Alternative Linear Disturbance Estimators in Linear Regression Analysis

Given the usual linear model, $Y=X\beta+\varepsilon$, Y being an $n\times1$ vector of values of a resultant variable, X a fixed $n\times p$ matrix $[x_{ij}]$ such that x_{ij} is the ith value of the jth explanatory variable, β is a $p\times1$ vector of parameters ε an $n\times1$ vector of errors $\varepsilon\sim N(0,\sigma^2 I)$ and X being of rank p, $p<n$. Then b, the OLS estimator of β is $b = (X'X)^{-1}X'Y$

Define M to be the matrix $I-X(X'X)^{-1}X'$, which is trivially idempotent

$$\text{tr}(M) = \text{tr}(I) - \text{tr}\,(X(X'X)^{-1}X')$$
$$\text{tr}(M) = \text{tr}(I_n) - \text{tr}(I_p) = n-p$$

So the trace of M is $n-p$, so its rank is $n-p$ and it has $n-p$ eigenvalues of unity, the rest being zero. The vector of residuals is

$$e = Y - Xb = Y - X(X'X)^{-1}X'Y = MY$$

$e = MY$ .So $e = MX\beta + M\varepsilon$

But $MX = 0$

$e = MY = M\varepsilon$

Regression analysis in common with much other statistical analysis (such as ANOVA) uses the residuals as estimators[1] of the true disturbances in a battery of testing procedures. Hence it is important to examine the properties of those residuals and, where they are in any sense unsatisfactory, to consider alternatives. This study is motivated by the fact that even given homoscedastic serially uncorrelated errors with a fixed dispersion matrix as in this model, the residuals follow a heteroskedastic, autocorrelated data dependent distribution as is seen from the result below

$e = M\varepsilon$

So e is normal and unbiased

$$E(ee') = E(M\varepsilon\varepsilon'M')$$
$$= \sigma^2 MM'$$

$$V(e) = \sigma^2 M$$

$e \sim N(0,\sigma^2 M)$ where M is data dependent and not generally scalar. In this paper lengthy proofs are not given but marked by an asterisk. The proofs have been submitted to the review and are available from the author or the editor.

However the least squares residual vector is the best linear (in Y) unbiased residual vector. Let j be any other residual vector AY

$$E(AY) = 0 \Rightarrow AX\beta + E(\varepsilon) = 0 \Rightarrow AX = 0 \text{ and } AY = A\varepsilon$$
Let $A = A^* + M$
$$\text{o}(\text{o}(\Sigma, I), I)(j) = \text{o}(\Sigma, I)[(j-\varepsilon)(j-\varepsilon)']\ [2]$$
$$= \text{o}(\Sigma, I)[(A-I)\varepsilon\varepsilon'(A-I)]$$
$$= \sigma^2(A-I)(A-I)'$$
$$= \sigma^2(A^*+M-I)(A^*+M-I)'$$
$$= \sigma^2(A^*+X(X'X)^{-1}X')(A^*+X(X'X)^{-1}X')'$$
$$= \sigma^2 A^{*\prime}A^* + \sigma^2 X(X'X)^{-1}X', \text{ because } A^*X = AX - MX = 0$$
$$\text{o}(\Sigma, I)(j) = \sigma^2 A^{*\prime}A^* + \sigma^2$$

For e, $A = M$ so $A^* = 0$ so $\text{o}(\Sigma, I)(e) = \sigma^2(I-M)$

So as $A^{*\prime}A^*$ is symmetric and hence positive semidefinite. $\text{o}(\Sigma, I)(e)$ exceeds

\o(Σ, l)(j) by a positive semidefinite matrix for any j and the result is hence proven.

The class of alternative disturbance estimators is the set of residuals which are linear in Y. unbiased and have scalar fixed dispersion matrices. Such vectors are written as CY where C is of order qXn. A necessary and sufficient condition for such a vector to be unbiased is that CX=O because

$$E[CY] = 0 \Rightarrow E[CX\beta + C\varepsilon] = 0 \Rightarrow CX = 0$$

Let D be the fixed scalar dispersion matrix. Without loss of generality let D = $\sigma^2 I$ (because this can be achieved by scalar multiplication of C). It is a necessary and sufficient condition for \o(Σ, l)[CY] to be $\sigma^2 I$, that CC' = I, because

$$E[CYY'C] = \sigma^2 I \Rightarrow C\sigma^2 C' = \sigma^2 I \Rightarrow CC' = I$$

An important result for LUS residuals is that CY = Cε - Ce as may thus be shown:

$$CY = C[X\beta+\varepsilon]$$
$$\Rightarrow CY = C\varepsilon$$
$$\Rightarrow MC' = [I - X(X'X)^{-1}X']C'$$
But X'C' = 0
$$\Rightarrow MC' = C'$$
$$\Rightarrow CM = CM' = C$$
$$\Rightarrow CY = CMY$$
$$\Rightarrow CY = C\varepsilon = Ce$$

The fact that CX = 0 implies that the n columns of C are subject to p linear dependencies. So

rank[C] = n-p
rank[C] = rank[CC']
rank[CC'] $\leq$ n-p
But CC' = I, of order qXq of full rank
rank[I] = rank[CC'] = q
q $\leq$ n-p

It is clearly desirable for the number of rows of C to be as close as possible to n-p. This result tells us that q is at most n-p, not that this maximum can be obtained. M has n-p eigenvalues of unity, the rest being zero. Let $q_1..q_{n-p}$ be the eigenvectors corresponding to the unit eigenvalues. If C is an n-pxp matrix, the rows of which are $q_1...q_{n-p}$, then CY is a LUS residual vector. This result hence shows that the maximum can be obtained. It is proven thus

$$(M-I)C' = 0$$
$$\Rightarrow X(X'X)^{-1}X'C' = 0$$
$$\Rightarrow CX = 0$$
CY is linear in Y
CY = Ce
V(CY) = V(Ce)
       =E[Cee'C']
$$V[CY] = \sigma^2 CMC'$$

But C orthogonally diagonalises M and the eigenvalues of M corresponding to the columns of C' are all unity.

$$V[CY] = \sigma^2 I$$

So CY is a vector of LUS residuals. The rows of C are clearly pairwise orthogonal and of length σ^2 in this case.

Since the maximum number of rows of C, n-p, can always be attained we always choose C to be of order n-pxn. For p observations there are no LUS residuals, this operation is defined as the base.

It is logical then to proceed to attempt to derive a LUS residual vector satisfying certain optimality conditions. One of the widely used classes of LUS residuals, the BLUS residuals of Theil (1965, 1968) satisfy such conditions. Consider the partitioning of X as $[X_0 | X_1]$ where X_0 represents those rows corresponding to observations in the base, e being correspondingly partitioned. Let the eigenvalues of $X_0(X'X)^{-1}X_0'$ be written in order of increasing magnitude as

$$d_1^2 \le d_2^2 \ldots \ldots \le d_p^2$$

and let q_i be the ith eigenvector of $(X_1 X_0^{-1})(X_1 X_0^{-1})'$. Then the BLUS residual vector is

$$u_1 = e_1 - X_1 X_0^{-1} \left(\sum_{i=1}^{h} (d_i, 1-d_i) \quad q_i q_i' \right) e_0$$

which, it may be proved[*]
(i) is a LUS residual vector.
(ii) is unique
(iii) minimises the expected sum of square errors, i.e. if v_1 is any estimator of ε_1, $E[(v_1 - \varepsilon_1)'(v_1 - \varepsilon_1)] \ge E[(u_1 - \varepsilon_1)'(u_1 - \varepsilon_1)]$, this property being known as weak optimality.

An attempt was made in Theil(1968) to show strong optimality, i.e. that $o(\Sigma, I)(v_1) \ge o(\Sigma, I)(u_1)$ for all v_1, an estimator of e_1. The proof is not generally correct; it entails an added constraint on v_1. Grossman and Styan (1970) disproved strong optimality for BLUS residuals. While weak optimality implies that

$$tr[o(\Sigma, I)(v_1)] \ge tr[o(\Sigma, I)(u_1)],$$

because strong optimality is not a property of BLUS residuals, we cannot even be sure that the diagonal elements of $o(\Sigma, I)(u_1)$ do not exceed the corresponding elements of $o(\Sigma, I)(v_1)$, for some v_1. If $\lambda_j[A]$ is the jth largest eigenvalue of A, Grossman and Styan (1970) did show that

$$\lambda_j[o(\Sigma, I)(v_1)] \ge \lambda_j[o(\Sigma, I)(u_1)] \quad *$$

but this is little more than a mathematical curiosum.
The weak optimality condition applies only, however, to LUS residuals. By the result given earlier, if a_1 is any vector of estimators of ε_1, $o(\Sigma, I)(a_1) \ge o(\Sigma, I)(e_1)$. So $ES(a_1) \ge ES(e_1)$. A convenient scalar measure of the difference between these two quantities is the ratio of $ES(e_1)$ to $ES(a_1)$ which Abrahamse and Koerts (1970)define as the efficiency of a_1. The efficiency of u_1 is

$$\frac{\sum_{i=1}^{h}(1-d_i^2)}{2\sum_{i=1}^{h}(1-d_i)} \quad , \text{ as may be shown thus}$$

$$o(\Sigma, I)(u_1) = 2\sigma^2 \sum_{i=1}^{h}(1-di) \; p_i p_i', \text{ writing in scalar format, since } d_i = 1 \text{ for all } i > h$$

$$ES(u_1) = tr(u_1) = 2\sigma^2 \sum_{i=1}^{h}(1-d_i)$$

$$o(\Sigma, I)(e_1) = \sigma^2 X_1(X'X)^{-1}X_1'$$

$$ES(e_1) = \sigma^2 tr[o(\Sigma, I)(e_1)]$$

$$= \sigma^2 tr[X_1(X'X)^{-1}X_1'] = \sigma^2 tr[(X'X)^{-1}X_1'X_1]$$

$$= \sigma^2 tr[(X'X)^{-1}(X'X - X'_0 X_0)]$$

$$= \sigma^2 p - \sigma^2 tr(N^{-1})$$

$$\Rightarrow ES(e_1) = \sigma^2 \sum_{i=1}^{h} 1-d_i^2$$

$\Rightarrow \backslash f(ES(e_1), ES(u_1)) = \backslash f(\backslash i \backslash su(i=1, h, 1-d_i^2), 2 \backslash i \backslash su(i=1, h, (1-d_i)))$

Implicit in any vector of disturbance estimators is a vector of coefficients estimators. For u_1, this is β^*_1 such that $u_1 = y_1 - X_1 \beta^*_1$

$$u_1 = e_1 - X_1 X_0^{-1}[\backslash i \backslash su \quad (i=1, h, \backslash f(d_i, 1-d_i)) \quad q_i q_i'] e_o$$

$$u_1 = Y_1 - X_1 b + X_1 X_0^{-1}[\backslash i \backslash su \quad (i=1, h, \backslash f(d_i, 1-d_i)) \quad q_i q_i'] e_o$$

$$X_1 \beta^*_1 = X_1 b + X_1 X_0^{-1}[\backslash i \backslash su \quad (i=1, h, \backslash f(d_i, 1-d_i)) \quad q_i q_i'] e_o$$

$$\Rightarrow \beta^*_1 = b + X_1 X_0^{-1}[\backslash i \backslash su \quad (i=1, h, \backslash f(d_i, 1-d_i)) \quad q_i q_i'] e_o$$

Trivially, $E(\beta^*_1) = \beta$ but $\backslash o(\Sigma, 1)(\beta^*_1) \geq \backslash o(\Sigma, 1)(b_1)$ by the Gauss-Markov Theorem.

All this analysis of BLUS residuals is however dependent on the choice of base. Ignoring the nonsingularity requirement for X_0, there are ${}^nC_k$ possible choices of base. The approach to the use of residuals suggested by Theil (1968) is to first of all identify the set of all possible bases and then to select that one which minimises the expected error sum of squares. If the set of all possible bases is the set of all bases for which X is nonsingular, then the computational cost of this will be very high, but for many applications, other constraints will exist (e.g. that the base must be in the centre of the ordered predictors). The large number of possible bases is one general objection to the BLUS procedure. Another fundamental one is the fact that the dispersion matrix of the BLUS residual vector exceeds that of the least squares one by a positive semidefinite matrix. Finally, p observations have no BLUS residuals. The second of these is probably the most serious because computational advances lessen the first problem and because n is typically much greater than p. A measure of the gravity of this problem is the efficiency of the BLUS residual vector.

The recursive residual vector is another member of the class of LUS residual vectors. Let $b_{(r)}, X_{(r)}, y_{(r)}$ be b, X and y with only the the first r observations, $r > p$.

The rth recursive residual is defined as

$$Pr = \backslash f(y_r - X'_{(r)} b_{(r-1)}, \backslash r(1 + x'_r (X'_{(r-1)} X_{(r-1)})^{-1} x_r)) \quad , r \varepsilon \{p+1 \ldots \ldots n\}$$

The computation of recursive residuals is greatly simplified by the result [proved in Harvey(1981)]

$$(X'_{(r)} X_{(r)})^{-1} = (X'_{(r-1)} X_{(r-1)})^{-1} - \backslash f((X'_{(r-1)} X_{(r-1)})^{-1} x_r x'_r (X'_{(r-1)} X_{(r-1)})^{-1}, 1 + x'_r (X'_{(r-1)} X_{(r-1)})^{-1} x_r)$$

which means that only one matrix inversion is needed to compute a whole set of recursive residuals. Hence, as is shown in Brown, Durbin and Evans(1975)

$$b_r = b_{r-1} + (X'_{(r)} X_{(r)})^{-1} x_r (y_r - x'_r b_{r-1}),$$ this being the explicit recursion formula used to compute $\{w_{p+1} \ldots \ldots w_n\}$

Parallel to the problems of BLUS residuals, recursive residual vectors have variance-covariance matrices exceeding those of the corresponding least squares ones by a positive semidefinite matrix and that the first p observations have no recursive residuals.

A generalisation of the LUS class of residual vectors is the LUF class. These are disturbance estimators which are linear in Y, unbiased and which have dispersion matrices $\sigma^2 \Omega$ which are not data-dependent and which are hence fixed. But it is clearly desirable that certain properties of $\sigma^2 \Omega$ mirror those of $\sigma^2 M$. Following Abrahamse and Koerts (1971), Ω is required to be idempotent and of rank n-p.

Let M(X) be the space spanned by the columns of X and let v be a LUF

residual vector, $\Psi'Y$, Ψ being of order n-pxn-p. So

$E(\Psi'Y) = 0 \Rightarrow \Psi'X = 0$

So the columns of Ψ' are elements of M(X)', the orthogonal complement of M(X). By the dimension theorem, if T is a linear transformation from the vector space V to the vector space W then the sum of the rank and nullity of T is equal to the dimension of V. Hence the sum of the rank of Ψ and the dimension of the solution space of $\Psi'X=0$ is n, so the dimension of M(X)' is n-p. Besides, clearly,

$E(vv') = E(\Psi'yy'\Psi) = \sigma^2\Psi'\Psi = \sigma^2\Omega \Rightarrow \Psi\Psi'=\Omega$

Let K be that nxn-p matrix the columns of which are the eigenvectors corresponding to the unit eigenvalues of Ω and let P be that nxn-p matrix the columns of which are the eigenvectors corresponding to the unit eigenvalues of M. Then Abrahamse and Koerts show that the residual vector

$K[K'MK]^{-\backslash f(1,2)} K'MY$

is a LUF vector satisfying weak optimality, and that it is unique.

Consider a reformulation of the OLS model in which A is an nxn-p matrix of rank n-p the columns of which span M(X)'. Xb must be an element of M(X) in OLS. But Xb is an element of M(X) iff A'Xb=0. Defining λ as a column vector of n-p Lagrange multipliers the estimation of b in OLS can be reformulated as the constrained optimisation of the minimisation with respect to β of $(y-X\beta)'(y-X\beta)$ subject to $A'X\beta=0$, equivalent to the Lagrangian minimisation of $(y-X\beta)'(y-X\beta)-2\lambda'A'X\beta$, the solution being b, $\backslash o(\lambda,^{\wedge})$. By Harrison and Keogh(1984)

$Xb= [I - A(A'A)^{-1}A']X\beta$, $\backslash o(\lambda,^{\wedge})= (A'A)^{-1}A'Y$, $\backslash o(\lambda,^{\wedge})\sim N(0,\sigma^2(A'A)^{-1})$

As $\backslash o(\lambda,^{\wedge})$ is linear in y, unbiased and has a fixed dispersion matrix, it has an interpretation as a set of LUF residuals. Letting $A=D(D'D)^{-1}$ $\backslash o(\backslash \lambda,^{\wedge})= ((D'D)^{-1}$ $D'D(D'D)^{-1})^{-1}$ $(D'D)^{-1}D'Y = D'D(D'D)^{-1} D'Y =D'$. So, any LUF residual vector may be generated in this way and so any LUS residual vector may too provided that A'A=I.

The process of finding LUF residual vectors differs very fundamentally from that for LUS residuals. The matrix P is data dependent. For generating LUF residuals in a particular set of analyses we must specify typical matrices such that for any one application this will be a good approximation. Since P is a basis for M(X), finding a typical P matrix is equivalent to finding a typical X matrix. Once the typical X matrix has been specified for any field, then given the nonsingularity of K'P, v is uniquely determined. The issue of choice of base does not even arise.

A typical X matrix is not necessarily given however. Intuitively, indeed this is likely to be quite unusual. But one case where it does exist is for slow trending time series which Theil and Nagar (1961) argue is quite common. X, in this case, can be approximated by the eigenvectors of the matrix A ,$[q_{ij}]$ which is of order nXn, where

q_{ii} = 2 for all i $\varepsilon\{2,.....n-1\}$, q_{ii} =1 for i $\varepsilon\{1,n\}$ and q_{ij}=-1 for all i,j such that $|i-j|=1$; all other elements of the matrix being zero.

This is because the eigenvectors of A have a "slowly changing" character, so X may behave like the eigenvectors of A. P is a matrix of eigenvectors of M corresponding to the unit eigenvalues of M, so it spans the orthogonal complement of the space spanned by X, because as has been shown KP'X=0. So P behaves like the eigenvectors corresponding to the n-p largest eigenvalues of A. Let the matrix, the columns of which are these eigenvectors be L, of order nxn-p. It follows that it is reasonable to choose L for the matrix K, and Ω is thus LL'.

Von Neumann (1941) showed that the eigenvectors of A are given by h_i =

$\f{1,c_i}$ [Cos\f{π(i-1),2n)}, Cos\f{3π(i-1),2n)}, ...,Cos\f{(2n-1)π(i-1),2n)}]

$c_1 = \r(n)$, $c_i = \r(\f{n,2})$ for all $i \, \varepsilon\{2....n\}$

corresponding to the eigen values $2[1 - Cos \, \f{\pi(i-1),n)}]$, $i \, \varepsilon\{1.....n\}$. L is hence $[h_{p+1}........h_n]$.*

The theory of LUF residuals however relies on Ω being determined *a priori* and so the typical X matrix must be too. This is the major limitation of LUF residuals. A typical X matrix may not exist in a given field of learning. But, more critically, the fact of the typical X matrix being determined *a priori* means that the analyst may not be influenced by the extent to which his X matrix satisfies the requirements to be "typical" in the given field of learning without undermining the statistical validity of the procedures. If he finds that the X matrix is completely different from that which he expected, *ex. ante*, if he modifies his beliefs about what X matrix is "typical" in that field of learning, then it can no longer be assumed that Ω is fixed and the theory of LUF residuals breaks down.

Given the central role of disturbance estimators in econometric specification and misspecification testing, alternative disturbance estimators can clearly be applied to a battery of testing procedures including tests for heteroskedasticity, serial correlation, concavity, convexity, general non-linearity or structural change over time, as well as to misspecification tests. Length being constrained, not all these applications can be examined.

Exact parametric tests of heteroskedasticity, using LUS residuals, were proposed by Theil (1971) and Phillips and Harvey (1973), the former using BLUS residuals, the latter employing recursive ones. If $\sigma^2 = Var(\varepsilon_j) = \sigma^2 f(j)$ then the null is that $f(j)$ is unity for all j and H_A is that there exists some j such that $f(j) \neq 1$. For power studies, writers have confined themselves to more restricted forms of heteroskedasticity.

Let the observations be ordered in non-decreasing values of σ^2_j, according to an alternative hypothesis. Let a regression be run on the first and last m observations, $2m \leq n$ [2m<n, for n odd], yielding residual vectors e_1 and e_2 respectively. The Goldfeld-Quandt test statistic is

$R = \f{e'_2 e_2, e'_1 e_1} = \f{e'E_2 e, e'E_1 e)}$, where $E_2 = \b(\a\co2(0,0,0,I_m))$, where $E_1 = \b(\a\co2(I_m,0,0,0))$, where I_m is the identity matrix of order mXm.

$\Rightarrow R = \f{\varepsilon M'E_2 M\varepsilon, \varepsilon'M'E_1 M\varepsilon)}$

But $M'E_2 M$ and $M'E_1 M$ are trivially idempotent

$\Rightarrow e'_2 e_2 \sim \sigma^2 \chi^2_m$, $e'_1 e_1 \sim \sigma^2 \chi^2_m$

$e'_1 e_1$ and $e'_2 e_2$ are independent

$R \sim F_{m,m}$

Testing may hence be conducted. Discarding central observations has been found to increase power, for example by Phillips and Harvey.

The procedures of Theil (1971) and Phillips and Harvey (1973) adapt this test for LUS residuals, represented by a vector w. As the discarding of the central observations increases power, it follows that, for n-p even the central p observations be used as the base and for n-p odd, one of the two possible sets of central p observations be used. In this way the disadvantage of LUS residuals (that observations in the base do not have corresponding residuals) can be mitigated. Phillips and Harvey (1973, 1974) allow for the possibility of discarding more observations than are in the base; thereby increasing the number of possible bases. In the context of computationally expensive BLUS residuals this seems unlikely to be desirable, so Theil (1968, 1971) does not allow for this: the more possible bases we consider the greater the computational cost. Theil then

suggests that a BLUS residual vector be computed for each [one or two] admissible base, the one with the lower expected error sum of squares being selected. In this paper, the approach of Theil is initially followed, so two regressions are run; one on the observatons before the base [p elements] and one on the observations after it, yielding LUS residual vectors w_1 and w_2 respectively. The ratio for either BLUS or recursive tests is then

$S = \f(w_2\text{'}w_2, w_1\text{'}w_1)$

depending on whether $\f(w_1, w_2)$ are BLUS or recursive residual vectors.

Clearly $w_2\text{'}w_2 \sim \sigma^2\chi^2_n, w_1\text{'}w_1 \sim \sigma^2\chi^2_m$, where w_1 and w_2 are of length m and n respectively.

$E[w_1\text{'}w_2]=0$

$w_2\text{'}w_2$ and $w_1\text{'}w_1$ are independent, because

$S \sim F_{n,m}$

and so testing may be conducted by comparison with the critical values of an $F_{n,m}$ variate.

Harvey and Phillips (1974) is an important discussion of the relative merits of these three tests. The BLUS residuals have the least expected sum of squared estimation errors, but there is no theoretical reason why this should imply that a test based on them will be most powerful. The recursive residuals, on the other hand, are easier to obtain. Harvey and Phillips claim that tests based on recursive or BLUS residuals are more flexible than the Goldfeld-Quandt test as recursive or BLUS residuals may be used with different orderings of the observations, while least squares residuals may not. They however provide no theoretical or empirical support for this claim. Changing the order of observations changes the set of admissable bases; but if one set of LUS residuals is used with different orderings of the bases, then only the base in the first test will necessarily be admissable. Even if other bases are, by chance, admissable, the BLUS residual vectors they yield will not necessarily satisfy weak optimality. It seems plausible, hence, that using different orderings without fresh bases may cause a loss of power, and the phenomenon may hence merit further study. Harvey and Phillips use the procedure of Imhof (1961) to compute the power of S under various alternative hypotheses. The procedure enables this to be done exactly, since S is a quadratic form in random normal variables. They drew the following key conclusions.

(i) For designing the most powerful test it is optimal to discard around max $(p, \f(n,3))$ central observations, this optimum being relatively flat.

(ii) The tests based on BLUS residuals dominate those based on recursive residuals in power, but the difference is very small and probably does not justify the extra computatonal cost of the former.

(iii) All conclusions are quite robust to the specification of the heteroskedasticity.

(iv) No overall conclusion may be drawn as to the relative power of those tests and the Goldfeld-Quandt test.

Ramsey (1969) discusses a number of specification tests involving LUS residuals. Only one is claimed to be sensitive to simple heteroskedasticity. This is essentially a special case of Bartlett's M Specification Error Test [BAMSET]. Defining $z_1...z_{n-p}$ to be the coordinates of a vector of LUS residuals, z, the set of n-p LUS residuals is divided into t non-intersecting subsets, such that the residuals in the jth set have a sum of squares L_j and N_j elements are in this set. Define

$$s^2_j = \f(L_j, N_j), \quad s^2 = \f(\i\su(j=1, t, N_j s_j^2), n-p), \quad j \in \{1.....t\}$$

The null specified by Ramsey is that $z^2_j \sim \sigma^2\chi^2_1, j\varepsilon\{1,......,n-p\}$ while H_A is

$z_j^2 \sim \sigma_j^2 \chi^2_1$, with σ_j^2 not equal for all j, He shows that the test statistic

$$n\text{-}p \log s^2 - \sum_{j=1,t} N_j \log s^2_j \sim \chi^2_{t\text{-}1},$$

and testing may hence be conducted. Unfortunately the power of this test may not be computed using the method of Imhof (1961). The specification of the null and alternative hypotheses is significantly different from that used by other writers and has the disadvantage that, as many misspecifications imply z^2_i to be distributed as non-central χ^2, the test is not defined for these cases. Ramsey (1969) notes that this is the case if there is functional misspecification or underspecification.

Brown, Durbin and Evans (1975) propose a test for structural change over time based on recursive residuals. Formally, the null is that β and σ^2 are constant over time, and H_A is that they are not (although a variety of one-sided procedures could be devised if desired). Let the set of recursive residuals be $\{v_t\}, t\varepsilon\{p_1....n\}$. Two alternative statistics may then be computed, known as the cusum and cusum squared test statistics,

$$w_j = \frac{\sum_{t=p+1,j} v_t}{\sum_{t=p+1,n} v_t^2} \quad \text{and} \quad s_j = \frac{\sum_{t=p+1,j} v_t^2}{\sum_{t=p+1,n} v_t^2}$$

respectively, for the jth observation. Consider a plot of w_j against j. Under the null, over time, it should be close to zero. Applying the theory of Brownian motion in statistical mechanics they showed that there is an $\alpha\%$ probability that some point will be above line A or below line B, under the null where line A goes through the points $(p,+f(\alpha)\sqrt{n\text{-}p})$ and $(n,3f(\alpha)\sqrt{n\text{-}p})$ and line B goes through the points $(p,\text{-}f(\alpha)\sqrt{n\text{-}p})$ and $(n,\text{-}3f(\alpha)\sqrt{n\text{-}p})$. Values for $f(\alpha)$ were computed; notably $f(0.05) = 0.948$. A diagnostic test may thus be devised which rejects the null if some portion of the plot is above line A or below line B. By simple geometry the equations of lines A and B may be deduced as $w_t = f(\alpha)\sqrt{n\text{-}p} + \frac{2f(\alpha)(t\text{-}p)}{\sqrt{n\text{-}p}}$ and $w_t = \text{-}f(\alpha)\sqrt{n\text{-}p} - \frac{2f(\alpha)(t\text{-}p)}{\sqrt{n\text{-}p}}$ respectively. The approach for the cusum of squares test is similar, but based on a slightly different rationale

$$\sum_{t=p+1,j} v_t^2 \sim \chi^2_{j\text{-}p} \Rightarrow \sum_{t=p+1,n} v_t^2 \sim \chi^2_{n\text{-}p}$$

$$\Rightarrow s_j = \frac{x_1}{x_1 x_2} \text{ such that } x_1 \sim \chi^2_{j\text{-}p}, \; x_2 \sim \chi^2_{n\text{-}j}$$

$$\Rightarrow s_j \sim \beta(j\text{-}p, n\text{-}j), \text{ following Rao (1973)}.$$

Rao observes that for a $\beta(p,q)$ variate, it has an expectation of $\frac{p}{p+q}$. It follows that $E(s_j) = \frac{j\text{-}p}{n\text{-}p}$. So, if we consider a plot of s_j against j it should be close to a line going through $\frac{j\text{-}p}{n\text{-}p}$ for all j, known as the mean value line. By simple geometry, the equation of the mean value line is, hence

$$s_t = \frac{t\text{-}p}{n\text{-}p}$$

Brown, Durbin and Evans (1975) show that for any required significance level, α, the probability that the plot crosses either or both lines

$$s_t = \frac{t\text{-}p}{n\text{-}p} \pm C_o(d)$$

is equal to α. They tabulate values of C_o and so a diagnostic test may hence be deduced. It could in principle be extended to other LUS residual vectors, the problem becoming statistically more difficult, however. Garbade (1977) in statistical simulations concluded that the cusum of squares test was more powerful than the cusum test and that not even the former is very powerful

against changes in β.

A final application of alternative disturbance estimators, which uses LUF residual vectors, is to computing the distribution of Von Neumann ratios. As the Von Neumann ratio is a linear transformation of the Durbin-Watson test statistic, this could either be regarded as an appplication to either misspecification or serial correlation testing. The exact distribution of the Von Neumann ratio for least squares residuals is data dependent although, given a particular X matrix, it may be computed exactly.

Given a vector of LUS residuals, the distribution of the corresponding Von Neumann ratio may be computed exactly. For n-p BLUS residuals the Von Neumann ratio is $Q = \f(\i\su(d=2,n-p,(u_{1\alpha}-u_{1\alpha-1})^2),(n-p-1)e_1'e_1) = \f(u'_1 Au_1,(n-p-1)e_1'e_1)$ where A is the appropriate elementary matrix. Press and

Brooks (1969) derived significance limits of Q for n-p$\varepsilon\{2...60\}$, and showed that asymptotically, $Q \sim N(2,\f(4,n-p))$, an asymptotic approximation valid for n-p at least around 60. Hence one may test for serial correlation as proposed by Theil (1971).

It is in this application that writers such as Abrahamse and Louter (1971) would tend to argue that LUF residual vectors are especially appropriate. The difficulty in computing Von Neumann ratios lies in data-dependency not in whether the dispersion matrix of disturbance estimators is scalar or not. Since the BLUS residual vectors satisfy weak optimality in a more restricted class than the LUF residual vectors of Abrahamse and Koerts (1971) it seems plausible that using LUF residual vectors will entail a gain in power, although the formal justification for this claim relies on Monte Carlo simulations. Besides, given a typical X matrix for a certain field, LUF residuals are uniquely determined, unlike BLUS residuals. Given a time series where differences are small by comparison to the level predictor variables, LUF residuals may be derived by the procedure referred to. In this case, where A is as defined in the section on LUF residuals the Von Neumann ratio is $Q' = \f(v'Av,v'v)$ Let $l_1.....l_{n-p}$ be the eigenvalues of K'AK. Then Koerts and Abrahamse (1969) show that $Q'=\f(\i\su(i=1,n-p,l_i\varsigma_i^2),\i\su(i=1,n-p,\varsigma_i^2))$ where $\varsigma = (\varsigma_1.....\varsigma_{n-p})$, $\varsigma \sim N(0,\sigma^2 I)$. Let Q be the orthogonal matrix the columns of which are the eigenvectors of A. By orthogonal diagonalisation

$D=diag[\partial_1.....\partial_n]=Q'AQ$, where $\partial_1.....\partial_n$ are the eigenvalues of A.

$\Rightarrow A=QDQ'$

$\Rightarrow K'AK=K'QDQ'K$

But Q is the matrix $[h_1...h_n]$, K is the matrix $[h_{p+1}.......h_n]$ and $h_1...h_n$ are of unit length and pairwise orthogonal. So $K'Q=[O^{(n-p)}_{(p)}I_{(n-p)}]$ where $O^{(n-p)}_{(p)}$ is of order n-pxp and $I_{(n-p)}$ is of order n-pxn-p, being zero and identity matrices respectively.

$\Rightarrow$ K'AK = $[O^{(n-p)}_{(p)}\ I_{(n-p)}]$ D $\B(\A(O^{(n-p)}_{(p)},I_{(n-p)}))$

$\Rightarrow$ K'AK = $\b(\A\CO2(O_{(p)},O_{(n-p)},O_{(n-p)},J))$ where J is diag$[\partial_{p+1}.....\partial_n]$

$\Rightarrow Q' = \f(\i\su(i=p+1,n,d_i\varsigma_i^2),\i\su(i=p+1,n,\varsigma_i^2))$

This is the Durbin-Watson upper bound d_u, apart from the multiplicative constant. Durbin-Watson tests may hence be conducted. This result also provides some explanation for why statistical simulations have tended to suggest that the actual distribution of the Durbin-Watson test statistic is closer to d_u than d_l.

Abrahamse and Koerts (1971) ran some statistical simulations to compare

the power of this procedure with the BLUS one in serial correlation with slow trending time series. The former dominated the latter in power. The strength of the argument for the LUF test that follows from this is clearly however greatly blunted by the need in this procedure to find a typical X matrix and hence a suitable Ω matrix. The procedure breaks down if one cannot be found. As the theory of LUF residuals is based on Ω being fixed a priori it follows that the typical X matrix must be fixed in advance. This may be a quite unreasonable requirement, and so the alternative approach based on BLUS residuals may be preferred.

Jonathan Wright

Notes
(1) The term estimator is used here in a nonstandard way, not to estimate a parameter in a probability distribution but to estimate a value taken on by a random vector, ε.

(2)The variance-covariance matrix of a, a vector of estimators of a random vector ε is defined as $\backslash o(\Sigma, I)(a)$, $E[(a-\varepsilon)(a-\varepsilon)']$. This is again a non-standard practice, following logically however from (i). The matrix $E[(a-E(a))(a-E(a))']$ is termed the dispersion matrix of a, $V(a)$.

Bibliography

Abrahamse A. and Koerts J "A Comparison Between the Power of the Durbin-Watson Test and the Power of the BLUS Test" *Journal of the American Statistical Association*, 64, 1969, pp933-941.

Abrahamse A. and Koerts J. "New Estimators of Disturbances in Regression Analysis", *Journal of the American Statistical Association*, 66, 1971, pp 71-74

Amir-Moez and Fass, "Elements of Linear Spaces", Pergamon Press, Oxford, 1962.

Anton, H., "Elementary Linear Algebra",Wiley,New York,1973.

Brown, Durbin and Evans, "Techniques for Testing for the Constancy of Regression Relationships Over Time", *Journal of the Royal Statistical Society*, Series B, 37, 1975, pp149-192.

Garbade, K., "Two Methods for Examining the Stability of Regression Coefficients", *Journal of the American Statistical Association*, 72, 1977, pp 54-63.

Grossman and Styan, "On Theil's BLUS Residuals", *Journal of the American Statistical Association*, 67, 1972, pp672-673

Harrison M. and Keogh G., "A Lagrange Multiplier Interpretation of Disturbance Estimators with an Application to Testing for Nonlinearity", *Journal of Econometrics*, 27, 1985, pp. 259-269.

Harvey, A, "Time Series Models", Phillip Allan Publishers Ltd., Oxford, 1981.

Harvey, A. and Phillips, G., "A Comparison of the Power of Some Tests for Heteroskedasticity in the General Linear Model", *Journal of Econometrics*, 2, 1974, pp 307-316.

Imhof, JP, "Computing the Distribution of Quadratic Forms in Normal Variables", *Biometrika*, 48, 1961, pp. 419-426

Noble and David "Elementary Linear Algebra and its Applications", Prentice Hall, London, 1981.

Phillips, G. and Harvey, A., "A Simple Test for Serial Correlation in Regression Analysis", *Journal of the American Statistical Association*, 69, 1974, pp. 935-939; first presented 1973

Press, S. and Brooks, R., "Test for Specification Errors in Classical Linear Least Squares Regression Analysis" ,*Journal of the Royal Statistical Society*, Series B ,31, 1967, pp. 360-371.

Rao, C., "Linear Statistical Inference and its Applications", Wiley, New York, 1973.

Theil, H. and Nagar, A, "Testing the Independence of Regression Disturbances", *Journal of the American Statistical Association,* 56, 1961, pp. 793-806.

Theil, H., "The Analysis of Disturbances in Regression Analysis", *Journal of the American Statistical Association,* 60, 1965, pp. 1067-1079.

Theil, H., "A Simplification of the BLUS Procedure for Analysing Regression Disturbances ,*Journal of the American Statistical Association,* 63, 1968, pp. 242-259.

Theil, H., "Principles of Econometrics", Wiley, New York, 1971.