Faculty Working Papers

WALLACE'S WEAK MEAN SQUARE ERROR CRITERION
FOR TESTING LINEAR RESTRICTIONS IN REGRESSION:
A TIGHTER BOUND

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

#88

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

February 20, 1973

WALLACE'S WEAK MEAN SQUARE ERROR CRITERION FOR TESTING LINEAR RESTRICTIONS IN REGRESSION: A TIGHTER BOUND

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

#88

WALLACE'S WEAK MEAN SQUARE ERROR CRITERION FOR TESTING LINEAR RESTRICTIONS IN REGRESSION: A TIGHTER BOUND

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

A sharper set of sufficient conditions for the criterion used for the weak mean square error test of Wallace is established to replace the sufficient conditions of his development.

1. Introduction

In a recent issue of this Journal, Wallace [4], building on some of his earlier work [1,3], with the K parameter classical linear statistical model

$$(1.1)\qquad \underline{y} = X\underline{\beta} + \underline{e},$$

under J linear restrictions or hypotheses

$$(1.2)\qquad R\underline{\beta} - \underline{r} = \underline{0},$$

presents a test for determining when the difference of the risk functions of the least squares unrestricted and restricted estimators is positive. Wallace's sufficient condition for the risk of the restricted estimator, $E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta})$ to be at least as small as that of the unrestricted estimator, $E(\underline{b}-\underline{\beta})'(\underline{b}-\underline{\beta})$ is that the magnitude of the specific errors, $\underline{\delta}$, in the null hypotheses, (1.2), be such that

$$(1.3)\qquad \lambda \le \frac{1}{2}\mu_k \operatorname{tr} S^{-1}R'(RS^{-1}R')^{-1}RS^{-1} = \lambda_o ,$$

where μ_k is the smallest characteristic root of $S=X'X$. The parameter λ equals $(R\underline{\beta}-\underline{r})'(RS^{-1}R')^{-1}(R\underline{\beta}-\underline{r})/2\sigma^2$ which equals $\underline{\delta}'(RS^{-1}R')^{-1}\underline{\delta}/2\sigma^2$ and is the noncentrality parameter in the $F_{(\lambda,J,T-K)}$ distribution of the test statistic, $u = (R\underline{b}-\underline{r})'(RS^{-1}R')^{-1}(R\underline{b}-\underline{r})/J\hat{\sigma}^2$.

The purpose of the present paper is to develop a sharper set of conditions for determining (i) where the risk function of the restricted estimator is less than that of the conventional estimator, (ii) when it is impossible to determine which risk function is smaller, and (iii) when the risk function of the conventional estimator is less than that of the restricted estimator.

A comparison of the results of Wallace and those of this paper appears in Section 2, and Section 3 summarizes our results and indicated directions for further work.

2. Reparameterization of the Model

Reparameterization of the model, (1.1) and (1.2), is a convenient device for clarifying the behavior of the risk function of $\hat{\underline{\beta}}$, the restricted estimator. Using a non-singular matrix W = QP such that $W\underline{\beta} = \underline{\theta}$ and $Z = XP^{-1}Q'$ the original model, (1.1) and (1.2), becomes

(2.1) $\quad y = Z\underline{\theta} + \underline{e}$

and

(2.2) $\quad [I_J 0]\underline{\theta} - \underline{r}_o = \underline{0}$

where P is such that $P^{-1'}(X'X)P^{-1} = I$, Q is an orthogonal matrix such that $QBQ' = \begin{pmatrix} I_J & 0 \\ 0 & 0 \end{pmatrix}$, since $B = P^{-1'}R'(RS^{-1}R')^{-1}RP^{-1}$ is idempotent, and $\underline{r}_o = G_J^{-1}\underline{r}$ from $RP^{-1}Q'QP^{-1'}R = RS^{-1}R' = G_J G_J'$ where $RP^{-1}Q' = [G_J 0]$. Of course, $Z'Z = I_K$ and any estimator $\tilde{\underline{\theta}}$ implies an estimator $W^{-1}\tilde{\underline{\theta}} = \tilde{\underline{\beta}}$ for $\underline{\beta}$.

The least squares estimator, restricted estimator, test statistic, and non-centrality of the reparametrized model are $\hat{\underline{\theta}} = Z'\underline{y} \sim N(\underline{\theta}, \sigma^2 I_K)$,

$\hat{\hat{\underline{\theta}}}' = [\underline{r}_o', \hat{\underline{\theta}}_{K-J}]$, $u = \dfrac{(\hat{\underline{\theta}}_J - \underline{r}_o)'(\hat{\underline{\theta}}_J - \underline{r}_o)}{J\hat{\sigma}^2}$, and $\lambda = (\underline{\theta}_J - \underline{r}_o)'(\underline{\theta}_J - \underline{r}_o) / 2\sigma^2$,

respectively, where $\underline{\theta}_{K-J} = [0 \ I_{K-J}] \ \hat{\underline{\theta}}$.

The weighted risk functions for the least squares unrestricted and restricted estimators, which are the same as the unweighted risk function of $\underline{b}$ and $\hat{\underline{\beta}}$, are

$$(2.3) \qquad \rho(\underline{b},\underline{\beta}) = E(\underline{b}-\underline{\beta})'(\underline{b}-\underline{\beta}) = E(\hat{\underline{\theta}}-\underline{\theta})'W^{-1'}W^{-1}(\hat{\underline{\theta}}-\underline{\theta}) = \sigma^2 \mathrm{tr} W^{-1'}W^{-1} = \sigma^2 \mathrm{tr} S^{-1}$$
$$= \sigma^2\left(\sum_{i=1}^{J} d_i + \sum_{i=J+1}^{K} d_i\right)$$

where $d_1,\cdots, d_J$ and $d_{J+1},\cdots, d_K$ are characteristic roots of $A = [I_J \ 0] \ W^{-1'}W^{-1}[I_J \ 0]'$ and $[0 \ I_{K-J}] \ W^{-1'}W^{-1}[0 \ I_{K-J}]'$ respectively, and

$$(2.4) \qquad \rho(\hat{\underline{\beta}},\underline{\beta}) = E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) = E(\hat{\hat{\underline{\theta}}}-\underline{\theta})'W^{-1'}W^{-1}(\hat{\hat{\underline{\theta}}}-\underline{\theta})$$
$$= E(\hat{\underline{\theta}}-\underline{\theta})'W^{-1'}W^{-1}(\hat{\underline{\theta}}-\underline{\theta}) + E(\underline{r}_o-\underline{\theta}_J)'A(\underline{r}_o-\underline{\theta}_J) - E(\hat{\underline{\theta}}_J-\underline{\theta}_J)'A(\hat{\underline{\theta}}_J-\underline{\theta}_J).$$

In order to determine conditions under which the risk function for $\underline{b}$ exceeds that for $\hat{\underline{\beta}}$, we subtract (2.4) from (2.3) giving

$$(2.5) \qquad E(\underline{b}-\underline{\beta})'(\underline{b}-\underline{\beta}) - E(\hat{\underline{\beta}}-\underline{\beta})'(\hat{\underline{\beta}}-\underline{\beta}) = -(\underline{\theta}_J-\underline{r}_o)'A(\underline{\theta}_J-\underline{r}_o) + E(\hat{\underline{\theta}}_J-\underline{\theta}_J)'A(\hat{\underline{\theta}}_J-\underline{\theta}_J)$$

which is non-negative if and only if

$$(2.6) \qquad E(\hat{\underline{\theta}}_J-\underline{\theta}_J)'A(\hat{\underline{\theta}}_J-\underline{\theta}_J) = \sigma^2 \mathrm{tr} A \geq (\underline{\theta}_J-\underline{r}_o)'A(\underline{\theta}_J-\underline{r}_o) .$$

From a theorem on extrema of quadratic forms [2,p.51],

$$\sup_{(\underline{\theta}_J-\underline{r}_o)} \frac{(\underline{\theta}_J-\underline{r}_o)'A(\underline{\theta}_J-\underline{r}_o)}{(\underline{\theta}_J-\underline{r}_o)'(\underline{\theta}_J-\underline{r}_o)} = d_L$$

where d_L is the largest characteristic root of A, we have

$$(2.7) \qquad \sigma^2 \mathrm{tr} A \geq (\underline{\theta}_J-\underline{r}_o)'A(\underline{\theta}_J-\underline{r}_o) \leq d_L(\underline{\theta}_J-\underline{r}_o)'(\underline{\theta}_J-\underline{r}_o) = d_L 2\lambda\sigma^2.$$

Furthermore, $S^{-1}R'(RS^{-1}R')^{-1}RS^{-1} = P^{-1}Q'QP^{-1'}R'(RS^{-1}R')^{-1}RP^{-1}Q'QP^{-1'}$ $= W^{-1}[I_J\ 0]'[I_J\ 0]W^{-1'}$, $S^{-1} = P^{-1}Q'QP^{-1'} = W^{-1}W^{-1'}$, $\mathrm{tr}W^{-1}[I_J\ 0]'[I_J\ 0]W^{-1'}$ $= \mathrm{tr}A$, and $\mathrm{tr}S^{-1} = \mathrm{tr}W^{-1'}W^{-1}$. The non-zero characteristic roots of $W^{-1}[I_J\ 0]'$ $[I_J\ 0]W^{-1'}$, $\begin{bmatrix} I_J & 0 \\ 0 & 0 \end{bmatrix} W^{-1'}W^{-1} \begin{bmatrix} I_J & 0 \\ 0 & 0 \end{bmatrix}$ and A are the same as are the characteristic roots of S^{-1} and $W^{-1'}W^{-1}$ since the order of matrices does not affect the values of the characteristic roots of their product. It is easy to see that d_L is less than the largest characteristic root of S^{-1}, $\sup_{\underset{\sim}{x}} \dfrac{\underset{\sim}{x}'W^{-1'}W^{-1}\underset{\sim}{x}}{\underset{\sim}{x}'\underset{\sim}{x}}$ for all $\underset{\sim}{x}$, since

$$\sup_{(\underset{\sim}{x}_J \underset{\sim}{0})} \frac{(\underset{\sim}{x}'_J 0')W^{-1'}W^{-1}(\underset{\sim}{x}_J 0)}{(\underset{\sim}{x}_J \underset{\sim}{0})'(\underset{\sim}{x}_J \underset{\sim}{0})} = d_L$$

, restricting $\underset{\sim}{x}_{K-J}$ to be $\underset{\sim}{0}$, must be at least as small as the supremum for all $\underset{\sim}{x}$. Hence, the largest root of S^{-1} is greater than or equal to that of $S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$, which is d_L.

Thus a sufficient condition, and the smallest one that will always hold, for the risk of $\underset{\sim}{b}$ to be at least as large as $\hat{\underset{\sim}{\beta}}$ is that

$$(2.8) \qquad \lambda \leq \frac{1}{2d_L}\mathrm{tr}A = \frac{1}{2d_L}\mathrm{tr}S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}.$$

This expression is the same as (1.3) with μ_k which is the reciprocal of the largest root of S^{-1}, replaced by $1/d_L$ so that the bound obtained for λ must be greater than or equal to that in the Wallace result.

To explore the composition of the size of the two risk functions (2.5), we use an orthogonal transformation C to obtain the characteristic roots of A so

$$(2.5a) \qquad E(\underset{\sim}{b}-\underset{\sim}{\beta})'(\underset{\sim}{b}-\underset{\sim}{\beta}) - E(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta})'(\hat{\underset{\sim}{\beta}}-\underset{\sim}{\beta}) = \sigma^2\mathrm{tr}A - (\underset{\sim}{\theta}_J-\underset{\sim}{r}_o)'C'CAC'C(\underset{\sim}{\theta}_J-\underset{\sim}{r}_o)$$

$$= \sigma^2\mathrm{tr}A - \underset{\sim}{\xi}'D\underset{\sim}{\xi} = \sigma^2\sum_{i=1}^{J} d_i - \sum_{i=1}^{J}\xi_i^2 d_i$$

where D is a diagonal matrix whose elements are the characteristic roots, d_i, of A and $\underset{\sim}{\xi}$ is a vector incorporating the specification error in the exact restrictions. One can perform the conceptual experiment of varying the values of the elements of $\underset{\sim}{\xi}$ so that only one of them is non zero while $\lambda = \dfrac{\Sigma\xi_i^2}{2\sigma^2}$ remains constant,

thus $2\lambda\sigma^2 = \xi_L^2$ and

$$\text{(2.5b)} \quad E(b-\beta)'(b-\beta) - E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) \leq \sigma^2 \sum_{i=1}^{J} d_i - 2\lambda\sigma^2 d_L .$$

Since we never know the values for ξ, the risk function of b is greater than equal to that of $\hat{\beta}$ if $\lambda \leq \sum_{i=1}^{J} d_i/d_L$ and without information on the values of ξ no tighter bound is possible. A similar conceptual experiment associated the smallest characteristic root of A, d_s, shows that the risk of $\hat{\beta}$ must exceed that of b whenever $\lambda \leq \sum_{i=1}^{J} d_i/2d_s$. Nothing can be said in a given problem situation about which risk function is larger when $\frac{\Sigma d_i}{2d_L} \leq \lambda \leq \frac{\Sigma d_i}{2d_s}$.

For the risk function of $\hat{\beta}$, the inequality

$$\text{(2.9)} \quad \sigma^2\left[\sum_{i=J+1}^{K} d_i + 2\lambda d_s\right] \leq E(\hat{\beta}-\beta)'(\hat{\beta}-\beta) \leq \sigma^2\left[\sum_{i=J+1}^{K} d_i + 2\lambda d_L\right]$$

holds for each value of λ with the equalities holding at $\lambda=0$ where the risk function of $\hat{\beta}$ is smallest. The terms in (2.9) are also equal if all the roots of S are equal as is the case for an idempotent matrix. Furthermore, since R and S are known, the characteristic roots of $S^{-1}R'(RS^{-1}R')^{-1}RS^{-1}$ are easy to obtain.

3. Concluding Remarks

By reparameterizing the model, a sharper result has been given for comparing the risks of the restricted and conventional least squares estimators. The relation of the risk functions of the least squares unrestricted and restricted estimators have been shown to involve a region for λ when $\rho(b,\beta) \geq \rho(\hat{\beta},\beta)$, a

region of λ where uncertainty exists about the comparison of $\rho(\underline{b},\underline{\beta})$ and $\rho(\hat{\underline{\beta}},\underline{\beta})$, and a region for λ where $\rho(\underline{b},\underline{\beta}) \leq \rho(\hat{\underline{\beta}},\underline{\beta})$. Without knowing the values of $\underline{\xi}$ there is no way to reduce the region of uncertainty further. The use of minimum risk as a criterion for the choice of an estimator is not new, nor is the abandonment of unbiasedness, but these changes in criteria have not been exploited very systematically in applied econometric work. The work of Wallace [4] and this extension provide a systematic basis for determining the statistical consequences in a risk context of making use of each of the estimators.

University of Illinois at Urbana-Champaign

References

[1] Goodnight, J. and T. D. Wallace (1971). "Operational Techniques and Tables for Making Weak MSE tests for Restrictions in Regression," Econometrica, forthcoming.

[2] Rao, C. R. (1965). Linear Statistical Inference and Its Applications, John Wiley and Sons, Inc., New York.

[3] 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.

[4] Wallace, T. D. (1971). "Weaker Criteria and Tests for Linear Restrictions in Regression," Econometrica, forthcoming.

[5] Wallace, T. D. and C. E. Toro-Vizcarrondo (1969). "Tables for the Mean Square Error Test for Exact Linear Restrictions in Regression," Journal of the American Statistical Association, Vol.64, pp. 1644-93.

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