No. 2003–18

ON THEIL'S ERRORS

By Jan R. Magnus and Ashoke K. Sinha

February 2003

ISSN 0924-7815

On Theil's errors

Jan R. Magnus*
CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

and

Ashoke K. Sinha
CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

February 24, 2003

Abstract

We take a fresh look at Theil's BLUS residuals and ask why they have gone out of fashion. All our simulation experiments indicate that tests based on BLUS residuals have higher power than those based on the more popular recursive residuals, even in those cases (structural breaks) where intuition would favour the recursive residuals.

JEL Classification: B23; C10; C13; C20

Keywords: BLUS residuals; Recursive residuals; Power comparisons

*Corresponding author. E-mail address: magnus@uvt.nl.

1 Introduction

This paper is not about Theil's errors. Even in the unlikely event that we had found errors in Theil's work, more courage than we possess would have been required to expose them. The paper does, however, concern Theil's *treatment* of errors (disturbances) in regression. In particular, it concerns Theil's treatment of the predicted errors, the so-called residuals.

Theil worried about the fact that, even if the disturbances are i.i.d., the residuals are neither independent nor identically distributed, thus making direct use of the residuals in testing homoskedasticity or serial independence impossible. Thus motivated, he introduced the BLUS residuals in a path-breaking paper in 1965. These residuals are linear unbiased, have a scalar variance matrix, and are also 'best' in a mean squared error sense. Thus they appear to be ideally suited for the task for which they were invented. In the five to ten years following Theil's publication, a number of refinements and improvements were published, by Theil himself, by former colleagues in Rotterdam (former, because Theil had by then moved to Chicago), and by others, but after that BLUS residuals went out of fashion.

Why did this happen? The main reason is the emergence of a competing set of residuals, namely the recursive residuals, originating with the paper by Brown, Durbin and Evans in 1975. These recursive residuals have a more intuitive appeal than the BLUS residuals, and are widely believed to be well-suited when dealing with the possibility of a structural break.[1] Modern econometric software contains recursive residuals routinely, but seldom BLUS residuals.

The BLUS and recursive residuals contain exactly the same information, because both are in one-to-one correspondence with the full set of OLS residuals. Thus the only way to compare them is through their power properties. The main purpose of the current paper is to demonstrate that BLUS residuals are not less powerful than recursive residuals; in fact, we claim they are more powerful. Thus we make a case for reinstating BLUS residuals into the mainstream of econometrics.

We employ two historical data sets (both of which we extend) to demonstrate our point: the original data used by Theil (1965) and the data used

[1] Schweder's (1976) paper on structural shifts does not even reference Theil's work.

by Quandt (1958).

In section 2 we introduce Theil's BLUS predictor and present its optimality and uniqueness properties (Theorem 1). In section 3, we pose the opposite question (Theorem 2), implying that the recursive residuals (and many other sets of residuals) have a BLUS optimality property: they are 'best' in the sense that they are as close as possible to a given linear combination of the disturbances. Durbin's recursive residuals are formally defined in section 4. In section 5 we use extensions of Theil's original data in order to compare the power of BLUS and recursive residuals against heteroskedasticity. BLUS appears to be superior, be it slightly. Then, in section 6, we use Quandt's data and the cusum and cusum-of-squares techniques to try and detect a structural break. Neither the BLUS nor the recursive residuals are successful, mostly because the number of observations is small. In section 7 we therefore extend our data and our analysis, leading to a proper comparison of the power properties of the BLUS and recursive residuals against structural breaks. We conclude that BLUS, again, is superior, in spite of the intuitive appeal of the recursive residuals. We offer some conclusions in section 8. An appendix contains the proofs of the two theorems.

2 Theil's BLUS predictor

In 1965 Theil's paper 'The analysis of disturbances in regression analysis' appeared. In this seminal contribution Theil considered the standard linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \mathrm{E}(\boldsymbol{\varepsilon}) = \boldsymbol{0}, \qquad \mathrm{E}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \boldsymbol{I}_n,$$

where $\boldsymbol{X}$ is a nonrandom $n \times k$ matrix of full rank k.[2] Normality is assumed only when desired to compute confidence intervals. Theil's principal concern was to test the assumptions on the disturbance vector $\boldsymbol{\varepsilon}$, in particular homoskedasticity and serial independence. Since $\boldsymbol{\varepsilon}$ is unobservable, Theil first tried to find an observable random vector, say $\boldsymbol{e}$, which approximates $\boldsymbol{\varepsilon}$ as closely as possible in the sense that it minimizes

$$\mathrm{E}(\boldsymbol{e} - \boldsymbol{\varepsilon})'(\boldsymbol{e} - \boldsymbol{\varepsilon})$$

[2]We adopt the notation proposed in Abadir and Magnus (2002).

subject to the constraints

$$\text{(i) } \boldsymbol{e} = \boldsymbol{A}\boldsymbol{y} \text{ for some square matrix } \boldsymbol{A} \text{ (linearity)},$$
$$\text{(ii) } \mathrm{E}(\boldsymbol{e} - \boldsymbol{\varepsilon}) = \boldsymbol{0} \text{ for all } \boldsymbol{\beta} \text{ (unbiasedness)}.$$

This leads to the best linear unbiased predictor of $\boldsymbol{\varepsilon}$,

$$\boldsymbol{e} = \boldsymbol{M}\boldsymbol{y}, \qquad \boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}', \tag{1}$$

which we recognize as the ordinary least-squares (OLS) residual vector.

Thus, the OLS residuals are best linear unbiased, but their variance matrix is nonscalar. In fact, $\mathrm{var}(\boldsymbol{e}) = \sigma^2\boldsymbol{M}$, whereas the variance matrix of $\boldsymbol{\varepsilon}$, which $\boldsymbol{e}$ hopes to resemble, is $\sigma^2\boldsymbol{I}_n$.

For this reason Theil wished to find a predictor of $\boldsymbol{\varepsilon}$ (more precisely, of $\boldsymbol{S}'\boldsymbol{\varepsilon}$) which, in addition to being linear and unbiased, has a scalar variance matrix. There is a whole class of such predictors. The 'best' in this class is Theil's BLUS predictor: best linear unbiased with scalar variance matrix.

Definition 1: Consider the linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Let $\boldsymbol{S}$ be a given $n \times (n-k)$ matrix. A random $(n-k) \times 1$ vector $\boldsymbol{w}$ is called a *BLUS predictor* of $\boldsymbol{S}'\boldsymbol{\varepsilon}$ if

$$\mathrm{E}(\boldsymbol{w} - \boldsymbol{S}'\boldsymbol{\varepsilon})'(\boldsymbol{w} - \boldsymbol{S}'\boldsymbol{\varepsilon})$$

is minimized subject to the constraints

$$\text{(i) } \boldsymbol{w} = \boldsymbol{A}'\boldsymbol{y} \text{ for some } n \times (n-k) \text{ matrix } \boldsymbol{A} \text{ (linearity)},$$
$$\text{(ii) } \mathrm{E}(\boldsymbol{w} - \boldsymbol{S}'\boldsymbol{\varepsilon}) = \boldsymbol{0} \text{ for all } \boldsymbol{\beta} \text{ (unbiasedness)},$$
$$\text{(iii) } \mathrm{var}(\boldsymbol{w}) = \sigma^2\boldsymbol{I}_{n-k} \text{ (scalar variance matrix)}.$$

The next theorem provides the unique solution to this problem.

Theorem 1: Consider the linear regression model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Let $\boldsymbol{S}$ be a given $n \times (n-k)$ matrix such that $\mathrm{rk}(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S}) = n-k$. Then the BLUS predictor of $\boldsymbol{S}'\boldsymbol{\varepsilon}$ is

$$\boldsymbol{w}^* = \boldsymbol{A}'\boldsymbol{y}, \qquad \boldsymbol{A} = \boldsymbol{M}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{-1/2},$$

where $(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{-1/2}$ is the positive definite square root of $(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{-1}$.

Theil's original proof is a little cumbersome. A much shorter proof, following Magnus and Neudecker (1988), is presented in the appendix.

In a follow-up paper, Theil (1968) showed that the BLUS predictor $\boldsymbol{w}^*$ satisfies a stronger optimality property, namely that

$$\mathrm{E}(\boldsymbol{w}^* - \boldsymbol{S}'\boldsymbol{\varepsilon})(\boldsymbol{w}^* - \boldsymbol{S}'\boldsymbol{\varepsilon})' \leq \mathrm{E}(\boldsymbol{w} - \boldsymbol{S}'\boldsymbol{\varepsilon})(\boldsymbol{w} - \boldsymbol{S}'\boldsymbol{\varepsilon})'$$

for any linear unbiased predictor $\boldsymbol{w}$ with scalar variance matrix. A whole chapter of the *Principles of Econometrics* (Theil, 1971), 43 pages in total, is devoted to BLUS residuals. Theil only considered the possibility that $\boldsymbol{S}$ is a selection matrix, so that a subset of $n-k$ of the disturbances is predicted, but Theorem 1 does not require this property of $\boldsymbol{S}$.

In using BLUS residuals in practice, one must choose a 'base' (in Theil's terminology), that is one must choose which k observations to disregard. Ideally, "which disturbances should be disregarded is largely a matter of power with respect to a specific alternative hypothesis" (Theil, 1965, p. 1070). Since maximizing power leads to considerable complications, Theil (1971, p. 217) adopted a more practical approach. When testing against heteroskedasticity, choose the middle k observations; when testing against first-order autocorrelation, choose the first k observations or the last k or a mixture of the two.

Improvements and extensions of Theil's work on BLUS residuals can be found in Koerts (1967), Putter (1967), Koerts and Abrahamse (1968), Abrahamse and Koerts (1971), and others.

3 The opposite question

In the previous sections we asked whether, given $\boldsymbol{S}$, we could find an optimal $\boldsymbol{A}$. We now raise the opposite question, that is, we ask if, given $\boldsymbol{A}$, we can find $\boldsymbol{S}$ such that $\boldsymbol{A}'\boldsymbol{y}$ is a BLUS predictor of $\boldsymbol{S}'\boldsymbol{\varepsilon}$. Such an $\boldsymbol{S}$ will not be unique.

Thus, suppose we are given an $n \times (n-k)$ matrix $\boldsymbol{A}$ satisfying $\mathrm{var}(\boldsymbol{A}'\boldsymbol{e}) = \sigma^2 \boldsymbol{I}_{n-k}$, that is

$$\boldsymbol{A}'\boldsymbol{M}\boldsymbol{A} = \boldsymbol{I}_{n-k}.$$

Without loss of generality we may assume that $\mathrm{col}(\boldsymbol{A}) \subseteq \mathrm{col}(\boldsymbol{M})$, that is $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{B}$ for some $n \times (n-k)$ matrix $\boldsymbol{B}$. Then, $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_{n-k}$, $\boldsymbol{M}\boldsymbol{A} = \boldsymbol{A}$ and $\mathrm{rk}(\boldsymbol{M}) = \mathrm{rk}(\boldsymbol{A})$, and hence $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{A}'$; see Magnus and Neudecker (1988, Theorem 2.8).

There always exists a matrix $\boldsymbol{S}$ such that $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{-1/2}$, for example, $\boldsymbol{S} = \boldsymbol{A}$. Theorem 2 provides the full class of matrices with this property.

Theorem 2: Let $\boldsymbol{A}$ be a given $n \times (n-k)$ matrix such that $\boldsymbol{A}'\boldsymbol{M}\boldsymbol{A} = \boldsymbol{I}_{n-k}$, and assume that $\text{col}(\boldsymbol{A}) \subseteq \text{col}(\boldsymbol{M})$. Then the class of matrices $\boldsymbol{S}$ satisfying $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{-1/2}$ is given by

$$\boldsymbol{S} = \boldsymbol{A}\boldsymbol{Q} + \boldsymbol{X}\boldsymbol{R},$$

where $\boldsymbol{Q}$ is positive definite (and symmetric) and $\boldsymbol{R}$ is arbitrary.

The consequence of Theorem 2 is that any predictor $\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{y} = \boldsymbol{A}'\boldsymbol{e}$ with $\text{var}(\boldsymbol{w}) = \sigma^2\boldsymbol{I}_{n-k}$ has an optimality property, namely that $\boldsymbol{w}$ is the BLUS predictor of $\boldsymbol{S}'\boldsymbol{\varepsilon}$, where $\boldsymbol{S}$ is given in Theorem 2. More specifically this implies that the recursive residuals have a BLUS interpretation and thus possess an optimality property.

4 Durbin's recursive residuals

The history of recursive residuals is more ambiguous. The idea of recursive residuals in econometrics was first presented by Durbin at the European Meeting on statistics, econometrics, and management science in Amsterdam, September 1968 (Brown and Durbin, 1968). After Brown's death in 1972, Durbin invited Evans to complete the calculations started by Brown. This lead to Brown, Durbin and Evans (1975), read before the Royal Statistical Society in December 1974. Apparently unaware of Durbin's work, Hedayat and Robson (1970) discussed recursive residuals (which they call stepwise residuals), as did Phillips and Harvey (1974). Durbin obtained the recursive residuals as a generalization of the well-known Helmert transformation, and he lectured on this material at the London School of Economics in the mid-1950s. It thus seems fair to attribute the application of recursive residuals in econometrics to Durbin.

Farebrother (1978) discovered that recursive residuals (including the link with Helmert's transformation) were already discussed by Pizzetti (1891). In fact, one may even trace the original idea back to Gauss (1821); see Plackett (1950) and Young (1984, Appendix 2).

The recursive residuals are defined as follows. Let $\boldsymbol{x}_1', \ldots, \boldsymbol{x}_n'$ denote the rows of $\boldsymbol{X}$, and $y_1, \ldots, y_n$ the components of $\boldsymbol{y}$. Now define $\boldsymbol{X}_{(r)}' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r)$ and $\boldsymbol{y}_{(r)}' = (y_1, \ldots, y_r)$, and let $\boldsymbol{b}_{(r)}$ denote the OLS estimator of $\boldsymbol{\beta}$ based on the first r observations, that is, $\boldsymbol{b}_{(r)} = (\boldsymbol{X}_{(r)}'\boldsymbol{X}_{(r)})^{-1}\boldsymbol{X}_{(r)}'\boldsymbol{y}_{(r)}$. The recursive residuals are then defined as

$$w_r = \frac{y_r - \boldsymbol{x}_r'\boldsymbol{b}_{(r-1)}}{\sqrt{1 + \boldsymbol{x}_r'(\boldsymbol{X}_{(r-1)}'\boldsymbol{X}_{(r-1)})^{-1}\boldsymbol{x}_r}}, \qquad r = k+1, \ldots, n.$$

The unbiasedness and linearity of w_r is obvious. The fact that y_r and $\boldsymbol{b}_{(r-1)}$ are uncorrelated implies that $\mathrm{var}(w_r) = \sigma^2$. The fact that w_r and w_s are uncorrelated for $r < s$ follows from $\mathrm{cov}(y_r - \boldsymbol{x}_r'\boldsymbol{b}_{(r-1)}, y_s - \boldsymbol{x}_s'\boldsymbol{b}_{(s-1)}) = 0$, which is easily seen by writing y_r and $\boldsymbol{b}_{(r-1)}$ as linear functions of the disturbances. We thus obtain an $(n-k)\times 1$ vector $\boldsymbol{w} = (w_{k+1}, \ldots, w_n)'$ satisfying $\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{y}$ such that $\boldsymbol{w} \sim (0, \sigma^2\boldsymbol{I}_{n-k})$.

We now have two sets of constructed residuals: the BLUS residuals, say $\boldsymbol{w}_1 = \boldsymbol{A}_1'\boldsymbol{y}$, and the recursive residuals, say $\boldsymbol{w}_2 = \boldsymbol{A}_2'\boldsymbol{y}$. Let $\boldsymbol{e} = \boldsymbol{M}\boldsymbol{y}$ denote the full set of residuals. Since $\boldsymbol{A}_1\boldsymbol{A}_1' = \boldsymbol{M}$ and $\boldsymbol{A}_1'\boldsymbol{A}_1 = \boldsymbol{I}_{n-k}$, we see that

$$\boldsymbol{A}_1\boldsymbol{w}_1 = \boldsymbol{e}, \qquad \boldsymbol{w}_1 = \boldsymbol{A}_1'\boldsymbol{e},$$

so that the BLUS residuals and the full set of OLS residuals are in one-to-one correspondence. In exactly the same way, the recursive residuals and the OLS residuals are in one-to-one correspondence. Hence, BLUS and recursive residuals are in one-to-one correspondence, in fact

$$\boldsymbol{w}_1 = \boldsymbol{A}_1'\boldsymbol{A}_2\boldsymbol{w}_2, \qquad \boldsymbol{w}_2 = \boldsymbol{A}_2'\boldsymbol{A}_1\boldsymbol{w}_1.$$

Since both sets of residuals contain exactly the same information, this immediately raises the question which residuals are 'better', that is, have higher power. To this question we now turn.

5 Power comparisons: Theil's data

In order to illustrate the use of BLUS residuals in practice, Theil (1965) (and also Theil (1971, pp. 215–216)) considered the simple example,

$$y_t = \beta_1 t + \beta_2 \sin(t/2) + \varepsilon_t, \qquad t = 1, \ldots, n,$$

where $\beta_1 = 1$, $\beta_2 = 10$, and the ε_t are i.i.d. $N(0,1)$. Taking $n = 20$ independent draws from the $N(0,1)$ distribution, and choosing the middle $k = 2$ observations (10 and 11) as the base, Theil calculated the $n - k = 18$ BLUS residuals w_j and computed

$$F = \frac{\sum_{j=1}^{9} w_j^2/9}{\sum_{j=10}^{18} w_j^2/9},$$

which follows an $F(9,9)$-distribution under the null hypothesis, and takes the value $F = 0.4042$ in this example. The associated two-sided p-value is 0.1934, so that the null hypothesis is not rejected at the 5% level.

The example is a little curious, because it does not tell us anything about the usefulness of the test. Since the data are generated under the null hypothesis of homoskedasticity, we know in advance that the probability of rejection is 5%. Surely it is more interesting to generate the data under the alternative hypothesis of heteroskedasticity. Thus, we assume that the disturbances ε_t are independently distributed as $N(0, t/2)$, so that their variance increases over time. Choosing the same model and parameter values as before, and letting the sample size n grow from 20 to 100, we repeat Theil's heteroskedasticity test. With $10,000$ replications for each of $n = 20, 25, 30, \ldots, 100$, we obtain good estimates of the power of the test, ranging from 31% when $n = 20$ to 95% when $n = 100$.

FIGURE 1

We can also use the recursive residuals instead of the BLUS residuals in order to perform the heteroskedasticity test. Using the same set-up, we see in Figure 1 that the power of the test based on recursive residuals is very similar but slightly lower than the test based on BLUS residuals.[3] These results confirm the power comparisons in Harvey and Phillips (1974).

Thus, so far there is no reason to believe that recursive residuals are better than BLUS residuals. In fact, if anything, the opposite is true. Of course, one could object that we have favoured BLUS by choosing an example (heteroskedasticity) for which BLUS was developed. Hence, we now consider an example (structural break) for which the recursive residuals seem *a priori* preferable.

[3]One may argue that a two-sided F-test is inappropriate here, and that one should perform a one-sided test. The resulting power curves are very similar to Figure 1 and lead to the same conclusions.

6 Cusum and cusum-of-squares: Quandt's data

To put the recursive residuals in the best possible light, we consider testing for a structural break. Since the data underlying the examples in Brown, Durbin and Evans (1975) are not available, we use the data studied by Quandt (1958). These data are generated by the process

$$y_t = \begin{cases} 2.5 + 0.7x_t + \varepsilon_t, & t = 1, \ldots, 12, \\ 5 + 0.5x_t + \varepsilon_t, & t = 13, \ldots, 20, \end{cases}$$

where the ε_t are i.i.d. $\mathrm{N}(0, 1)$ distributed. The $\{x_t\}$ are the numbers 1 to 20, but randomized.[4] The technique described in Quandt (1958, 1960) is appropriate if we know that there is one break, but we do not know where the break occurred. For each r from $r = k+1$ to $r = n-k-1$ (that is from 3 to 17) Quandt calculates

$$\lambda_r = \log\left(\frac{\text{max likelihood given } \mathrm{H}_0}{\text{max likelihood given } \mathrm{H}_1}\right),$$

where H_0 is the hypothesis of no structural break, and H_1 the hypothesis that the observations in the period $t \leq r$ come from a different regression than those in the period $t \geq r+1$. It is easy to show that

$$\lambda_r = \frac{r}{2}\log\hat{\sigma}_1^2 + \frac{n-r}{2}\log\hat{\sigma}_2^2 - \frac{n}{2}\log\hat{\sigma}^2,$$

where $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, and $\hat{\sigma}^2$ represent the usual estimates of σ^2 based on the first r observations, the last $n-r$ observations, and all observations, respectively.[5] The value of r where λ_r attains a minimum is then an estimate of the switch-point.

FIGURE 2

Figure 2 shows that r is correctly estimated at $r = 12$ in this case. There exists, however, no formal test, because the distribution of $\min(\lambda_r)$ under H_0 is unknown.

[4] The purpose of the randomization is not entirely clear. The resulting x_t's are 'independent', but there is nothing that requires them to be.

[5] In this case, BLUS, recursive, and OLS residuals all produce an identical value of λ_r.

A formal test along different lines was developed in Brown, Durbin and Evans (1975) in the form of cusum and cusum-of-squares plots. Suppose that $w_{k+1}, \ldots, w_n$ is a set of recursive residuals, distributed i.i.d. $\mathrm{N}(0, \sigma^2)$ under the null hypothesis. Let $\hat{\sigma}^2$ be the usual estimate of σ^2. Then we define the cusum W_r and the cusum-of-squares s_r as

$$W_r = \frac{1}{\hat{\sigma}} \sum_{j=1}^{r} w_{k+j}, \qquad s_r = \frac{1}{(n-k)\hat{\sigma}^2} \sum_{j=1}^{r} w_{k+j}^2, \qquad r = 1, \ldots, n-k.$$

Under the null hypothesis of no structural break, W_r and s_r should not cross certain bounds, which are provided in Durbin (1969) and Brown, Durbin and Evans (1975).

FIGURES 3 AND 4

We see from Figures 3 and 4 that W_r and s_r (the dash-dotted lines) do not cross the bounds, indicating that neither cusum nor cusum-of-squares indicate that a structural break has occurred. Hence, even though there is a structural break and the null hypothesis is false, the tests do not reject the null hypothesis.

Instead of using the recursive residuals we can also use the BLUS residuals for this purpose.[6] The resulting plots for the Quandt data are also provided in Figures 3 and 4 (solid lines). The conclusions are the same, although the cusum-of-squares plot almost crosses the bound at $r = 10$.

The failure to identify a structural break is possibly due to the particular data set or to the small sample size. A more complete treatment of the power properties of these tests is therefore required. This discussion is provided in the next section.

7 Power comparisons: Quandt's simulated data

To gain further insight in the possible power differences between BLUS and recursive residuals, we extend Quandt's set-up as follows:

$$y_t = \begin{cases} 2.5 + 0.7t + \varepsilon_t, & t = a_n, \ldots, 12, \\ 5 + 0.5t + \varepsilon_t, & t = 13, \ldots, b_n, \end{cases}$$

[6] When using BLUS to test against structural breaks we always select the first and last observations as our base. Other choices of base have been considered too, but do not alter the conclusions.

where, as before, the ε_t are i.i.d. $N(0,1)$ distributed, and

$$a_n = 1 - 3\left(\frac{n-20}{5}\right), \qquad b_n = 20 + 2\left(\frac{n-20}{5}\right), \qquad n = 20, 25, \ldots, 100.$$

In the new set-up the break continues to be at $t = 12$, and the ratio of observations before the break and after the break continues to be $3 : 2$. There are only two differences between Quandt's data and the current data. First, we have more data: $20 \leq n \leq 100$. Secondly, we do not randomize the x_t, so that $x_t = t$. For each draw, the test either (correctly) rejects the null hypothesis or not. With $10,000$ draws the average number of rejections is an accurate estimate of the power of the test.

FIGURES 5 and 6

We first consider the BLUS residuals. We see from Figure 5 that the cusum test has uniformly much better power than the cusum-of-squares test, which has rather poor power. The same is true for the recursive residuals (Figure 6), although the power of the cusum-of-squares test is not quite so poor as for the BLUS residuals. We conclude that — if we are testing against a shift in the mean (the β's) — then cusum should be used.[7]

FIGURE 7

The most relevant comparison are the cusum plots of the BLUS and recursive residuals. These are plotted in Figure 7. The difference in power is small, but the BLUS residuals (again) have slightly higher power. This is remarkable, because we are now comparing the BLUS and recursive residuals in a situation (structural breaks) where the recursive residuals should have an advantage. Apparently, they do not.

So far we have only considered structural breaks in the mean. It is rather intuitive that the cusum test (which is linear) should have higher power in this situation than the cusum-of-squares test (which is quadratic), and this

[7]We sometimes find (with cusum, not with cusum-of-squares) boundary crossings at the very beginning or end of our data. This is somewhat unsatisfactory, so we also plotted the power curves when boundary crossings at the 5% tails of the data were ignored (one observation at each end for $n \leq 60$, two for $n > 60$). This made very little difference when working with BLUS residuals, but much more difference with the recursive residuals. This means that boundary crossings at the extremes occur regularly with the recursive residuals, another property where BLUS has the advantage.

is confirmed by our simulations. It is quite possible, however, that if we consider a structural break in the variance, the cusum-of-squares test will have higher power than the cusum test.[8] To investigate this possibility we extend the Quandt data in a different direction, and consider

$$y_t = 2.5 + 0.7t + \varepsilon_t, \qquad t = a_n, \ldots, b_n,$$

where $\varepsilon_t \sim \mathrm{N}(0,1)$ when $t \leq 12$ and $\varepsilon_t \sim \mathrm{N}(0,2)$ when $t > 12$. In this case the power of the cusum test is always lower than 20%, also for large n. But the power of the cusum-of-squares test increases more or less linearly over the interval $20 \leq n \leq 100$. The power of the BLUS procedure is (again) somewhat higher than that of the recursive residuals procedure.

In a practical situation where one is uncertain whether to test against a structural break in the mean or in the variance, one typically performs both the cusum and the cusum-of-squares tests. If there is a structural break in the variance, then cusum will in all probability not be significant, but cusum-of-squares might be, and BLUS gives you a (slightly) better chance of detecting the break than the recursive residuals. If there is a structural break in the mean, then cusum will in all probability be significant (especially when using BLUS residuals). Another look at Figures 5 and 6 now shows that the cusum-of-squares test will probably not be significant for BLUS, but may be significant for the recursive residuals. The low power of cusum-of-squares in Figure 5 can thus be used to advantage!

8 Conclusion

In this paper we have tried to show that the BLUS residuals, invented by Theil in 1965, are still a mighty weapon and should be thought of as one of Theil's main contributions to econometrics. The fact that BLUS has gone out of fashion and has been replaced by recursive residuals does not appear to be justified. All our simulation results point to the superiority of BLUS. We hope that our results will lead to a return of BLUS residuals into the mainstream of econometrics and will become available in econometric software packages.

[8] The same intuition was also formulated by Brown, Durbin and Evans (1975, p. 159).

Acknowledgements

We thank A.P.J. Abrahamse, J.-M. Dufour, J. Durbin, A.C. Harvey, R.E. Quandt, and G.P.H. Styan for useful discussions.

Appendix: Proofs

Proof of Theorem 1: We seek a linear predictor $\boldsymbol{w}$ of $\boldsymbol{S}'\boldsymbol{\varepsilon}$, that is a predictor of the form $\boldsymbol{w} = \boldsymbol{A}'\boldsymbol{y}$, where $\boldsymbol{A}$ is a constant $n \times (n-k)$ matrix. Unbiasedness of the prediction error requires

$$\mathbf{0} = \mathrm{E}(\boldsymbol{A}'\boldsymbol{y} - \boldsymbol{S}'\boldsymbol{\varepsilon}) = \boldsymbol{A}'\boldsymbol{X}\boldsymbol{\beta} \qquad \text{for all } \boldsymbol{\beta} \text{ in } \mathbb{R}^k,$$

which yields

$$\boldsymbol{A}'\boldsymbol{X} = \mathbf{O}. \tag{2}$$

The variance matrix of $\boldsymbol{w}$ is $\mathrm{E}(\boldsymbol{w}\boldsymbol{w}') = \sigma^2\boldsymbol{A}'\boldsymbol{A}$. In order to satisfy condition (iii) of Definition 1, we thus require

$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_{n-k}. \tag{3}$$

Under the constraints (2) and (3), the prediction error variance is

$$\mathrm{var}(\boldsymbol{A}'\boldsymbol{y} - \boldsymbol{S}'\boldsymbol{\varepsilon}) = \sigma^2(\boldsymbol{I} + \boldsymbol{S}'\boldsymbol{S} - \boldsymbol{A}'\boldsymbol{S} - \boldsymbol{S}'\boldsymbol{A}). \tag{4}$$

Hence the BLUS predictor of $\boldsymbol{S}'\boldsymbol{\varepsilon}$ is obtained by minimizing the trace of (4) with respect to $\boldsymbol{A}$ subject to the constraints (2) and (3). This amounts to solving the problem

$$\begin{array}{ll} \text{maximize} & \mathrm{tr}(\boldsymbol{A}'\boldsymbol{S}) \\ \text{subject to} & \boldsymbol{A}'\boldsymbol{X} = \mathbf{O} \quad \text{and} \quad \boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_{n-k}. \end{array}$$

We define the Lagrangian function

$$\psi(\boldsymbol{A}) = \mathrm{tr}\,\boldsymbol{A}'\boldsymbol{S} - \mathrm{tr}\,\boldsymbol{L}_1'\boldsymbol{A}'\boldsymbol{X} - \frac{1}{2}\,\mathrm{tr}\,\boldsymbol{L}_2(\boldsymbol{A}'\boldsymbol{A} - \boldsymbol{I}_{n-k}),$$

where $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ are matrices of Lagrange multipliers and $\boldsymbol{L}_2$ is symmetric. Differentiating ψ with respect to $\boldsymbol{A}$ yields

$$\begin{aligned} \mathrm{d}\psi &= \mathrm{tr}(\mathrm{d}\boldsymbol{A})'\boldsymbol{S} - \mathrm{tr}\,\boldsymbol{L}_1'(\mathrm{d}\boldsymbol{A})'\boldsymbol{X} - \frac{1}{2}\,\mathrm{tr}\,\boldsymbol{L}_2(\mathrm{d}\boldsymbol{A})'\boldsymbol{A} - \frac{1}{2}\,\mathrm{tr}\,\boldsymbol{L}_2\boldsymbol{A}'\,\mathrm{d}\boldsymbol{A} \\ &= \mathrm{tr}\,\boldsymbol{S}'\,\mathrm{d}\boldsymbol{A} - \mathrm{tr}\,\boldsymbol{L}_1\boldsymbol{X}'\,\mathrm{d}\boldsymbol{A} - \mathrm{tr}\,\boldsymbol{L}_2\boldsymbol{A}'\,\mathrm{d}\boldsymbol{A}. \end{aligned}$$

The first-order conditions are

$$\boldsymbol{S} = \boldsymbol{X}\boldsymbol{L}_1' + \boldsymbol{A}\boldsymbol{L}_2 \tag{5}$$
$$\boldsymbol{A}'\boldsymbol{X} = \mathbf{O} \tag{6}$$
$$\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_{n-k}. \tag{7}$$

Pre-multiplying (5) with $\boldsymbol{X}'$ yields

$$\boldsymbol{L}_1' = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{S}, \tag{8}$$

because $\boldsymbol{X}'\boldsymbol{A} = \mathbf{O}$ in view of (6). Inserting (8) in (5) gives

$$\boldsymbol{M}\boldsymbol{S} = \boldsymbol{A}\boldsymbol{L}_2. \tag{9}$$

Also, pre-multiplying (5) with $\boldsymbol{A}'$ gives

$$\boldsymbol{A}'\boldsymbol{S} = \boldsymbol{S}'\boldsymbol{A} = \boldsymbol{L}_2 \tag{10}$$

in view of (6) and (7) and the symmetry of $\boldsymbol{L}_2$. Pre-multiplying (9) with $\boldsymbol{S}'$ and using (10) we find $\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S} = \boldsymbol{L}_2^2$, and hence

$$\boldsymbol{L}_2 = (\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{1/2}. \tag{11}$$

Since we wish to maximize $\mathrm{tr}(\boldsymbol{A}'\boldsymbol{S})$, it follows from (10) that we need to maximize the trace of $\boldsymbol{L}_2$. Therefore we must choose in (11) the *positive definite* square root of $\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S}$. Inserting (11) in (9) yields $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{-1/2}$.

Proof of Theorem 2: Since $(\boldsymbol{A} : \boldsymbol{X})$ is a nonsingular $n \times n$ matrix, we can always write $\boldsymbol{S} = \boldsymbol{A}\boldsymbol{Q} + \boldsymbol{X}\boldsymbol{R}$ for some $\boldsymbol{Q}$ and $\boldsymbol{R}$. Using $\boldsymbol{M} = \boldsymbol{A}\boldsymbol{A}'$ and $\boldsymbol{A}'\boldsymbol{S} = \boldsymbol{Q}$, we then obtain

$$\boldsymbol{A} = \boldsymbol{M}\boldsymbol{S}(\boldsymbol{S}'\boldsymbol{M}\boldsymbol{S})^{-1/2} = \boldsymbol{A}\boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{Q})^{-1/2}.$$

Premultiplying by $\boldsymbol{A}'$ gives $\boldsymbol{I}_{n-k} = \boldsymbol{Q}(\boldsymbol{Q}'\boldsymbol{Q})^{-1/2}$, so that

$$\boldsymbol{Q} = (\boldsymbol{Q}'\boldsymbol{Q})^{1/2}.$$

It is now clear that $\boldsymbol{Q}$ must be symmetric and positive definite, thus completing the proof.

References

Abadir, K.M., Magnus, J.R. (2002). Notation in econometrics: a proposal for a standard. Econometrics Journal 5, 76–90.

Abrahamse, A.P.J., Koerts, J. (1971). New estimators of disturbances in regression analysis. Journal of the American Statistical Association 66, 71–74.

Brown, R.L., Durbin, J. (1968). Methods of investigating whether a regression relationship is constant over time. Selected Statistical Papers, European Meeting, Mathematical Centre Tracts 26. Mathematisch Centrum, Amsterdam.

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

Durbin, J. (1969). Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. Biometrika 56, 1–15.

Farebrother, R.W. (1978). An historical note on recursive residuals. Journal of the Royal Statistical Society, Series B 40, 373–375.

Gauss, C.F. (1821, collected works 1873). Theoria Combinationis Observationum Erroribus Minimis Obnoxiae. Werke 4, Göttingen.

Harvey, A.C., Phillips, G.D.A. (1974). A comparison of the power of some tests for heteroskedasticity in the general linear model. Journal of Econometrics 2, 307–316.

Hedayat, A., Robson, D.S. (1970). Independent stepwise residuals for testing homoscedasticity. Journal of the American Statistical Association 65, 1573–1581.

Koerts, J. (1967). Some further notes on disturbance estimates in regression analysis. Journal of the American Statistical Association 62, 169–183.

Koerts, J., Abrahamse, A.P.J. (1968). On the power of the BLUS procedure. Journal of the American Statistical Association 63, 1227–1236.

Magnus, J.R., Neudecker, H. (1988, 1999), Matrix Differential Calculus with Applications in Statistics and Econometrics, revised edition. John Wiley, Chichester/New York.

Phillips, G.D.A., Harvey, A.C. (1974). A simple test for serial correlation in regression analysis. Journal of the American Statistical Association 69, 935–939.

Pizzetti, P. (1891). I Fondamenti Matematici per la Critica dei Risultati Sperimentali, Genoa. Reprinted in: Atti della Università di Genova, 1892.

Plackett, R.L. (1950). Some theorems in least squares. Biometrika 37, 149–157.

Putter, J. (1967). Orthonormal bases of error spaces and their use for investigating the normality and variances of residuals. Journal of the American Statistical Association 62, 1023–1036.

Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. Journal of the American Statistical Association 53, 873–880.

Quandt, R.E. (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. Journal of the American Statistical Association 55, 324–330.

Schweder, T. (1976). Some "optimal" methods to detect structural shift or outliers in regression. Journal of the American Statistical Association 71, 491–501.

Theil, H. (1965). The analysis of disturbances in regression analysis. Journal of the American Statistical Association 60, 1067–1079.

Theil, H. (1968). A simplification of the BLUS procedure for analyzing regression disturbances. Journal of the American Statistical Association 63, 242–251.

Theil, H. (1971). Principles of Econometrics. North-Holland, Amsterdam.

Young, P. (1984). Recursive Estimation and Time-Series Analysis: An Introduction. Springer-Verlag, Berlin.

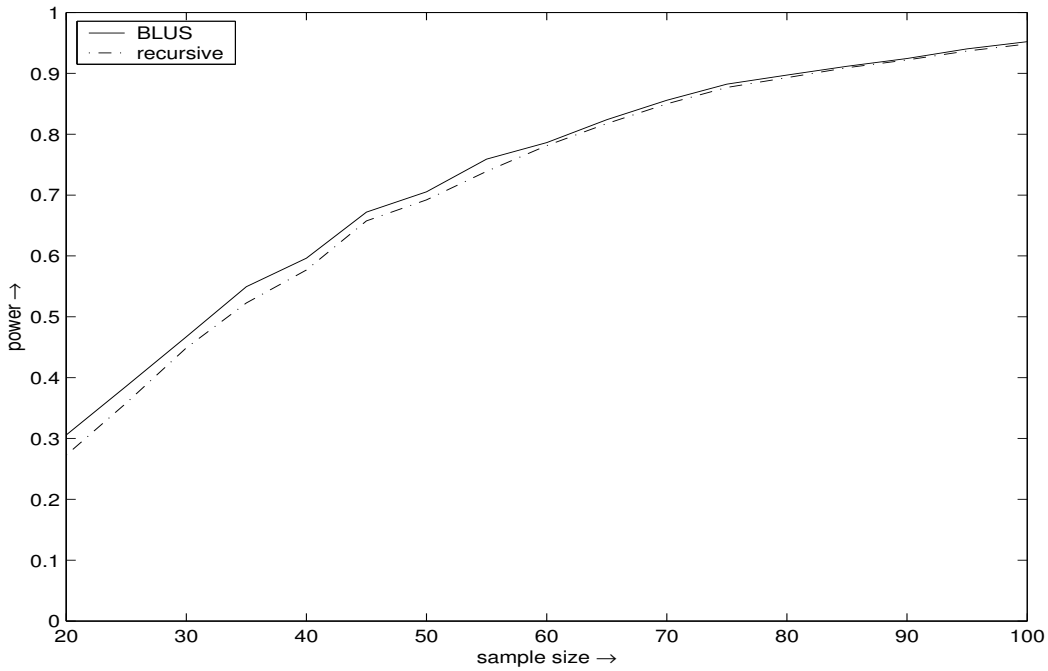

Fig. 1. Power of two-sided F-test: Theil's simulated data.

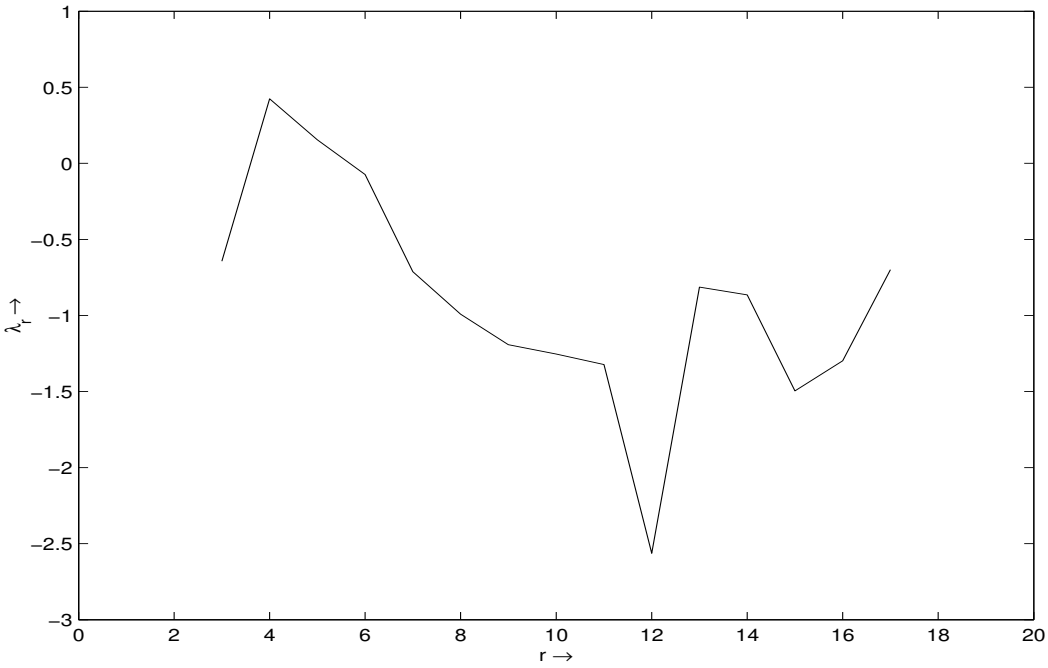

Fig. 2. Quandt's likelihood ratio λ_r.

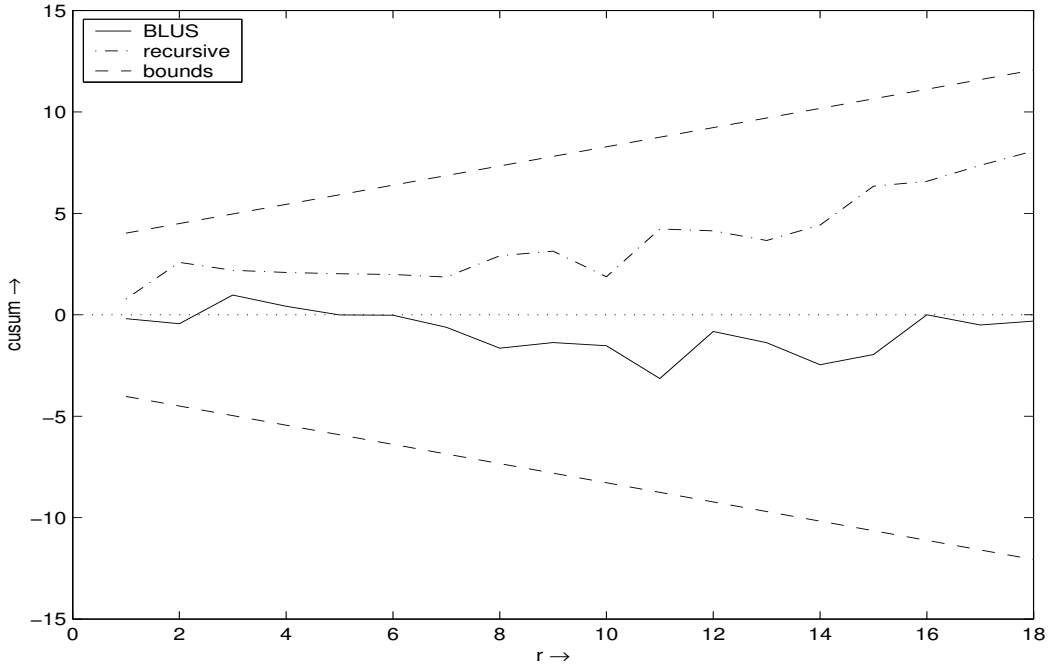

Fig. 3. Cusum plots: Quandt's data.

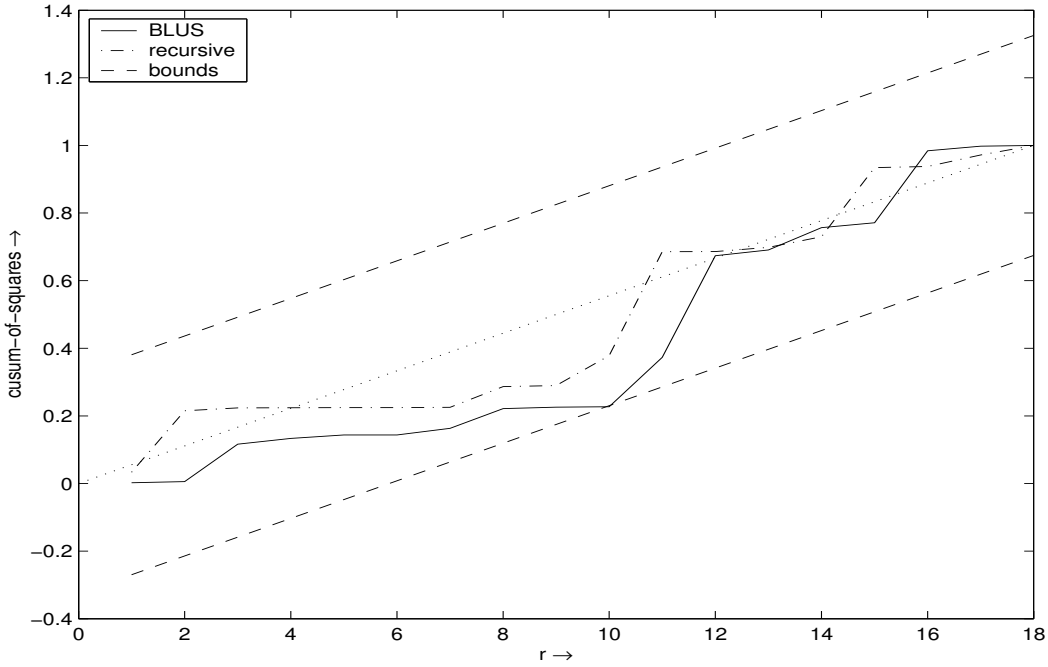

Fig. 4. Cusum-of-squares plots: Quandt's data.

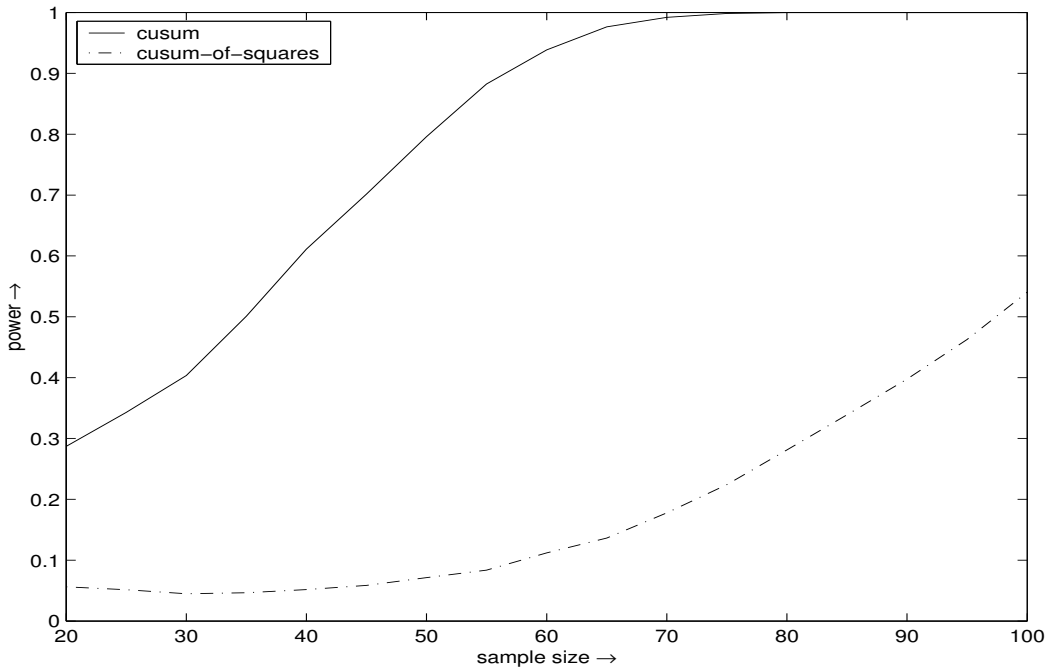

Fig. 5. Power of BLUS residuals: Quandt's simulated data.

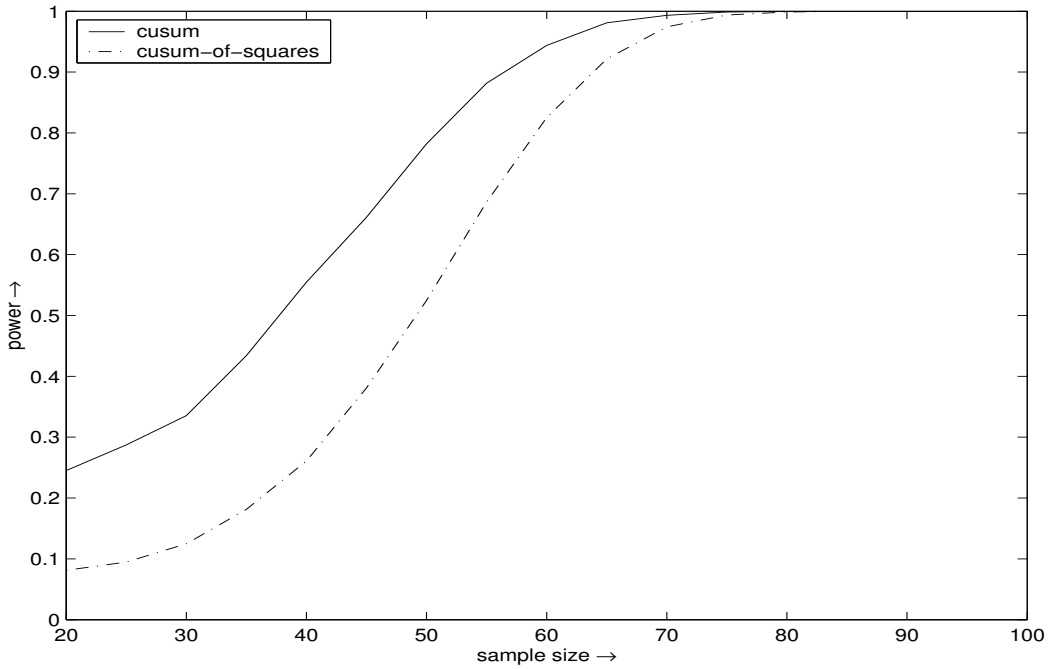

Fig. 6. Power of recursive residuals: Quandt's simulated data.

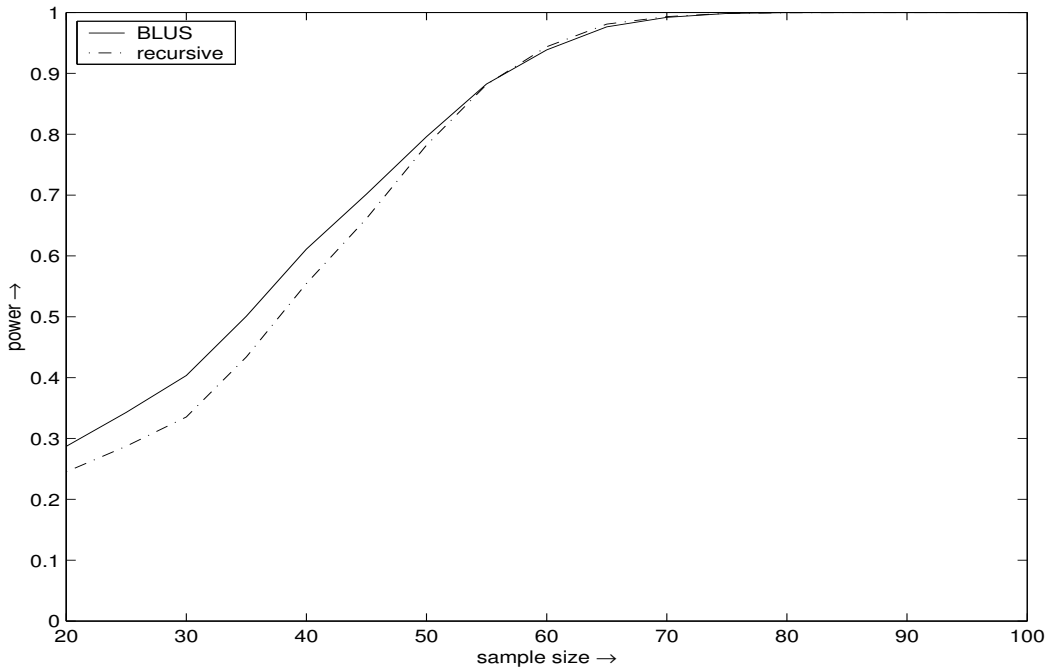

Fig. 7. BLUS and recursive residuals compared: Quandt's simulated data.