BEBR

FACULTY WORKING
PAPER NO. 1187

THE LIBRARY OF THE

NOV 1 9 1985

UNIVERSITY OF ILLINOIS
URBANA-CHAMPAIGN

Additivity and Separability of the Lagrange Multiplier, Likelihood Ratio and Wald Tests

A. K. Bera
C. R. McKenzie

College of Commerce and Business Administration
Bureau of Economic and Business Research
University of Illinois, Urbana-Champaign

BEBR

FACULTY WORKING PAPER NO. 1187

College of Commerce and Business Administration

University of Illinois at Urbana-Champaign

September, 1985

Additivity and Separability of the Lagrange Multiplier, Likelihood Ratio and Wald Tests

A. K. Bera, Assistant Professor
Department of Economics

C. R. McKenzie
Australian National University

ADDITIVITY AND SEPARABILITY OF THE LAGRANGE MULTIPLIER, LIKELIHOOD RATIO AND WALD TESTS

by

A.K. Bera
University of Illinois

and

C.R. McKenzie*
Australian National University

ABSTRACT: Conditions for the additivity of the Lagrange multiplier, likelihood ratio and Wald tests are developed. The relationship between these conditions and the conditions for separability are also established.

KEYWORDS: Lagrange multiplier test, likelihood ratio test, Wald test, additivity, separability.

*The first author acknowledges the research support of the Research Board and the Bureau of Economic and Business Research of the University of Illinois. The second author gratefully acknowledges the financial support of the Japanese Government's Department of Education and the research support provided by the Japanese Government's Economic Planning Agency. The authors would like to thank John Geweke and Adrian Pagan for their helpful comments on an earlier draft.

Postal Address:

Colin R. McKenzie,
Department of Statistics,
Faculty of Economics and Commerce,
The Australian National University, GPO Box 4,
Canberra ACT 2601,
AUSTRALIA.

Digitized by the Internet Archive
in 2011 with funding from
University of Illinois Urbana-Champaign

http://www.archive.org/details/additivitysepara1187bera

1. Introduction

In the statistical literature, three basic principles are available for hypothesis testing, namely the likelihood ratio (LR), Wald (W) and Lagrange multiplier (LM) (or score) principles. Their asymptotic equivalence under the null hypothesis and under local alternatives is well known. The purpose of this paper is to examine the additivity and separability properties of these tests.

Additivity focuses on the optimal way of combining tests of different hypotheses and indicates a joint test statistic can sometimes be obtained by adding up the component statistics. Alternatively, rather than applying a joint test, the individual tests can be applied separately and the overall significance level can be calculated. An interesting feature of LM statistics is that they are sometimes additive, that is the LM test for testing a joint hypothesis is the sum of LM statistics testing the components of the null hypothesis separately. This kind of additivity was first noted by Pesaran (1979). He found that the LM test for testing the dynamic specification of the *deterministic* and *stochastic* parts (of the linear regression model) simultaneously can be decomposed into two independent parts. This holds even for more complicated cases; for example, the tests developed in Bera and Jarque (1982) for different combinations of normality (N), homoscedasticity (H), serial independence (I) of the regression disturbances and functional form (F) were found to be additive. There are some cases where additivity fails, e.g., if a lagged dependent variable is introduced into the Bera and Jarque (1982) framework, the tests will not be additive nor can the LM test derived in Jarque and Bera (1982) for testing H_o: $u \sim NH$ against non-normality ($\bar{N}$) and heteroscedasticity ($\bar{H}$), where u is the disturbance term in a limited dependent variable model, be decomposed into independent parts. In this paper, we provide the necessary and sufficient conditions for LM tests to be additive in this sense and also examine the additivity properties of the

Aitchison (1962) introduced the concept of separability which is a useful piece of information because it may mean that for large samples the computations required for hypothesis testing can be considerably reduced. If two hypotheses are separable and the sample is large, while testing one hypothesis we may be able to ignore the other hypothesis, that is, the test is robust to whether the other hypothesis is true or not. We relate the concept of separability to additivity in the context of these three testing principles.

2. Additivity and Separability

Let $\ell_i(\theta)$ denote the log-density function for the ith observation, where θ is a $p \times 1$ parameter vector. Say we have N independent observations, then the log-likelihood function is $\ell \equiv \ell(\theta) = \sum_{i=1}^{N} \ell_i(\theta)$. Assume the hypothesis to be tested is H_o: $h(\theta) = 0$ where $h(\theta)$ is an $r \times 1$ vector function of θ and it is assumed that $H \equiv H(\theta) = \partial h(\theta)/\partial\theta$ has full column rank i.e. rank(H) = r. The LM statistic can be written as [see Breusch and Pagan (1980, p.240)]

$$LM = \tilde{d}'\tilde{I}^{-1}\tilde{d} = \tilde{\lambda}'\tilde{H}'\tilde{I}^{-1}\tilde{H}\tilde{\lambda} \qquad (1)$$

where $d \equiv d(\theta) = \partial\ell/\partial\theta$ is the efficient score vector, $I \equiv I(\theta) = E(-\partial^2\ell/\partial\theta\partial\theta')$ is the information matrix[1], $\tilde{\lambda}$ are the Lagrangian multipliers satisfying the equation $\tilde{d} + \tilde{H}\tilde{\lambda} = 0$ and "~" indicates the quantities have been evaluated at the restricted maximum likelihood estimate of θ say $\tilde{\theta}$.

If we partition H_o into H_A: $h_1(\theta) = 0$ H_B: $h_2(\theta) = 0$ with a similar partition for $H = [H_1 : H_2]$, then the LM test for testing H_o will be additive between the two hypotheses H_A and H_B iff $\tilde{H}'\tilde{I}^{-1}\tilde{H}$ is block

---

[1] As pointed out in Davidson & McKinnon (1983) and Bera & McKenzie (1984) a number of alternative asymptotically equivalent forms of the information matrix are available. For these alternative forms, additivity will only be asymptotic.

diagonal between the Lagrange multipliers for the two hypotheses, that is $\tilde{H}_1'\tilde{I}^{-1}\tilde{H}_2 = 0$.

If we partition the parameter vector θ into $\theta = (\theta_1', \theta_2')'$ and consider testing hypotheses of the form H_o: $\theta_2 = \theta_2^o$ then the LM test for testing H_o will be additive between all the individual hypotheses if $\tilde{I}^{22}$ is block diagonal between the testing parameters of the two hypotheses, where $\tilde{I}^{22}$ refers to the (2,2) block of $\tilde{I}^{-1}$. This can easily be seen by writing the LM test as [see Breusch and Pagan (1980, p.241)]

$$LM = \tilde{d}_2'[\tilde{I}_{22} - \tilde{I}_{21}\tilde{I}_{11}^{-1}\tilde{I}_{12}]^{-1}\tilde{d}_2$$

$$= \tilde{d}_2'\tilde{I}^{22}\tilde{d}_2 \quad \text{(say)}$$

where $d_2 = \partial\ell(\theta)/\partial\theta_2$ and $I_{jk} = E\left[\Sigma_i \left(\frac{\partial \ell_i}{\partial \theta_j}\right)\left(\frac{\partial \ell_i}{\partial \theta_k}\right)'\right]$ for $j,k = 1,2$. Obviously, the necessary and sufficient condition for additivity is the block diagonality of $\tilde{I}^{22}$. Under H_o and appropriate regularity conditions [see e.g. Serfling (1980, p.144-5)], $\partial\tilde{\ell}/\partial\theta_2$ asymptotically follows a multivariate normal distribution with mean zero and variance-covariance matrix $I_{22} - I_{21}I_{11}^{-1}I_{12}$. Therefore, the block diagonality of $I_{22} - I_{21}I_{11}^{-1}I_{12}$, i.e. zero covariance between the two components of $\partial\tilde{\ell}/\partial\theta_2$ corresponding to the two hypotheses, implies asymptotic independence of the different components of $\partial\tilde{\ell}/\partial\theta_2$ and the LM test is based on this vector.

Additivity of the LM test can easily be related to Aitchison's (1962) concept of separability. In our example, separability means that, for example, tests of H_o: $h_1(\theta) = 0$ against H: $h_1(\theta) \neq 0 | h_2(\theta) = 0$ and H_o: $h_2(\theta) = 0$ against H: $h_2(\theta) \neq 0 | h_1(\theta) = 0$ use the same critical regions as tests of H_o: $h_1(\theta) = 0$ against H: $h_1(\theta) \neq 0$ and H_o: $h_2(\theta) = 0$ against H: $h_2(\theta) \neq 0$ respectively. Aitchison (1962) provided a sufficient condition for two hypotheses to be separable with respect to all the other

statistics - the LR, W and LM. Suppose we want to examine whether H_A: $h_1(\theta) = 0$ and H_B: $h_2(\theta) = 0$ are separable. A sufficient condition is that $H_1' I^{-1} H_2 = 0$ for all θ satisfying $h_1(\theta) = 0$ and $h_2(\theta) = 0$. This condition is identical[2] to the necessary and sufficient condition for the LM tests to be additive. Therefore additivity of the LM test implies separability of the LM test and since Aitchison's result applies to all the three test principles this also implies separability with respect to the LR and W tests.

Given the additivity of the LM test, it is interesting to investigate whether the LR and W tests share this property. Let ℓ_{AB}, $\ell_{A\bar{B}}$, $\ell_{\bar{A}B}$ and $\ell_{\bar{A}\bar{B}}$ be the log-likelihood values when both H_A and H_B restrictions, only H_A restrictions, only H_B restrictions and no restrictions are imposed respectively. If LR_{AB} is the joint LR test of both restrictions, LR_A the LR test of H_A restrictions and LR_B the LR test of H_B restrictions then

$$LR_{AB} = 2[\ell_{\bar{A}\bar{B}} - \ell_{AB}]$$

$$LR_A = 2[\ell_{\bar{A}B} - \ell_{AB}]$$

$$LR_B = 2[\ell_{A\bar{B}} - \ell_{AB}] .$$

Now

$$LR_{AB} = LR_A + LR_B$$

iff

$$\ell_{\bar{A}\bar{B}} = \ell_{\bar{A}B} + \ell_{A\bar{B}} - \ell_{AB} .$$

In general, the above relation is not true in either finite or large samples.

---

[2] Strictly speaking the conditions are not identical. For additivity we need $\tilde{H}_1' \tilde{I}^{-1} \tilde{H}_2 = 0$ which is implied by $H_1' I^{-1} H_2 = 0$ for all θ satisfying $h(\theta) = 0$, but not vice-versa. However, without loss of generality, we can assume that the parameter space over which $\tilde{H}_1 \tilde{I}^{-1} \tilde{H}_2 = 0$ has measure zero. Then we can say that the conditions are almost surely identical.

But if we rewrite it as

$$[\ell_{\bar{A}\bar{B}} - \ell_{A\bar{B}}] = [\ell_{\bar{A}B} - \ell_{AB}]$$

i.e., $$LR_{A\bar{B}} = LR_{A}$$

where $LR_{A\bar{B}}$ is the test of the H_A restriction without imposing the H_B restrictions. Separability implies $LR_{A\bar{B}} \overset{a}{=} LR_{A}$, where $\overset{a}{=}$ denotes asymptotic equivalence. Therefore separability of the LR test implies the LR test will be additive in an asymptotic sense.

Turning to the question of the additivity of W it is easy to show that, given separability, a sufficient condition for W to be additive is that $\hat{H}_1'\hat{I}^{-1}\hat{H}_2 = 0$ where "^" indicates the quantities have been evaluated at the unrestricted maximum likelihood estimate of θ say $\hat{\theta}$.

$$\begin{aligned} W_{AB} &= h(\hat{\theta})'[\hat{H}'\hat{I}^{-1}\hat{H}]^{-1}h(\hat{\theta}) \\ &= h_1(\hat{\theta})'(\hat{H}_1'\hat{I}^{-1}\hat{H}_1)^{-1}h_1(\hat{\theta}) + h_2(\hat{\theta})'(\hat{H}_2'\hat{I}^{-1}\hat{H}_2)h_2(\hat{\theta}) \\ &\qquad\qquad \text{given } \hat{H}_1'\hat{I}^{-1}\hat{H}_2 = 0 \\ &= h_1(\dot{\theta})'(\dot{H}_1'\dot{I}^{-1}\dot{H}_1)^{-1}h_1(\dot{\theta}) + h_2(\ddot{\theta})'(\ddot{H}_2'\ddot{I}^{-1}\ddot{H}_2)h_2(\ddot{\theta}) \\ &\qquad\qquad \text{by separability} \\ &= W_A + W_B \end{aligned}$$

where "$\cdot$" and "$\ddot{}$" denote the quantities have been evaluated at the restricted maximum likelihood estimates with the restrictions $h_2(\theta) = 0$ and $h_1(\theta) = 0$ imposed respectively. If we partition the parameter vector θ as before and consider testing restrictions of the form $H_o: \theta_2 = \theta_2^*$, where θ_2^* is a vector of fixed constants, then the W test will be additive if $\hat{I}^{22}$ is block diagonal with respect to the testing parameters, where $\hat{I}^{22}$ denotes the (2,2) block of $\hat{I}^{-1}$. In the next section, we provide some examples of the additivity and non-additivity of the LR, W and LM tests.

3. Some Examples

Consider the following linear regression problem

$$y_i = X_i'\beta + u_i \qquad i = 1,2,\ldots,N$$

where X_i is a $k \times 1$ vector representing the i^{th} observation on k fixed regressors, β is a $k \times 1$ vector of fixed unknown parameters, u_i are assumed to be serially correlated $(\bar{I})$ and generated by a first order autoregressive (AR) process $u_i = \rho u_{i-1} + \varepsilon_i$, $|\rho| < 1$ where ε_i are assumed to be normally and independently distributed but heteroscedastic $(\bar{H})$ with the form $V(\varepsilon_i) = \sigma_i^2 = \sigma^2 + \alpha' Z_i$ where Z_i is an $\ell \times 1$ vector representing the i^{th} observation on ℓ fixed variables and α an $\ell \times 1$ vector of fixed unknown parameters. If we let H_o: $u \sim HI$ and denote $LM_{H\bar{I}}$, $LM_{\bar{H}}$ and $LM_{\bar{I}}$ to be the LM statistics for testing H_o against H: $u \sim \bar{H}\bar{I}$, H: $u \sim \bar{H}I$ and H: $u \sim H\bar{I}$ respectively then Bera and Jarque (1982) have shown that $LM_{H\bar{I}} = LM_{\bar{H}} + LM_{\bar{I}}$. Our results indicate that the LR test will also be additive. For this example, when $u \sim N\bar{H}\bar{I}$ we have [calculated from the derivatives (A.2)-(A.5) given in Appendix A]

$$I = \begin{bmatrix} \Sigma_i \dfrac{\bar{X}_i\bar{X}_i'}{\sigma_i^2} & 0 & 0 & 0 \\ 0 & \tfrac{1}{2}\Sigma_i\left(\dfrac{1}{\sigma_i^4}\right) & \tfrac{1}{2}\Sigma_i\left(\dfrac{Z_i'}{\sigma_i^4}\right) & 0 \\ 0 & \tfrac{1}{2}\Sigma_i\left(\dfrac{Z_i}{\sigma_i^4}\right) & \tfrac{1}{2}\Sigma_i\left(\dfrac{Z_iZ_i'}{\sigma_i^4}\right) & 0 \\ 0 & 0 & 0 & N(1+\sum_{j=1}^{\infty}\rho^{2j}) \end{bmatrix}$$

where $\bar{X}_i = (X_i - \rho X_{i-1})/\sigma_i$. From the above expression, it is easily seen that $I^{\bar{I}\bar{I}} = 0$ so that W also will be additive asymptotically.

If the regressor set includes a lagged dependent variable, say y_{i-1}, then $I_{\beta\sigma} \neq 0$, but the information matrix is still block diagonal between (β,ρ) and (σ^2,α) so that the inverse will also be block diagonal. Hence the tests for heteroscedasticity and serial correlation will still be additive[3]. The introduction of a lagged dependent variable, y_{i-1}, into Z_i will not alter the structure of the information matrix nor invalidate the additivity result.

The assumption of normality or, more importantly, the assumption that $E(\varepsilon_i^3) = 0$ is however critical. If this assumption is relaxed then $I_{\beta\sigma^2}$, $I_{\beta\alpha} \neq 0$ but block diagonality between (β,σ^2,α) and ρ holds. If, in addition, y_{i-1} is introduced into the regressor set then $I_{\beta\rho}$ is also non-zero and block diagonality is lost. Similarly if, instead of appearing in the regressor set, y_{i-1} appears in Z_i then $I_{\alpha\rho}$ is also non zero and block diagonality is lost[4]. In both these cases, additivity no longer holds.

From the previous example with only fixed regressors we can see that if $h_1(\theta)$: $R\beta = 0$, $h_2(\theta)$: $\alpha = 0$ or $h_1(\theta)$: $R\beta = 0$, $h_2(\theta)$: $\rho = 0$ that these hypotheses will be additive and separable since $H_1' I^{-1} H_2 = 0 \ \forall \ \theta$. This implies that, under normality, *if the sample is large* while testing the restrictions $R\beta = 0$ we can ignore the presence of autocorrelation or heteroscedasticity[5]. Also the different test statistics can simply be added to form a joint test. This additivity will disappear for the first hypothesis if the regressor set includes y_{i-1} and in the second case if $E(\varepsilon_i^3) \neq 0$.

---

[3] For any y_{i-j}, $j > 1$, $I_{\beta\rho} = 0$.

[4] Block diagonality is not lost if the lagged dependent variable appearing in the regressor set or Z_i is y_{i-j}, $j > 1$.

[5] For example, Phillips and McCabe (1983) have shown the independence of the common tests for stability of regression coefficients (linear restriction) and of the disturbance variance (heteroscedasticity) in linear

In the following example, the LM and LR tests are additive but the W test is not necessarily. Consider testing $u \sim N(0,1)$ in the following framework

$$\frac{d\ell_i(c_1,c_2)}{du_i} = \frac{c_1-u_i}{1-c_1u_i+c_2u_i^2}$$

where one would test H_o: $c_1 = c_2 = 0$. It is shown in Bera and Jarque (1981) that $LM_{c_1c_2} = LM_{c_1} + LM_{c_2}$ so that LR will also be additive. However $\hat{I}^{c_1c_2}$ will not in general be zero so that W may not be additive.

The last example is the case where none of the three tests are additive. This is the case of testing the null hypothesis that the disturbance term in a limited dependent variable model is normally distributed and homoscedastic against the alternative hypothesis of non-normality and heteroscedasticity [see Jarque and Bera (1982)].

4. Conclusion

Additivity of the LM test implies asymptotic additivity of the LR test but not in general additivity of the W test. This shows another computational advantage of the LM test. After carrying out one-directional LM tests, a joint test can be obtained when additivity applies simply by adding up the component statistics or a number of test statistics can be combined to form a joint test. For the LR (and sometimes for W) tests such an operation is valid only for large samples. Here we should also mention that since all three statistics are asymptotically equivalent under the null hypothesis and for local alternatives, additivity of the LM test implies asymptotic additivity of both the W and LR tests under the null hypothesis and for local alternatives.

Pagan and Hall (1983) claim that a major disadvantage of examining the additivity properties through the information matrix is that the calculation of the information matrix is dependent on certain distributional assumptions, e.g. symmetry of the disturbances, and that additivity of the tests may merely reflect this fact. One of our examples in section 3 illustrated the importance of the distributional assumptions and that account can be taken of them in the information matrix based approach.

APPENDIX

For our model $\theta = (\beta', \sigma^2, \alpha', \rho)'$, $H_0: \alpha = 0, \rho = 0$ and the log-likelihood function $\ell(\theta)$ is given by

$$\ell(\theta) = \sum_{i=1}^{N} \ell_i(\theta) = -(N/2) \ln 2\pi - \tfrac{1}{2} \sum_{i=1}^{N} \ln \sigma_i^2 - \tfrac{1}{2} \sum_{i=1}^{N} \frac{\varepsilon_i^2}{\sigma_i^2} \tag{A.1}$$

where $\sigma_i^2 = \sigma^2 + \alpha' Z_i$ and $\varepsilon_i = u_i - \rho u_{i-1}$ with $u_i = y_i - X_i'\beta$. From the above equation following first order derivatives are easily obtained

$$\frac{\partial \ell_i(\theta)}{\partial \beta} = \frac{1}{\sigma_i^2} (X_i - \rho X_{i-1}) \varepsilon_i \tag{A.2}$$

$$\frac{\partial \ell_i(\theta)}{\partial \sigma^2} = - \frac{1}{2\sigma_i^2} + \frac{\varepsilon_i^2}{2\sigma_i^4} \tag{A.3}$$

$$\frac{\partial \ell_i(\theta)}{\partial \alpha} = - \frac{Z_i}{2\sigma_i^2} + \frac{\varepsilon_i^2 Z_i}{2\sigma_i^4} \tag{A.4}$$

and

$$\frac{\partial \ell_i(\theta)}{\partial \rho} = \frac{1}{\sigma_i^2} (y_{i-1} - X_{i-1}'\beta) \varepsilon_i \, . \tag{A.5}$$

Taking cross-products of the derivatives and then taking expectations, we obtain the information matrix.

REFERENCES

Aitchison, J. (1962), "Large-sample restricted parametric test", *Journal of the Royal Statistical Society (Series B)*, 24, 234-250.

Bera, A.K. and C.M. Jarque (1981), "An efficient large-sample test for normality of observations and regression residuals", Working Papers in Economics and Econometrics, No. 040, The Australian National University.

Bera, A.K. and C.M. Jarque (1982), "Model specification tests : A simultaneous approach", *Journal of Econometrics*, 20, 59-82.

Bera, A.K. and C.R. McKenzie (1984), "Alternative forms and properties of the Lagrange multiplier test", *Journal of Applied Statistics*, (forthcoming

Davidson, R. and J.G. MacKinnon (1983), "Small sample properties of alternative forms of the Lagrange multiplier test", *Economics Letters*, 12, 269-275.

Jarque, C.M. and A.K. Bera (1982), "Efficient specification tests for limited dependent variable models", *Economics Letters*, 9, 153-160.

Pagan, A.R. and A.D. Hall (1983), "Diagnostic tests as residual analysis", *Econometric Reviews*, 2, 159-218.

Pesaran, M.H. (1979), "Diagnostic testing and exact maximum likelihood estimation of dynamic models", in *Proceedings of the Econometric Society European Meeting*, 1979, ed. E.G. Charatsis, (Amsterdam: North Holland), pp. 63-87.

Phillips, G.D.A. and B.P. McCabe (1983), "The independence of tests for structural change in regression models", *Economics Letters*, 12, 283-287.

Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, (New York: John Wiley and Sons).

UNIVERSITY OF ILLINOIS-URBANA
3 0112 037627640