THE INFORMATION MATRIX TEST FOR THE LINEAR MODEL

A.R. Hall*

University of Warwick

No. 250

February, 1984.

* This research was undertaken during the tenure of a Social Science Research Council studentship. I am extremely grateful to John Aldrych, Trevor Breusch, Grayham Mizon, Ray O'Brien, Mark Salmon and Ken Wallis for very valuable discussions on various aspects of the work.

An earlier version of this paper was presented at the 1983 SSRC Econometric Study Group Conference at Warwick. I am grateful to the participants for some very useful comments.

Paper presented at the 1983 European Meeting of the Econometric Society held at Pisa.

ABSTRACT

We derive the information matrix test, suggested by White (1982), for the normal fixed regressor linear model, and show that the statistic decomposes asymptotically into the sum of three independent quadratic forms. One of these is White's (1980) general test for heteroscedasticity and the remaining two components are quadratic forms in the third and fourth powers of the residuals respectively. Our results show that the test will fail to detect serial correlation and never be asymptotically optimal against heteroscedasticity, skewness and non-normal kurtosis. The information matrix test is contrasted with the test procedures of Bera and Jarque (1983) and Godfrey and Wickens (1982), who construct a composite statistic from asymptotically optimal and independent tests against particular alternatives. Our results suggest that this alternative strategy is likely to be a more fruitful source of a general regression diagnostic.

I. INTRODUCTION

Two types of statistical test can be constructed to assess model adequacy: pure significance tests, for which a significant statistic only implies that the model should be rejected, and constructive tests, which require the specification of a particular alternative hypothesis. The familiar Wald (W), Likelihood Ratio (LR) and Lagrange Multiplier (LM) tests fall into this second class. The advantage of constructive tests is that, by design, they are powerful against the chosen alternative, but their weakness is that this power may be at the expense of poor performance against other misspecifications. In practice it is often desired to test the model's adequacy against a variety of alternatives. However to avoid the problem of induced tests encountered with a sequence of dependent tests because the actual and nominal significance levels differ due to the statistical dependence of each test on those preceeding it in the sequence, it is desirable to construct one general test of misspecification. This can be achieved in two possible ways. We can either seek a more general testing principle than the W, LR and LM to produce a pure significance test, or, as suggested by Godfrey and Wickens (1982), try to design a composite statistic from the existing constructive tests.

White (1982) considers the properties of maximum likelihood estimators and the appropriate form of the W, LR and LM test statistics when the model is misspecified. The latter provides the motivation for this information matrix test principle. This is based on the fact that if a model is correctly specified, then the following 'information matrix' identity holds:

$$E\left[-\frac{\partial^2 L}{\partial \underset{\sim}{\theta} \partial \underset{\sim}{\theta}'}\right] = E\left[\frac{\partial L}{\partial \underset{\sim}{\theta}} \frac{\partial L}{\partial \underset{\sim}{\theta}'}\right]$$

where L is the log likelihood function and $\underset{\sim}{\theta}$ the parameters vector. The right-hand side of the identify is the variance-covariance matrix of the score vector and the left-hand side the information matrix. The familiar form of the W, LM and LR tests are derived under the assumption that this identity holds. As White (1982) observes the information matrix test can therefore be used as a preliminary test before conducting inference with the usual tests.

In this paper we consider the information matrix test as a pure significance test. Our interpretation can be justified since the derivation of the identity is based on the assumption that the model is correctly specified. Consequently violation of this identity for the model with which we are working can be interpreted as evidence of misspecification. As with the LM test, the model only requires estimation under the null hypothesis, but, unlike the LM, rejection of the null only implies model inadequacy and not its potential cause. In this paper we examine the information matrix test in the fixed regressor normal linear model. We assess exactly which misspecifications will be detected by this version of theinformation matrix test and the extent to which it may be considered a comprehensive test of model adequacy.

Our results show that the test statistic derived under the null hypothesis that the model is correctly specified is the sum of three independent components. The first of these is White's direct test for heteroscedasticity (White, 1980). The other two are quadratic forms in the third and fourth powers of the residuals respectively. The information matrix test is therefore the sum of three constructive tests. Our analysis shows that the misspecifications to which these three are sensitive, are not sufficiently varied for the test to be considered a comprehensive test of model adequacy in this case.

Section 2 briefly restates the information matrix test with slight adaptation to allow for independently but not identically distributed random variables. In Section 3 the test statistic for our model is derived. Finally in Section 4 the potential causes of misspecification to which the test is sensitive are discussed.

2. THE INFORMATION MATRIX TEST

White (1982) follows the usual maximum likelihood approach in deriving his results for i.i.d. variables. The case being considered here is

$$y_t \sim IN(\underset{\sim}{x}_t'\beta,\ \sigma^2 \underset{\sim}{I})$$

i.e. independent but not identically distributed random variables. We therefore have to extend White's work to cover this more general situation. The assumptions required to ensure that his theory translates to our model are given below. In the following we give the definitions appropriate to the i.n.i.d. case, but follow White's line of argument.

Let the assumed p.d.f. of y_t be $f_t(y_t, \underset{\sim}{\theta})$ where $\underset{\sim}{\theta}$ is a p x 1 parameter vector. The parameter estimates obtained from maximising the assumed log likelihood are $\hat{\underset{\sim}{\theta}}_n$. If the model is correctly specified then the following information matrix identity holds:-

$$\underset{\sim}{A}(\underset{\sim}{\theta}) + \underset{\sim}{B}(\underset{\sim}{\theta}) = \underset{\sim}{0}\ ;$$

where $$\underset{\sim}{A}(\underset{\sim}{\theta}) = \lim_{n\to\infty} \{n^{-1} \sum_{t=1}^{n} E\left[\frac{\partial^2 \log f_t(y_t, \underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j}\right]\}$$

and the variance covariance matrix of the score vector is

$$\underset{\sim}{B}(\underset{\sim}{\theta}) = \lim_{n\to\infty} \{n^{-1} \sum_{t=1}^{n} E\left[\frac{\partial \log f_t(y_t, \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}}\right]\left[\frac{\partial \log f_t(y_t, \underset{\sim}{\theta})}{\partial \underset{\sim}{\theta}'}\right]\}$$

The matrices $\underset{\sim}{A}(\underset{\sim}{\theta})$ and $\underset{\sim}{B}(\underset{\sim}{\theta})$ are unobservable, but can be consistently estimated by $\underset{\sim}{A}_n(\underset{\sim}{\theta})$ and $\underset{\sim}{B}_n(\underset{\sim}{\theta})$ respectively where

$$\underset{\sim}{A}_n(\underset{\sim}{\theta}) = \{n^{-1} \sum_{t=1}^{n} \frac{\partial^2 \log f_t(y_t,\underset{\sim}{\theta})}{\partial\theta_i \partial\theta_j}\},$$

$$\underset{\sim}{B}_n(\underset{\sim}{\theta}) = \{n^{-1} \sum_{t=1}^{n} \frac{\partial \log f_t(y_t,\underset{\sim}{\theta})}{\partial\theta_i} \frac{\partial \log f_t(y_t,\underset{\sim}{\theta})}{\partial\theta_j}\}$$

Subject to certain regularity conditions (see White, 1982). the elements of $\sqrt{n}\,(\underset{\sim}{A}_n(\hat{\theta}_n) + \underset{\sim}{B}_n(\hat{\underset{\sim}{\theta}}_n))$ are distributed asymptotically jointly normal with mean zero if the model is correctly specified. Due to the symmetry of these matrices, the test of misspecification need only be based on the $p(p+1)/2$ indicator vector $\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$ whose elements consist of the lower triangular elements of $(\underset{\sim}{A}_n(\hat{\underset{\sim}{\theta}}_n) + \underset{\sim}{B}_n(\hat{\underset{\sim}{\theta}}_n))$. Accordingly we define the indicator vector as follows:

Let

$$d_{kt} = \partial \log f_t(y_t,\underset{\sim}{\theta})/\partial\theta_i \cdot \partial \log f_t(y_t,\underset{\sim}{\theta})/\partial\theta_j + \partial^2 \log f_t(y_t,\underset{\sim}{\theta})/\partial\theta_i \partial\theta_j ,$$

where $k = 1, \ldots, p(p+1)/2$; $i=1, \ldots, p$; $j=1, \ldots, p$, $i \geq j$,

with $k = (j-1)(p-j/2) + i$

The indicator vector is then $\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$ whose k^{th} element is

$$D_{kn}(\hat{\underset{\sim}{\theta}}_n) = n^{-1} \sum_{t=1}^{n} d_{kt}(y_t,\hat{\underset{\sim}{\theta}}_n).$$

The test can be based on the full set of $p(p+1)/2$ indicators or a subset of q of them. We therefore define $\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$ to be a $q \times 1$ vector with $q \leq p(p+1)/2$.

The variance covariance matrix of $\sqrt{n}\, \underset{\sim}{D}_n(\underset{\sim}{\theta}_n)$ can be shown to be $\underset{\sim}{V}(\underset{\sim}{\theta})$ evaluated at $\underset{\sim}{\theta}=\underset{\sim}{\theta}_o$ the true parameter values. $\underset{\sim}{V}(\underset{\sim}{\theta})$ is defined by

$$\underset{\sim}{V}(\underset{\sim}{\theta}) = \lim_{n\to\infty} \{n^{-1} \sum_{t=1}^{n} E\left[\underset{\sim}{\alpha}_t \underset{\sim}{\alpha}_t'\right]\}$$

where $\underset{\sim}{\alpha}_t = \underset{\sim}{d}_t(y_t,\underset{\sim}{\theta}) - \underset{\sim}{\nabla D}(\underset{\sim}{\theta})\underset{\sim}{A}(\underset{\sim}{\theta})^{-1}\underset{\sim}{\nabla}\log \underset{\sim}{f}_t(y_t,\underset{\sim}{\theta})$;

$$\underset{\sim}{\nabla D}(\underset{\sim}{\theta}) = \lim_{n\to\infty} \{n^{-1} \sum_{t=1}^{n} E \frac{\partial d_{kt}(y_t,\underset{\sim}{\theta})}{\partial \theta_i}\}$$

and $\underset{\sim}{\nabla}\log \underset{\sim}{f}_t(y_t,\underset{\sim}{\theta})$ is the score vector for one observation.

The matrix $\underset{\sim}{V}(\underset{\sim}{\theta}_o)$ is unobservable, but can be consistently estimated by $\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$, defined as follows:

$$\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n) = n^{-1} \sum_{t=1}^{n} \underset{\sim}{a}_t \underset{\sim}{a}_t',$$

$$\underset{\sim}{a}_t = \underset{\sim}{d}_t(y_t,\underset{\sim}{\theta}) - \underset{\sim}{\nabla D}_n(\hat{\underset{\sim}{\theta}}_n)\underset{\sim}{A}_n(\hat{\underset{\sim}{\theta}}_n)^{-1}\underset{\sim}{\nabla}\log \underset{\sim}{f}_t(y_t,\underset{\sim}{\theta}) ,$$

$$\underset{\sim}{\nabla D}_n(\hat{\underset{\sim}{\theta}}_n) = \{n^{-1}\sum_{t=1}^{n} \partial\frac{d_{kt}(y_t,\hat{\underset{\sim}{\theta}}_n)}{\partial\theta_i}\}$$

In White's derivation of the information matrix test statistic (see appendix of White, 1982), his proof rests on five key results:

1) the convergence in strong probability of $\hat{\underset{\sim}{\theta}}{}^*_n = (\hat{\underset{\sim}{\beta}}', \hat{\sigma}^2)$ to $\underset{\sim}{\theta}_o' = (\underset{\sim}{\beta}', \sigma^2)$.
2) $\underset{\sim}{\nabla D}(\underset{\sim}{\theta}_o)$ is finite.
3) the mean value theorem can be applied to $\sqrt{n}\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$.
4) $\hat{\underset{\sim}{\theta}} - \underset{\sim}{\theta}_o$ has a well defined asymptotic distribution.

5) A central limit theorem can be applied to the score vector in the i.n.i.d. case .

The assumptions typically made in the fixed regressor model ensure that these five points hold. In particular it is necessary to assume:-

(i) plim $\underset{\sim}{X}'\underset{\sim}{X}/n$ is finite and positive definite

and (ii) plim $(\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{x}_t\underset{\sim}{x}_t' = \underset{\sim}{0} \quad \forall t$

Under the null hypothesis that the model is correctly specified (i.e. $f_t(y_t,\underset{\sim}{\theta}_o)$ is the true distribution) and given the above assumptions, the information matrix test statistic is then

$$n\,\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)'\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)^{-1}\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n),$$

which will be asymptotically distributed as χ^2_q on the null.

3. DERIVATION OF THE TEST STATISTIC FOR THE FIXED REGRESSOR NORMAL LINEAR MODEL.

The assumed model specification is:

$$\underset{\sim}{y} = \underset{\sim}{X}\underset{\sim}{\beta} + \underset{\sim}{u} \qquad \underset{\sim}{u} \sim N(\underset{\sim}{0},\sigma^2 \underset{\sim}{I}_n).$$

Let $\lim_{n\to\infty} \frac{\underset{\sim}{X}'\underset{\sim}{X}}{n} = \underset{\sim}{M}$: a KxK positive definite, finite matrix whose $(i,j)^{th}$ element is denoted m_{ij}.

Let the log likelihood of one observation be LLF :

$$LLF = k - \tfrac{1}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}(y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta})^2$$

The first order partial derivatives with respect to $\underset{\sim}{\beta}$ and σ^2 are:-

$$\frac{\partial LLF}{\partial \underset{\sim}{\beta}} = \frac{1}{\sigma^2}(y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta})\underset{\sim}{x}_t ,$$

$$\frac{\partial LLF}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4}(y_t - \underset{\sim}{x}_t'\underset{\sim}{\beta})^2 .$$

The second order partial derivatives are then:

$$\frac{\partial^2 LLF}{\partial \underset{\sim}{\beta}\partial \underset{\sim}{\beta}'} = -\frac{1}{\sigma^2}\underset{\sim}{x}_t\underset{\sim}{x}_t' ,$$

$$\frac{\partial^2 LLF}{\partial \underset{\sim}{\beta}\partial \sigma^2} = -\frac{1}{\sigma^4}u_t\underset{\sim}{x}_t ,$$

$$\frac{\partial^2 LLF}{\partial (\sigma^2)^2} = \frac{1}{2\sigma^4} - \frac{1}{\sigma^6}u_t^2 .$$

With our assumed specification the parameter estimates are

$$\hat{\underset{\sim}{\theta}}_n = \begin{bmatrix} \hat{\underset{\sim}{\beta}} \\ \hat{\sigma}^2 \end{bmatrix} = \begin{bmatrix} (\underset{\sim}{X}'\underset{\sim}{X})^{-1}\underset{\sim}{X}'\underset{\sim}{y} \\ \hat{\underset{\sim}{u}}'\hat{\underset{\sim}{u}}/n \end{bmatrix}$$

This implies that

$$\underset{\sim}{B}_n(\hat{\underset{\sim}{\theta}}_n) = n^{-1} \sum_{t=1}^{n} \begin{bmatrix} \frac{1}{\hat{\sigma}^4} \hat{u}_t^2 \underset{\sim}{x}_t \underset{\sim}{x}_t' & \\ \frac{-1}{2\hat{\sigma}^4} \hat{u}_t \underset{\sim}{x}_t' + \frac{1}{2\hat{\sigma}^6} \hat{u}_t^3 \underset{\sim}{x}_t' \, , & \frac{1}{4\hat{\sigma}^4} - \frac{\hat{u}_t^2}{2\hat{\sigma}^6} + \frac{\hat{u}_t^4}{4\hat{\sigma}^8} \end{bmatrix}$$

$$\underset{\sim}{A}_n(\hat{\underset{\sim}{\theta}}_n) = n^{-1} \sum_{t=1}^{n} \begin{bmatrix} -\frac{1}{\hat{\sigma}^2} \underset{\sim}{x}_t \underset{\sim}{x}_t' & \\ -\frac{\hat{u}_t \underset{\sim}{x}_t'}{\hat{\sigma}^4} & \frac{1}{2\hat{\sigma}^4} - \frac{\hat{u}_t^2}{\hat{\sigma}^6} \end{bmatrix}$$

Note that under H_o: $\sum_{t=1}^{n} \hat{u}_t \underset{\sim}{x}_t = \underset{\sim}{0}$.

The indicator vector therefore comprises three subvectors. The first $K(K+1)/2$ elements of $\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$ are the subvector

$$\underset{\sim}{\Delta}_1 = n^{-1} \sum_{t=1}^{n} \frac{1}{\hat{\sigma}^4} \begin{bmatrix} (\hat{u}_t^2 - \hat{\sigma}^2) x_{1t}^2 \\ \vdots \\ (\hat{u}_t^2 - \hat{\sigma}^2) x_{Kt}^2 \end{bmatrix} ,$$

whose sth element is of the form $n^{-1} \sum_{t=1}^{n} (\hat{u}_t^2 - \hat{\sigma}^2)\, x_{it} x_{jt} / \hat{\sigma}^4$ where $s = (j-1)(K-j/2)+i \quad i \geq j \; , \; i \; , \; j = 1, \, .. \, , K$. This subvector examines the discrepancy between the two estimators of the variance-covariance matrix of the $\underset{\sim}{\beta}$ coefficients. The next K elements are the subvector

$$\underset{\sim}{\Delta}_2 = n^{-1} \sum_{t=1}^{n} \frac{1}{2\hat{\sigma}^6} \hat{u}_t^3 \underset{\sim}{x}_t' .$$

This vector measures the discrepancy in the estimators of the covariance of $\underset{\sim}{\beta}$ and σ^2 .

The last element of $\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$ is Δ_3 which compares the estimators of the variance of $\hat{\sigma}^2$.

$$\Delta_3 = n^{-1} \sum \frac{\hat{u}_t^4}{4\hat{\sigma}^8} - \frac{3}{4\hat{\sigma}^4}$$

The indicator vector is then given by

$$\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n) = \begin{bmatrix} \underset{\sim}{\Delta}_1 \\ \underset{\sim}{\Delta}_2 \\ \Delta_3 \end{bmatrix} .$$

We now need a consistent estimate of the covariance matrix of $\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$, $\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$, which we will construct from the sample moments of the subvectors $\underset{\sim}{\Delta}_i$. Let $n^{-1} \sum_{t=1}^{n} \hat{\underset{\sim}{\Delta}}_t^{ij}$ be the sample covariance between $\underset{\sim}{\Delta}_i$ and $\underset{\sim}{\Delta}_j$. The matrix $\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$ can then be written as, giving lower triangular elements only:-

$$\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n) = n^{-1} \sum_{t=1}^{n} \begin{bmatrix} \hat{\underset{\sim}{\Delta}}_t^{11} & & \\ \hat{\underset{\sim}{\Delta}}_t^{21} & \hat{\underset{\sim}{\Delta}}_t^{22} & \\ \hat{\underset{\sim}{\Delta}}_t^{31} & \hat{\underset{\sim}{\Delta}}_t^{32} & \hat{\underset{\sim}{\Delta}}_t^{33} \end{bmatrix}$$

The appropriate form of the sample covariance matrices is given in the table below. The derivation of the matrix is left to an appendix. To simplify the notation let :-

$\xi_{ijt} = x_{it}x_{jt} - (\underset{\sim}{x}_i'\underset{\sim}{x}_j)/n$ and $\underset{\sim}{\xi}_t$ is the vector whose sth element is ξ_{ijt} where $s = (j-1)(K-j/2)+i$ $i \geq j$, $i,j = 1, \ldots, K$

and $z = \dfrac{3}{4\sigma^4} - \dfrac{3u_t^2}{2\sigma^6} + \dfrac{u_t^4}{4\sigma^8}$.

Submatrix	dimension	typical element
$\hat{\Delta}_t^{11}$	$K(K+1)/2 \times K(K+1)/2$	$(\hat{u}_t^2 - \hat{\sigma}^2)^2 \xi_{mpt}\xi_{rst}/\hat{\sigma}^8$: (i,j)th element given by $i=(p-1)(K-p/2)+m$, $j=(s-1)(K-s/2)+r$
$\hat{\Delta}_t^{21}$	$K \times (K+1)K/2$	$(\hat{u}_t^2 - \hat{\sigma}^2)(\hat{u}_t^2 - 3\hat{\sigma}^2)\hat{u}_t x_{it}\xi_{rst}/2\hat{\sigma}^{10}$ $(i,j)^{th}$ element given by $i=i$, $j=(s-1)(K-s/2)+r$
$\hat{\Delta}_t^{31}$	$1 \times K(K+1)/2$	$(\hat{u}_t^2 - \hat{\sigma}^2)\hat{z}\xi_{rst}/\hat{\sigma}^4$: j^{th} element given by $j=(s-1)(K-s/2)+r$
$\hat{\Delta}_t^{22}$	$K \times K$	$\frac{1}{4}(\hat{u}_t^2 - 3\hat{\sigma}^2)^2\hat{u}_t^2 x_{it}x_{jt}/\hat{\sigma}^{12}$: $(i,j)^{th}$ element
$\hat{\Delta}_t^{32}$	$1 \times K$	$(\hat{u}_t^2 - 3\hat{\sigma}^2)\hat{u}_t\hat{z}x_{it}/2\hat{\sigma}^6$: i^{th} element
$\hat{\Delta}_t^{33}$	1×1	$\hat{z}^2$

It is possible to simplify this expression for $\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$ by using the fact that is asymptotically equivalent to $\underset{\sim}{V}(\underset{\sim}{\theta}_o)$. It can be shown that under H_o $\underset{\sim}{V}(\underset{\sim}{\theta}_o)$ is block diagonal, and so can be consistently estimated by a block diagonal $\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$. This result is derived in the appendix, and depends only on two properties of the model. Firstly that under H_o u_t and $\underset{\sim}{x}_t'$ are independent, and secondly that u_t is assumed to have a symmetric distribution about zero, and finite moments up to and including order eight. For the present we assume u_t to be distributed normally, and give a more robust version of the test in section 4.

$\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$ is then given by:-

$$\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n) = n^{-1}\sum_{t=1}^{n}\begin{bmatrix} 2\hat{\sigma}^{-4}\underset{\sim}{\xi}_t\underset{\sim}{\xi}_t' & & \\ & \frac{3}{2}\hat{\sigma}^{-6}\underset{\sim}{x}_t\underset{\sim}{x}_t' & \\ & & \frac{3}{2}\hat{\sigma}^{-8} \end{bmatrix}$$

The test statistic is then $n\,\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)'\left[\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)\right]^{-1}D_n(\hat{\underset{\sim}{\theta}}_n)$ which is distributed asymptotically central χ^2 with $(K+1)(K+2)/2$ degrees of freedom.

Therefore under H_o, the information matrix test is the sum of three independent components, the first of which is easily recognised to be asymptotically equivalent to White (1980)'s direct test of heteroscedasticity. This enables the calculation of the test statistic to be considerably simplified, as it is the sum of the following three statistics.

(1) nR^2 from the regression of $\hat{u}_t^2$ on a constant and $\sum_{j=1}^{K}\sum_{k=j}^{K}\alpha_s x_{jt}x_{kt}$ (where R^2 is the constant adjusted coefficient of multiple correlation). White (1980) has shown that this corresponds to the statistic for the first $K(K+1)/_2$ elements of $\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$. If $x_{it} = 1, \forall t$, then x_{1t}^2 would be omitted from this regression, as indeed would the indicator involving x_{1t}^2 be omitted from the test, thereby reducing the degrees of freedom by 1.

(2) $n \frac{1}{6} \hat{\sigma}^{-6}$ ESS : where ESS is the explained sum of squares from the regression of $\hat{u}_t^3$ on $\underset{\sim}{x}_t'$, $\hat{\sigma}$ coming from original OLS regression of y_t on $\underset{\sim}{x}_t'$. This in turn under H_o is nR^2 from the regression of $\frac{\sqrt{2}}{\sqrt{5}} \hat{u}_t^3$ on $\underset{\sim}{x}_t'$ which equals nR^2 from the regression of $\hat{u}_t^3$ on $\underset{\sim}{x}_t'$.

(3) $\frac{n}{24\sigma^8} (n^{-1}\Sigma \hat{u}_t^4 - 3\hat{\sigma}^4)^2$.

The analysis in Section 4 is concerned with the fixed regressor model, but it is important to note that these conditions are sufficient but not necessary for the above decomposition of the test. The result requires only the independence of regressors and error. It therefore extends to the stochastic regressor model and, for example, applies to the dynamic regression model, provided that the assumptions made in the derivation in the appendix concerning the moments of x_t are satisfied.

4. ANALYSIS OF THE INFORMATION MATRIX TEST AS A TEST OF MISSPECIFICATION

We have proposed the information matrix test as a pure significance test of model adequacy. Rejection of the null hypothesis implies only that the model is inadequate and it is therefore interesting to examine the misspecifications that may cause a significant statistic. This enables us to assess whether the complete information matrix test provides a comprehensive test, and to suggest possible reductions of the indicator vector that deliver more powerful tests against specific hypotheses.

As Hausman (1978) notes most econometric specification tests for the linear model are designed to pick up violations of (i) the orthogonality assumption: $E(\underset{\sim}{u}|\underset{\sim}{X}) = \underset{\sim}{0}$ or plim $\frac{1}{T}\underset{\sim}{X}'\underset{\sim}{u} = 0$ and (ii) the sphericality assumption : $V(\underset{\sim}{u}|\underset{\sim}{X}) = \sigma^2\underset{\sim}{I}_n$. We therefore start by considering the extent to which such misspecifications are detected by the present test.

Violation of the orthoganlity assumption will not be detected directly, but this in turn will cause heteroscedasticity which, it is argued below will be detected. The original vector $\underset{\sim}{D}_n(\underset{\sim}{\theta})$ contains terms in $\underset{\sim}{X}'\underset{\sim}{u}$ but these disappear when evaluated at $\hat{\underset{\sim}{\theta}}_n$ due to the orthogonality of the fixed regressor matrix to the OLS residuals. However the existence of systematic components in the error process, say due to an omitted variable, would cause the expected and actual properties of higher moments of the error to differ, and we now examine the extent to which this is detected by the test.

The sphericality assumption can be violated by heteroscedasticity and/or serial correlation in the errors. We consider these in turn.

It was remarked earlier that one component of the information matrix test was asymptotically equivalent to White's direct test under normality. This in turn is easily seen to be the LM test (Breusch and Pagan, 1979) against heteroscedasticity of the form $\sigma_t^2 = h\left(\sum_{ij}^{K} \alpha_s x_{it} x_{jt} + \alpha_o\right)$. Therefore whilst his direct test will have asymptotically unit power, as White notes (p.826), his test will only be locally optimal if this implicit H_1 coincides with the true σ_t^2 specification. However the choice of the correct variables for the H_1 specification requires the additional information, the very absence of which, in most cases, led White to suggest this general test. As one would expect, the greater the discrepancy between this implicit H_1 and the true specification the greater the loss of power, (see Hall (1982)).

This effect is accentuated by the introduction of the remaining full information matrix test indicators. Koenker (1981) has shown that the Breusch Pagan LM statistic is distributed asymptotically non central χ^2 under local H_1 of the form $\sigma_t^2 = (1+g(\underset{\sim}{z}_t'\underset{\sim}{\gamma}_o/\sqrt{n})\sigma^2$. This has been extended in Hall (1982), using similar arguments to those used in Koenker's derivation, to the case where the variables in the implicit H_1 and true specification do not coincide. The noncentrality parameter is the only factor affected by this deviation. White's direct test statistic will therefore be a noncentral χ^2 under local H_1 (of the above formulation) regardless of the composition of $\underset{\sim}{Z}$. It can also be shown by using Amemiya's (1978) residual decomposition and similar arguments to Koenker (1981) (see Hall (1982)), that the remaining quadratic forms from the information matrix test will be distributed asymptotically central χ^2 with K and 1 degrees of freedom respectively. Further these three quadratic forms will be independent under this H_1. Therefore by the additive property of χ^2 random variables, namely

$$\Sigma \chi^2_{p_i}(m_i) = \chi^2_{\Sigma p_i}(\Sigma m_i),$$

the effect of introducing the remaining indicators is purely to increase the degrees of freedom of the test statistic. This of course, causes a loss of power that will increase with the numbers of regressors.

Whilst some of the indicators are sensitive to heteroscedasticity, none of them will pick up serial correlation. This is easily seen by examining the principle behind White's direct test. It looks at the discrepancy between two estimators of the standard errors of OLS estimator. One involves using $n^{-1}\Sigma\, \underset{\sim}{x}_t \underset{\sim}{x}_t' \hat{u}_t^2$ as an estimator of $\underset{\sim}{X}'\underset{\sim}{V}\underset{\sim}{X}/n$ and the other $\hat{\sigma}^2(\underset{\sim}{X}'\underset{\sim}{X}/n)$. The two coincide when the errors are homoscedastic, but only the first of these estimators is consistent under heteroscedasticity. If the errors are serially correlated, although both biased and inconsistent, the two estimators of $\underset{\sim}{X}'\underset{\sim}{V}\underset{\sim}{X}/n$ will be asymptotically equivalent. The remaining indicators will be similarly insensitive to serial correlation, and so the information matrix test will have power equal to its size against this mispecification.[1/]

1/ Following White (1982) we have considered a sample of observations on a univariate endogeneous variable. We could have considered a sample of one on a (Tx1) random vector y. However due to the orthogonality of the regressors and least squares residuals, this alternative version of the test will only be meaningful if we impose the structure of the covariance of u under the alternative on $B_n(\hat{\theta}_n)$. This is clearly not in the spirit of the way in which the test has been set up.

It is worth noting that although the derivation of our statistic is unaltered by the presence of lagged dependent variables in x_t, the above argument would not apply. If u_t is modelled as white noise, but is in fact serially correlated, then plim $n^{-1}\sum_{t=1}^{n} u_t^2 x_t x_t'$ does not equal plim $\hat{\sigma}^2 n^{-1}\sum_{t=1}^{n} x_t x_t'$ provided the lagged dependent variable is not of lag length less than the orders of serial correlation. However we will again encounter a loss of power from the inclusion of unnecessary indicators as those two probability units are the same for the fixed regressor elements of x_t.

Our test will therefore be sensitive to heteroscedasticity and consequently would detect implicitly any other misspecifications that cause this effect. In this class would be, for instance, parameter variation, omitted variables (unless they are orthogonal to the regressors) and incorrect functional form.

The remaining indicators involving $\hat{u}_t^3$ and $\hat{u}_t^4 - 3\hat{\sigma}^4$ will be sensitive to skewness and nonnormal kurtosis of the errors respectively, although the appropriate form of the critical region requires investigation. If $\underset{\sim}{x}_t$ contains a constant term then the information matrix test principle could be used to derive the following independent statistics:

$$n\left[n^{-1}\sum_{t=1}^{n}\frac{1}{6\hat{\sigma}^6}\hat{u}_t^3\right]^2$$

and $$n\left[n^{-1}\sum_{t=1}^{n}\frac{\hat{u}_t^4 - 3\hat{\sigma}^4}{24\hat{\sigma}^8}\right]^2$$

These are asymptotically equivalent under the null hypothesis to the tests for skewness and kurtosis suggested by Bowman and Shenton (1975) and the LM test for normality based on (i) the Edgeworth expansion (Keifer and Salmon, 1983) and (ii) the Pearson Family of distributions (Bera and Jarque, 1980). In fact all three components of the information matrix test will be sensitive to non normality. Koenker (1981) notes that the Breusch-Pagan LM statistic is sensitive to non normal kurtosis, and suggests 'studentising' the test by dividing the quadratic form by $\frac{1}{n}\sum_{u=1}^{n}(\hat{u}_t^2 - \hat{\sigma}^2)^2$ rather than $2\hat{\sigma}^4$. This ensures robustness against non normal errors. White's original direct test therefore corresponds to the studentised LM test of homscedastic errors against the alternative: $\sigma_t^2 = h(\alpha_o + \sum_{i,j}\alpha_s x_{it}x_{jt})$.

To reduce this sensitivity to non normality, the information matrix test can be studentised in a similar fashion. The above arguments for the block diagonality of $\underset{\sim}{V}(\underset{\sim}{\theta})$ rested purely on the requirement that

$$E(u_t^\alpha) = \begin{cases} 0 & \alpha = 1,3,5,7 \\ \text{finite} & \alpha = 2,4,6,8 \end{cases}$$

for which normality is sufficient. The studentised version of the information matrix test consists of the same indicator vector as before, with the following variance-covariance matrix:-

$$\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n) = \left[\begin{array}{c|c|c} (\hat{\mu}_4-\hat{\sigma}^4)\, n^{-1}\sum_{t=1}^{n} \underset{\sim}{\xi}\,\underset{\sim}{\xi}' & \underset{\sim}{0} & \underset{\sim}{0} \\ \hline \underset{\sim}{0} & (\hat{\mu}_6-6\hat{\mu}^4\hat{\sigma}^2+9\hat{\sigma}^6)\dfrac{\underset{\sim}{X}'\underset{\sim}{X}}{n} & \underset{\sim}{0} \\ \hline \underset{\sim}{0} & \underset{\sim}{0} & \dfrac{1}{16\hat{\sigma}^{16}}\quad \hat{\mu}_8-12\hat{\mu}_6\hat{\sigma}^2+42\hat{\mu}_4\hat{\sigma}^4-27\hat{\sigma}^2 \end{array}\right]$$

where $\hat{\mu}_i = n^{-1}\sum_{t=1}^{n}\hat{u}_t^i$.

This amended test is sensitive to heteroscedasticity, skewness and non normal kurtosis. All the arguments given above apply to this case as well, the only change being in the scaling factor on the non-centrality parameters. Within this framework, the White direct test corresponds to the quadratic form in the subvector consisting of all the indicators sensitive to heteroscedasticity derived from the information matrix test.

Our analysis has been concerned with the complete information matrix test. As noted earlier in White's (1982) derivation of the test there is no need to restrict attention to this case alone, and we can reduce the number of indicators to gain power against particular alternatives. The information matrix test principle can therefore be used to derive White's (1980) direct test of heteroscedasticity and tests of skewness and kurtosis asymptotically equivalent under the null to those already familiar in the literature.

We have proposed the complete information matrix test as a general test of mode adequacy, but our results suggest that it should not be considered as such for our model. The seemingly pure significance test decomposed into three constructive tests. When viewed as this combination, the test has two drawbacks, firstly the insensitivity of all three components to serial correlation is a serious problem with time series data. Secondly within the framework several misspecifications may cause heteroscedasticity, and whilst this may be detected by the first component we have no guide to the source of this effect. The implicit alternative may also not be appropriate for the case in hand, and the heteroscedasticity inducing misspecifications may be better detected by more constructive tests. These can be overcome by adopting the alternative strategy suggested by Godfrey and Wickens (1982) of combining asymptotically independent constructive tests. Our results suggest that this is likely to be a more fruitful way of deriving a general test of model adequacy, that avoids the problems of induced tests. This agrees with the testing strategy suggested by Bera and Jarque (1982). They derive a series of LM tests of homoscedasticity, serial independence, normality and correct functional form in turn, which are asymptotically independent within their model. Whilst the LM test of the joint null hypothesis is the sum of the individual tests taken together they form a comprehensive test of model adequacy. The advantage of their method compared to calculating the complete information matrix test is that by testing each hypothesis in turn the cause of misspecification can be identified and the loss of power encountered by the inclusion of unnecessary indicators avoided.

We can partially offset these problems with the complete information matrix test by taking advantage of the decomposition demonstrated above and by introducing a fourth quadratic form designed to test for serial correlation. For instance we could introduce the following LM test statistic of H_o: serially uncorrelated errors against H_1: $u_t = \rho u_{t-1} + \varepsilon_t$ is white noise:-

$$\frac{T \sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=1}^{T} \hat{u}_t^2}$$

This statistic is asymptotically independent of the original three under the null hypothesis. It is worth noting however that this composite statistic cannot be derived from the fixed regressor normal linear model with first order autoregressive errors using the information matrix test principle. For this model we have three additional indicator subvectors comparing the estimators of the covariance of $\hat{\rho}$ and $\hat{\beta}$, $\hat{\rho}$ and $\hat{\sigma}^2$ and the variance of $\hat{\rho}$, within this model we also lose the extreme form of the block diagonality of the covariance matrix present in our original model. For within this more general model the covariance between any two non identical subvectors is not always zero. We can therefore show that whilst the decomposition of our original test applies to the stochastic regressor case, given independence of regressors and error, our test does not similarly decompose into the sum of quadratic forms in each of the subvectors for the fixed regressor normal linear model with serially correlated errors.

Appendix.

(i) We begin by deriving the correct expression for $\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$. It was stated earlier that:-

$$\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n) = n^{-1}\sum_{t=1}^{n} \underset{\sim}{a}_t \, \underset{\sim}{a}_t'$$

where $\underset{\sim}{a}_t = \underset{\sim}{d}_t(y_t\underset{\sim}{\theta}) - \nabla\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)\underset{\sim}{A}_n(\hat{\underset{\sim}{\theta}}_n)^{-1}\underset{\sim}{\nabla}\log \underset{\sim}{f}_t(y_t,\hat{\underset{\sim}{\theta}}_n)$.

The first step is therefore to calculate $\underset{\sim}{a}_t$. We begin by deriving $\nabla\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$. Our line of argument will be that because $\nabla\underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)$ and $\nabla\underset{\sim}{D}(\underset{\sim}{\theta}_o)$ are asymptotically equivalent, it is possible to simplify the former by examining the latter under H_o.

$$\underset{\sim}{d}_t(y_t,\underset{\sim}{\theta}) = \begin{bmatrix} \frac{1}{\sigma^4}(u_t^2 - \sigma^2)x_{1t}^2 \\ \vdots \\ \frac{1}{\sigma^4}(u_t^2 - \sigma^2)x_{Kt}^2 \\ \frac{1}{2\sigma^6}(u_t^2 - 3\sigma^2)u_t\underset{\sim}{x}_t \\ \frac{3}{4\sigma^4} - \frac{3u_t^2}{2\sigma^6} + \frac{u_t^4}{4\sigma^8} \end{bmatrix}$$

This implies that:-

$$
\underset{\sim}{\nabla}\underset{\sim}{D}(\underset{\sim}{\theta}) = \lim n^{-1} \Sigma E
\left[\begin{array}{c:c}
\frac{-2}{\sigma^4}\, u_t x_{1t}^2 \underset{\sim}{x}_t' & -\frac{(2u_t^2-\sigma^2)}{\sigma^6}\, x_{1t}^2 \\
\vdots & \vdots \\
\frac{-2}{\sigma^4}\, u_t x_{Kt}^2 \underset{\sim}{x}_t' & -\frac{(2u_t^2-\sigma^2)}{\sigma^6}\, x_{Kt}^2 \\
\hdashline
\left[\frac{-3u_t^2}{2\sigma^6} x_{1t} + \frac{3\sigma^2 x_{1t}}{2\sigma^6}\right] \underset{\sim}{x}_t' & \frac{-3u_t^3 x_{1t}}{2\sigma^8} + \frac{3u_t x_{1t}}{\sigma^6} \\
\vdots & \vdots \\
\left[\frac{-3}{2\sigma^6}(u_t^2-\sigma^2) x_{Kt}\right] \underset{\sim}{x}_t' & \frac{-3u_t^3 x_{Kt}}{2\sigma^8} + \frac{3u_t x_{Kt}}{\sigma^6} \\
\hdashline
\left(\frac{3u_t}{\sigma^6} - \frac{u_t^3}{\sigma^8}\right) \underset{\sim}{x}_t' & \frac{-3}{2\sigma^6} + \frac{9u_t^2}{2\sigma^8} - \frac{u_t^4}{\sigma^{10}}
\end{array}\right]
\begin{array}{l}
\left.\begin{array}{c} \\ \\ \\ \end{array}\right\} K(K+1)/2 \\
\left.\begin{array}{c} \\ \\ \\ \end{array}\right\} K \\
\left.\begin{array}{c} \\ \end{array}\right\} 1
\end{array}
$$

$$\underbrace{\qquad\qquad\qquad}_{K} \qquad\qquad \underbrace{\qquad\qquad}_{1}$$

Under H_o $E(u_t^{\alpha} x_{it} x_{jt}) = E(u_t^{\alpha}) E(x_{it} x_{jt})$ since u_t and $\underset{\sim}{x}_t'$ are independent, $E(u_t) = E(u_t^3) = 0$, and $E(u_t^2) = \sigma^2$ and $E(u_t^4) = 3\sigma^4$,

thus

$$
\underset{\sim}{\nabla}\underset{\sim}{D}(\underset{\sim}{\theta}_o) =
\left[\begin{array}{c:c}
0 & \begin{array}{c} \frac{-1}{\sigma^4}\, m_{11} \\ \vdots \\ \frac{-1}{\sigma^4}\, m_{KK} \end{array} \\
\hdashline
0 & 0 \\
\hdashline
0 & 0
\end{array}\right]
\quad \text{where} \quad \underset{\sim}{\theta}_o = [\underset{\sim}{\beta}, \sigma^2]
$$

This can be consistently estimated by the matrix $\nabla \underset{\sim}{D}_n(\underset{\sim}{\theta}_n)$ whose only non zero elements occur in the top right hand block, namely

$$\begin{bmatrix} \frac{-1}{\hat{\sigma}^4} & \underset{\sim}{x}_1'\underset{\sim}{x}_1/n \\ & \vdots \\ \frac{-1}{\hat{\sigma}^4} & \underset{\sim}{x}_K'\underset{\sim}{x}_K/n \end{bmatrix}$$

where $\underset{\sim}{x}_i' = (x_{i1}x_{i2} \ldots x_{in})$ as opposed to $\underset{\sim}{x}_t' = (x_{it} \ldots x_{Kt})$

We have already shown that:-

$$\underset{\sim}{A}_n(\hat{\underset{\sim}{\theta}}_n) = \begin{bmatrix} \frac{-1}{\hat{\sigma}^2}\frac{\underset{\sim}{X}'\underset{\sim}{X}}{n} & 0 \\ 0 & -\frac{1}{2\hat{\sigma}^4} \end{bmatrix}$$

$$\therefore \quad \nabla \underset{\sim}{D}_n(\hat{\underset{\sim}{\theta}}_n)\, A_n(\hat{\underset{\sim}{\theta}}_n)^{-1} = \left[\begin{array}{c|c} 0 & \frac{2}{n}\underset{\sim}{x}_1'\underset{\sim}{x}_1 \\ & \vdots \\ & \frac{2}{n}\underset{\sim}{x}_K'\underset{\sim}{x}_K \\ \hline 0 & 0 \\ \hline 0 & 0 \end{array}\right]$$

Finally $\nabla \log f_t(u_t, \hat{\theta}_n) = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} \hat{u}_t x_t \\ \\ \frac{-1}{2\hat{\sigma}^2} + \frac{\hat{u}_t^2}{2\hat{\sigma}^4} \end{bmatrix}$

$$\therefore \; a_t = d_t(y_t, \hat{\theta}_n) - \nabla D_n(\hat{\theta}_n) A_n(\hat{\theta}_n)^{-1} \nabla \log f_t(y_t, \hat{\theta}_n) =$$

$$= d_t(y_t, \hat{\theta}_n) - \begin{bmatrix} \frac{1}{\hat{\sigma}^4} [\hat{u}_t^2 - \hat{\sigma}^2] \frac{x_1' x_1}{n} \\ \vdots \\ \frac{1}{\hat{\sigma}^4} [\hat{u}_t^2 - \hat{\sigma}^2] \frac{x_K' x_K}{n} \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{\hat{\sigma}^4} (\hat{u}_t^2 - \hat{\sigma}^2)(x_{1t}^2 - x_1' x_1 /n) \\ \vdots \\ \frac{1}{\hat{\sigma}^4} (\hat{u}_t^2 - \hat{\sigma}^2)(x_{Kt}^2 - x_K' x_K /n) \\ \frac{1}{2\hat{\sigma}^6} (\hat{u}_t^2 - 3\hat{\sigma}^2) \hat{u}_t x_t \\ \frac{3}{4\hat{\sigma}^4} + \frac{\hat{u}_t^4}{4\hat{\sigma}^8} - \frac{3\hat{u}_t^2}{2\hat{\sigma}^6} \end{bmatrix}$$

$V_n(\hat{\theta}_n)$ is then equal to $n^{-1} \sum_{t=1}^{n} a_t a_t'$, which is result stated in the main text.

(ii) We now establish the block diagonality of $\underset{\sim}{V}(\underset{\sim}{\theta}_o)$ under H_o

From the definitions already given, it can be seen that $\underset{\sim}{V}(\underset{\sim}{\theta}_o)$, when partioned in a similar way to $\underset{\sim}{V}_n(\hat{\underset{\sim}{\theta}}_n)$, is given by

$$\underset{\sim}{V}(\underset{\sim}{\theta}_o) = \begin{bmatrix} \underset{\sim}{\Delta}_o^{11} & & \\ \underset{\sim}{\Delta}_o^{21} & \underset{\sim}{\Delta}_o^{22} & \\ \underset{\sim}{\Delta}_o^{31} & \underset{\sim}{\Delta}_o^{32} & \underset{\sim}{\Delta}_o^{33} \end{bmatrix}$$

where $\Delta_o^{ij} = \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E\left[\underset{\sim}{\Delta}_t^{ij}\right]$

where $\underset{\sim}{\Delta}_t^{ij}$ corresponds to the expression for $\hat{\underset{\sim}{\Delta}}_t^{ij}$, given in section 3, except that $\underset{\sim}{\theta}$ is evaluated at $\underset{\sim}{\theta}_o$ instead of $\hat{\underset{\sim}{\theta}}_n$ and so $\hat{u}_t^{\alpha}$ and σ^2 must be replaced by u_t^{α} and σ^2 respectively, and $\xi_{ijt} = x_{it}x_{jt} - m_{ij}$

To evaluate each block we use the following results (i) u_t and $\underset{\sim}{x}_t'$ are independent under H_o and (ii) given normality: $E(u_t^{\alpha}) = 0$ for $\alpha = 1,3,5,7$ and $E(u_t^2) = \sigma^2$, $E(u_t^4) = 3\sigma^4$, $E(u_t^6) = 15\sigma^6$, $E(u_t^8) = 105\sigma^8$.

Block $\underset{\sim}{\Delta}_o^{11}$: $\underset{\sim}{\Delta}_o^{11} = 2\sigma^{-4} \lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} \underset{\sim}{\xi}_t \underset{\sim}{\xi}_t'$.

This matrix consists of elements of the type.

$$2\sigma^{-4} \lim_{n\to\infty} \sum_{t=1}^{n} x_{it}x_{jt}x_{pt}x_{mt} - M_{ij}M_{pm}$$

and so it is necessary to make the additional assumption that $\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} x_{it}x_{jt}x_{pt}x_{m}t$ is finite $\forall, i, j, p, m$.

Block $\underset{\sim}{\Delta}_o^{21}$ under H_o, u_t and $\underline{x}_t'$ are independent, and $\underset{\sim}{\Delta}_o^{21}$ only contains terms in u_t^{α} where α is an odd positive number. In addition we assume $n^{-1}\,\Sigma x_{it}x_{jt}x_{kt}$ is bounded.

$$\therefore \quad \underset{\sim}{\Delta}_o^{21} = \underset{\sim}{0} \; .$$

Block $\underset{\sim}{\Delta}_o^{32}$: Examining the terms in u_t gives .

$$E\left[(u_t^2 - \sigma^2)\left(\frac{3}{4\sigma^4} - \frac{3u_t^2}{2\sigma^6} + \frac{u_t^4}{4\sigma^8}\right)\right] = 0$$

Note also that the terms in $\underline{x}_t'$ are $\lim_{n\to\infty} n^{-1}\,\Sigma\, x_{it}x_{jt} - m_{ij} = 0$,

$$\therefore \quad \underset{\sim}{\Delta}_o^{31} = \underset{\sim}{0}$$

Block $\underset{\sim}{\Delta}_o^{22}$: The terms in u_t are :

$$w = u_t^6 - 6u_t^4\sigma^2 + 9\sigma^4 u_t^2 \; .$$

Under normality $E(w) = 6\sigma^6$ and so

$$\underset{\sim}{\Delta}_o^{22} = \frac{3}{2}\,\sigma^{-6}\,\underset{\sim}{M} \; .$$

Block $\underset{\sim}{\Delta}_o^{32}$: By similar arguments to block $\underset{\sim}{\Delta}_o^{21}$, $\underset{\sim}{\Delta}_o^{32} = \underset{\sim}{0}$

Block $\underset{\sim}{\Delta}_o^{33}$: $E(z^2) = \frac{3}{2}\,\sigma^{-8}$

Therefore the matrix $\underset{\sim}{V}(\underset{\sim}{\theta}_o)$ is block diagonal under H_o as stated in the text.

REFERENCES:

Amemiya, T. (1977), 'A note on the heteroscedastic model', Journal of Econometrics, 6, 365-370.

Bera, A.K. and Jarque, C.M. , (1980) 'An efficient large-sample test for normality of observations and regression residuals', Mimeo, Australian National University.

Bera, A.K. and Jarque, C.M. (1982), 'Model specification tests: A simultaneous approach', Paper presented at Third Latin American Meeting of the Econometric Society, Mexico City.

Bowman, K.O. and Shenton, L.R., (1975) 'Omnibus contours for departures from normality based on b_1 and b_2', Biometrika 62 (1975) 243-250.

Breusch, T.S. and Pagan, A.R. (1979), 'A simple test for heteroscedasticity and random coefficient variation', Econometrica, 47, 1287-1294.

Godfrey, L.G. and Wickens, M.R. (1982), 'Tests of Misspecification using locally equivalent alternative -odels', pp.71-104 in Evaluating the reliability of macroeconomic models, Chow, G.C. and Corsi, P. (eds.), London, Wiley.

Hall, A.R. (1982), 'The Information Matrix test for the General Linear Model', unpublished MSc dissertation, University of Southampton.

Hausman, J.A. (1978), 'Specification tests in econometrics'. Econometrica, 46, 1251-1271.

Keifer, N.M. and Salmon, M., (1983) 'Testing normality in econometric models', Economics Letters, 11, pp.123-127.

Koenker, R., (1981), 'A note on studentising a test of heteroscedasticity', Journal of Econometrics, 17, 107-112.

White, H. (1980), 'A heteroscedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity', Econometrica, 48, 817-838.

White, H. (1982), 'Maximum Likelihood estimation of misspecified models', Econometrica, 50, 1-26.