</sup> We recognize that James-Stein type estimators exist that can "clip a corner" of the set represented by the shaded area in Figure 1. See, e.g., Sclove, et al. [9], Judge, et al. [7], and Bock, et al. [1].

<sup>4</sup> As noted by a referee, the shaded area in Figure 1 is not a function of the critical value; hence, minimizing the area between  $R(X\beta^*)$  and the shaded area is, in a sense, the same as minimizing the total area below  $R(X\beta^*)$ . Subtracting the shaded area has the convenient virtue of making the resultant integral bounded.

<sup>5</sup> Essentially the same criterion has been employed in optimal pre-test procedures for situations involving uncertainty in model specifications by several authors; e.g. by Huntsburger [6] in estimating a mean and by Toyoda and Wallace [11] in estimating variance.

$\lambda = \infty$  does not yield a minimum so that in what follows we concentrate on the range  $\lambda \in [0, +\infty)$ .

Let  $f(F'; m^*, T - k, \theta)$  be  $(m^*/(T - k))f'(F'; m^*, T - k, \theta)$  where  $f'$  is the non-central  $F$  density with  $m^*, T - k$  degrees of freedom and noncentrality  $\theta$ , and  ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$  represent the Gauss hypergeometric function. Then, the necessary condition for minimizing  $G(\lambda)$  is the following.<sup>6</sup>

$$\begin{aligned}
 (14) \quad \frac{dG(\lambda)}{d\lambda} &= -\left(\frac{m}{T-k}\right) \left[ m \int_0^\infty f\left(\frac{m\lambda}{t-k}; m+2, T-k, \theta\right) d\theta \right. \\
 &\quad - 4 \int_0^\infty \theta f\left(\frac{m\lambda}{T-k}; m+2, T-k, \theta\right) d\theta \\
 &\quad \left. + 2 \int_0^\infty \theta f\left(\frac{m\lambda}{T-k}; m+4, T-k, \theta\right) d\theta \right] \\
 &= -\frac{m}{T-k} \left[ m \sum_{i=0}^\infty \frac{\{m\lambda/(T-k)\}^{m/2+i}}{i! B\left(\frac{m}{2} + 1 + i, \frac{T-k}{2}\right) \{1 + m\lambda/(T-k)\}^{(m+T-k)/2+1+i}} \right. \\
 &\quad \times \int_0^\infty \theta^i e^{-\theta} d\theta \\
 &\quad - 4 \sum_{i=0}^\infty \frac{\{m\lambda/(T-k)\}^{m/2+i}}{i! B\left(\frac{m}{2} + 1 + i, \frac{T-k}{2}\right) \{1 + m\lambda/(T-k)\}^{(m+T-k)/2+1+i}} \int_0^\infty \theta^{i+1} e^{-\theta} d\theta \\
 &\quad \left. + 2 \sum_{i=0}^\infty \frac{\{m\lambda/(T-k)\}^{m/2+1+i}}{i! B\left(\frac{m}{2} + 2 + i, \frac{T-k}{2}\right) \{1 + m\lambda/(T-k)\}^{(m+T-k)/2+2+i}} \int_0^\infty \theta^{i+1} e^{-\theta} d\theta \right] \\
 &= \frac{m\Gamma\left(\frac{m+T-k}{2} + 1\right) \{m\lambda/(T-k)\}^{m/2}}{(T-k)\Gamma\left(\frac{T-k}{2}\right)\Gamma\left(\frac{m}{2} + 1\right) \{1 + m\lambda/(T-k)\}^{(m+T-k)/2+1}} \\
 &\quad \times \left[ m {}_2F_1\left(1, \frac{m+T-k}{2} + 1; \frac{m}{2} + 1; \frac{m\lambda}{m\lambda + T-k}\right) \right. \\
 &\quad - 4 {}_2F_1\left(2, \frac{m+T-k}{2} + 1; \frac{m}{2} + 1; \frac{m\lambda}{m\lambda + T-k}\right) \\
 &\quad \left. + \frac{2m(m+T-k+2)\lambda}{(m+2)(m\lambda + T-k)} {}_2F_1\left(2, \frac{m+T-k}{2} + 2; \frac{m}{2} + 2; \frac{m\lambda}{m\lambda + T-k}\right) \right] \\
 &= 0.
 \end{aligned}$$

<sup>6</sup> As the integrand of  $G(\lambda)$  and its derivative with respect to  $\lambda$  are continuous, we can interchange the order of differentiation and integration.

That is, the minimum is attained when

$$(15) \quad \lambda = 0$$

or

$$(16) \quad [ \ ] = 0.$$

Using some properties of the hypergeometric function, the inside of the bracket of (16) reduces to

$$(17) \quad y = \phi(y)$$

where

$$(18) \quad \phi(y) = \frac{m - 4}{2} \frac{{}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right)}{d\left[{}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right)\right] / dy}$$

$$= \frac{mI_y\left(\frac{m}{2} + 1, \frac{T - k}{2} + 1\right) + (m - 4)I_y\left(\frac{m}{2}, \frac{T - k}{2} + 1\right)}{(T - k + 2m - 2)I_y\left(\frac{m}{2}, \frac{T - k}{2} + 1\right)}$$

and

$$(19) \quad y \equiv m\lambda / (m\lambda + T - k),$$

provided that  $\lambda \neq 0$  (i.e.,  $y \neq 0$ ) and  $m \geq 5$ . Here,  $I_y(\cdot, \cdot)$  denotes the incomplete beta function ratio; i.e.,

$$I_y(\alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^y t^{\alpha-1}(1 - t)^{\beta-1} dt.$$

It is also seen that the optimal value of  $\lambda$  for  $m \leq 4$  is

$$(20) \quad \lambda = 0 \quad (\text{i.e., } y = 0).$$

Proofs leading to the results (18) and (20) are given in Appendix 2.

The solutions of equation (17) are fixed points. It is intractable to express the fixed points as simple functions of the numerator and denominator degrees of freedom. Therefore, using an iterative search procedure, we have computed the fixed point solutions, differing from zero, to three decimal places for various selected degrees of freedom. By calculating the corresponding values of  $G(\lambda)$  and comparing them with  $G(0)$ , we have found that the minimum of  $G(\lambda)$  is attained when  $\lambda \neq 0$  for  $m \geq 5$ .<sup>7</sup> The results of the minimum values of  $G(\lambda)$ , the

<sup>7</sup> By transforming the non-central  $F$ -distribution function to the incomplete beta function ratio [12, p. 947] and interchanging the order of integrations and summations of the result by virtue of a theorem in Halmos [5, p. 112],  $G(\lambda)$  can be rewritten as

$$G(\lambda) = -m \sum_{j=0}^{\infty} I_y\left(\frac{m}{2} + 1 + j, \frac{T - k}{2}\right) + 2 \sum_{j=0}^{\infty} (j + 3)I_y\left(\frac{m}{2} + 2 + j, \frac{T - k}{2}\right) + 4I_y\left(\frac{m}{2} + 1, \frac{T - k}{2}\right) + \frac{m^2}{4}.$$

Since the infinite series in the above formula converge fairly fast, we have used it in computing approximate values of  $G(\lambda)$ , setting an accuracy restriction. Further information on the method of numerical computations of  $G(\lambda)$  can be obtained from either of the authors upon request.

**TABLE I**  
OPTIMAL CRITICAL VALUES, THE CORRESPONDING AVERAGE RELATIVE RISK, AND THE CORRESPONDING LEVELS OF SIGNIFICANCE

$m^a$	$T - k^b$	$\lambda^{*c}$	$G(\lambda^*)^d$	$\alpha^e$	$\alpha' \left( \frac{1}{2} \right)^f$	$\alpha' \left( \frac{m}{2} \right)^g$
1	any value	0	.250	1.000	1.000	1.000
2	any value	0	1.000	1.000	1.000	1.000
3	any value	0	2.250	1.000	1.000	1.000
4	any value	0	4.000	1.000	1.000	1.000
5	5	.266	6.241	.914	.939	.985
	6	.291	6.239	.902	.931	.983
	7	.311	6.238	.892	.924	.981
	8	.330	6.236	.881	.916	.980
	16	.416	6.229	.831	.879	.970
	24	.457	6.225	.804	.860	.964
	60	.512	6.218	.766	.832	.957
	120	.538	6.215	.747	.817	.952
	1000	.600	6.213	.700	.780	.940
6	6	.475	8.912	.807	.852	.963
	7	.510	8.900	.785	.835	.958
	8	.539	8.889	.766	.821	.954
	16	.675	8.829	.672	.745	.932
	24	.734	8.796	.627	.708	.921
	60	.822	8.739	.557	.650	.903
	120	.854	8.715	.531	.628	.896
	1000	.833	8.696	.544	.641	.904
	7	7	.647	11.941	.710	.766
8		.682	11.908	.686	.746	.934
16		.850	11.721	.564	.641	.902
24		.922	11.618	.507	.592	.886
60		1.029	11.444	.421	.514	.859
120		1.073	11.367	.385	.481	.847
1000		1.143	11.299	.334	.431	.823
8	8	.789	15.267	.627	.687	.917
	16	.976	14.868	.488	.563	.877
	24	1.059	14.648	.423	.503	.857
	60	1.170	14.275	.332	.418	.826
	120	1.215	14.113	.296	.382	.812
	1000	1.363	13.949	.209	.289	.752
16	16	1.381	48.004	.263	.303	.790
	24	1.482	44.917	.187	.224	.756
	60	1.624	39.781	.090	.119	.697
	120	1.668	37.594	.062	.087	.677
	1000	1.731	35.462	.036	.055	.642
24	24	1.611	84.987	.125	.146	.720
	60	1.745	69.406	.042	.055	.658
	120	1.784	62.905	.022	.031	.637
	1000	1.825	56.585	.009	.014	.611
60	60	1.865	252.742	.009	.010	.619
	120	1.899	200.903	.002	.002	.595
	1000	1.933	152.728	.00004	.00007	.565
120	120	1.933	517.261	.00018	.00021	.603
	1000	1.967	323.787	.00000	.00000	.548

<sup>a</sup>  $m$  is the number of restrictions.

<sup>b</sup>  $T$  is the number of points in the sample space;  $k$  is the number of regressors.

<sup>c</sup>  $\lambda^*$  is the optimal critical value of the pre-test of estimation.

<sup>d</sup>  $G(\lambda^*)$  is the minimum average relative risk corresponding to  $\lambda^*$ .

<sup>e</sup>  $\alpha$  is the probability of type one error for the central  $F$  corresponding to  $\lambda^*$ .

<sup>f</sup>  $\alpha'(\frac{1}{2})$  is the probability of type one error for the non-central  $F$  with non-centrality  $\frac{1}{2}$  corresponding to  $\lambda^*$ .

<sup>g</sup>  $\alpha'(m/2)$  is the probability of type one error for the non-central  $F$  with non-centrality  $m/2$  corresponding to  $\lambda^*$ .

corresponding optimal values of  $\lambda$ , and also the corresponding optimal levels of significance are shown in Table I.

#### 4. A COMPARISON OF PRE-TEST CRITICAL VALUES FOR ALTERNATIVE CRITERIA

Significance levels for both the central  $F$  and the non-central  $F$  with non-centrality parameters of  $1/2$  and  $m/2$  have been calculated and are given in Table I. The central  $F$  values apply when the null hypothesis of the pre-test is that the restrictions are valid. The non-central  $F$  values correspond to the strong  $MSE$  test derived by Toro-Vizcarrondo and Wallace [10], and the second weak  $MSE$  test developed by Wallace [13].

The sharpest differences in critical values between our approach of minimizing the average relative risk and the minimax regret approach employed by Sawa and Hiromatsu and Brook take place when the number of restrictions,  $m$ , is very small. In particular, if  $m$  is less than or equal to 4, our critical value for the pre-test is zero so that the corresponding probability of type one errors in any sense is always unity; i.e., we should always reject the hypothesis and use the unrestricted least squares. On the other hand, the optimal critical values found by Sawa and Hiromatsu<sup>8</sup> and Brook take very stable values between 1.8 and 2.1 even for  $1 \leq m \leq 4$ . Therefore, the corresponding type one errors depart considerably from unity, being less than the 40 per cent level. As  $m$  increases beyond 4, the critical values ( $\lambda^*$ ) take positive values and gradually increase for any given denominator degrees of freedom ( $T - k$ ). However, up to around  $m = 16$  the values of  $\lambda^*$  are considerably less than the values of  $\lambda'$ , implying larger critical regions than the minimax regret criterion for these cases. As  $m$  increases beyond 20, and particularly if  $T - k$  also increases, the difference between  $\lambda^*$  and  $\lambda'$  becomes less and less.

In statistical decision problems it often happens that "optimal" decisions sharply differ depending on different decision criteria adopted. Considering this fact, it is interesting that both criteria compared here give very close values of the optimal significance level if both degrees of freedom are large enough. However, if  $m$  is small and if  $T - k$  also is small, the optimal critical values derived in this paper are much smaller than the alternative minimax regret case. That is, our approach is more "conservative" in favor of the ordinary, least squares estimators. This is understandable from Figure 1. For given denominator degrees of freedom the critical area to the left of  $m/2$  becomes larger and the area to the right becomes smaller as the values of  $\lambda$  decrease. However, the area to the left is a "relatively" fixed area because the total minimized area runs over the whole range of  $\theta$ . Therefore, it is clear that if  $m$  takes small values, then choosing small values also for  $\lambda$  is better; hence, forcing the area to the right of  $m/2$  to zero gives the optimal decision for  $m \leq 4$ . On the other hand, the minimax regret approach searches for a value of  $\lambda$  which balances the regrets of both sides,  $0 \leq \theta \leq m/2$  and  $m/2 \leq \theta < +\infty$ . Therefore, the minimax regret approach disregards the

<sup>8</sup> Sawa and Hiromatsu calculated the critical values ( $\lambda'$ ) for only the  $m = 1$  case.



broad area of the right tail, resulting in the range of about 1.8 to 2.1 for minimax critical values. (See Brook [3, pp. 59–60] for those values.)

### 5. SOME CONCLUDING REMARKS

For small degrees of freedom the critical values for pre-testing derived in this paper are disposed toward unrestricted estimators relative to either the Sawa and Hiromatsu and Brook critical values and common practice of using the one or five per cent significance points in standard  $F$  tables. This result is insured by the choice of risk function and criterion. The risk functions focus on the conditional expected value of the dependent variable and the criterion attaches a cost to extreme values of the non-centrality. In applied research, the investigator is quite often interested in the elements of the  $\beta$  vector per se and variances of unrestricted estimators are quite sensitive to the condition of the design matrix. And the very fact that the investigator considers a set of potential restriction often reflects something other than a diffuse prior.<sup>9</sup> Considering the innocence attached to priors by the criterion considered here and in Sawa and Hiromatsu and the absence of design consideration in the risk function, perhaps the most striking feature of the minimax regret and minimum average risk critical values is that restrictions are given some non-zero chance of being adopted, at least in certain cases.

*Kobe University*  
and  
*Duke University*

*Manuscript received June, 1974; revision received January, 1975.*

### APPENDIX 1: PROOF OF THE CONVERGENCE OF THE IMPROPER INTEGRAL (13)

We use the following two lemmas.

LEMMA 1 (Brook):  $s(\theta, \lambda) \geq r(\theta, \lambda)$  for  $\theta \in [0, +\infty)$  and  $\lambda \in [0, +\infty)$ , where the equality holds if and only if  $\theta$  and/or  $\lambda$  equals zero.

PROOF: See Brook [3, pp. 33–34].

LEMMA 2 (Gun):  $\theta^\mu [1 - r(\theta, \lambda)] \rightarrow 0$  as  $\theta \rightarrow \infty$  for  $\mu > 0$ .

PROOF. See Gun [4, pp. 52–54] and/or Brook [3, p. 37].

From Lemmas 1 and 2, it holds that

$$(A.1) \quad 1 - r(\theta, \lambda) = O(1/\theta^\mu)$$

and

$$(A.2) \quad \theta[2\{1 - r(\theta, \lambda)\} - \{1 - s(\theta, \lambda)\}] = O(1/\theta^\mu)$$

for  $\mu > 1$ . Hence, the sufficient condition for the convergence of the improper integral (13) in the text is satisfied. See, e.g., Widder [14, p. 330] for the sufficient condition. Q.E.D.

<sup>9</sup> Sometimes not. Dropping variables routinely based on  $t$  tests is the obverse.

APPENDIX 2: PROOF OF (18) AND (20)

Using the following differentiation formulas [12, p. 557],

$$\frac{d[{}_2F_1(a, b; c; y)]}{dy} = \frac{ab}{c} {}_2F_1(a + 1, b + 1; c + 1; y),$$

the inside of the bracket of (16) reduces to

$$\frac{d}{{}dy} [y {}_2F_1(a, b; c; y)] = {}_2F_1(a + 1, b; c; y),$$

and rearranging the result, the inside of the bracket of (16) reduces to

$$(A.3) \quad y = \frac{m - 4}{2} \frac{{}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right)}{d\left[{}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right)\right] / dy}$$

which is the first equation of (18). However, noting that the variable  $y$  and all parameters in the present hypergeometric function are positive and recalling the definition of the hypergeometric function, it is clear that

$${}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right) > 0$$

and

$$d\left[{}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right)\right] / dy > 0.$$

This means (A.3) holds if and only if  $m \geq 5$  for  $y \in (0, 1)$ , whence, from (15),  $\lambda = 0$  when  $m \leq 4$ .

Now, using a linear transformation formula [12, p. 559]

$${}_2F_1(a, b; c; y) = (1 - y)^{c-a-b} {}_2F_1(c - a, c - b; c; y)$$

and a relation between the hypergeometric function and the incomplete beta function [12, p. 945]

$${}_2F_1(a, 1 - b; a + 1; y) = (a/y^a) B_y(a, b)$$

where

$$B_y(a, b) = \int_0^y t^{a-1} (1 - t)^{b-1} dt,$$

we obtain

$$(A.4) \quad \begin{aligned} {}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right) &= {}_2F_1\left(\frac{m}{2}, \frac{T - k}{2}; \frac{m}{2} + 1; y\right) / (1 - y)^{(T-k)/2 + 1} \\ &= m B_y\left(\frac{m}{2}, \frac{T - k}{2} + 1\right) [2y^{m/2} (1 - y)^{(T-k)/2 + 1}]. \end{aligned}$$

Differentiating (A.4) with respect to  $y$ , we get

$$(A.5) \quad \begin{aligned} &\frac{d\left[{}_2F_1\left(1, \frac{m + T - k}{2} + 1; \frac{m}{2} + 1; y\right)\right]}{dy} \\ &= \frac{m B_y\left(\frac{m}{2}, \frac{T - k}{2} + 1\right)}{4y^{m/2} (1 - y)^{(T-k)/2 + 1}} \left(\frac{(T - k + 2)y - m(1 - y)}{y(1 - y)}\right) + \frac{m}{2y(1 - y)}. \end{aligned}$$

Using (A.4) and (A.5), and rearranging the result, we obtain

(A.6)

$$y = \left[ (m-2)B_y \left( \frac{m}{2}, \frac{T-k}{2} + 1 \right) - y^{m/2}(1-y)^{(T-k)/2+1} \right] / \left[ \left( \frac{T-k}{2} + m-1 \right) B_y \left( \frac{m}{2}, \frac{T-k}{2} + 1 \right) \right].$$

Further, using a recurrence relation of the incomplete beta function ratio [12, p. 944]

$$I_y(a, b) = \frac{\Gamma(a+b)}{\Gamma(a+1)\Gamma(b)} y^a(1-y)^b + I_y(a+1, b),$$

(A.6) leads to the second equation of (18) in the text.

#### REFERENCES

- [1] BOCK, M. E., G. G. JUDGE, AND T. A. YANCEY: "Some Comments on Estimation in Regression after Preliminary Tests of Significance," *Journal of Econometrics*, 1 (1973), 191-200.
- [2] BOCK, M. E., T. A. YANCEY, AND G. G. JUDGE: "The Statistical Consequences of the Preliminary Test Estimators in Regression," *Journal of the American Statistical Association*, 68 (1973), 109-116.
- [3] BROOK, R. J.: "On the Use of Minimax Regret Function to Set Significance Points in Prior Tests of Estimation." Ph.D. dissertation, North Carolina State University at Raleigh (1972). University Microfilms, Ann Arbor, Michigan.
- [4] GUN, A.: "The Use of a Preliminary Test for Interactions in the Estimation of Factorial Means." Ph.D. dissertation, North Carolina State University at Raleigh (1965). University Microfilms, Ann Arbor, Michigan.
- [5] HALMOS, P. R.: *Measure Theory*. New York: Van Nostrand, 1950.
- [6] HUNTSBERGER, D. V.: "A Generalization of a Preliminary Testing Procedure for Pooling Data," *Annals of Mathematical Statistics*, 26 (1955), 734-743.
- [7] JUDGE, G. G., T. A. YANCEY, AND M. E. BOCK: "Properties of Estimators after Preliminary Tests of Significance when Stochastic Restrictions are Used in Regression," *Journal of Econometrics*, 1 (1973), 29-48.
- [8] SAWA, T., AND T. HIROMATSU: "Minimax Regret Significance Points for a Preliminary Test in Regression Analysis," *Econometrica*, 6 (1973), 1093-1102.
- [9] SCLOVE, S. L., C. MORRIS, AND R. RADHAKRISHNAN: "Non-optimality of Preliminary-test Estimators for Means of a Multivariate Normal Distribution," *Annals of Mathematical Statistics*, 43 (1972), 1481-1490.
- [10] TORO-VIZCARRONDO, C., AND T. D. WALLACE: "A Test of the Mean Square Error Criterion for Restrictions in Linear Regression," *Journal of the American Statistical Association*, 63 (1968), 558-572.
- [11] TOYODA, T., AND T. D. WALLACE: "Estimation of Variance after a Preliminary Test of Homogeneity and Optimal Levels of Significance for the Pre-test." Discussion paper, North Carolina State University (1974).
- [12] UNITED STATES DEPARTMENT OF COMMERCE, NATIONAL BUREAU OF STANDARDS: *Handbook of Mathematical Functions*. Applied Mathematics Series, no. 55. Washington, D.C.: Government Printing Office, 1964.
- [13] WALLACE, T. D.: "Weaker Criteria and Tests for Linear Restrictions in Regression," *Econometrica*, 40 (1972), 689-698.
- [14] WIDDER, D. V.: *Advanced Calculus*, 2nd ed. Englewood Cliffs, N.J.: Prentice-Hall, 1961.