1985

On Robustness of Tests of Linear Restrictions in Regression Models with Elliptical Error Distributions

Victoria Zinde-Walsh

Aman Ullah

THE CENTRE FOR DECISION SCIENCES AND ECONOMETRICS

On Robustness of Tests of Linear Restrictions in Regression Models with Elliptical Error Distributions

Victoria Zinde-Walsh
Aman Ullah

TECHNICAL REPORT NO.8
NOVEMBER 1985

Centre For Decision Sciences And Econometrics
Social Science Centre
The University of Western Ontario
London, Ontario N6A 5C2

ON ROBUSTNESS OF TESTS OF LINEAR RESTRICTIONS IN REGRESSION MODELS WITH ELLIPTICAL ERROR DISTRIBUTIONS

by

Victoria Zinde-Walsh and Aman Ullah*

University of Western Ontario

*The authors gratefully acknowledge research support from SSHRC and NSERC, respectively. They are also thankful to B. K. Sinha and the members of the U.W.O. Econometrics Workshop for useful comments and discussions on the subject matter of this paper.

ON ROBUSTNESS OF TESTS OF LINEAR RESTRICTIONS IN REGRESSION MODELS WITH ELLIPTICAL ERROR DISTRIBUTIONS

by

Victoria Zinde-Walsh and Aman Ullah

1. Introduction

Testing a set of linear restrictions in a regression model is usually performed with the help of the F-statistic, or the statistic based on the likelihood ratio (LR). More recently two other procedures, the Lagrangian Multiplier or Rao Score (RS) test due to Rao (1947) and Silvey (1959), and the Wald (W) test (1943), have become popular with econometricians; see, for example, Breusch and Pagan (1980) and Evans and Savin (1982).

A statistic can be called numercially robust over a class of error distributions if its values are independent of the specific error distribution from that class. If the statistic is such that no matter which error distribution from the class of distributions is considered, the test criterion remains unchanged then the statistic is inferentially robust over that class.

If the statistics F, LR, RS and W are constructed based on the assumption of the spherical normal error distribution (normal error with the covariance matrix $\sigma^2 I$), then F and LR are numerically robust against the class of all monotonically decreasing continuous spherical distributions, but RS and W are not. However, all these statistics are inferentially robust over this class, thus the test conclusions reached under the assumption of normality will not be overturned if the error distribution is spherical. These results are derived by Ullah and Zinde-Walsh (1984), (1985).

In this paper we consider the issues of numerical and inferential robustness of F, LR, RS and W tests, based on the assumption of spherical normality, against the general class of elliptical error distributions (errors with the nonscalar covariance matrix Σ). We provide the necessary and sufficient conditions of numerical robustness for the class of covariance matrices often used in econometrics, for example, autoregressive (AR), moving average (MA) and heteroscedasticity. Our investigation shows that for these covariance matrices the numerical robustness of test statistics under consideration is rare. Our results are more general than those given in Ghosh and Sinha (1980) and Sinha and Mukhopadhyaya (1981) who consider only intra-class covariance structure. Also, while Khatri (1981) gave conditions for numerical robustness in terms of pairs of data and covariance matrices, robustness over classes of covariance matrices considered here has not been examined in his paper.

Our investigation also showed the limitations of exact inferential robustness. We therefore looked into the robustness of tests by developing bounds for critical values which will ensure that the conclusions based on the usual tests are not affected against a particular class of distributions. Bounds for critical values of test statistics for t and F-tests for first-order AR, MA and ARMA processes have been tabulated (for normal errors) by Vinod (1976), Vinod and Ullah (1981) and Kiviet (1980). Their calculations involve the knowledge of all the eigenvalues of the matrices which characterize these processes and are quite complex. The situation becomes more complicated for higher order ARMA processes. Our method offers bounds which are cruder for the specific processes considered by Vinod and Ullah and Kiviet, but they have the advantage of calculational simplicity and generality, i.e., they provide critical values that guarantee robustness of the test conclusions, for any Σ matrix, over wide classes of error distributions, and would utilize only the highest and lowest eigenvalues of the covariance matrix.

The plan of the paper is as follows. Section 2 develops the notation and definitions. Section 3 deals with the problem of numerical robustness and some applications. In Section 4 we examine the question of robustness of test conclusions and provide our bounds on the critical values of statistics. Finally, the proofs of the lemmas and theorems are presented in Section 5.

2. Definitions and Notation

We consider the general linear regression model

(2.1) $$y = X\beta + u$$

where y is an nX1 vector of observations, X is an nXp known matrix of rank $p < n$, β is a pX1 unknown parameter vector and u is an nX1 disturbance vector such that

(2.2) $$\phi(u) = \sigma^{-n} |\Sigma|^{-½} f\left(\frac{u'\Sigma^{-1}u}{\sigma^2}\right)$$

with a monotonically decreasing f, and positive definite Σ. If $\Sigma = I$, the distribution (2.2) reduces to a spherically symmetric distribution.

Our problem is to test a set of r linear restrictions $H_0{:}R\beta = 0$ against $H_1{:}R\beta{:} \neq 0$, where R is an rXp known matrix. Under this hypothesis, following

Ghosh and Sinha (1980, p. 338), we can rewrite the model as

(2.3) $$y = X_o\beta_o + u,$$

where X_o is an nXm known matrix of rank $m < p$.

We denote by F, LR, RS, and W the values of the statistics calculated according to the usual formulae under the assumption of multivariate normality of u. F can be written as

(2.4) $$F = \left[\frac{(y-X_o\hat{\beta}_o)'(y-X_o\hat{\beta}_o)}{(y-X\hat{\beta})'(y-X\hat{\beta})} - 1\right]\frac{q}{r}, \qquad q = n-p$$

where $\hat{\beta}_o$ and $\hat{\beta}$ are the respective least squares estimators of β_o and β, and LR, RS and W can be expressed through F, respectively, as

(2.5) $$LR = n \log(1 + \frac{r}{q} F), \quad RS = n \frac{r}{q} F/(1+ \frac{r}{q} F), \quad W = n \frac{r}{q} F.$$

We introduce the following projection matrices:

(2.6) $$P = X(X'X)^{-1}X'; \; A = I-P; \; P_o = X_o(X_o'X_o)^{-1}X_o'; \; A_o = I-P_o.$$

where I is the identity matrix. The following properties can be easily verified:

(2.7) rank P > rank P_o; rank A < rank A_o; $PP_o = P_o$;

$AA_o = A$; $PX = X$; $PX_o = X_o$.

Using (2.6) and noting that $(1 + \frac{r}{q} F) = \frac{u'A_o u}{u'Au}$ we can rewrite (2.4) and (2.5) as

(2.8)

$$F = \frac{u'(A_o-A)u}{u'Au}\frac{q}{r}; \qquad LR = n \log \frac{u'A_o u}{u'Au};$$

$$RS = n \frac{u'(A_o-A)u}{u'A_o u}; \qquad W = n \frac{u'(A_o - A)u}{u'Au}.$$

If u is in fact distributed as spherical normal, all the statistics have the known distributions. If the error distribution is spherical, i.e. is given by (2.2) with $\Sigma = I$, we denote the values of the appropriate corresponding statistics by F_ϕ, LR_ϕ, RS_ϕ and W_ϕ.

If the distribution of u is elliptically normal, we denote the statistics by F_Σ, LR_Σ, RS_Σ and W_Σ.

It is known that

$$(2.9)\qquad F_\Sigma = \left[\frac{(y-X_o\hat{\beta}_{o\Sigma})'\Sigma^{-1}(y-X_o\hat{\beta}_{o\Sigma})}{(y-X\hat{\beta}_\Sigma)'\Sigma^{-1}(y-X\hat{\beta}_\Sigma)} - 1\right]\frac{q}{r}$$

is the familiar F-statistic for testing H_o, with

$$(2.10)\qquad \hat{\beta}_{o\Sigma} = (X_o'\Sigma^{-1}X_o)^{-1}X_o'\Sigma^{-1}y; \quad \hat{\beta}_\Sigma = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y.$$

Further, as in (2.5) we have

$$(2.11)\qquad LR_\Sigma = n\log(1 + \frac{r}{q}F_\Sigma); \; RS_\Sigma = n\frac{r}{q}F_\Sigma/(1 + \frac{r}{q}F_\Sigma); \; W_\Sigma = n\frac{r}{q}F_\Sigma.$$

For a general elliptical distribution in (2.2) denote the appropriate statistics by $F_{\phi,\Sigma}$, $LR_{\phi,\Sigma}$, $RS_{\phi,\Sigma}$, $W_{\phi,\Sigma}$. Ullah and Zinde-Walsh (1984, 1985) have analyzed the numerical robustness of LR, RS and W tests against spherically symmetric distributions. In particular, they have shown that

$$(2.12)\qquad LR = LR_\phi; \; RS = \psi_\phi^{-1}RS_\phi; \; W = \rho_\phi^{-1}W_\phi,$$

where ψ_ϕ and ρ_ϕ are constants which depend on the spherical distribution $\phi(u)$. Thus LR is numerically robust but RS and W are not.

The elliptical distribution (2.2) can be transformed into a spherical by the substitution $u = \Sigma^{1/2}v$. Thus for this case, from (2.12) we easily obtain

$$(2.13)\qquad LR_\Sigma = LR_{\phi,\Sigma}; \; RS_\Sigma = \Psi_\phi^{-1}RS_{\phi,\Sigma}; \; W_\Sigma = \rho_\phi^{-1}W_{\phi,\Sigma}.$$

Here LR_Σ is numerically robust against non-normality, but RS_Σ and W_Σ are not .

3. Main Results on Numerical Robustness

It was mentioned in Section 2 that Ullah and Zinde-Walsh (1984, 1985) analyzed the numerical robustness of F, LR, RS and W against spherically symmetric distributions and of F_Σ, LR_Σ, RS_Σ and W_Σ against elliptically symmetric distributions. Here we look into robustness of F, LR, RS and W (under spherical normal) against elliptical normal distributions by comparing the values of these statistics with the values F_Σ, LR_Σ, RS_Σ and W_Σ as in (2.9)-(2.11).

Conclusions about robustness against general elliptical distributions will follow in view of the relationships (2.13). We also note that we derive the results for parametric classes of Σ matrices often used in economic literature.

For deriving the conditions under which F_Σ (or LR_Σ) is numerically robust over some class of Σ given the data matrix X, we consider

$$(3.1) \qquad \ell_\Sigma = 1 + \frac{r}{q} F_\Sigma$$

and examine the conditions under which $\ell_\Sigma = \ell_o$ for $\ell_o = 1 + \frac{r}{q} F$.

Consider a class $\Omega\rho$ of matrices Σ with $\Sigma^{-1} = (I-H_\rho)$ where H_ρ is some symmetric matrix over some parameter space $B \in R^k$, $\rho = (\rho_1,\ldots,\rho_k) \in B$ with $H_o = 0$ for $\rho_1 = \ldots = \rho_k = 0$ and $|y'H_\rho y| < y'y$ for all possible y, $\rho \in B$.

We now state the following lemmas which are used in the proof of Theorem 1.

Lemma 1

Over all $\Sigma \in \Omega_\rho$, ℓ_Σ can be represented as

$$(3.2) \qquad \ell_\Sigma = \frac{y'A_o y + y'A_o T_\rho^o A_o y}{y'Ay + y'AT_\rho Ay}$$

where

$$(3.3) \qquad T_\rho = \sum_i^\infty \Big[\sum_{i_1 + \ldots + i_{m+1} = i} (-1)^{m+1} H_\rho^{i} A H^{i_1} \ldots A H^{i_{m+1}} \Big]$$

and T_ρ has a similar representation with A_o replacing A in (3.3), with A and A_o defined by (2.6).

Lemma 2

Suppose that for some symmetrix matrix T

$$(3.4) \qquad (y'A_o T A_o y) \cdot (y'Ay) = (y'ATAy)(y'A_o y).$$

Then $A_o T A_o = \theta A_o$; $TA = \theta A$, where θ is some constant.

Theorem 1

Suppose that H_ρ is a polynomial or a convergent series in the parameters $\rho_1, \ldots, \rho_k$ with symmetric matrices as coefficients. Then if $T(r_1, \ldots, r_k)$ is the coefficient of $\rho^{r_1}, \ldots, \rho^{r_k}$ in $\sum_{r=1}^m H_\rho^r$, $m = \Sigma r_i$ and if $\ell_\Sigma = \ell_o$ it follows that

$$(3.5) \qquad A_o T(r_1, \ldots, r_k) A_o = \theta(r_1, \ldots, r_k) A_o$$

for some constant $\theta(r_1, \ldots, r_k)$.

For proofs of the lemmas and Theorem 1 see Section 5.

<u>Remark 1</u> <u>Suppose that</u> $A_o H_\rho^k A_o = \theta_{k,\rho} A_o$, k=1,2,..., <u>where</u> $\theta_{k,\rho}$ <u>is a scalar function of</u> $\rho_1, \ldots, \rho_k$. <u>Then</u> $\ell_\Sigma \equiv \ell_o$.

To prove the above statement one only needs to note that

$$AH_\rho^k A = AA_o H_\rho^k A_o A = A\theta_{k,\rho} A_o A = \theta_{k,\rho} A$$

and to substitute into (3.3), (3.2).

Theorem 1 and Remark 1 give the necessary and sufficient conditions for the constancy of ℓ_Σ and therefore for the numerical robustness of F, RS, LR and W statistics against elliptically normal errors that can be described by a variance-covariance matrix $\Sigma \in \Omega_\rho$, $\Omega_\rho = \{\Sigma | \Sigma^{-1} = I - H_\rho$ with H_ρ being a polynomial or convergent series$\}$.

The stringency of these conditions makes numerical robustness an exception rather than the rule. No process with non-trivial H_ρ gives rise to robust statistics for all possible X and X_o, therefore numerical robustness has applications mainly for experimental design. Also, of course, one can always check if the observation matrix X just happened to lead to statistics numerically robust against a particular process in the errors, but if so it would be strictly a matter of luck. We show that our results generalize those on experimental design with intraclass covariance structure by Ghosh and Sinha (1980) and examine the possibilities for numerical robustness over heteroscedastic and ARMA processes.

3.1 <u>Implications for Intraclass Covariance Structures</u>

The result of Ghosh and Sinha (1980, Theorem 3.1) follows as a special case of our theorem. Indeed, they considered $\Sigma = (1-\rho)I + \rho\, 1_n \times 1_n'$, $-\frac{1}{n-1} < \rho < 1$, where 1_n is a column vector of ones, and hence

$1_n \times 1_n' = nQ$, where Q is a projector of rank 1 onto the subspace spanned by 1_n. Here $\Sigma^{-1} = \frac{1}{1-\rho}[I - \frac{\rho n}{(1+(n-1)\rho)} Q]$. Direct application of Theorem 1 to Σ^{-1} implies that $A_o Q A_o = \theta A_o$. Since rank $A_o >$ rank $A \geq 1$, it follows that $\theta = 0$, $A_o 1 \times 1' A_o = 0$, $A_o 1 = 0$, therefore 1_n is the eigenvector of both P_o and P as stated in the result of Ghosh and Sinha. It is just as easy to verify that Theorem 3.2 of the same paper follows from our results.

3.2 Implications for Heteroscedastic Errors

Theorem 1 also provides a characterization of the class of heteroscedastic Σ for a given A and restriction R over which $\ell_\Sigma = \ell_o$: it is required that $\Sigma = \gamma_o I + \Lambda$, where the diagonal matrix Λ is such that $\Lambda A_o = 0$, this implies that the A matrix has a block-diagonal structure with a block of zeros.

3.3 Implications of Theorem 1 for Autoregressive (AR) Error Structures

The matrix Σ^{-1} is known for autoregressive processes of order k,AR(k). If we set all but one of the parameters of AR(k) – $\rho_1, \rho_2, \ldots, \rho_k$ equal to zero: $\rho_k \neq 0$, $\rho_i = 0$ for $i \neq k$, then Σ^{-1} reduces to the matrix $I + \rho C_{1k} + \rho^2 C_{2k}$. Here C_{1k} is the matrix with elements $(C_{1k})_{ij}$ equal to -1 if $|i-j| = k$ and 0 otherwise, and C_{2k} is a diagonal matrix with elements $(C_{2k})_{ij}$ equal to -1 if $k < i = j < n - k$ and 0 otherwise. We shall denote this process AR(k,0). A necessary condition for constancy of ℓ_Σ for AR(k,0) is that $A_o C_{1k} A_o = \theta A_o$, where A_o is a projector of rank no less than 2. This implies that C_{1k} should have at least two identical eigenvalues, which is true only if $k \geq \frac{n}{2} + 1$. If $k \geq \frac{n}{2} + 1$ then C_{1k} has a kernel of dimension $n - 2(n-k) = 2k - n \geq 2$. In this case ℓ_Σ is constant for all A_o that project into the intersection of the kernel of the matrix C_{1k} and of either the image or the kernel of C_{2k}. Then, of course, $\theta = 0$ and $A_o C_{2k} A_o = \gamma A_o$ with $\gamma = 0$ or 1. It is not hard to check that for these k and A_o this suffices for constancy of ℓ_Σ.

3.4 Implications for the Moving Average Error Structure

The class of MA(k,0) error structures where all but the parameters of order ℓ are zeros is represented by $\Sigma = \sigma^2[(1+w^2)I + wC_{1k}]$, where C_{1k} is the same as for the AR(k) structure. We denote $\sigma^2(1 + w^2)$ by γ and $w/(1+w^2)$ by ρ. Theorem 1 can be directly applied to this class of Σ after noting that $\Sigma^{-1} = (1+\gamma^2)(I+\rho C_{1k}+\rho^2 C_{1k}^2+\ldots)$. For this Σ, ℓ_Σ is robust only if $A_o C_{1k} A_o = \theta A_o$. It can be easily seen that for the particular C_{1k} in the MA error structure $|w| \leq 2$ always. Therefore the theorem applies to all ρ such that $|\rho| < 1/2$. Since $\rho = w/(1+w^2)$, this condition on ρ is satisfied for any w. So the condition $A_o C_{1k} A_o = \theta A_o$ is necessary and sufficient for robustness over all possible MA error structures.

Here again if $k \geq \frac{n}{2} + 1$, matrices A_o which result in robust test statistics, exist. Such an A_o would project onto the kernel of C_{1k}.

Thus we conclude that there are some data structures that produce statistics that are robust over AR and MA error processes of sufficiently high orders (which do not include lower order components). We also notice that the higher the order of the error process the larger the class of data matrices that give robust statistics. This is hardly surprising since in the limiting case processes of order higher than the dimensions of the data will not affect the statistics at all.

We also note that in general the larger the number of equal eigenvalues, including zeros, of H (or the larger the dimension of any projector in the canonical representation of the symmetric matrix H) the more possibilities for numerically robust statistics.

Note that if $\ell_\Sigma = \ell_o$, then $F = F_{\phi,\Sigma}$ and $LR = LR_{\phi,\Sigma}$, but unless the distribution is elliptical normal $RS \neq RS_{\phi,\Sigma}$ and $W \neq W_{\phi,\Sigma}$.

4. Inferential Robustness and Bounds on Critical Values

If two test statistics are such that one is a monotonic function of the other, then any probabilistic statement about one implies a similar statement about the other. Thus if one falls beyond a critical value for some level of the test, so does the other. Therefore, as was stated in Ullah and Zinde-Walsh (1985) (and can be seen immediately from (2.12)) RS and W are inferentially robust over the class of all spherical monotonic error distributions.

Here we examine the inferential robustness of the test statistics F, LR, RS and W, calculated under the assumption of spherical normality, for general elliptic distributions. To emphasize this we denote the statistics by $F(\Sigma)$, $LR(\Sigma)$, $RS(\Sigma)$ and $W(\Sigma)$. Since the test statistics are inferentially robust against spherical distributions it will not make any difference to our conclusions whether the statistics bear the subscript ϕ or not.

Consider the variate $S(\Sigma) = F(\Sigma)\frac{r}{q}$, where

$$(4.1) \qquad S(\Sigma) = \frac{u'A_1u}{u'A_2u}$$

with $A_1 = A_o - A$, $A_2 = A$ as defined in (2.6). The critical values for $S(\Sigma)$ depend on the matrix Σ. Indeed, consider the transformation $u = \Sigma^{1/2}v$, then

$$(4.2) \qquad S(\Sigma) = \frac{v'\Sigma^{1/2}A_1\Sigma^{1/2}v}{v'\Sigma^{1/2}A_2\Sigma^{1/2}v}$$

where v is spherically symmetric. Denote S(I) by S.

We observe that as long as $S(\Sigma)$ is inferentially robust over a class Ω of Σ matrices all the statistics $F(\Sigma)$, $LR(\Sigma)$, $RS(\Sigma)$ and $W(\Sigma)$ are inferentially robust over Ω as well. We assume that $I \in \Omega$.

Denote by $O(n)$ the group of orthogonal nXn matrices in the Euclidean space R^n. For any $T \in O(n)$ the distribution of S and of $S(\Sigma)$ in (4.2) is invariant with respect to the transformation $T: R^n \to R^n$.

Lemma 3:

For a positive definite matrix $\Sigma^{1/2}$ and any two mutually orthogonal idempotent matrices A_1, A_2 there exists $T \in O(n)$ such that $\Lambda_i = T'\Sigma^{1/2}A_i\Sigma^{1/2}T$ is a diagonal matrix for $i=1,2$.

Proof: See Section 5.

This lemma allows us to rewrite $S(\Sigma)$ by substituting $w = Tv$ as

$$(4.3) \qquad S(\Sigma) = \frac{w'\Lambda_1 w}{w'\Lambda_2 w},$$

where we can write

$$(4.4) \qquad \begin{aligned} \Lambda_1 &= \mathrm{diag}(\mu_1,\dots,\mu_k, 0,..0), \qquad \mu_1 \le \dots \le \mu_k, \qquad k=p-m \\ \Lambda_2 &= \mathrm{diag}(0,..0, \mu_{p+1},\dots,\mu_n) \qquad \mu_{p+1} \le \dots \le \mu_n \end{aligned}$$

where diag(...) denotes a diagonal matrix with given diagonal elements.

A similar transformation for S yields

$$(4.5) \qquad S = \frac{w'Q_1 w}{w'Q_2 w}$$

with $Q_1 = \mathrm{diag}(1,..1,0,..0)$, where the first k elements equal 1, and $Q_2 = \mathrm{diag}(0,..,0,1,..1)$ where the last n-p elements equal 1. Note that the transformation of S may be performed with a matrix from O(n) different from

T, but the distributions of $S(\Sigma)$ and S are not affected by an orthogonal transformation of the spherical variable.

Clearly the following inequality holds

$$\frac{\mu_1}{\mu_n} S \leq S(\Sigma) \leq \frac{\mu_k}{\mu_{p+1}} S.$$

It follows from (4.3) that all the values for $S(\Sigma)$ within the bounds given by (4.6) are realized for some w. Therefore a sufficient condition for inferential robustness is that $\mu_1/\mu_2 = \mu_k/\mu_p$.

However, this type of condition is hardly less restrictive than those demanded for numerical robustness.

We thus seek bounds on the critical values of the statistics $F(\Sigma)$, $LR(\Sigma)$, $RS(\Sigma)$, $W(\Sigma)$ which will assure the test conclusions over some class $\Omega(\ni I)$ as long as the respective values calculated according to (2.4) and (2.5) are outside of these bounds.

Since A_1, A_2 are projectors with eigenvalues equal to 0 or 1 the eigenvalues of $A\Sigma^{\frac{1}{2}}$ are bounded by the eigenvalues of $\Sigma^{\frac{1}{2}}$. Denote λ_{max} the highest and λ_{min} the lowest eigenvalue of Σ, by δ_Σ the ratio $\lambda_{max}/\lambda_{min}$. Clearly

(4.7) $\quad \lambda_{min} w'Q_i w \leq w'\Lambda_i w \leq \lambda_{max} w'Q_i w.$

Therefore

(4.8) $\quad \delta_\Sigma^{-1} S \leq S(\Sigma) \leq \delta_\Sigma S.$

This inequality holds irrespective of A_1, A_2 and the particular Σ, and only reflects one characteristic of Σ--the ratio of the highest to lowest eigenvalues. The bigger δ_Σ is the more Σ is distinguished from I for which $\delta = 1$.

For any two statistics S_1 and S_2 with $S_1 \le S_2$ everywhere their cumulative distributions $G_i(x) = \text{Prob}(S_i < x)$ i=1,2 are related:

$$G_1(x) \ge G_2(x),$$

and therefore for some level of the test the critical values satisfy:

$$S_1^{cr} \le S_2^{cr}.$$

From this observation and (4.8) we obtain the following theorem.

Theorem 2:

The critical values $F_{cr}(\Sigma)$ are located within the following intervals dependent on δ—the ratio of highest to lowest eigenvalues of Σ:

$$\delta^{-1} F_{cr} \le F_{cr}(\Sigma) \le \delta F_{cr}. \tag{4.9}$$

Corollary

$$LR_{cr} + u \log(\delta^{-1} - (1-\delta^{-1})/\ell_{cr}) \le LR(\Sigma) \le LR_{cr} + n \log(\delta - (\delta-1)/\ell_{cr}),$$

$$\ell_{cr} = \exp(LR_{cr}/n)$$

(4.10)

$$\frac{\delta^{-1} RS_{cr}\, n}{n-(1-\delta^{-1})RS_{cr}} \le RS_{cr}(\Sigma) \frac{\delta\, RS_{cr}\, n}{n+(\delta-1)RS_{cr}}, \quad \delta^{-1} W_{cr} \le W_{cr}(\Sigma) \le \delta\, W_{cr}.$$

The inequalities (4.10) are derived easily from (4.8) and (2.5).

4.1 Discussion of the Results

The relationship (4.9) has the following immediate interpretations for the F-test. Firstly, if a class Ω of Σ matrices is such that the biggest ratio of the maximum to minimum eigenvalues of $\Sigma \in \Omega$ is limited by some δ, then the test conclusions are the same for any Σ as long as either $F/F_{cr} > \delta$ or $F_{cr}/F > \delta$, where F and F_{cr} are, respectively, the value of the test statistic according to (2.4) and the critical value for the hypothesis test under the spherical normal. Secondly, if $F > F_{cr}$ ($F < F_{cr}$) then the test conclusions are robust over the class Ω of Σ matrices with δ the ratios of maximum to minimum eigenvalues, such that $\delta(\Sigma) < F/F_{cr}$ ($\delta(\Sigma) < F_{cr}/F$).

Since the relationship for W in (4.10) is similar to (4.9) for F the same conclusions apply. A simple examination of (4.10) shows that the bounds on the critical values for $RS(\Sigma)$ are inside the interval $[\delta^{-1}RS_{cr}, \delta RS_{cr}]$, thus the conclusions made above hold for RS as well.

The following example demonstrates how our bounds compare to those obtained by Vinod (1976) and Vinod and Ullah (1981) for the t statistic under an AR(1) process. Suppose that $\rho = .5$. Then the eigenvalues of the variance-covariance matrix are contained between the asymptotic ($T \to \infty$) maximum and minimum eigenvalue, $1 + \rho^2 + 2\rho = 2.25$ and $1 + \rho^2 - 2\rho = .25$. Thus the bounds on the critical value of the t statistic can be calculated based on the square root of the ratio $\sqrt{2.25/.25} = 3$. The critical values given by Vinod and Ullah are tabulated according to the number of restrictions p and sample size n. If n=50, p=5, for instance, their Table 4.1 gives 1.14 and 3.93 as the lower and upper bounds, respectively, at the 5% level, whereas our calculation, which involves only dividing and multiplying by 3 of the standard critical value, gives .671 and 6.042 as the lower and upper bounds, respectively.

However, there are three ways in which our results are an improvement. Firstly, they relate to any Σ matrices, not just those generated by an AR(1) or MA(1) process. Secondly, they require the calculation of the maximum and minimum eigenvalues of Σ only, whereas Vinod and Ullah utilized all the eigenvalues in a much more complicated calculation. Thirdly, our bounds are independent of A and A_o matrices.

Note that if the bounds on the positive eigenvalues of $A\Sigma A$ and $(A_o-A)\ \Sigma\ (A_o-A)$ can be established they will provide more accurate intervals for critical values as can be seen from (4.6) and the fact that

$$\frac{\mu_1}{\mu_n} > \delta_\Sigma^{-1} \quad \text{and} \quad \frac{\mu_k}{\mu_p} \leq \delta_\Sigma .$$

Recall that under the multivariate normal error distributions Evans and Savin (1982) have derived for $|W/n| < 1$ the following relationship

$$(4.11) \qquad W - LR \simeq LR - RS \simeq W^2/2n$$

generalizing the known inequality

$$(4.12) \qquad W \geq LR \geq RS.$$

Ullah and Zinde-Walsh (1984) have shown that a more complex relationship exists between the statistics W_ϕ, LR_ϕ, RS_ϕ when the distribution is spherical but non-normal. Here, once again, straightforward inequalities relating the bounds on the statistics can be derived.

For any of the statistics F, LR, RS or W, denote by an upper or lower bar the upper or lower bound, correspondingly, given by (4.9) and (4.10). Next we denote

$$(4.13) \qquad F_U = (\bar{F}-\underline{F})/\bar{F} \quad \text{and} \quad F_L = (\bar{F}-\underline{F})/\underline{F}.$$

Similarly, we define LR_U and LR_L, RS_U and RS_L, W_U and W_L. These ratios show the length of the interval between the bounds in relation to its upper and lower point, respectively. Thus they measure the "tightness" of the bounds on the critical values of the statistics and the following theorem establishes a ranking of the statistics with respect to this characteristic.

Theorem 3:

The following relationships hold

(4.14) $F_U = W_U \geq LR_U \geq RS_U$; $F_L = W_L \geq LR_L \geq RS_L$.

Proof:

See Section 5.

This theorem demonstrates the relative robustness of the bounds on critical values for the different statistics. The bounds are the tightest for RS and are the worst when the F test or the W test is used.

5. Proofs of the Lemmas and Theorems

Proof of Lemma 1

From (2.9) and (3.1) we can write

$$(5.1) \qquad \ell_\Sigma = \frac{(y-X_o\hat{\beta}_{o\Sigma})'\Sigma^{-1}(y-X_o\hat{\beta}_{o\Sigma})}{(y-X\hat{\beta}_\Sigma)'\Sigma^{-1}(y-X\hat{\beta}_\Sigma)} .$$

We transform the denominator of (5.1) by substituting

$\hat{\beta}_\Sigma = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}y$ and get

$$(5.2) \qquad y'\Sigma^{-1}[I-X(X'\Sigma^{-1}X)^{-1}X\Sigma^{-1}]y.$$

Consider

$$(5.3) \qquad \Sigma^{-1} - \Sigma^{-1} X(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}$$

where $\Sigma^{-1} = (I-H)$ with $|y'H y| < y'y$.

We can expand part of (5.3) into a geometric series:

$$X[X'(I-H)X]^{-1}X' = X(X'X)^{-\frac{1}{2}}[I-(X'X)^{-\frac{1}{2}}X'HX(X'X)^{-\frac{1}{2}}]^{-1}(X'X)^{-\frac{1}{2}}X'$$
$$= P + PHP + PHPHP + ... + P(HP)^k + ...$$

where P is defined by (2.6). Substituting into (5.3) we get

$$I - H - P - PHP - PHP - ... - P(HP)^k - ... + HP + HPHP + ...$$
$$+ (HP)^k + ... - HPH - HPHPH - ... - H(PH)^k - ... + PH + PHPH + ...$$
$$+ (PH)^k + ... = A - AHA - AHPHA - ... - AH(PH)^k A - ...$$

Indeed the last term is obtained in the following way:

$$-P(HP)^k HP + (HP)^k HP - H(PH)^k + P(HP)^k H = -A\, H(PH)^k A$$

(we have substituted $P = I-A$). We therefore get that (5.3) can be represented as

(5.4) $$A - AHA - AHPHA - \ldots - AH(PH)^k A - \ldots$$

Further we can replace P by I-A everywhere in (5.4) to get

(5.6) $$A-AH\,A-AH^2 A+AHAHA+\ldots+\sum_{i_1+\ldots+i_{k+1}=i} (-1)^{k+1} AH^{i_1} AH^{i_2} A\ldots H^{i_{k+1}} A+\ldots$$

This formula can be easily verified by substitution of P=I-A into (5.4). The numerator of ℓ_Σ can be transformed in an analogous manner. This concludes the proof of Lemma 1.

<u>Proof of Lemma 2</u>

For a given vector y we define mutually orthogonal unitary vectors y_o, y_1 and y_2, such that $y_i'y_i = 1$, $i=0,1,2$; $y_i'y_j = 0$ for all $0 \le i < j \le 2$;

$$Ay = \alpha y_1;\ (A_o - A)y = \beta y_o;\ P_o y = \gamma y_2.$$

By substituting into (3.4) we get

(5.7) $$\alpha^2(\alpha^2 y_1'Ty_1 + \alpha\beta y_1'Ty_o + \alpha\beta y_o'Ty_1 + \beta^2 y_o'Ty_o) = (\alpha^2+\beta^2)(\alpha^2 y_1'Ty_1).$$

We equate the coefficients of all the monomials in α and β in (5.7) and get

(5.8) $$y_1'Ty_1 = y_o'Ty_o,$$

(5.9) $$y_1'Ty_o + y_o'Ty_1 = 0.$$

Since T is symmetric, (5.9) implies

(5.10) $$y_1'Ty_o = y_o'Ty_1 = 0.$$

Conditions (5.8) and (5.10) hold for any y. We can denote $y_1'Ty_1$ by θ, where θ is a constant. For any y,

$$y_1' T y_1 = y'ATAy/\alpha^2 = \theta,$$

where $\alpha^2 = y'Ay$, and we have $y'ATAy = \theta y'Ay$.

Similarly, using (5.10) in addition to (5.8), we can show that

$$y'A_o T A_o y = \theta y' A_o y.$$

Proof of Theorem 1

Consider the expression (3.2) for ℓ_Σ. We can write it as

$$\ell_\Sigma = \ell_o \frac{1 + y'A_o \overset{o}{T}_\rho A_o y/y'A_o y}{1 + y'AT_\rho Ay/y'Ay} \tag{5.11}$$

If $\ell_\Sigma = \ell_o$ it follows that

$$(y'A_o \overset{o}{T}_\rho A_o y)y'Ay = (y'AT_\rho Ay)y'A_o y. \tag{5.12}$$

The expressions on each side of (5.12) are series in the parameters $\rho_1 \ldots \rho_k$ of H_ρ. If the two series of the right and left sides in (5.12) are to be identical all the coefficients of monomials have to coincide. Consider ρ_1 and the coefficient of the lowest power of ρ_1 in T_ρ, it is some symmetric matrix T_1 (it is also the coefficient of ρ_1 in H_ρ).

We get for T_1

$$(y'A_o T_1 A_o y)y'Ay = (y'AT_1 Ay)y'A_o y$$

thus by Lemma 2,

$$A_o T_1 A_o = \theta_1 A_o \quad (\text{and } AT_1 A = \theta_1 A).$$

Similar equalities hold for all coefficients of lowest powers in T_ρ.

Any coefficient of a monomial $\rho_1^{r_1},\dots,\rho_k^{r_k}$ in T_ρ^o can be represented as the sum of such a coefficient in $\Sigma_{r=1}^{r_1+r_2+\dots+r_k} H_\rho^r$ denoted by $T(r_1,\dots,r_k)$, and products of coefficients of lower power monomials with A_o in between. We can now use induction to show that

$$A_o T(r_1,\dots,r_k) A_o = \theta(r_1,\dots,r_k) A_o. \tag{5.13}$$

If (5.13) holds for all coefficients of monomials of lower power, we can replace such $A_o T A_o$ by the appropriate θA_o and will arrive at (3.4).

This concludes the proof of Theorem 1.

<u>Proof of Lemma 3</u>:

Let $T_i \in O(n)$ be a matrix that diagonalizes $\Sigma^{1/2}A_i\Sigma^{1/2}$: $T_i'\Sigma^{1/2}A_i\Sigma^{1/2}T_i = \Lambda_i$, where Λ_i is diagonal. Denote by Q_i the orthogonal projector onto the space of non-zero eigenvectors of $\Sigma^{1/2}A_i\Sigma^{1/2}$.

We show that Q_1 and Q_2 are mutually orthogonal, thus each T_i can be represented by the same matrix $M_1 + M_2 + M_3$ with M_i mutually orthogonal, the columns of M_1 (M_2) formed by the orthonormal system of non-zero eigenvectors of $\Sigma^{1/2}A_1\Sigma^{1/2}$ ($\Sigma^{1/2}A_2\Sigma^{1/2}$). Suppose that for some vector ξ, $\Sigma^{1/2}A_i\Sigma^{1/2}\xi = \lambda\xi$ with $\lambda \neq 0$, then $Q_i\xi = \xi$. We have $A_i\Sigma^{1/2}\xi = \lambda\Sigma^{-1/2}\xi = A_i^2\Sigma^{1/2}\xi = \lambda A_i\Sigma^{-1/2}\xi$, therefore $A_i\Sigma^{-1/2}\xi = \Sigma^{-1/2}\xi$. Clearly then for any η such that $\eta = Q_i\eta$ we have $A_i\Sigma^{-1/2}\eta = \Sigma^{-1/2}Q_i\eta$. For a vector η for which $Q_1\eta = Q_2\eta = \eta$ one would have $\eta = \Sigma^{1/2}A_2\Sigma^{-1/2}\eta = \Sigma^{1/2}A_1\Sigma^{-1/2}\eta = (\Sigma^{1/2}A_2\Sigma^{-1/2})(\Sigma^{1/2}A_1\Sigma^{-1/2})\eta = \eta = 0$ since $A_2A_1 = 0$. Therefore Q_1 and Q_2 are orthogonal projectors.

This concludes the proof.

<u>Proof of Theorem 3</u>

Represent all the bounds as functions of $S = S_{cr}$, by combining (2.11), (4.9) and (4.10).

$$\bar{F}=(q/r)\delta S;\ \bar{W}=n\delta S;\ \overline{LR}=n\log(1+\delta S);\ \overline{RS}=n\delta S/(1+\delta S);$$

(5.14)

$$\underline{F}=(q/r)\delta^{-1}S;\ W=n\delta^{-1}S;\ \underline{LR}=n\log(1+\delta^{-1}S);\ \underline{RS}=n\delta^{-1}S/(1+\delta^{-1}S).$$

Next we derive directly that

(5.15) $F_U = W_U = 1-\delta^{-2};\ F_L = W_L = \delta^2-1;$

(5.16) $W_U-LR_U = \ell n(1+\delta^{-1}S)/\ell n(1+\delta S) - \delta^2$

(5.17) $W_L-LR_L = \delta^2 - \ell n(1+\delta S)/\ell n(1+\delta^{-1}S);$

(5.18) $LR_U-RS_U = \delta^{-2}(1+\delta S)/(1+\delta^{-1}S) - \ell n(1+\delta^{-1}S)/\ell n(1+\delta S)$

(5.19) $LR_L-RS_L = \ell n(1+\delta S)/\ell n(1+\delta^{-1}S) - \delta^2(1+\delta^{-1}S)/(1+\delta S).$

It immediately follows from (5.15) that whatever conclusions will be proved to hold here with respect to W will apply to F as well.

Examine (5.16). The expression $\ell n(1+\delta^{-1}S) - \delta^{-2}\,\ell n(1+\delta S)$ is always non-negative since it equals zero for $S = 0$ and its derivative with respect to S is $\dfrac{\delta^{-1}(\delta S-\delta^{-1}S)}{(1+\delta^{-1})(1+\delta S)}$ and is thus positive. This proves that (5.16) is positive for positive S. Similarly, we show that $\delta^2\ell n(1+\delta^{-1}S) - \ell n(1+\delta S)$ is positive for $S > 0$, thus (5.17) is positive.

Next, consider the expression

$$\delta^{-2}(1+\delta S)\ell n(1+\delta S) - (1+\delta^{-1}S)\ell n(1+\delta^{-1}S)$$

related to (5.18). It is zero for $S = 0$, its derivative is equal to $\delta^{-1}\ell n\dfrac{1+\delta S}{1+\delta^{-1}S}$ and positive. Therefore (5.18) is positive. Similarly from $(1+\delta S)\ell n(1+\delta S) - \delta^2(1+\delta^{-1}S)\ell n(1+\delta^{-1}S)$ being positive (identical proof) it follows that (5.19) is positive.

This concludes the proof of Theorem 3.

REFERENCES

Evans, G. B. A. and N. E. Savin: "Conflict Among the Criteria Revisited: The W. LR, and LM Tests," Econometrica 50 (1982), 737-748.

Ghosh, M. and B. K. Sinha: "On the Robustness of Least Squares Procedures in Regression Models," Journal of Multivariate Analysis 10 (1980), 332-342.

Khatri, C. G.: "Study of F-Tests Under Dependent Models," The Indian Journal of Statistics 43 (1981), A, 107-110.

Kiviet, J. F.: "Effects of ARMA Errors on Tests for Regression Coefficients: Comments on Vinod's Article; Improved and Additional Results," Journal of the American Statistical Association 75 (1980).

Rao, C. R.: "Large Sample Tests of Statistical Hypotheses Concerning Several Parameters with Applications to Problems of Estimation," Proceedings of the Cambridge Philosophical Society 44 (1974), 50-57.

Savin, N. E.: "Conflict Among Testing Procedures in a Linear Regression Model With Augoregressive Disturbances," Econometrica 44 (1976), 1303-1315.

Silvey, S. D.: "The Lagrange Multiplier Test," Annals of Mathematical Statistics 30 (1959), 389-407.

Sinha, B. K. and B. B. Mukhopodhyay: "A Note on a result of Ghosh and Sinha," Calcutta Statistical Association Bulletin (1981), 169-177.

Ullah, A. and V. Zinde-Walsh: "On the Robustness of LM, LR and W Tests in Regression Models," Econometrica 52 (1984), 1055-1066.

________: "Estimation and Testing in Regression Models with Spherically Symmetric Errors," Economics Letters 17 (1985), 127-132.

Vinod, H. D.: "Effects of ARMA Errors on the Significance Tests for Regression Coefficients," Journal of the American Statistical Association 71 (1976), 929-933.

Vinod, H. D. and A. Ullah: Recent Advances in Regression Methods. New York: Marcel Dekker, 1981.

Wald, A.: "Tests of Hypotheses Concerning Several Parameters When the Number of Observations is Large," Transactions of the American Mathematical Society 54 (1943), 426-482.