Journal of Econometrics 6 (1977) 103–119. © North-Holland Publishing Company

# TESTING FOR FUNCTIONAL MISSPECIFICATION IN REGRESSION ANALYSIS*

Andrew C. HARVEY

*University of Kent, Canterbury, England*

Patrick COLLIER

*Department of Employment, London, England*

Received March 1975, final version received October 1976

Recursive residuals may be used to detect functional misspecification in a regression equation. A simple $t$-statistic and a related Sign test may be constructed from the residuals. The powers of these tests compare favourably with the Durbin–Watson and other tests commonly used to detect functional misspecification from residuals. In addition the tests are relatively robust to serial correlation in an otherwise correctly specified model, and this is a further point in their favour.

## 1. Introduction

A problem which arises frequently in regression analysis is that of testing whether the functional form employed for the $k$ regressors, $X_1, \ldots, X_k$, is appropriate. Quite often, however, the functional form of only one of the variables, say the $k$th, is in doubt, and it is this case which we are primarily interested in here.

There are two basic approaches to testing for functional misspecification. The first is to run a new regression in which an extra term in $X_k^2$ is included. A test of significance on the coefficient of this variable then provides a test on the specification of $X_k$. Such a test will usually have a relatively high power, even when the form of misspecification is something other than an omitted quadratic term. Despite this, however, the second approach, which employs test statistics based on residuals from the original regression, has considerable appeal and is widely used; see, for example, Theil (1971, pp. 222–225). Part of the appeal of this approach lies in the fact that valuable information on the form of misspecification may be obtained simply by looking at the residuals – what one might refer to as 'data analysis'. The ordinary least squares (OLS) residuals are

*We should like to thank Richard Quandt and the referee for their comments on an earlier draft of this paper, and Garry Phillips for his helpful advice at all stages in the research. However, we are solely responsible for any errors. A preliminary version of this paper was presented at the 3rd World Congress of the Econometric Society, Toronto, August 1975.

not the only residuals which may be used for this purpose, and in this article we show that an alternative set of residuals, the 'recursive residuals', can provide an important complementary tool to the OLS residuals.

A number of procedures based on residuals have been proposed for testing against functional misspecification. Perhaps the most widely used are the tests which have been developed essentially in connection with serial correlation. The procedure proposed by Durbin and Watson uses OLS residuals. Alternatively Theil's BLUS residuals [Theil (1971, ch. 5)] may, because of their independence under the null hypothesis, be used directly in the von Neumann Ratio. The recursive residuals, described in Brown, Durbin and Evans (1975) and Phillips and Harvey (1974) have similar properties to the BLUS residuals and so may also be employed to give an exact test with the von Neumann Ratio.

Apart from having a simpler interpretation than the BLUS residuals, the recursive residuals have the additional advantage that they are much easier to compute; in fact, if the Gentleman routine, which is a more accurate way of computing OLS estimates, is used, the recursive residuals emerge as a by-product; see Farebrother (1976). Ease of computation, however, is not the main attraction of the recursive residuals in this context. Their importance lies in the fact that they may exhibit a very different pattern of behaviour to the OLS residuals under functional misspecification. If the form of the misspecification is such that the 'correct' functional form for the misspecified variable is a concave or a convex function of the variable actually included in the regression, this pattern may be exploited to construct an exact test based on a statistic which has a $t$-distribution under the null hypothesis. The results presented in the later sections of the paper indicate this test is, in such circumstances, more powerful than the Durbin–Watson and von Neumann Ratio tests. In addition it has the important advantage of being relatively robust to serial correlation.

The plan of the paper is as follows. The recursive residuals, and the $t$- statistic based on them, are described in section 2. In section 3 some observations are made on the patterns of residuals likely to emerge under different types of misspecification. Following a discussion of the method of computing the powers of the tests in section 4, some empirical comparisons of relative powers are made for a number of different models. Section 6 introduces the Sign and 'Geary' tests, while the problem of detecting misspecification when more than one variable is misspecified is considered in section 7. In section 8 the robustness of the $t$-statistic with respect to serial correlation is examined.

## 2. The recursive residuals and the $t$-statistic

Consider the general linear (regression) model

$$y_j = x_j'\beta + u_j, \qquad j = 1, \ldots, n, \tag{1}$$

where $x_j$ is $k \times 1$ vector of (fixed) observations on the independent variables, $\beta$ is a $k \times 1$ vector of coefficients, $y_j$ is the $j$th observation on the dependent variable, and $u_j$ is a disturbance term. The disturbances are assumed to be normally and independently distributed with zero expectation and constant variance.

The $n-k$ recursive residuals are defined by

$$\tilde{u}_j = \frac{y_j - x_j' b_{j-1}}{(1 + x_j'(X_{j-1}' X_{j-1})^{-1} x_j)^{\frac{1}{2}}}, \qquad j = k+1, \ldots, n, \tag{2}$$

where $b_j$ is the least-squares estimate of $\beta$ obtained from the first $j$ observations, and $X_j$ is a $j \times k$ matrix of full rank consisting of the first $j$ sets of observations on the independent variables. These residuals are easily computed since simple recursive formulae exist for updating $b_j$ and $(X_j' X_j)^{-1}$. However, as already noted the residuals also emerge as a by-product of the Gentleman routine.

Under the null hypothesis that model (1) holds, the recursive residuals have the same properties as the true disturbances; see, for example, Phillips and Harvey (1974). Hence, if $\bar{u}$ is the arithmetic mean of the recursive residuals, the statistic

$$\psi = \left[ (n-k-1)^{-1} \sum_{j=k+1}^{n} (\tilde{u}_j - \bar{u})^2 \right]^{-\frac{1}{2}} (n-k)^{-\frac{1}{2}} \sum_{j=k+1}^{n} \tilde{u}_j \tag{3}$$

follows a $t$-distribution with $(n-k-1)$ degrees of freedom under the null hypothesis.

The test procedure (the '$\psi$-test') consists simply of arranging the observations in ascending (or descending) order according to the variable which is to be tested for functional misspecification. The least squares coefficients are calculated recursively and the statistic (3) is formed from the resulting recursive residuals. A two-sided $t$-test is then normally carried out although, as will be clear from the discussion below, a one-sided test may sometimes be appropriate if additional information is available.

In order to see why this statistic is likely to provide an effective test against functional misspecification, consider the interpretation of the recursive residuals in a simple bivariate regression situation in which a linear relationship is assumed between the dependent variable and the explanatory variable, $X$, when, in fact, the true relationship is non-linear, being of the form $E[y_j] = \delta + f(X_j)$. The forecasting errors, $w_j = y_j - x_j' b_{j-1}$, for this case are shown in fig. 1. Since the recursive residuals are just the forecasting errors standardised, it is clear that when $f(X)$ is a convex function of $X$, as it is in fig. 1, the recursive residuals will tend to be positive. Conversely when $f(X)$ is concave they will tend to be negative, but in either case $\psi$ will tend to be large in absolute value. The same heuristic argument obviously extends to the general multiple regression case, provided that the variables not being tested are correctly specified.

Besides calculating the $\psi$-statistic it is also useful simply to plot the recursive residuals, as well as the OLS residuals, against the appropriate independent variable. The cumulative sum (CUSUM) of recursive residuals, suggested by Brown, Durbin and Evans (1975), may also be plotted. Such graphs are likely to be particularly valuable when $f(X_k)$ is neither convex nor concave for then the $\psi$-test may be ineffective as positive and negative recursive residuals will tend to cancel each other out. Note that if a formal CUSUM test is to be carried out, the variance of the disturbance term should be estimated by $\sum_{j=k+1}^{n}(\tilde{u}_j-\bar{\tilde{u}})^2/(n-k-1)$, rather than by $\sum_{j=k+1}^{n}\tilde{u}_j^2/(n-k)$, since this is likely to give a more powerful test; see the comments by Harvey in Brown, Durbin and Evans (1975, p. 179).

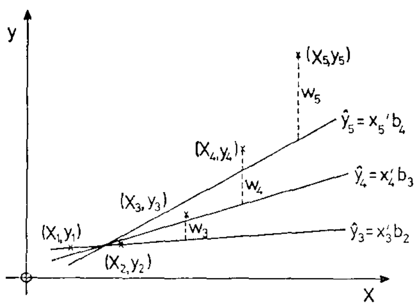


Fig. 1. Derivation of recursive residuals for a misspecified bivariate regression model.

## 3. Properties of residuals in the presence of functional misspecification

The assumed model (1) may be written in matrix terms as

$$y = X\beta+u, \qquad u \sim N(0,\sigma^2 I), \tag{4}$$

where $y$ and $u$ are $n\times 1$ vectors, $X$ is an $n\times k$ matrix, and $\beta$ is a $k\times 1$ vector of parameters. A general form for the correct specification of the model is

$$y = X_r\delta_r+f+v, \qquad v \sim N(0,\sigma^2 I), \tag{5}$$

where $\delta_r$ is an $r\times 1$ vector of parameters, $v$ is an $n\times 1$ vector of disturbances, and $X_r$ is an $n\times r$ matrix whose columns are a subset of the columns of $X$, corresponding to correctly specified variables. Each element in the $n\times 1$ vector $f$ is a function of the corresponding elements in the remaining columns of $X$.

The $n \times 1$ vector of OLS residuals obtained under the assumption that (4) is the appropriate model may be written

$$e = Mf + Mv, \tag{6}$$

where $M = I - X(X'X)^{-1}X'$. It follows therefore that these residuals are normally distributed with

$$E[e] = Mf \quad \text{and} \quad V(e) = \sigma^2 M. \tag{7}$$

For the $(n-k) \times 1$ vector of recursive residuals,

$$\tilde{u} = Cf + Cv, \tag{8}$$

where $C$ is an $(n-k) \times n$ matrix such that $Cy = \tilde{u}$; see Phillips and Harvey (1974). In this case

$$E[\tilde{u}] = Cf \quad \text{and} \quad V(\tilde{u}) = \sigma^2 I_{n-k}. \tag{9}$$

Similar results hold for the BLUS residuals.

It will be observed that the expectations of all three residual vectors are proportional to the corresponding residual vectors obtained from regressing $f$ on $X$. Thus if a quadratic term for the $k$th explanatory variable has been omitted, so that the elements of $f$ are equal to $\delta X_k + \gamma X_k^2$, where $\delta$ and $\gamma$ are parameters, the expectations of the residual vectors will be equal to the corresponding residual vectors obtained from regressing $X_k^2$ on the included explanatory variables,[1] multiplied by $\gamma$.

The patterns of the residuals, and hence the relationship between $f$ and $X$, are crucial in determining the powers of the various tests. This is brought out formally in the next section. When more than one explanatory variable is misspecified, however, all three sets of residuals, BLUS, recursive and OLS, may emerge in such a way that the misspecification is difficult to detect. Suppose, for example, that two variables, $X_{k-1}$ and $X_k$, are misspecified because they should really enter the equation in logarithmic form. In this case the elements of $f$ will be equal to $\gamma_{k-1} \log X_{k-1} + \gamma_k \log X_k$, and the expectations of the residuals will be weighted sums of the corresponding residuals obtained from regressing $\log X_{k-1}$ and $\log X_k$ on the included variables. The weights are equal to $\gamma_{k-1}$ and $\gamma_k$ and the values of these parameters may be such that the misspecification is not apparent in any of the three sets of residuals, i.e., there is a

[1]Any columns of $X$ which appear in $f$ disappear from the expression $Cf$ and $Mf$ as the columns of $X$ are orthogonal to the $C$ and $M$ matrices. In this case, therefore, the term in $X_k$ may be disregarded in computing $Cf$ and $Mf$.

'cancelling out' effect. In cases like this the powers of tests based directly on residuals will tend to be low. The results in section 7 illustrate this point.

## 4. Powers of the tests

All of the test statistics discussed so far may be written as ratios of quadratic forms in independent normal variables, and this enables their powers to be computed by the method of Imhof (1961).

The square of the $\psi$-statistic, which follows an $F$-distribution under the null hypothesis, can be written in the form

$$\psi^2 = \frac{(n-k-1)}{(n-k)} \cdot \frac{\tilde{u}'T\tilde{u}}{\tilde{u}'E\tilde{u}}, \tag{10}$$

where $T$ is a square matrix of order $(n-k)$ in which each element is equal to unity, $E = I_{n-k}-(n-k)^{-1}T$, and the vector $\tilde{u}$ has the properties set out in (9).

If $q_\alpha$ is the appropriate significance point for a (one-tailed) $F$-test with a Type I error of size $\alpha$ and $q_\alpha^* = q_\alpha(n-k)/(n-k-1)$, the power of the $\psi$-test under the alternative hypothesis is

$$1-\text{Prob. } [\tilde{u}'(T-q_\alpha^* E)\tilde{u} \leqq 0]. \tag{11}$$

Now let the $i$th characteristic root of $(T-q_\alpha^* E)$ be denoted by $\lambda_i$ and let $P$ be an orthogonal matrix of corresponding characteristic vectors. Denote the $i$th element of $\sigma^{-1}P'E(\tilde{u})$ by $\phi_i$. Expression (11) may then be written

$$1-\text{Prob.}\left[\sum_{i=1}^{n-k} \lambda_i w_i^2 \leqq 0\right], \tag{12}$$

where the $w_i$'s are independent non-central chi-square variates with one degree of freedom and non-centrality parameter, $\phi^2$; see Koerts and Abrahamse (1969, pp. 81–82).

For a given misspecified model $E(\tilde{u})$ may be obtained directly from (9) and the probability (12) may then be computed by the method of Imhof. Our calculations were performed using the FORTRAN programmes given in Koerts and Abrahamse (1969, pp. 155–160). Note that one of the roots of $T-q_\alpha^* E$ is equal to $n-k$, while the other $n-k-1$ are all equal to $-q_\alpha^*$.

The von Neumann Ratio (for recursive residuals) is

$$Q = \frac{\sum_{j=k+2}^{n} (\tilde{u}_j-\tilde{u}_{j-1})^2}{\sum_{j=k+1}^{n} (\tilde{u}_j-\bar{u})^2} = \frac{\tilde{u}'A\tilde{u}}{\tilde{u}'E\tilde{u}}, \tag{13}$$

where $A$ is a square matrix of order $(n-k)$ defined in Koerts and Abrahamse (1969, p. 68). The power of the von Neumann Ratio test may therefore be computed, for BLUS as well as recursive residuals, by a similar method to that employed for the $\psi$-test.

The Durbin–Watson '$d$-statistic' is

$$d = \frac{\sum_{j=2}^{n} (e_j - e_{j-1})^2}{\sum_{j=1}^{n} e_j^2} = \frac{e'Ae}{e'e}, \tag{14}$$

where $A$ is an $n \times n$ matrix, similar in form to the matrix in (13) above. From (6) we have $e = M(f+v) = M\zeta$. Expression (14) may therefore be written

$$d = \frac{\zeta' MAM\zeta}{\zeta' M\zeta}, \tag{15}$$

and so its power may also be evaluated by Imhof's method as $\zeta \sim N(f, \sigma^2 I_n)$.

## 5. Empirical results

The powers of the tests described in the preceding sections were evaluated for a number of different models. All tests were carried out at the 5% level of significance and in all cases the observations were arranged in such a way that the values of the explanatory variables being tested for functional misspecification were in ascending order. This ordering is necessary for the Durbin–Watson and von Neumann Ratio tests, as well as the $\psi$-test, to be effective.

The von Neumann Ratio test was carried out with both BLUS and recursive residuals. Since there are $k+1$ possible 'bases' for both sets of residuals in serial correlation tests [Theil (1971, p. 223)], some criterion is necessary for selecting a base which yields a test with a relatively high power. After some preliminary experimentation it was decided to select the BLUS base by the method proposed by Ramsey (1969, p. 354). The base for the recursive residuals was formed from the sets of observations containing the smallest $k/2$ and largest $k/2$ values of the misspecified explanatory variable; cf. the 'split basis' criteria suggested in Phillips and Harvey (1974).

The figures presented for the Durbin–Watson Bounds test are such that an entry of, for example, '0.18–0.37' indicates that the probabilities of rejecting the null hypothesis are 0.18 and 0.37 when the lower bound, $d_L$, and the upper bound, $d_U$, respectively, are taken as critical values. In the discussion of our results below we distinguish carefully between the 'Bounds test' and the 'full Durbin–Watson procedure'. The latter requires that an approximation to the exact significance point be calculated when the '$d$-statistic' falls in the inconclusive region; see Durbin and Watson (1971). In some cases, however, no calculations are necessary since a valid approximation can be made simply by taking $d_U$ to be the appropriate significance point; see Hannan and Terrell (1968).

A two-sided $\psi$-test was employed. However, it should be noted that if one is prepared to specify either convexity or concavity for the alternative functional form, a relatively more powerful one-sided test can be used.

All the regression models considered contained a constant term. The explanatory variables were obtained from

$$X_{2i} = 1+\sin(i/2), \qquad i = 1, \ldots, 15,$$

and

$$X_{3i} = 1+\cos(i/2), \qquad i = 1, \ldots, 15,$$

and appropriately ordered in each case. This data set may be found listed in Koerts and Abrahamse (1969, p. 153), and the correlation coefficient between $X_2$ and $X_3$ is 0.108. The independent variable

$$X_{4i} = i, \qquad i = 1, \ldots, n,$$

was also used in some of the models.

The tests were evaluated for four basic models, which are set out below. In all cases, the disturbance terms, $v_j$, are independently and normally distributed with mean zero and unit variance.

Model I(a) is

$$y_j = \delta_1+\delta_2 X_{2j}+\gamma X_{2j}^2+v_j, \qquad j = 1, \ldots, 15, \tag{16}$$

where the $X_{2j}$'s are in ascending order. Model I(b) contains $X_4$ in place of $X_2$.

Model II(a) is

$$y_j = \delta+\gamma \log X_{2j}+v_j, \qquad j = 1, \ldots, 15, \tag{17}$$

where the $X_{2j}$'s are in ascending order. Model II(b) replaces $X_2$ by $X_4$.

For both (16) and (17) functional misspecification arises from assuming a model of the form

$$y_j = \beta_1+\beta_2 X_{2j}+u_j. \tag{18}$$

Table 1

Powers of tests for Models I and II.

| Test | $n = 15$ | | | | | | $n = 30$ | |
|---|---|---|---|---|---|---|---|---|
| | I(a) | | | I(b) | II(a) | II(b) | I(b) | II(b) |
| | $\gamma = 2$ | $\gamma = 3$ | $\gamma = 4$ | $\gamma = 0.05$ | $\gamma = -1.0$ | $\gamma = -3.0$ | $\gamma = 0.01$ | $\gamma = -1.5$ |
| $\psi$ | 0.47 | 0.77 | 0.93 | 0.54 | 0.51 | 0.75 | 0.72 | 0.71 |
| VNR (BLUS) | 0.35 | 0.66 | 0.89 | 0.25 | 0.32 | 0.20 | 0.40 | 0.18 |
| VNR (REC) | 0.32 | 0.62 | 0.85 | 0.34 | 0.25 | 0.36 | 0.49 | 0.27 |
| D–W | 0.19–0.39 | 0.44–0.72 | 0.78–0.93 | 0.33–0.55 | 0.11–0.26 | 0.19–0.42 | 0.40–0.53 | 0.15–0.26 |

Powers[2] of the tests are set out in table 1. The $\psi$-test is clearly more powerful than the von Neumann Ratio tests, the superiority being even more marked in Model II than Model I. With regard to the Bounds test, it should be noted that there is only one explanatory variable in the model, the observations on which are arranged in ascending order and hence can be considered to 'slowly changing'. Under the circumstances, $d_U$ can be taken to be approximately equal to the true critical value of the $d$-statistic and so the higher 'power' for the Bounds test can be taken as representative of the power of the full Durbin–Watson procedure.

Results are also presented for Models I(b) and II(b) with 30, rather than 15, observations. The relative power superiority of the $\psi$-test appears, if anything, to be even greater than before.

On figs. 2 and 3 we have plotted the expected values of the recursive and BLUS residuals, obtained from (9), for bases (1, 2) and (1, 15) of Model I(a)

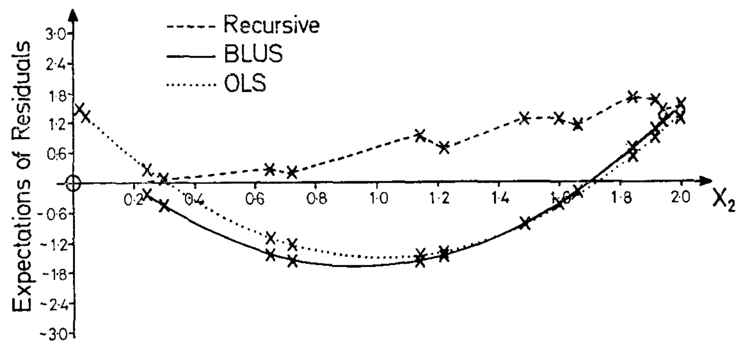


Fig. 2. Expectations of residuals for Model I(a): base (1, 2).

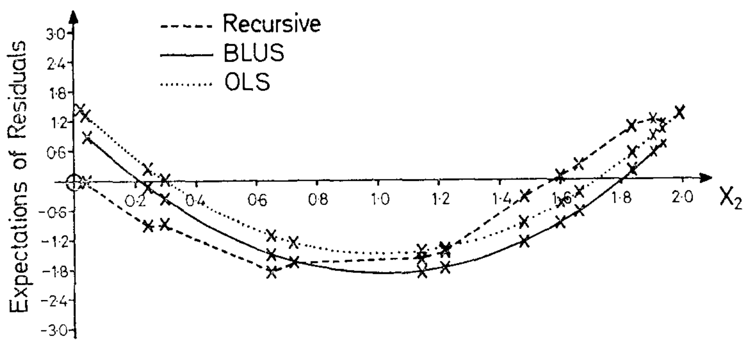


Fig. 3. Expectations of residuals for Model I(a): base (1, 15).

[2]Note that the powers of the tests are independent of the $\delta$ parameters in the true model; see section 4.

with $\gamma = 3$. [Referring to a base as (1, 2) simply indicates that it was formed from the sets of observations containing the two smallest values of the misspecified explanatory variable. Because of the assumed ordering of the data, these sets of observations are indexed 1 and 2.] By joining up the points on these figures, it is possible to gain a much clearer visual impression of the pattern of residuals likely to energe. For example, it will be seen that the expected values of the BLUS residuals follow much the same pattern as the expected values of the OLS residuals, particularly for the base (1, 2) which is 'optimal' according to Ramsey's criterion. Similarly, the expectations of the recursive residuals are close to those of the OLS residuals for base (1, 15), thus making it clear why a 'split base' is used to obtain the recursive residuals in the von Neumann Ratio tests. On base (1, 2), however, the recursive residuals all have positive expected values. This is exactly what the reasoning of section 2 suggests and it illustrates why the $\psi$-test is effective in this case.

We now consider the power of the tests for models containing two explanatory variables (plus a constant term). Only one explanatory variable (the one under test) is subject to functional misspecification, however; the other is correctly specified.

Model III(a) is

$$y_j = \delta_1+\delta_2 X_{2j}+\delta_3 X_{3j}+3X_{3j}^2+v_j, \qquad j = 1, \ldots, 15, \tag{19}$$

where the observations are arranged such that the $X_{3j}$'s are in ascending order. Model III(b) replaces $X_3$ by $X_4$ and sets the coefficient of $X_4^2$ equal to 0.05.

Model IV(a) is

$$y_j = \delta_1+\delta_2 X_{2j}+1.5 \log_e X_{3j}+v_j, \qquad j = 1, \ldots, 15, \tag{20}$$

where the observations are ordered as in (19). Model IV(b) is obtained by replacing $X_3$ by $X_4$ and setting the coefficient of $\log X_4$ equal to 3.

The powers of the tests, computed under the assumption that the model is of the form

$$y_j = \beta_1+\beta_2 X_{2j}+\beta_3 X_{3j}+u_j, \tag{21}$$

are set out[3] in table 2. It will be seen that their relative magnitudes are similar to those presented in table 1. Note that since only one of the regressors in the model is 'slowly changing' it is no longer appropriate to take $d_U$ as the critical

[3]The 'powers' given for the recursive residual von Neumann Ratio are based on the average of the powers for the two 'split bases' (1, 2, 15) and (1, 14, 15).

value in the Durbin–Watson test. In fact, we calculated that for both models the use of $d_U$ in this situation would result in a Type I error of 0.15, and so the 'upper bound' figures in table 2 cannot be regarded as being a good indicator of the power of the full Durbin–Watson procedure.

Table 2
Powers of tests for Models III and IV ($n = 15$).

| Test | III(a) | III(b) | IV(a) | IV(b) |
|---|---|---|---|---|
| $\psi$ | 0.81 | 0.56 | 0.96 | 0.76 |
| VNR (BLUS) | 0.71 | 0.20 | 0.77 | 0.20 |
| VNR (REC) | 0.57 | 0.26 | 0.63 | 0.29 |
| D–W | 0.46–0.89 | 0.25–0.76 | 0.33–0.87 | 0.18–0.67 |

Finally, there is very little evidence in either of the tables to suggest that the power of the von Neumann Ratio tests are higher when BLUS residuals, rather than recursive residuals, are employed; cf. Harvey and Phillips (1974), and Phillips and Harvey (1974). However, our conclusions on this topic should be regarded as being somewhat tentative since the whole question is complicated by the problem of base selection.

## 6. The Sign and 'Geary' tests

Two simple tests are evaluated in this section. The first, suggested by Geary (1970), is closely related to the Runs test. This (one-sided) test was carried out with OLS, BLUS and recursive residuals, the choice of base in the last two cases being the same as that employed in the von Neumann Ratio tests. The second test is a Sign test [see, for example, Siegal (1956, pp. 68–75)], employed on recursive residuals which have been calculated using the first basis. This test is carried out simply by adding up the number of positive recursive residuals. A two-sided test is then performed using binomial tables, since, under the null hypothesis, the probability that a particular recursive residual will be positive is one-half. The rationale for this test is exactly the same as for the $\psi$-test; see fig. 1.

The powers of the Geary and Sign tests were estimated by Monte Carlo methods for the models described in the previous section. Four hundred independent replications were used in each case and so the sampling error of $p$ the estimated power, is $0.05[p(1-p)]^{\frac{1}{2}}$. The Type I errors cannot, in general, be set exactly equal to 0.05 for these tests. However, the results presented in table 3 are for tests with Type I errors fairly close to 0.05, and so a comparison of their powers with the corresponding figures in table 2 is reasonably meaningful. These results also give a fair impression of the results obtained for the other models; these have not been presented here but are available on request from the authors.

For all the models considered the power of the Sign test was, as expected, somewhat lower than the $\psi$-test. However, it was more powerful than the various Geary tests and in many cases it was even superior to the von Neumann Ratio tests. The Geary tests themselves tended, in general, to be less powerful than the

Table 3
Powers of Sign and Geary tests for Models III and IV.

| Test | Type I error | III(a) | III(b) | IV(a) | IV(b) |
|---|---|---|---|---|---|
| Sign | 0.038 | 0.59 | 0.36 | 0.80 | 0.53 |
| Geary (BLUS) | 0.033 | 0.34 | 0.29 | 0.28 | 0.28 |
| Geary (REC) | 0.033 | 0.35 | 0.27 | 0.26 | 0.22 |
| Geary (OLS) | 0.029 | 0.26 | 0.11 | 0.12 | 0.07 |

von Neumann Ratio tests. In many cases the difference in power was very marked, although in certain models [e.g., III(b) and IV(b)] the powers were of a similar order. The Geary test based on OLS residuals was, on the whole, inferior to the other tests, although it must be remembered that the Type I error for this test is only nominal.

As a final point it should be noted that neither the Sign test nor the Geary test is non-parametric when used in this context. Even an assumption that the disturbances are identically and symmetrically distributed is not sufficient to ensure that the tests are exact, since only under normality will the BLUS and recursive residuals be independent rather than merely uncorrelated. Despite this, however, it seems reasonable to conjecture that the Sign test will be more robust to departures from normality than the $\psi$-test.

## 7. Functional misspecification of more than one explanatory variable

So far we have assumed that all the explanatory variables in the regressi o are, apart from the one under test, correctly specified in terms of functional form. If this assumption does not hold, the power of the various non-linearity tests may be relatively low. In the case of the $\psi$-test, the heuristic reasoning of section 2 indicates that this test will still be effective when all the misspecified variables are positively correlated with each other and the true variables are either all concave or all convex functions of these variables. In general, it will not be possible to justify the use of the $\psi$-test in this way. However, the performance of the other tests is also likely to be erratic since, as with the $\psi$-test, their justification lies in the fact that certain distinctive patterns of residuals can be assumed to occur under the alternative hypothesis. As was pointed out in section 3, these patterns will not necessarily emerge once we allow more than one explanatory variable to be misspecified. The extent to which the powers of the tests can vary in different situations is indicated by the results below.

Model V(a) is

$$y_j = \beta_1 + 2 \log X_{4j} + 1.5 \log X_{3j} + v_j, \qquad j = 1, \ldots, 15, \tag{22}$$

where the observations are arranged such that the $X_{3j}$'s are in ascending order. The assumed regression equation is of the form (21). We are therefore testing $X_3$ for functional misspecification [cf. Model IV(a)] but with the additional complication that $X_4$ is incorrectly specified also. However, the observations on $X_4$ are in ascending order and so there will be a strong positive correlation between $X_4$ and $X_3$. Since in addition both log $X_3$ and log $X_4$ are concave functions of the assumed variables $X_3$ and $X_4$, the argument elaborated earlier suggests that the $\psi$-test is likely to be fairly effective. The results set out in table 4 indicate that this is indeed the case. However, it will be seen that it is no longer true to say that the $\psi$-test is more powerful than the other tests.

Table 4
Powers of tests for Models V and VI ($n$ = 15).

| Test | V(a) | V(b) | VI(a) | VI(b) |
|---|---|---|---|---|
| $\psi$ | 0.74 | 0.27 | 0.02 | 1.00 |
| VNR (BLUS) | 0.69 | 0.16 | 0.10 | 0.93 |
| VNR (REC) | 0.78 | 0.21 | 0.14 | 0.83 |
| D–W | 0.47–0.97 | 0.03–0.36 | 0.01–0.24 | 0.79–1.00 |
| Sign | 0.51 | 0.15 | 0.02 | 0.98 |
| Geary (BLUS) | 0.44 | 0.13 | 0.12 | 0.65 |
| Geary (REC) | 0.32 | 0.10 | 0.08 | 0.59 |
| Geary (OLS) | 0.12 | 0.02 | 0.02 | 0.27 |

In Model V(b) the sign of the parameter of log $X_4$ has been changed and the results indicate that there is a considerable loss of power[4] for all tests compared with Model V(a). Model VI(a) is obtained by replacing $X_4$ by $X_2$ in eq. (22), while Model VI(b) gives the parameter of log $X_2$ a negative sign. Again the assumed model is (21) and so the main difference between Models V and VI is

[4]The results in sections 3 and 4 show that the loss in power of the $\psi$-test, von Neumann Ratio and Bounds tests would have been the same if the coefficient of log $X_3$, rather than log $X_4$, had been made negative. Similarly, if both coefficients were negative, the situation with regard to power would be exactly the same as in Model V(a).

that in the latter the two misspecified regressors are only weakly correlated. However, it will be seen that the power implications are very different, for now the tests are, on the whole, powerful when the coefficients of the explanatory variables are of different sign and weak when the coefficients have the same sign.

From the point of view of this study, the main conclusion of this section is that when more than one variable is misspecified, the $\psi$-test may lose much of its power advantage vis-à-vis the other tests. Similar conclusions hold for the Sign test as its behaviour closely mirrors that of the $\psi$-test. However in the cases where these tests have very low power, the powers of the other tests are low as well. Furthermore some simulation results carried out on the RESET test introduced by Ramsey (1969), which is a general misspecification test, showed that even this had a similar erratic performance in these circumstances.

## 8. Robustness of $\psi$-test to serial correlation

The Durbin–Watson and von Neumann Ratio tests are, of course, relatively powerful when used to test against serial correlation in an otherwise correctly specified model. In economic time series it will often be the case that the observations on a variable we wish to test for functional misspecification are ordered (approximately) in the same way whether they are arranged according to time or according to increasing value. In such cases the Durbin–Watson and von Neumann Ratio tests will not be able to distinguish between functional misspecification and serial correlation. On the other hand, the form of the $\psi$-test suggests that it will be relatively insensitive to serial correlation.

In order to test the above conjecture the following model (Model VII) was considered:

$$y_j = \beta_1 + \beta_2 X_{4j} + \omega_j, \qquad j = 1, \ldots, 20, \tag{23}$$

where $\omega_j = \rho\omega_{j-1} + \varepsilon_j$, the $\varepsilon_j$'s being normally and independently distributed with zero mean and constant variance, and $|\rho| < 1$. Under the assumption that the variables in (23) are correctly specified, the probabilities that the von Neumann Ratio, Bounds and $\psi$-tests reject the null hypothesis may be computed by a straightforward extension of Imhof's method; cf. Koerts and Abrahamse (1969). The results presented in table 5 are for tests carried out at the 5% level of significance. The power of the $\psi$-test is low compared with the other two tests and this would seem to be a point in its favour as a test for functional misspecification.

As regards the tests described in section 6, the Geary version of the Runs test is primarily designed to detect serial correlation. Indeed, its power has been studied in this context by Habibagahi and Pratschke (1972). On the other hand, the Sign test is similar in construction to the $\psi$-test. One would therefore expect it to be relatively robust to departures from the assumption of serial independence of the disturbances, which is implicit in the maintained hypothesis (1).

Table 5
Powers of tests for Model VII.

| Test | $\rho = 0.5$ | $\rho = 0.9$ |
|---|---|---|
| $\psi$ | 0.15 | 0.34 |
| VNR (REC) | 0.50 | 0.86 |
| D–W | 0.37–0.57 | 0.79–0.89 |

## 9. Conclusion

This article has shown how recursive residuals may be used to detect functional misspecification in a regression equation, the technique being particularly useful when it is felt that only one variable may be misspecified. Much is to be gained simply by looking at these residuals, although a $t$-statistic may easily be constructed from them. When the form of misspecification is such that a concave or a convex function of the incorrectly specified variable should appear in the 'correct' equation, this $t$-statistic provides a test which has a relatively high power compared with the Durbin–Watson and von Neumann Ratio tests which are commonly used in this situation. Furthermore the $t$-statistic is relatively robust to serial correlation and this is clearly a useful property in this context. Similar properties hold for the closely related Sign test which is perhaps even simpler that the $t$-test. While this test is less powerful than the $\psi$-test it may possibly have the compensating advantage of being more robust to departures from normality in the disturbances.

## References

Brown, R.L., J. Durbin and J.M. Evans, 1975, Techniques for testing the constancy of regression relationships over time (with discussion), Journal of Royal Statistical Society B 37, 149–192.

Durbin, J. and G.S. Watson, 1971, Testing for serial correlation in least squares regression III, Biometrika 58, 1–19.

Farebrother, R.W., 1976, A remark on algorithm AS 75: Basic procedures for large, sparse or weighted linear least squares problems, Applied Statistics 25, 323–324.

Geary, R.C., 1970, Relative efficiency of count and sign changes for assessing residual autoregression in least squares regression, Biometrika 57, 123–127.

Habibagahi, H. and J.L. Pratschke, 1972, A comparison of the power of the von Neumann ratio, Durbin–Watson and Geary tests, Review of Economics and Statistics 54, 179–185.

Hannan, E.J. and R.D. Terrell, 1968, Testing for serial correlation after least squares regression, Econometrica 36, 133–150.

Harvey, A.C. and G.D.A. Phillips, 1974, A comparison of the power of some tests for heteroscedasticity in the general linear model, Journal of Econometrics 2, 307–316.
Imhof, P.J., 1961, Computing the distribution of quadratic forms in normal variables, Biometrika 48, 419–426.
Koerts, J. and A.P.J. Abrahamse, 1969, On the theory and application of the general linear model (Universitaire Pers Rotterdam, Rotterdam).
Phillips, G.D.A. and A.C. Harvey, 1974, A simple test for serial correlation in regression analysis, Journal of the American Statistical Association 69, 935–939.
Ramsey, J.B., 1969, Tests for specification errors in classical linear least-squares regression analysis, Journal of the Royal Statistical Society B 31, 350–371.
Siegal, S., 1956, Nonparametric statistics for the behavioural sciences (McGraw–Hill, New York).
Theil, H., 1971, Principles of econometrics (North-Holland, Amsterdam).