Open access • Journal Article • DOI:10.2307/1912962

Weaker criteria and tests for linear restrictions in regression — Source link

T. D. Wallace

Published on: 01 Jul 1972 - Econometrica (Econometrica)

Econometrica, Vol. 40, No. 4 (July, 1972)

WEAKER CRITERIA AND TESTS FOR LINEAR RESTRICTIONS IN REGRESSION

BY T. D. WALLACE[1]

The standard F test for linear restrictions in regression is relevant as a criterion but fails to capture the notion of tradeoff between bias and variance. Average squared distance criteria yield operational tests that are more appropriate, depending upon objectives. In the present paper two alternative criteria are developed. The first allows testing of the hypothesis that the average squared distance of a restricted estimator from the parameter point in k space is less than the average squared distance of the unrestricted, ordinary least squares estimator from the same parameter point. The second sets up a test of betterness of the restricted estimator over the unrestricted estimator of $E(Y|X)$, where betterness is again defined in average squared distance.

1.

THE STANDARD PROCEDURE for deciding whether to impose linear restrictions in regression has been the Snedecor F test. To summarize, one begins with a linear model

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

where Y is $N \times 1$, X is $N \times k$ and fixed, at least conditionally, β is $k \times 1$, and ε is an $N \times 1$ vector of random variables. A general linear reduction in the parameter space from k to $k - m$ can be written

$$H'\beta = h \tag{2}$$

where H' is $m \times k$ and known and the rank of H is m, and h is an $m \times 1$ known vector.

Defining $\text{SSE}(\hat{\beta})$ as the error sum of squares for the least squares estimators constrained by equation (2), $\text{SSE}(b)$ as the error sum of squares for the unrestricted ordinary least squares estimators, and $\hat{\sigma}^2$ as the least squares estimate of σ^2 in the unrestricted case.

$$u = \frac{\text{SSE}(\hat{\beta}) - \text{SSE}(b)}{m\hat{\sigma}^2} \tag{3}$$

is distributed as a noncentral F with parameters m, $N - k$, and λ. The noncentrality parameter, λ, is

$$\lambda = \frac{(H'\beta - h)'(H'S^{-1}H)^{-1}(H'\beta - h)}{2\sigma^2}, \tag{4}$$

where $S = X'X$.

[1] My thanks to the Economics Department, the University of Washington at Seattle, for a visiting appointment during the summer, 1970, during which time a first draft of this paper was completed. I am grateful to Allan Seheult and Burt Holland for directing my attention to the possibility of this line of investigation and to Burt Holland and an unknown referee for helpful suggestions on an earlier draft.

Thus, the test that $\lambda = 0$ tests truth of the restriction $H'\beta = h$, and the distribution of u becomes the central F under the null hypothesis. Such results can be found in Chipman and M. Rao [2] or in C. R. Rao [**10**]. The test is uniformly most powerful.

For example, to implement the test for pooling data from different sources one may proceed as in Chow [**3**] to specify

$$(5) \qquad \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} \sim N(0, \sigma^2 I),$$

where the partitions represent, say, two time periods, and there are N observations in each cross section and k parameters in each cross section. Choosing H' as $(I_k, -I_k)$ and h null puts the pooling problem into the more general framework.

The general testing procedure has considerable flexibility for handling many such questions and is at least approximately correct under a wide range of distributional assumptions other than the ones stated here. There are, however, several complaints that one might air about the procedure. For example:

(i) Those who use the test applied to the pooling question find it a rare event when the test statistic is smaller than the tabulated central F at the commonly tabulated .05, .01 levels of type one error.

(ii) The null hypothesis can ultimately be rejected if the sample size gets large enough—i.e., if $x \sim N(\mu, \sigma^2)$, and μ unknown, the chance that we pick μ precisely out of the real numbers as a null hypothesis is zero and for $\mu_0 \neq \mu$, we can ultimately reject $H_0 : \mu = \mu_0$ by reducing $\sigma_{\bar{x}}$.

(iii) We may want to use the constrained estimators even when the restriction is not valid. Looking at the distribution of b and $\hat{\beta}$, the ordinary least squares and restricted least squares estimators, one finds,

$$(b - \beta) \sim N(0, \sigma^2 S^{-1}),$$

and

$$(\hat{\beta} - \beta) \sim N\{-S^{-1}H(H'S^{-1}H)^{-1}(H'\beta - h), \sigma^2 S^{-1} - \sigma^2 S^{-1}H(H'S^{-1}H)^{-1}H'S^{-1}\}.$$

Thus, if the restrictions are valid, the bias of $\hat{\beta}$ is zero and the variances smaller ($\sigma^2 S^{-1}H(H'S^{-1}H)^{-1}H'S^{-1}$ is non-negative definite). But even if the restriction is strictly invalid, the constrained estimators have smaller variances and one might be willing to make a tradeoff, accepting some bias in order to reduce variances. The truth of $H'\beta = h$ is relevant but overstrong as a criterion.

(iv) Using the test as a decision rule gives rise to sequential estimators whose properties are formidable and little understood.[2]

The remainder of this paper suggests three alternative tests within the classical framework that avoid all except the last of the complaints. The first of these tests has been presented elsewhere and is only reviewed here. The second is not easy to

[2] For example, see Bancroft [**1**].

use because considerable computation is required and the test statistic is not amenable to simple tabulation. The third test requires no more computation than provided in standard computer routines and a tabulation of the test statistic is provided in a companion paper [**4**].

2.

As stated in Section 1, if the restriction ($H'\beta = h$) is not valid, the restricted estimators, $\hat{\beta}$, are biased but have smaller variances. The possibility of tradeoff can be captured via the concept of mean squared error. For example, suppose we accept $\hat{\beta}$ as better than b, the ordinary least squares estimator, if and only if

$$\text{(6)} \qquad \mathrm{MSE}(l'\hat{\beta}) \leqslant \mathrm{MSE}(l'b) \quad \text{for every} \quad l \neq 0,$$

l being any $k \times 1$ constant vector.[3] It is convenient to define the inequality (6) as the strong MSE criterion.

It has been shown [**11**] that each linear combination of $\hat{\beta}$ has smaller mean squared error than the same linear combination of b if and only if

$$\text{(7)} \qquad \lambda \leqslant \tfrac{1}{2},$$

where λ is the noncentrality parameter defined in equation (4) above. Thus, a uniformly most powerful test exists to test whether the restricted estimator $\hat{\beta}$ is better in strong MSE. The relevant hypothesis is that the noncentrality parameter be less than one-half against the alternative that it is greater than one-half. Tabulation of the $F_{(m,N-k,\lambda)}$ for $\lambda = \frac{1}{2}$ may be found in [**12**], along with examples, including the pooling problem of equation (5) above.

3.

Even though the strong MSE testing procedure overcomes some of the objections raised by the older testing via the central F, one is still struck by the strength of the null hypothesis. To have the mean squared error of every linear combination of $\hat{\beta}$ better than each corresponding linear combination of b is quite persuasive that $\hat{\beta}$ is better.[4] But the question arises whether some weaker but still acceptable criterion would lead to another test.

A weaker mean squared error concept in vector estimation is

$$\text{(8)} \qquad \Delta^2 = E(\hat{\theta} - \theta)'(\hat{\theta} - \theta) = \operatorname{tr}[\Sigma_{\hat{\theta}\hat{\theta}} + (\mathrm{BIAS}\,\hat{\theta})(\mathrm{BIAS}\,\hat{\theta})'] = \sum_i \mathrm{MSE}(\hat{\theta}_i).$$

Geometrically, Δ^2 is the "average" squared Euclidian distance of the point $\hat{\theta}$ from θ, whatever the dimension of the parameter space.

[3] $\mathrm{MSE}(\hat{\theta}_i) = E(\hat{\theta}_i - \theta_i)^2 = \operatorname{var}\hat{\theta}_i + (\mathrm{BIAS}\,\hat{\theta}_i)^2$ for scalar estimation. Geometrically, the mean squared error is the average squared distance of $\hat{\theta}_i$ to θ_i. For the vector $\hat{\theta}$, MSE $l'\hat{\theta}$ is the quadratic form $l'[\Sigma_{\hat{\theta}\hat{\theta}} + (\mathrm{BIAS}\,\hat{\theta})(\mathrm{BIAS}\,\hat{\theta})']l$, where $\Sigma_{\hat{\theta}\hat{\theta}}$ is the variance-covariance matrix of $\hat{\theta}$ and BIAS $\hat{\theta}$ is the vector that displays biases. The criterion in equation (6) is equivalent to a requirement that $E(b - \beta)(b - \beta)' - E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$ be non-negative definite.

[4] To have such betterness equivalent to $\lambda \leqslant \frac{1}{2}$ rather dramatically illustrates the strength of the requirement that $\lambda = 0$.

The unrestricted least squares estimator, b, is unbiased, distributed as a multivariate normal, with mean squared error matrix

$$\text{MSE}_{bb} = \sigma^2 S^{-1} \tag{9}$$

where $S = X'X$. The average squared distance of b from β is the trace of $\sigma^2 S^{-1}$.

Recalling the algebra of restricted least squares [11], one can write the restricted estimator, $\hat{\beta}$, of β as

$$\hat{\beta} = b - S^{-1}H(H'S^{-1}H)^{-1}(H'b - h) \tag{10}$$

where b is the unrestricted ordinary least squares estimator. It follows from (10) and the distributional assumptions of the model that $\hat{\beta}$ is multivariate normal:

$$\hat{\beta} \sim N(\beta + \text{BIAS}\,\hat{\beta}, \Sigma_{\hat{\beta}\hat{\beta}}) \tag{11}$$

where the bias vector and covariance matrix are

$$\text{BIAS}\,\hat{\beta} = -S^{-1}H(H'S^{-1}H)^{-1}(H'\beta - h) \tag{12}$$

and

$$\Sigma_{\hat{\beta}\hat{\beta}} = \sigma^2 S^{-1} - \sigma^2 S^{-1}H(H'S^{-1}H)^{-1}H'S^{-1}. \tag{13}$$

Thus the mean squared error matrix for $\hat{\beta}$ is

$$\begin{aligned}\text{MSE}_{\hat{\beta}\hat{\beta}} = {} & \sigma^2 S^{-1} - \sigma^2 S^{-1}H(H'S^{-1}H)^{-1}H'S^{-1} \\ & + S^{-1}H(H'S^{-1}H)^{-1}(H'\beta - h)(H'\beta - h)'(H'S^{-1}H)^{-1}H'S^{-1},\end{aligned} \tag{14}$$

and the average squared distance of $\hat{\beta}$ from β is the trace of the matrix of equation (14).

To have the restricted estimator better in average squared distance means that the trace of equation (14) must be smaller than the trace of equation (9). A sufficient condition for this to be the case is to have the noncentrality parameter λ smaller than one-half, i.e., to have the inequality (7) hold, because if the difference between two matrices is positive semi-definite, one is assured that the difference of their traces is nonnegative. The relevant question is whether some weaker condition exists that insures $\hat{\beta}$ to be better in average squared distance.

Hence, taking the trace of the difference of equations (9) and (14), interest centers on the condition(s) under which

$$\begin{aligned}&\operatorname{tr} S^{-1}H(H'S^{-1}H)^{-1}(H'\beta - h)(H'\beta - h)'(H'S^{-1}H)^{-1}H'S^{-1} \\ &\quad \leqslant \sigma^2 \operatorname{tr} S^{-1}H(H'S^{-1}H)^{-1}H'S^{-1}.\end{aligned} \tag{15}$$

Using the fact that $\operatorname{tr} AB = \operatorname{tr} BA$ where AB and BA and their traces are defined, the left hand side of inequality (15) can be written as the scalar

$$[(H'\beta - h)'(H'S^{-1}H)^{-1}H'S^{-1}][S^{-1}H(H'S^{-1}H)^{-1}(H'\beta - h)] = Z'Z, \tag{16}$$

where Z is the $m \times 1$ vector indicated by the brackets.

Referring to the definition of λ in equation (4), one can see that it is a simple trick to re-express λ as

$$(17) \qquad \lambda = \frac{Z'SZ}{2\sigma^2} \quad \text{or} \quad 2\sigma^2\lambda = Z'SZ.$$

This allows use of an inequality to bound the $Z'Z$ of (16); namely,

$$(18) \qquad \frac{2\sigma^2\lambda}{\mu_1} \leqslant Z'Z \leqslant \frac{2\sigma^2\lambda}{\mu_k},$$

where μ_1 and μ_k are, respectively, the largest and smallest characteristic roots of the S matrix.[5]

Thus, an alternative sufficient condition that the restricted estimator $\hat{\beta}$ be better in average squared distance is

$$(19) \qquad \lambda \leqslant \theta, \quad \text{where} \quad \theta = \tfrac{1}{2}\mu_k \operatorname{tr} S^{-1}H(H'S^{-1}H)^{-1}H'S^{-1}.$$

Since θ can be calculated from the nonstochastic variables and whatever set of restrictions are in question, and λ is a parameter in a well-defined distribution, the inequality in (19) can be used for testing purposes. That the criterion can be less demanding than the one for which λ must be smaller than $\frac{1}{2}$ is illustrated by taking both S and H' as a kth order identity. Such an extreme but admissable case asks only that λ be no greater than $k/2$. However, one should note that there are cases for which $\theta < \frac{1}{2}$. For example, take

$$S = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and

$$H' = (1, 0).$$

For $\rho > 0$, the least root of S is $1 - \rho$ and θ is $\frac{1}{2}(1 + \rho^2)/(1 + \rho) < \frac{1}{2}$.

Computation of θ in a particular problem setting poses only slight difficulty. However, critical points for the noncentral F are not available over a wide range of the noncentrality parameter, nor would it be economic to do the necessary numerical integration as each case presented itself. Nor, as seen above, is it always the case that $\theta < \frac{1}{2}$. Even though an approximation is made available in the companion paper [4]. for the noncentral $F(\theta)$ for various θ, the usefulness of a simpler criterion is apparent.

4.

For ease of subsequent discussion, consider a restatement of the model as

$$(20) \qquad Y = XD^{-1}D\beta + \varepsilon, \qquad \varepsilon \sim N(0, \sigma^2 I), \qquad H'D^{-1}D\beta = h,$$

[5] See Graybill [6, p. 309] for the relevant theorem and proof.

where $D = \operatorname{diag}\{\sqrt{\Sigma X_1^2}, \sqrt{\Sigma X_2^2}, \ldots, \sqrt{\Sigma X_k^2}\}$. Letting $D\beta = \beta^*$, $XD^{-1} = X^*$, and $H'D^{-1} = G'$, the model of equation (20) can be restated:

$$(21) \qquad Y = X^*\beta^* + \varepsilon,$$
$$G'\beta^* = h.$$

The following results are immediate:

$$(22) \qquad b^* = Db, \qquad b^* \sim N(\beta^*, \sigma^2 R^{-1}),$$

and

$$(23) \qquad \hat{\beta}^* = D\hat{\beta},$$

where b^* and $\hat{\beta}^*$ are, respectively, unrestricted and restricted least squares estimators of β^*, and R is the sample correlation matrix of the regressors.

Since the u statistic of equation (3) can be written as

$$(24) \qquad u = [H'b - h][H'S^{-1}H]^{-1}[H'b - h]/m\hat{\sigma}^2,$$

it follows that u is invariant to the reparameterization from β and H to β^* and G. Invariance to the reparameterization also holds for the noncentrality parameter λ. (See equation (4).) Hence, without loss of generality, the question of whether to adopt $H'\beta = h$ can be recast into the question of whether to force $G'\beta^* = h$ in the restated model of equation (21).

In the reparameterized form,

$$(25) \qquad E(b^* - \beta^*)'(b^* - \beta^*) = \sigma^2 \sum_{i=1}^{k} \frac{1}{\gamma_i}$$

where the γ_i are the characteristic roots of the correlation matrix R.[6]

It is known that

$$(26) \qquad \sum_{i=1}^{k} \gamma_i = \operatorname{tr} R = k,$$

so that the minimum of (25) over the γ_i is for each γ_i to be unity. This is equivalent to orthogonality of the X and orthonormality of the X^* matrices. Hence the minimum average squared distance of b^* from β^* is $k\sigma^2$, the lower bound being achieved in the case of orthogonal X. Thus

$$(27) \qquad k\sigma^2 = \sum_{i=1}^{k} \gamma_i E(b_i^* - \beta_i^*) = E(b^* - \beta^*)'R(b^* - \beta^*)$$
$$\leqslant E(b^* - \beta^*)'(b^* - \beta^*).$$

Since R is a positive definite matrix that can be written as $C'\Gamma^{\frac{1}{2}}\Gamma^{\frac{1}{2}}C$, where C is orthogonal and $\Gamma^{\frac{1}{2}}$ is a diagonal matrix displaying the positive square roots of

[6] See Hoerl and Kennard [7].

the eigen values of R, the rotated average squared distance of (27) can be characterized as the mean of a $\sigma^2\chi^2_{(k)}$ variate, i.e.,

$$(28) \qquad b^* \sim N(\beta^*, \sigma^2 R^{-1}).$$

Hence,

$$(29) \qquad W = \Gamma^{\frac{1}{2}}C(b^* - \beta^*) \sim N(0, \sigma^2 I)$$

and

$$(30) \qquad v = W'W \sim \sigma^2\chi^2_{(k)} \quad \text{and} \quad Ev = k\sigma^2.$$

The same transformation of $\hat{\beta}^*$ yields a noncentral χ^2. By analogy from equations (11), (12), and (13),

$$(31) \qquad (\hat{\beta}^* - \beta^*) \sim N(-R^{-1}G(G'R^{-1}G)^{-1}[G'\beta^* - h], \\ \sigma^2 R^{-1} - \sigma^2 R^{-1}G(G'R^{-1}G)^{-1}G'R^{-1}),$$

and

$$(32) \qquad \frac{1}{\sigma}\Gamma^{\frac{1}{2}}C(\hat{\beta}^* - \beta^*) \sim N(-\Gamma^{\frac{1}{2}}CG(G'RG)^{-1}[G'\beta^* - h], \\ I - \Gamma^{\frac{1}{2}}CG(G'RG)^{-1}G'C\Gamma^{\frac{1}{2}}).$$

Letting

$$(33) \qquad W^* = \Gamma^{\frac{1}{2}}C(\hat{\beta}^* - \beta^*),$$

the quadratic form $W^{*\prime}W^*/\sigma^2$ is noncentral chi-square if and only if the covariance matrix of $(1/\sigma)W^*$ is idempotent.[7] And the condition holds in the present case as can be seen by squaring the covariance matrix of equation (32). The degrees of freedom of the noncentral χ^2 distribution of $W^{*\prime}W^*/\sigma^2$ is the trace of the covariance matrix which is $k - m$. The noncentrality parameter is the same λ as that of equation (4), and the average squared distance of $\Gamma^{\frac{1}{2}}C\hat{\beta}^*$ from $\Gamma^{\frac{1}{2}}C\beta^*$ is the expectation:

$$(34) \qquad EW^{*\prime}W^* = E(\hat{\beta}^* - \beta^*)'R(\hat{\beta}^* - \beta^*) = (k - m)\sigma^2 + 2\lambda\sigma^2.$$

The foregoing analysis motivates the following definition:
The restricted estimator, $\hat{\beta}$, is better in weak mean squared error if and only if

$$(35) \qquad E(\hat{\beta}^* - \beta^*)'R(\hat{\beta}^* - \beta^*) \leqslant E(b^* - \beta^*)'R(b^* - \beta^*)$$

or equivalently, if and only if

$$(36) \qquad E(\hat{\beta} - \beta)'S(\hat{\beta} - \beta) \leqslant E(b - \beta)'S(b - \beta).$$

[7] See Graybill [5, p. 84].

And from the results of equations (27) and (34), a necessary and sufficient condition for $\hat{\beta}$ to be better in weak mean squared error is

$$(37) \qquad \lambda \leqslant \frac{m}{2}.$$

This is a particularly nice result because the test statistic can be easily tabulated and to make the test the investigator need only make computations provided by most computer routines. The companion paper [**4**] provides a tabulation.

Note that for a single restriction the weak and strong mean squared error criteria coincide.

5.

Perhaps the main justification for defining betterness in weak MSE in terms of average squared distance after rotation by the design matrix is the simplicity of the criterion. However, the rotated average squared distance falls within the class of positive definite quadratic loss functions. And some people may be willing to subscribe to the notion that there is really no such thing as an "unweighted" average. Even so, it would be nice to claim that inequalities (36) or (37) insure that the unweighted average squared distance for $\hat{\beta}$ is smaller than the unweighted average squared distance for b. Such is not the case. However, for those who prefer the unweighted criterion function there is some solace in the following:

(i) The criterion $\lambda \leqslant \frac{1}{2}$ is sufficient for $\hat{\beta}$ to be better in unweighted average squared distance and a tabulation of the relevant test statistic is available [**12**].

(ii) The right-hand side of the potentially weaker sufficient condition of inequality (19) can be calculated in any real problem setting. And even though critical points corresponding to all λ values are not available in tabular form, there are approximations to the noncentral F that can be used. Such an approximation formula is provided in the companion paper [**4**], along with some evaluation of how it performs.

(iii) One can derive the weak MSE criterion of inequality (37) by concentrating on an alternative parameter space. Suppose interest is centered on estimating $X\beta$, the conditional mean of Y given X. Consider the two alternative estimators Xb and $X\hat{\beta}$ and define $X\hat{\beta}$ to be the better estimator of $E(Y|X)$ if and only if

$$(38) \qquad E(X\beta - X\hat{\beta})'(X\beta - X\hat{\beta}) \leqslant E(X\beta - Xb)'(X\beta - Xb).$$

The inequality in (38) reduces immediately to

$$(39) \qquad E(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) \leqslant E(b - \beta)'X'X(b - \beta)$$

which is equivalent to the inequalities in (36) and (37).[8] Hence, if interest centers on conditional mean forecasting, rather than the β vector per se, the criterion $\lambda \leqslant m/2$ is entirely appropriate.

[8] My thanks to Paul R. Johnson for a reference to a paper by Howard L. Jones [**8**] leading to this comment. Jones [**8**] also shows that in the MSE sense, $Y - Xb$ is always a better estimator of ε than is $Y - X\hat{\beta}$.

6.

Using the notation of this paper, Table I summarizes the various criteria and tests of linear restrictions suggested here in relation to previous practices.[9]

TABLE I

ALTERNATIVE CRITERIA AND TESTS FOR RESTRICTIONS IN LINEAR REGRESSION

Criterion	Critical value of λ	Test: Compute u (equation (3)) and compare it to the critical value of:
The set of restrictions, $H'\beta = h$, is true	$\lambda = 0$	The usual central F distribution
$\mathrm{MSE}(l'b) \geqslant \mathrm{MSE}(l'\hat{\beta})$ for all $l \neq 0$. Or equivalently, $E(b-\beta)(b-\beta)' - E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'$ = a non-negative definite matrix. (Strong MSE criterion.)	$\lambda \leqslant \frac{1}{2}$	Noncentral $F(\frac{1}{2})$. Tabulated in reference [**12**].
$E(b-\beta)'(b-\beta) - E(\hat{\beta}-\beta)'(\hat{\beta}-\beta)$ is a positive scalar. (First weak MSE criterion.)	$\lambda \leqslant \theta$	Noncentral $F(\theta)$. (Compute θ: the right hand side of (19) above. Compute probability of a larger F from approximation given in the companion paper [**4**].)[a]
$E(b-\beta)'X'X(b-\beta) - E(\hat{\beta}-\beta)'X'X \times (\hat{\beta}-\beta)$ is a positive scalar. (Second weak MSE criterion.)	$\lambda \leqslant (m/2)$	Noncentral $F(m/2)$. Tabulated in the companion paper [**4**].

[a] If $\theta < \frac{1}{2}$, use noncentral $F(\frac{1}{2})$, tabulated in Reference [**12**].

In conclusion, there is a methodological implication of the formal material presented in this paper favoring simplicity of model construction. Intuition would lead one to believe that there should be a cost to overspecification as well as underspecification of models. Indeed there is, and the cost takes the form of larger variances. It is the purpose of theory not only to count but to avoid double counting. The concentration of this paper on statistical testing is not meant to suggest that all potential restrictions ought to be tested. On the contrary—the results presented here can be interpreted to be in support of simpler specification and bold use of priors.

North Carolina State University

Manuscript received September, 1970; revision received February, 1971.

[9] My thanks to an unknown referee for suggesting the format for Table I.

REFERENCES

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

[2] Chipman, J. S., and M. M. Rao: "The Treatment of Linear Restrictions in Regression Analysis," *Econometrica*, 32 (1964), 198–209.

[3] Chow, Gregory C.: "Tests of Equality Between Sets of Coefficients in Two Regressions," *Econometrica*, 28 (1960), 591–605.

[4] Goodnight, James H., and T. D. Wallace: "Operational Techniques and Tables for Making Weak MSE Tests for Restrictions in Regressions," *Econometrica*, 40 (1972), 699–709.

[5] Graybill, Franklin A.: *An Introduction to Linear Statistical Models, Vol. I.* New York: McGraw-Hill Book Co., Inc., 1961.

[6] ———: *Introduction to Matrices with Applications in Statistics.* Belmont, California: Wadsworth Publishing Company, 1969.

[7] Hoerl, Arthur E., and Robert W. Kennard: "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics*, 12 (1970), 55–67.

[8] Jones, Howard L.: "Linear Regression Functions with Neglected Variables," *Journal of the American Statistical Association*, (1946), 356–369.

[9] Lehmann, E. L.: *Testing Statistical Hypotheses.* New York: John Wiley and Sons, Inc., 1959.

[10] Rao, C. Radhakrishna: *Linear Statistical Inference and Its Applications.* New York: John Wiley and Sons, Inc., 1965.

[11] Toro-Vizcarrondo, Carlos, 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.

[12] Wallace, T. D., and Carlos Toro-Vizcarrondo: "Tables for the Mean Square Error Test for Exact Linear Restrictions in Regression," *Journal of the American Statistical Association*, 64 (1969), 1649–1663.