

WEAKER MSE CRITERIA AND TESTS FOR LINEAR RESTRICTIONS IN REGRESSION MODELS WITH NON-SPHERICAL DISTURBANCES

Marjorie B. McELROY*

Duke University, Durham, NC 27706, USA

Received October 1974, final version received February 1977

This paper extends, in an asymptotic sense, the strong and the weaker mean square error criteria and corresponding tests to linear models with non-spherical disturbances where the error covariance matrix is unknown but a consistent estimator for it is available. The mean square error tests of Toro-Vizcarrondo and Wallace (1968) and Wallace (1972) test for the superiority of restricted over unrestricted linear estimators in a least squares context. This generalization of these tests makes them available for use with GLS, Zellner's SUR, 2SLS, 3SLS, tests of over identification, and so forth.

Wallace (1972) and Toro-Vizcarrondo and Wallace (1968) recently proposed three mean square error criteria and associated tests for the superiority of restricted over unrestricted least squares estimators. This paper extends these tests in an asymptotic sense to linear models with non-spherical disturbances where the error covariance matrix is unknown, but a consistent estimator for it is available. This setup would arise, for example, in the context of seemingly unrelated regressions where one was considering constraining the coefficients to be equal across equations. Except where noted the motivation and interpretation for these criteria are the same for non-spherical as for spherical disturbances. Likewise, the statistical consequences of using parameter estimates chosen sequentially on the basis of these criteria may be serious. For a discussion of the latter problems in the spherical case, see, for example, Judge, Bock and Yancey (1974) and Bock, Yancey and Judge (1973).

The linear model with nonspherical disturbances may be written as

$$y = X\beta + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2 \Sigma). \quad (1)$$

The q linear restrictions on β under consideration are

$$R\beta = r. \quad (2)$$

*The author wishes to express thanks to A.R. Gallant and T.D. Wallace for useful discussions, and two unknown referees for helpful criticism.

Here y is $n \times 1$, X is $n \times k$ and fixed at least conditionally, β is $k \times 1$ fixed and unknown, and ε is an $n \times 1$ vector of jointly normal disturbances with mean vector zero and positive definite covariance matrix Σ . R is $q \times k$, known and of rank q , and r is $q \times 1$ and known. There exists a non-singular P such that $P'P = \Sigma^{-1}$, and premultiplication of (1) by P gives, say

$$y^* = X^*\beta + \varepsilon^*, \quad \varepsilon^* = P\varepsilon \sim N(0, \sigma^2 I). \quad (3)$$

The unrestricted and restricted LS estimators of β in (3) are identical to the unrestricted and restricted Aitken estimators of β in (1). Thus when Σ is known, all of the MSE tests for LS estimators in Toro-Vizcarrondo and Wallace (1968) and Wallace (1972) have their immediate counterpart for Aitken estimators.

The remainder of this paper is devoted to showing that when Σ is unknown, but a consistent estimator of Σ is available,¹ all of the MSE tests may be used approximately for finite samples, the justification being based on the limiting distributions. It is convenient to first summarize the results for finite n and known Σ .

Defining $V_n = n(X'\Sigma^{-1}X)^{-1}$ for all n , the unrestricted Aitken estimator b for β in (1), its distribution, and its MSE are given, respectively, by

$$b = V_n \frac{X'\Sigma^{-1}y}{n}, \quad (4a)$$

$$\sqrt{n} \cdot (b - \beta) \sim N(0, \sigma^2 V_n), \quad (4b)$$

$$\text{MSE}(\sqrt{n} \cdot b) = \sigma^2 V_n, \quad (4c)$$

where b minimizes the sum of squared transformed residuals $P(y - Xb)$, i.e.,

$$\text{SSE}(b) = (y - Xb)' \Sigma^{-1} (y - Xb), \quad (5)$$

with distribution given by

$$\frac{1}{\sigma^2} \text{SSE}(b) \sim \chi^2(n - k). \quad (6)$$

Defining $\delta_n = \sqrt{n} \cdot (R\beta - r)$, the restricted Aitken estimator $\hat{\beta}$ for β in (1), its distribution and MSE are given, respectively, by

$$\hat{\beta} = b - V_n R'(RV_n R')^{-1} (Rb - r), \quad (7a)$$

¹For a summary of useful cases see Theil (1971, ch. 8.6).

$$\begin{aligned} \sqrt{n} \cdot (\hat{\beta} - \beta) &\sim N(-V_n R' (R V_n R')^{-1} \delta_n, \\ &\sigma^2 V_n [I - R' (R V_n R')^{-1} R V_n]), \end{aligned} \tag{7b}$$

$$\begin{aligned} \text{MSE}(\sqrt{n} \cdot \hat{\beta}) &= V_n R' (R V_n R')^{-1} \delta_n \delta_n' (R V_n R')^{-1} R V_n \\ &+ \sigma^2 V_n [I - R' (R V_n R')^{-1} R V_n], \end{aligned} \tag{7c}$$

where, subject to $\delta_n = 0$, $\hat{\beta}$ minimizes the sum of squared transformed residuals,

$$\text{SSE}(\hat{\beta}) = (y - X\hat{\beta})' \Sigma^{-1} (y - X\hat{\beta}). \tag{8}$$

Let $\lambda_n = \delta_n' (R V_n R')^{-1} \delta_n / 2\sigma^2$. The four null hypotheses of interest are: (i) $\sqrt{n} \cdot (R\beta - r) = 0$ is 'true' which holds iff $\lambda_n = 0$; (ii) $\hat{\beta}$ is better than b by the *strong MSE criterion* defined here as for all non-null $q \times 1$ vectors l ,

$$\begin{aligned} l' [\text{MSE}(\sqrt{n} \cdot b) - \text{MSE}(\sqrt{n} \cdot \hat{\beta})] \\ \equiv l' [V_n R' (R V_n R')^{-1} (\sigma^2 R V_n R' - \delta_n \delta_n') (R V_n R')^{-1} R V_n] l \geq 0, \end{aligned} \tag{9}$$

which holds iff $\lambda_n \leq 1/2$; (iii) $\hat{\beta}$ is better than b by the *first weak MSE criterion* defined here as

$$\begin{aligned} \text{tr} [\text{MSE}(\sqrt{n} \cdot b) - \text{MSE}(\sqrt{n} \cdot \hat{\beta})] \\ \equiv \text{tr} [\sigma^2 V_n R' (R V_n R')^{-1} R V_n - V_n R' (R V_n R')^{-1} \delta_n \delta_n' (R V_n R')^{-1} R V_n] \geq 0, \end{aligned} \tag{10}$$

which holds whenever

$$\lambda_n \leq \frac{1}{2} \mu_n \text{tr} [V_n R' (R V_n R')^{-1} R V_n] = \theta_n,$$

say, where μ_n is the smallest characteristic root² of V_n^{-1} , (iv) and $\hat{\beta}$ is better than b by the *second weak MSE criterion* defined here as

$$E[n(b - \beta)' V_n^{-1} (b - \beta) - n(\hat{\beta} - \beta)' V_n^{-1} (\hat{\beta} - \beta)] \geq 0, \tag{11}$$

which holds iff $\lambda_n \leq q/2$. In each case the alternative hypothesis is simply the negation of the null hypothesis. Note that for generalized LS this last criterion compares $\hat{\beta}$ to b not in terms of the average squared distance of the predicted y to the conditional mean of y (that is to $X\beta$) as in the case of LS, but rather

²For simplicity of exposition we neglect the tighter bound found by Yancey, Bock and Judge (1973), although it presents no difficulty for the non-spherical case.

compares the quadratic form in these vectors – the matrix of the quadratic form being Σ^{-1} .

Defining $Q(\Sigma) = \text{SSE}(\hat{\beta}) - \text{SSE}(b)$, the relevant test statistic is

$$\begin{aligned}
 U_n(\Sigma) &= \frac{Q(\Sigma)/q}{\text{SSE}(b)/(n-k)} \\
 &= \frac{n(Rb-r)'(RV_nR')^{-1}(Rb-r)/q}{\text{SSE}(b)/(n-k)} \sim F(q, n-k, \lambda_n), \tag{12}
 \end{aligned}$$

where the non-centrality parameter is given by

$$\lambda_n = \frac{1}{2\sigma^2} \delta_n'(RV_nR')^{-1} \delta_n. \tag{13}$$

Rejecting the null when $U_n(\Sigma)$ is too big provides a UMP test for $\lambda \leq \lambda^*$ against $\lambda > \lambda^*$ where $\lambda^* = \frac{1}{2}, \theta_n$, or $q/2$ according to the chosen MSE criterion.³

To examine the limiting distributions we assume that

$$\lim_{n \rightarrow \infty} n(X' \Sigma^{-1} X)^{-1} = \lim_{n \rightarrow \infty} V_n = V, \tag{14}$$

where V is positive definite and also that

$$\lim_{n \rightarrow \infty} \delta_n = \lim_{n \rightarrow \infty} \sqrt{n} \cdot (R\beta - r_n) = \delta, \tag{15}$$

where δ_n is now defined by $\delta_n = \sqrt{n} \cdot (R\beta - r_n)$, and δ is a fixed finite vector. This in turn insures that

$$\lim_{n \rightarrow \infty} \lambda_n = \frac{1}{2\sigma^2} \delta'(RVR')^{-1} \delta = \lambda, \text{ say.} \tag{16}$$

Since V is positive definite, λ is non-negative and equal to zero iff $\delta = 0$.

To motivate assumption (15) we note that little is revealed about the asymptotic performance of the tests when the restriction $R\beta - r = 0$ remains unchanged

³As noted in Toro-Vizcarrondo and Wallace (1968), for the spherical case the family of non-central F densities has the monotone likelihood ratio property, and thus the test of $\lambda \leq \frac{1}{2}$ is UMP against $\lambda > \frac{1}{2}$. As pointed out in conversation by Ron Gallant, this makes the strong MSE test UMP invariant rather than UMP. By the same token, the test of $\lambda \leq \lambda^*$ is UMP invariant. The first weak MSE criterion does not have this property because $\lambda \leq \theta$ is only sufficient for the criterion and thus the size of the test will be greater than α . All of these results for the spherical case carry over in a straightforward manner to the non-spherical case with known Σ .

as the sample size increases. As $n \rightarrow \infty$ the covariance matrices for both $\sqrt{n} \cdot \hat{\beta}$ and $\sqrt{n} \cdot b$ converge to constant matrices and, for a fixed but 'untrue' restriction, δ_n and hence the bias of $\sqrt{n} \cdot \hat{\beta}$ increases without bound while $\sqrt{n} \cdot b$ remains unbiased – see (7). Thus, for each of the three MSE criteria, there is a sufficiently large n to make $\hat{\beta}$ worse than b . This is reflected in the distribution of the test statistic U_n in that the non-centrality parameter λ_n increases without bound. Consequently, the power of a test of $\lambda_n \leq \lambda^*$ tends to one, i.e., the tests of hypotheses (i) through (iii) are consistent. Since these results hold for $\hat{\beta}_n$ corresponding to any fixed but 'untrue' restriction, in the limit these tests are unable to distinguish restricted estimators based on one 'untrue' restriction from those based on any other 'untrue' restriction. Hence, the asymptotic properties of the tests are best revealed by considering sequences of restricted estimators which these tests can, in the limit, distinguish – those corresponding to sequences of restrictions for which r_n converges to $R\beta$ at rate $1/\sqrt{n}$, i.e., under assumption (15).⁴

Under the assumptions (14) and (15), the limiting distributions of b and $\hat{\beta}$, the MSE's of these limiting distributions, as well as the three MSE criteria for these limiting distributions may be written down by replacing V_n , δ_n , and λ_n with V , δ , and λ , respectively, in eqs. (4), (7), and (9) through (13).

When Σ is replaced by a consistent estimator, say S , on the right-hand sides of (4a) and (7a), we obtain, say, \tilde{b} and $\tilde{\beta}$ – the unrestricted and restricted consistent generalized least squares estimators, respectively. It is well known that $\sqrt{n} \cdot (\tilde{b} - \beta)$ and $\sqrt{n} \cdot (b - \beta)$ have identical limiting distributions and therefore their limiting distributions have the same moments and MSE's. Thus asymptotically,

$$\sqrt{n} \cdot (\tilde{b} - \beta) \sim N(0, \sigma^2 V), \quad \text{MSE}(\sqrt{n} \cdot \tilde{b}) = \sigma^2 V. \tag{17}$$

Similar results obtain for the restricted estimators: $\sqrt{n} \cdot (\tilde{\beta} - \beta)$ converges in distribution to $\sqrt{n} \cdot (\hat{\beta} - \beta)$ and asymptotically,

$$\begin{aligned} \sqrt{n} \cdot (\tilde{\beta} - \beta) &\sim N(-VR'(RVR')^{-1}\delta, \\ &\sigma^2 V[I - R'(RVR')^{-1}RV]), \end{aligned}$$

and

$$\begin{aligned} \text{MSE}(\sqrt{n} \cdot \tilde{\beta}) &= VR'(RVR')^{-1}\delta\delta'(RVR')^{-1}RV \\ &+ \sigma^2 V[I - R'(RVR')^{-1}RV]. \end{aligned} \tag{18}$$

⁴This rate insures the power of the tests neither tends to α for $\lambda^* = 0$ (and to numbers even less than α for $\lambda^* = \frac{1}{2}, \theta_n$, or $q/2$) because convergence is too fast (i.e., $\lambda \rightarrow 0$), nor tends to one because convergence is too slow (i.e., $\lambda \rightarrow \infty$). Assumption (15) is similar to the usual ones found in discussions of the asymptotic properties of the power of a test [e.g., Lehmann (1959, p. 305) and Noether (1955)].

It follows that in terms of the limiting distributions the strong, first weak, and second weak MSE criteria may be stated as hypotheses about λ by again replacing Σ by S and also by replacing V_n , δ_n , and λ_n by V , δ , and λ , respectively in (9), (10), and (11) and in the paragraph in which these equations appear. Likewise, in (12), $U_n(S)$ has the same limiting distribution as $U_n(\Sigma)$, namely $q^{-1}\chi^2(q, \lambda)$. For any given large but finite sample, if we identify r with r_n and λ with λ_n , these results suggest the approximate test, reject $\lambda \leq \lambda^*$ when

$$U(S) = \frac{n-k}{q} \frac{(y - X\tilde{\beta})' S^{-1} (y - X\tilde{\beta}) - (y - X\tilde{b})' S^{-1} (y - X\tilde{b})}{(y - X\tilde{b})' S^{-1} (y - X\tilde{b})} \geq u_\alpha, \quad (19)$$

where, as in the spherical case, u_α is chosen in accordance with the particular λ^* in such a way as to give, in the limit, a size α critical region.

Thus when Σ is known the UMP property obtains for all three tests on λ . When Σ is replaced by S the tests proposed above are asymptotically UMP against $\lambda > \lambda^*$ in the sense that in the limit the power of each is the same as the power of the UMP test for known Σ and, therefore, in the limit no penalty is paid for having to estimate Σ .⁵ This in turn makes the proposed strong and second weak MSE tests asymptotically UMP invariant.

⁵Note that this is a different concept of asymptotically most powerful from that of Wald [Kendall and Stuart (1967, pp. 263–265)].

References

- Bock, M.E., T.A. Yancey and G.G. Judge, 1973, Statistical consequences of preliminary test estimators in regression. *Journal of American Statistical Association* 68, 109–116.
- Judge, G.G., M.E. Bock and T.A. Yancey, 1974, Post data model evaluation, *Review of Economics and Statistics* 56, 145–253.
- Kendall, M.G. and A. Stuart, 1967, *The advanced theory of statistics*, 2nd ed., vol. 2 (Charles Griffen, London).
- Lehmann, E.L., 1959, *Testing statistical hypotheses* (Wiley, New York).
- Noether, G.E., 1955, On a theorem of Pitman, *Annals of Mathematical Statistics* 26, 64–68.
- Theil, H., 1971, *Principles of econometrics* (Wiley, New York).
- 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* 63, 558–572.
- Wallace, T.D., 1972, Weaker criteria and tests for linear restrictions in regression, *Econometrica* 40, 689–698.
- Yancey, T.A., G.G. Judge and M.E. Bock, 1973, Wallace's weak mean square error criterion for testing linear restrictions in regression: A tighter bound, *Econometrica* 41, 1203–1206.