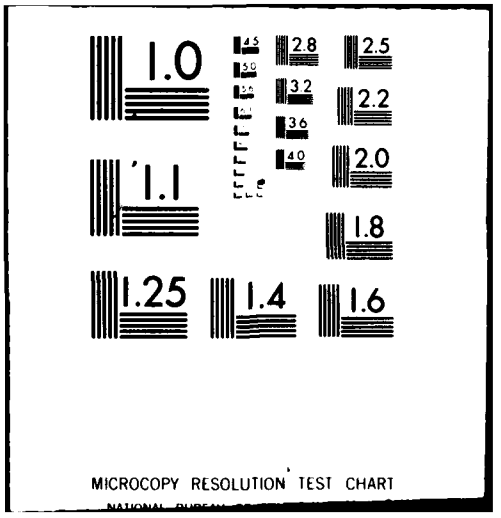
AD-A113 441 RAND CORP SANTA MONICA CA F/G 12/1
CONFLICT AMONG TESTING PROCEDURES IN A LINEAR REGRESSION MODEL.(U)
OCT 81 D F KOHLER
UNCLASSIFIED RAND/P-6691 NL

END
DATE FILMED
5-82
DTIC


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0
MICROCOPY RESOLUTION TEST CHART

AD A113441

CONFLICT AMONG TESTING PROCEDURES IN A LINEAR REGRESSION MODEL

Daniel F. Kohler

October 1981

DTIC FILE COPY

DTIC
ELECTED
MAR 1 3 1982
S
A

This document has been approved for public release and sale; its distribution is unlimited.

P-6691

82 04 12 120

CONFLICT AMONG TESTING PROCEDURES IN A LINEAR REGRESSION MODEL

1. INTRODUCTION

Savin [6] and Berndt and Savin [2] have shown that an inequality relation exists between different test statistics used for testing hypotheses of the form $r - R\beta = 0$. They found that the value of the likelihood ratio test statistic ($LR = -2\log\lambda$), the Wald test statistic (W), and the Lagrange multiplier test statistic (LM) are always such that

$$W \geqslant LR \geqslant LM \tag{1}$$

This result has been generalized by Breusch [3] who showed that the only necessary assumption for this inequality to hold is, that the disturbances follow a distribution which allows maximum-likelihood estimation. However, neither Breusch nor any of the authors before him were able to conclude anything about the power of the different tests. In this paper it will be shown that for finite but large samples a similar inequality relation to (1) exists between the powers of the three tests. The Wald test is uniformly more powerful than either of the other two tests, and the likelihood ratio test is more powerful than the Lagrange multiplier test for very large samples and for moderate-to-large differences between the null hypothesis and the true value of the tested parameters.

DTIC COPY INSPECTED 2

The assumption of a scalar covariance matrix is made to simplify the exposition. The results can probably be generalized to hold for any disturbance vector which allows maximum-likelihood estimation.

2. THE MODEL

Consider the model:

$$Y = X\beta + \varepsilon \tag{2}$$

$$\varepsilon \sim N(0, \sigma^2 I) \tag{3}$$

Let $\hat{\beta}$ and $\hat{\sigma}^2$ be the maximum-likelihood estimators obtained by unconstrained maximization of the likelihood function, and $\tilde{\beta}$ and $\tilde{\sigma}^2$ the corresponding constrained estimators. Furthermore, we shall need an estimator for the Lagrange multiplier ($\hat{\mu}$) and the ratio of the constrained to unconstrained maxima of the likelihood function (λ). The three test statistics can be written as:

$$LR = -2\log\lambda \tag{4}$$

$$W = n\left[(r-R\hat{\beta})'(R(X'X)^{-1}R')^{-1}(r-R\hat{\beta})\right]\Big/\hat{\sigma}^2 \tag{5}$$

$$LM = n\left[\hat{\mu}(R(X'X)^{-1}R')\hat{\mu}\right]\Big/\tilde{\sigma}^2 \tag{6}$$

To simplify the notation let $A = R\left(\frac{1}{n}X'X\right)^{-1}R'$. From the first-order conditions for maximizing the likelihood function subject to the constraint we can obtain an expression for $\hat{\mu}$:

$$(7) \qquad \hat{\mu} = \left(R(X'X)^{-1}R'\right)^{-1}(r-R\hat{\beta})$$

We can now rewrite (5) and (6) as

$$(8) \qquad W = n\left[(r-R\hat{\beta})' A^{-1}(r-R\hat{\beta})\right] \Big/ \hat{\sigma}^2$$

$$(9) \qquad LM = n\left[(r-R\hat{\beta})' A^{-1}(r-R\hat{\beta})\right] \Big/ \tilde{\sigma}^2$$

If the null hypothesis is true and the disturbances are normally distributed, we have

$$(10) \qquad \sqrt{n}(r-R\hat{\beta}) \sim N(0,\sigma^2 A)$$

Furthermore, $\tilde{\sigma}^2$ is a consistent estimator for σ^2 if the null hypothesis holds. $\hat{\sigma}^2$ is a consistent estimator regardless of whether the null hypothesis is true or not.

It follows that under the null hypothesis, W and LM converge to the same limiting Chi-square distribution with k degrees of freedom, where k is the rank of R. Independently it can be shown that LR also converges to the same Chi-square distribution. Combined with the inequality relation (1) this gives rise to possibly conflicting test results.[1]

3. APPROXIMATE RELATIVE POWER OF THE THREE TESTS

It is possible to evaluate the power functions in the case of finite but large samples.[2] When the null hypothesis is not true, we have:

$$(11) \qquad r - R\beta = a$$

therefore,

(12) $\sqrt{n}(r-R\hat{\beta}) \sim N\{\sqrt{n}\,a, \sigma^2 A\}$

For large samples we have approximately

(13) $\sqrt{n}(r-R\hat{\beta}) / \hat{\sigma} \simeq N(\sqrt{n}\,a/\sigma, A)$

If a is different from zero and n is large, the quadratic form W (see 8) follows approximately a noncentral Chi-square distribution[3] with k degrees of freedom and non-centrality parameter

(14) $c(n) = \frac{n}{2}[a/\sigma]' A^{-1} [a/\sigma]$ [4]

Let $f(W)$ be the density function of the Wald test statistic and under the alternative hypothesis and $\chi^2_\alpha\{0,k\}$ the critical value based on a central Chi-square distribution with k degrees of freedom. The power of the Wald test is then:

(15) $P(W) = \int_{\chi^2_\alpha\{0,k\}}^{\infty} f(W)dW$

For large samples we can approximate the density function $f(W)$ by a non-central Chi-square density function with non-centrality parameter $c(n)$. As n increases, $c(n)$ grows without limit, and the distribution of W explodes. From (15) we see that the asymptotic power of the Wald test is therefore equal to one. A similar expression can be constructed for the likelihood ratio test (21) and the Lagrange multiplier test (18). Their asymptotic power is also equal to one.

This is a common feature of all point hypotheses tests based on a consistent parameter estimate. It is, though comforting to the theoretician,

of little help to the practitioner trying to decide which test to use. It says, if anything, that if we had a truly infinite sample, it would not matter which test we applied, as long as it was based on a consistent parameter estimate.[5] It definitely does not imply that the tests are of equal power for finite but large samples, which is, alas, the best we can do in the real world. From (8) and (9) we can obtain an expression for LM .

$$(16) \qquad LM = (\hat{\sigma}^2 / \tilde{\sigma}^2)\, W$$

And

$$(17) \qquad f(LM) = f\left(\frac{\hat{\sigma}^2}{\tilde{\sigma}^2} \cdot W\right)$$

Given our definition of "large but finite samples," we can set $\hat{\sigma}^2/\tilde{\sigma}^2 = \sigma^2/\text{plim}\,\tilde{\sigma}^2 = p \leq 1$. It follows that the power of the Lagrange multiplier test is equal to

$$(18) \qquad P(LM) = \int_{\chi^2_\alpha(0,k)}^{\infty} f(p \cdot W) \cdot dpW$$

$$= \int_{p^{-1} \cdot \chi^2_\alpha\{0,k\}}^{\infty} f(W) \cdot dW$$

Given that $p^{-1} > 1$ if the null hypothesis is not true, we can conclude from (15) and (18) that $P(LM) < P(W)$.

A similar argument can be made for the relative power of the Wald and likelihood ratio tests. Consider $\tilde{\sigma}^2 = (Y-X\tilde{\beta})'(Y-X\tilde{\beta})/n$ and $\tilde{\beta} = \tilde{\beta} + n(X'X)^{-1} A^{-1} (r-R\hat{\beta})$. Then, letting $e = y - X\hat{\beta}$,we get by substitution:

$$\tilde{\sigma}^2 = 1/n\{e'e - 2[X(1/n\ X'X)^{-1}A^{-1}]'(r-R\hat{\beta}) + (r-R\hat{\beta})'A^{-1}n[R(X'X)^{-1}X'X(X'X)^{-1}R']A^{-1}(r-R\hat{\beta})\}$$

$$= \hat{\sigma}^2 + (r-R\hat{\beta})'\ A^{-1}(r-R\hat{\beta})$$

We can, therefore, express (8) as

$$(19) \qquad W = n(\tilde{\sigma}^2-\hat{\sigma}^2)/\hat{\sigma}^2 = n(\tilde{\sigma}^2/\hat{\sigma}^2-1)$$

Making use of the fact that $\lambda = (\tilde{\sigma}^2/\hat{\sigma}^2)^{-n/2}$, we can write

$$(22) \qquad W = n(\lambda^{-2/n}-1)$$

$$LR = -2\log\lambda = n\cdot\log[(W/n)+1)]$$

The power of the likelihood ratio test is

$$(21) \qquad P(LR) = \int_{\chi^2_\alpha\{0,k\}}^{\infty} f(LR)\cdot dLR$$

$$= \int_{\chi^2_\alpha\{0,k\}}^{\infty} f[n\cdot\log(W/n+1)]\cdot d[n\cdot\log(W/n+1)]$$

$$= \int_{n[-1+\exp(\chi^2_\alpha\{0,k\}/n)]}^{\infty} f(W)dW$$

Consider the lower limit of integration in (21):

$$(22) \qquad n\left[-1+\exp(\chi^2_\alpha\{0,k\}/n)\right] = n\left[-1+\sum_{i=0}^{\infty}(\chi^2_\alpha/n)^i/i!\right]$$

$$= n\left[\chi^2_\alpha/n+\sum_{i=2}^{\infty}(\chi^2_\alpha/n)^i/i!\right]$$

$$= \chi^2_\alpha + o(n^{-1})$$

From (22), (21) and (15) it follows that $P(LR) < P(W)$. Also note from (22) that the difference between $P(LR)$ and $P(W)$ decreases, as the sample size increases.

Consider the relative power of the LR and LM tests. From (21) and (18) we see that the only difference in power between the two tests must come from the difference in the lower limit of integration. For the likelihood ratio test this limit is:

$$(23) \qquad \chi^2_\alpha + n \sum_{i=2}^{\infty} (\chi^2_\alpha / n)^i / i! = \chi^2_\alpha \left[1 + \sum_{i=1}^{\infty} (\chi^2_\alpha / n)^i / (i+1)! \right]$$

the corresponding limit for the Lagrange multiplier test (see 18) is:

$$(24) \qquad (\text{plim}\, \tilde{\sigma}^2 / \sigma^2) \chi^2_\alpha = \left[1 + a' A^{-1} a / \sigma^2 \right] \chi^2_\alpha$$

Where we have made use of the fact that $\tilde{\sigma}^2 = \hat{\sigma}^2 + (r - R\hat{\beta})' A^{-1} (r - R\hat{\beta})$ and $\text{plim}\,(r - R\hat{\beta}) = a$. For a given sample size there exists an a , say a^0 , for which the Lagrange multiplier test and the likelihood ratio test are equally powerful, i.e., (23) = (24).

$$(25) \qquad \chi^2_\alpha \left[1 + \sum_{i=1}^{\infty} (\chi^2_\alpha / n)^i / (i+1)! \right] = \chi^2_\alpha [1 + a^{0\prime} A^{-1} a^0 / \sigma^2]$$

$$\sum_{i=1}^{\infty} (\chi^2_\alpha / n)^i / (i+1)! = a^{0\prime} A^{-1} a^0 / \sigma^2$$

From (25) we see that the lower limit of integration in (18) becomes relatively larger than the corresponding limit in (21) when a increases above a^0 . This means that the likelihood ratio test becomes more powerful,

as the difference between the null hypothesis and the true value of the tested parameter grows. In other words, the power function of the likelihood ratio test must intersect the power function of the Lagrange multiplier test from below, as depicted in Figure 1.

This result, however, is only an outflow of our incongruous assumption about the sample size. Note that the left-hand side of (25) is of order $1/n$, so that for very large samples it must be practically zero. This implies that a^0 must be practically zero as well. In the limit it must be exactly zero.

We can also consider (25) under a somewhat different aspect. Holding a^0 constant we can determine the effect of an increase in the sample size. If n is large enough so that $[X'X]n^{-1}$ has already attained it's limit, or is close to it, an increase in n would not have any noticeable effect on the right-hand side of (25). The left-hand side would become smaller however, leading us to conclude, that the likelihood ratio test would become more powerful.

4. RELAXING THE ASSUMPTIONS ABOUT SAMPLE SIZE

The comparison of the relative power of the Lagrange multiplier test and the Wald test in (18) depended heavily on our convenient definition of large but finite samples. But it is not certain that there exists a sample size for which the conditions of our definition hold. The difference between $\hat{\sigma}^2$ and σ^2 is of order $1/n$ and $c(n)$ is of order n, so that strictly speaking $\hat{\sigma}^2$ approaches σ^2 at the same speed as $c(n)$ approaches infinity. If we are prepared to set $\hat{\sigma}^2 = \sigma^2$ for a very large n, which is what we are doing when we use the asymptotic distribution to test hypotheses based on finite samples, we should apply the same standards and set $c(n)$ equal to infinity. Under these circumstances all the tests are of power one.

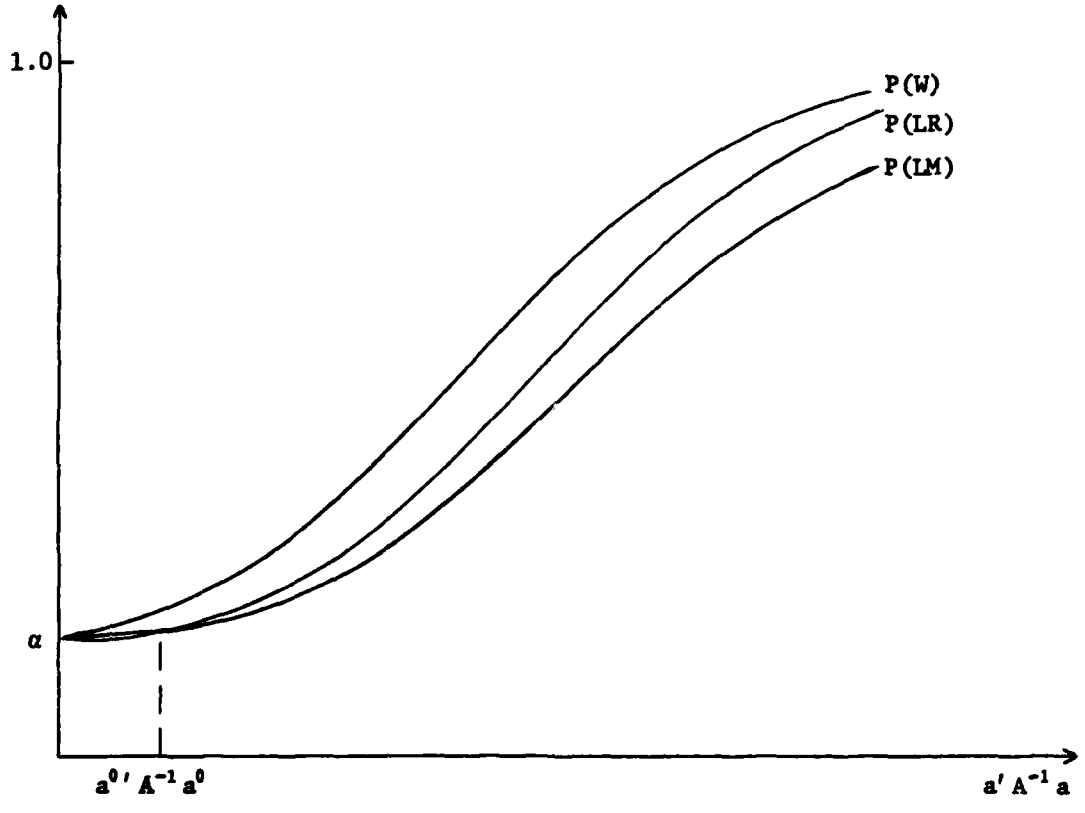


FIGURE 1 -- Relative Power of the Three Tests for a Finite Sample

In order to be able to say something about the relative power of the Lagrange multiplier test and the Wald test for finite samples we must treat $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ as stochastic. We can define a critical value $d_\gamma > 0$ of the random variable $\tilde{\sigma}^2/\hat{\sigma}^2$ such that

$$\Pr\{\tilde{\sigma}^2/\hat{\sigma}^2 \geqslant 1+d_\gamma\} = \gamma \tag{26}$$

Then we can rewrite (18) as

$$\Pr\left\{P(LM) \leqslant \int_{(1+d_\gamma)\chi^2_\alpha\{0,k\}}^{\infty} f(W)\cdot dW = P(W)\right\} = \gamma \tag{27}$$

As γ approaches unity, d_γ approaches zero so that we have

$$\Pr\left\{P(LM) \leqslant \int_{\chi^2_\alpha\{0,k\}}^{\infty} dW = P(W)\right\} = 1 \tag{28}$$

or $P(LM) \leqslant P(W)$.

This comparison of $P(LM)$ and $P(W)$ no longer relies on any special assumptions about the sample size. It applies to any sample size for which the Chi-square distribution adequately describes the distribution of W , i.e., a Chi-square is justified.

Qualitatively our results have changed little. Because $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ are stochastic,we have to allow for the possibility that they may be equal by a fluke, even if the null hypothesis is not true. In this case all three test statistics take on the value zero and commit a type two error. This is the main reason why the strict inequality derived from (18) has to be changed to $\leqslant$. However,with any probability of less than one, d_γ in (27) is > 0 and the strict inequality still holds.

5. A NUMERICAL EXAMPLE

A small Monte-Carlo experiment was conducted to determine the extent to which the power of the different tests varies. The model was simply $Y = a+\varepsilon$ with ε distributed normally and independently with mean zero and variance one. The null hypothesis was that $a = 0$. Table I summarizes the results for the tests using 100 samples of size 100 each.

The results are not very surprising. Indeed, the Wald test is more powerful than the other two tests. As a increases, the power of all three tests goes to one. The relative advantage of the Wald test is strongest for small a's.

6. CONCLUSIONS

It is true that both the likelihood ratio test and the Lagrange multiplier test have asymptotically the same power as the Wald test. For the likelihood ratio test this is primarily due to the fact that the difference in power between the two tests vanishes, as n goes to infinity (see (21)). However, the Lagrange multiplier test has asymptotically the same power as the Wald test, only due to the fact that the limiting distribution of both tests is not defined. The difference between the lower limits of integration in (18) and (15) converges to a number different from zero. Loosely speaking, one could state that if the limiting distribution of W under the alternative hypothesis was defined, the likelihood ratio test would still be asymptotically of equal power as the Wald test, while the Lagrange multiplier test would be uniformly less powerful.

TABLE I

NUMBER OF REJECTIONS PER HUNDRED

True Parameter Value	Wald Test	LR Test	LM Test
$a = .2$	25	21	20
$a = .3$	51	50	48
$a = .4$	88	86	85
$a = .5$	98	98	98
$a = 1$	100	100	100

Significance level: 5%.

It is trivial to note that two tests have the same asymptotic power, if this statement depends on the fact that the limiting distribution explodes. We must instead consider the relative power for sample sizes short of infinity. In this paper it was shown that there exists an inequality relation similar to (1) between the power of the three tests for large but finite samples. The Wald test is uniformly more powerful than either of the other two tests if we accept the stated definition of "large but finite" samples. If we are not able to set $\hat{\sigma}^2 = \sigma^2$ and $\tilde{\sigma}^2 = \text{plim}\,\tilde{\sigma}^2$ for a given sample size, the strict inequality relation has to be modified to $P(LM) \leqslant P(W)$. However, $P(LR) < P(W)$ apparently still holds, because this comparison does not depend on setting $\hat{\sigma}^2 = \sigma^2$.

It follows that for finite samples the Wald test is at least as powerful or more powerful than the other two tests. If it is no more difficult or costly to compute, then there appears to be little justification for using either the Lagrange multiplier test or the likelihood ratio test for the purpose of testing linear restrictions in a linear regression model.[6]

We could view the problem as one of misspecification of the critical region. If the different tests are proportional to each other (e.g., $LM = [(\hat{\sigma}^2/\tilde{\sigma}^2)\cdot W]$ (see expression (16)), they should not have the same critical values. If we define the critical value of LM test as $(\hat{\sigma}^2/\tilde{\sigma}^2)$ times the critical value of the W test, the two tests are, of course, equally powerful for all sample sizes. A similar argument applies for the critical value of the LR test.

FOOTNOTES

[1] It has to be pointed out, that, of course, no conflict arises if the Wald test accepts the null hypothesis or the Lagrange multiplier test rejects it.

[2] "Finite but large" samples are small enough so that $1/n$ is still different from zero, but the limiting distribution is an adequate approximation to the sampling distribution. This is the way in which we always use limiting results in practice.

[3] For small samples, the appropriate statistic to use would, of course, be $F = (\hat{\sigma}^2/ks^2)\cdot W$, where s^2 is the (unbiased) least squares estimator for σ^2

[4] See Kendall and Stuart [5], p. 237-41.

[5] Under these circumstances, hypothesis testing becomes trivial, because in the limit there are no 'unknown parameters'.

[6] There exist situations when the unrestricted model cannot be estimated, but we nevertheless wish to test linear restrictions (e.g., identifying restrictions). Under these circumstances, the LM test is the only feasible procedure.

REFERENCES

[1] Aitchison, J., and S. D. Silvey: "Maximum-Likelihood Estimation of Parameters Subject to Restraints," Annals of Mathematical Statistics. 29 (1958). 813-828.

[2] Berndt, E., and N. E. Savin: "Conflict Among Criteria for Testing Hypotheses in the Multivariate Linear Regression Model," Econometrica. 45 (1977). 1263-1278.

[3] Breusch, T. S.: "Conflict Among Criteria for Testing Hypothesis: Extension and Comments," Econometrica. 47 (1979). 203-207.

[4] Hardy, G. H., J. E. Littlewood and G. Polya: Inequalities. Cambridge University Press, 1967.

[5] Kendall, M. G., and A. Stuart: The Advanced Theory of Statistics, Vol. 2. 3rd edition, Hafner, New York, 1973.

[6] Savin, N. E.: "Conflict Among Testing Procedures in a Linear Regression Model with Autoregressive Disturbances," Econometrica. 44 (1976). 1303-1315.

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

[8] 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.