
Techniques for Testing the Constancy of Regression Relationships over Time

Author(s): R. L. Brown, J. Durbin and J. M. Evans

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 37, No. 2 (1975), pp. 149-192

Published by: Wiley for the Royal Statistical Society

Stable URL: <https://www.jstor.org/stable/2984889>

Accessed: 16-10-2019 14:40 UTC

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



Royal Statistical Society, Wiley are collaborating with JSTOR to digitize, preserve and extend access to *Journal of the Royal Statistical Society. Series B (Methodological)*

Techniques for Testing the Constancy of Regression Relationships over Time

By R. L. BROWN

J. DURBIN

and

J. M. EVANS

Central Statistical Office

*London School of Economics
and Political Science*

Central Statistical Office

[Read before the ROYAL STATISTICAL SOCIETY at a meeting organized by the RESEARCH SECTION
on Wednesday, December 4th, 1974, Professor R. L. PLACKETT in the Chair]

SUMMARY

Methods for studying the stability over time of regression relationships are considered. Recursive residuals, defined to be uncorrelated with zero means and constant variance, are introduced and tests based on the cusum and cusum of squares of recursive residuals are developed. Further techniques based on moving regressions, in which the regression model is fitted from a segment of data which is moved along the series, and on regression models whose coefficients are polynomials in time are studied. The Quandt log-likelihood ratio statistic is considered. Emphasis is placed on the use of graphical methods. The techniques proposed have been embodied in a comprehensive computer program, TIMVAR. Use of the techniques is illustrated by applying them to three sets of data.

Keywords: CUSUM; REGRESSION RESIDUALS; RECURSIVE RESIDUALS

1. INTRODUCTION

THIS paper describes and exemplifies a set of techniques for detecting departures from constancy of regression relationships over time when regression analysis is applied to time-series data. All the techniques described have been embodied in a computer program, TIMVAR. Enquiries about the availability of this program should be addressed to the Computer Development for Statistics Unit of the Central Statistical Office. A "User's Guide" to the program (Evans, 1973) is available from the Central Statistical Office. In what follows, the name TIMVAR will be used indifferently to describe either the set of methods used or the computer program written to implement them.

The theory underlying the paper was developed jointly by Brown and Durbin who gave a preliminary account of it in Brown and Durbin (1968). The original version of the program was written by C. E. Rogers and later work on it was done by R. P. Bayes and Evans. Mr Brown unfortunately died in 1972 so the actual writing of the paper was done by Durbin and Evans who accept full responsibility for the final version. However, since they have made substantial use of material left by Mr Brown they feel that he should be regarded as a co-author of the paper.

Regression analysis of time-series data is usually based on the assumption that the regression relationship is constant over time. In some applications, particularly in the social and economic fields, the validity of this assumption is open to question, and it is often desirable to examine it critically, particularly if the model is to be used for forecasting.

TIMVAR includes formal significance tests but its philosophy is basically that of data analysis as expounded by Tukey (1962). Essentially, the techniques are designed to bring out departures from constancy in a graphic way instead of parametrizing

particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives. From this point of view the significance tests suggested should be regarded as yardsticks for the interpretation of data rather than leading to hard and fast decisions.

The problem we consider is a special case of a general class of problems concerned with the detection of changes of model structure over time, but we shall not attempt to review here the extensive literature dealing with the whole range of problems. Apart from citing references which have specific relevance to our own treatment we merely call attention to two papers of special importance, those of Chernoff and Zacks (1964) and Hinkley (1972).

The next section begins by specifying the basic regression model and the null hypothesis under consideration. It goes on to show how this hypothesis can be investigated by constructing plots of cumulative sums and sums of squares of the so-called recursive residuals. These are the standardized residuals from the regression of each observation y_t , the regression coefficients being calculated from the observations y_1, \dots, y_{t-1} for $t = k+1, \dots, T$, where k is the number of regressors and T is the number of observations. It is shown that on the null hypothesis the recursive residuals are uncorrelated with zero mean and constant variance and are therefore independent under the normality assumption. Suitable formulae for carrying out the recursive calculations in a highly economical way are presented.

Other methods of transforming least-squares residuals to independent $N(0, \sigma^2)$ variables have been given by several authors including Theil (1965, 1968) and Durbin (1970). However, the recursive residuals seem preferable for detecting the change of model over time since until a change takes place the recursive residuals behave exactly as on the null hypothesis. When the change does occur, one hopes that signs of it will soon be apparent. With the other methods one would normally expect the effects of the change to be spread over the full set of transformed residuals.

In section 2.5 further techniques based on plotting the coefficients obtained by fitting the model to a segment of n successive observations and moving this segment along the series are presented. The plots are supplemented by a homogeneity test based on the analysis of variance. Section 2.6 considers the fitting and testing of time-trending regressions in which each coefficient is represented as a polynomial in time. The final technique considered is the plotting of Quandt's log-likelihood ratio statistic, intended to detect the single time-point, if any, at which there is a discontinuous change from one constant set of regression parameters to another.

In Section 3 the techniques developed are applied to three sets of real data taken from the field of economics. These examples illustrate how TIMVAR can be used in the model-building process to investigate the stability of models over time. The first example refers to data from the Post Office on the growth in the number of local telephone calls, the second uses data from the International Monetary Fund on the demand for money and the third deals with a model for forecasting manpower requirements using data provided by the Civil Service Department. Section 4 outlines the structure of the TIMVAR program and indicates the options available.

2. THE TECHNIQUES PROPOSED

2.1. *The Regression Model under Study*

The basic regression model we consider is

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_t + u_t, \quad t = 1, \dots, T, \quad (1)$$

where at time t , y_t is the observation on the dependent variable and \mathbf{x}_t is the column vector of observations on k regressors. The first regressor, x_{1t} , will be taken to equal unity for all values of t if the model contains a constant. The other regressors are assumed to be non-stochastic so auto-regressive models are excluded from consideration. The column vector of parameters, β_t , is written with the subscript t to indicate that it may vary with time. We assume that the error terms, u_t , are independent and normally distributed with means zero and variances σ_t^2 , $t = 1, \dots, T$. The hypothesis of constancy over time, which will be denoted by H_0 , is

$$\begin{aligned}\beta_1 &= \beta_2 = \dots = \beta_T = \beta, \\ \sigma_1^2 &= \sigma_2^2 = \dots = \sigma_T^2 = \sigma^2.\end{aligned}$$

We shall be more concerned with detecting differences among the β 's than among the σ 's though we do give a procedure which permits the investigation of variance changes. We have not considered the effects of serial correlation in the u 's on the performance of the tests proposed.

It is natural to look at residuals to investigate departures from model specification, and a variety of procedures for doing this have been proposed in the literature (see, for example, Anscombe, 1961; Anscombe and Tukey, 1963). However, experience has shown that in the present situation the plot of the ordinary least-squares residuals, or the plot of their squares, against time is not a very sensitive indicator of small or gradual changes in the β 's. In this respect the problem resembles that of detecting changes in the mean in industrial quality control for which the cumulative sum or cusum technique, introduced by Page (1954) and discussed further by Barnard (1959) and by Woodward and Goldsmith (1964), has been found to be a more effective tool for detecting small changes than the ordinary control chart in some circumstances.

This suggests that instead of plotting out the individual least-squares residuals z_t , the cusums $Z_r = \hat{\sigma}^{-1} \sum_1^r z_t$, $r = 1, \dots, T$, should be plotted, where we have divided by the estimated standard deviation $\hat{\sigma}$ to eliminate the irrelevant scale factor. The difficulty about this suggestion is that there seems no way of assessing the significance of the departure of the observed graph of Z_r against r from the mean-value line $E(Z_r) = 0$. The intractability of the problem arises from the fact that in general the covariance function $E(Z_r Z_s)$ does not reduce to a form that is manageable by standard Gaussian process techniques (cf. Mehr and McFadden, 1965). For instance, for the simple case of regression on a linear time trend with zero intercept, the covariance function is asymptotically $r - 3r^2 s^2 / 4T^3$ ($r < s$), which is an unmanageable form.

An alternative is to consider the standardized cusum of squares, $\hat{\sigma}^{-2} \sum_1^r z_t^2$. Although more tractable, this is still difficult to deal with. Instead of considering it we prefer to make the transformation to recursive residuals given in the following section which enables us to treat the problem in terms of standardized cusums and cusums of squares of independent $N(0, \sigma^2)$ variables.

2.2. The Recursive Residuals and their Properties

Assuming H_0 to be true, let \mathbf{b}_r be the least-squares estimate of β based on the first r observations, i.e. $\mathbf{b}_r = (\mathbf{X}'_r \mathbf{X}_r)^{-1} \mathbf{X}'_r \mathbf{Y}_r$ where the matrix $\mathbf{X}'_r \mathbf{X}_r$ is assumed to be non-singular, and let

w is a normalized residual at time r
(standardized prediction error)

$$w_r = \frac{y_r - \mathbf{x}'_r \mathbf{b}_{r-1}}{\sqrt{(1 + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r)}}, \quad r = k+1, \dots, T, \quad (2)$$

where $\mathbf{X}'_{r-1} = [\mathbf{x}_1, \dots, \mathbf{x}_{r-1}]$ and $\mathbf{Y}'_r = [y_1, \dots, y_r]$.

Lemma 1. Under H_0 , w_{k+1}, \dots, w_T are independent, $N(0, \sigma^2)$.

Proof. The unbiasedness of w_r is obvious and the assertion $V(w_r) = \sigma^2$ follows immediately from the independence of y_r and b_{r-1} . Also,

$$w_r = \left(u_r - \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \sum_{j=1}^{r-1} \mathbf{x}_j u_j \right) (1 + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r)^{-\frac{1}{2}}.$$

Since each w_r is a linear combination of the normal variates u_j , the w_j 's are jointly normally distributed. Now

$$\begin{aligned} E & \left[\left\{ u_r - \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \sum_{j=1}^{r-1} \mathbf{x}_j u_j \right\} \left\{ u_s - \mathbf{x}'_s (\mathbf{X}'_{s-1} \mathbf{X}_{s-1})^{-1} \sum_{j=1}^{s-1} \mathbf{x}_j u_j \right\} \right] \\ & = \sigma^2 [0 - 0 - \mathbf{x}'_s (\mathbf{X}'_{s-1} \mathbf{X}_{s-1}) \mathbf{x}_r + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} (\mathbf{X}'_{r-1} \mathbf{X}_{r-1}) (\mathbf{X}'_{s-1} \mathbf{X}_{s-1})^{-1} \mathbf{x}'_s] = 0 \quad (r < s). \end{aligned}$$

It follows that w_{k+1}, \dots, w_T are uncorrelated and are therefore independent in view of their joint normality. The transformation from the u_r 's to the w_r 's is a generalized form of the Helmert transformation (Kendall and Stuart, 1969, p. 250).

Let S_r be the residual sum of squares after fitting the model to the first r observations assuming H_0 true, i.e., $S_r = (\mathbf{Y}_r - \mathbf{X}_r \mathbf{b}_r)'(\mathbf{Y}_r - \mathbf{X}_r \mathbf{b}_r)$.

Lemma 2.

$$(\mathbf{X}'_r \mathbf{X}_r)^{-1} = (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} - \frac{(\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1}}{1 + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r}, \quad (3)$$

$$\mathbf{b}_r = \mathbf{b}_{r-1} + (\mathbf{X}'_r \mathbf{X}_r)^{-1} \mathbf{x}_r (y_r - \mathbf{x}'_r \mathbf{b}_{r-1}), \quad (4)$$

$$S_r = S_{r-1} + w_r^2, \quad r = k+1, \dots, T. \quad (5)$$

The relation (3) was given by Plackett (1950) and Bartlett (1951). It is used in the program to avoid having to invert the matrix $(\mathbf{X}'_r \mathbf{X}_r)$ directly at each stage of the calculations. It is proved by multiplying the left-hand side by $\mathbf{X}'_r \mathbf{X}_r$ and the right-hand side by $\mathbf{X}'_{r-1} \mathbf{X}_{r-1} + \mathbf{x}_r \mathbf{x}'_r = \mathbf{X}'_r \mathbf{X}_r$.

Proof of (4). Since b_r is the least-squares estimate it satisfies

$$\begin{aligned} \mathbf{X}'_r \mathbf{X}_r \mathbf{b}_r &= \mathbf{X}'_r \mathbf{Y}_r = \mathbf{X}'_{r-1} \mathbf{Y}_{r-1} + \mathbf{x}_r y_r = \mathbf{X}'_{r-1} \mathbf{X}_{r-1} \mathbf{b}_{r-1} + \mathbf{x}_r y_r \\ &= \mathbf{X}'_r \mathbf{X}_r \mathbf{b}_{r-1} + \mathbf{x}_r (y_r - \mathbf{x}'_r \mathbf{b}_{r-1}). \end{aligned}$$

Proof of (5).

$$\begin{aligned} S_r &= (\mathbf{Y}_r - \mathbf{X}_r \mathbf{b}_r)'(\mathbf{Y}_r - \mathbf{X}_r \mathbf{b}_r) \\ &= (\mathbf{Y}_r - \mathbf{X}_r \mathbf{b}_{r-1})'(\mathbf{Y}_r - \mathbf{X}_r \mathbf{b}_{r-1}) - (\mathbf{b}_r - \mathbf{b}_{r-1})' \mathbf{X}'_r \mathbf{X}_r (\mathbf{b}_r - \mathbf{b}_{r-1}) \\ &= S_{r-1} + (y_r - \mathbf{x}'_r \mathbf{b}_{r-1})^2 - \mathbf{x}'_r (\mathbf{X}'_r \mathbf{X}_r)^{-1} \mathbf{x}_r (y_r - \mathbf{x}'_r \mathbf{b}_{r-1})^2 \end{aligned}$$

which gives (5) on substituting for $(\mathbf{X}'_r \mathbf{X}_r)^{-1}$ from (3).

An alternative proof of (3), (4) and (5) may be derived from the results of Heyadat and Robson (1970) since their quantities f_r are multiples of our quantities w_r . Note that w_r is the standardized prediction error of y_r when predicted from y_1, \dots, y_{r-1} .

A situation arising frequently in practice is one where the regression model contains a constant and one of the regressor variables is itself constant for the first r_1 observations, where $r_1 \geq k$. Even though, in this case, the recursive residuals cannot be calculated from direct application of (2) above because of multicollinearity, it is possible to derive them in the following manner. The basic idea is to drop the initially

constant regressor (which must be supplied to the program last) at the beginning of the recursions, reducing the number of regressors to $k-1$. Then recursive residuals $w_k^0, \dots, w_{r_1}^0$ are derived from estimates $b_{k-1}^0, \dots, b_{r_1-1}^0$ of the regressor vector β^0 , the 0 denoting the fact that only $k-1$ regressors have been used. (The first component of each of these estimates will of course be an estimate of $(\beta_1 + \beta_k d)$ where d is the initial value of the last regressor.) When the last regressor has changed it is brought into the regression and recursive residuals are calculated from then on by formula (2). If, for $r = k+1, \dots, r_1$, w_r is defined to be w_{r-1}^0 then we again have a set of $T-k$ values w_{k+1}, \dots, w_r , which may be shown to be independent $N(0, \sigma^2)$ as before. When the k th regressor is brought into the regression the extra degree of freedom absorbed means that there is no increase in the residual sum of squares, i.e. $S_{r_1+1} = S_{r_1}^0$; moreover, apart from the constant term, the first $k-1$ components of $b_{r_1+1}^0$ are equal to the corresponding components of $b_{r_1}^0$. Having made the transition from $k-1$ to k regressors, the recursion proceeds as in the standard case.

2.3. The Cusum Test

If β_t is constant up to time $t = t_0$ and differs from this constant value from then on, the w_r 's will have zero means for r up to t_0 but in general will have non-zero means subsequently. This suggests examination of plots intended to reveal departures of the means of the w_r 's from zero as one travels along the series through time.

The first plot we consider is the plot of the cusum quantity

$$W_r = \frac{1}{\hat{\sigma}} \sum_{k+1}^r w_j,$$

against r for $r = k+1, \dots, T$, where $\hat{\sigma}$ denotes the estimated standard deviation determined by $\hat{\sigma}^2 = S_T/(T-k)$. We require a method of testing the significance of the departure of the sample path of W_r from its mean value line $E(W_r) = 0$. A suitable procedure is to find a pair of lines lying symmetrically above and below the line $W_r = 0$ such that the probability of crossing one or both lines is α , the required significance level.

From the properties of the w_r 's, under H_0 the sequence W_{k+1}, \dots, W_r is a sequence of approximately normal variables such that

$$E(W_r) = 0, \quad V(W_r) = r - k \quad \text{and} \quad C(W_r, W_s) = \min(r, s) - k,$$

to a good approximation. To derive the test, W_r is approximated by the continuous Gaussian process $\{Z_t, k \leq t \leq T\}$ with these mean and covariance functions. This is in fact the Brownian motion process starting from zero at time $t = k$. The form of straight line to choose was decided in two stages. The standard deviation of Z_t is $\sqrt{t-k}$. Consequently, if we wished to find a curve such that under H_0 the probability that the sample path lies above the curve at any point between $t = k$ and $t = T$ is constant, we should choose curves of the form $\pm \lambda \sqrt{t-k}$ where λ is constant. However, since we wish to limit ourselves to straight lines, the crossing probability cannot be constant for all t and the procedure adopted is to choose the family of lines tangent to the curves $\pm \lambda \sqrt{t-k}$ at the points halfway between $t = k$ and $t = T$. This leads to the family of pairs of straight lines through the points $\{k, \pm a \sqrt{T-k}\}, \{T, \pm 3a \sqrt{T-k}\}$, where a is the parameter. For any given line in this family, the probability that the point (r, W_r) lies outside the line is a maximum for r halfway between $r = k$ and $r = T$. We want to find a member of this family such that the probability that a sample path Z_t crosses it is $\frac{1}{2}\alpha$. Known results in Brownian

motion theory give for the probability that a sample path Z_t crosses the line $y = d + c(t-k)$ for some t in (k, T) the value

$$Q\left(\frac{d+c(T-k)}{\sqrt{T-k}}\right) + \exp(-2dc) Q\left(\frac{d-c(T-k)}{\sqrt{T-k}}\right),$$

where

$$Q(z) = \frac{1}{\sqrt{(2\pi)}} \int_z^\infty \exp(-\frac{1}{2}u^2) du$$

(see, for example, Durbin, 1971, Lemma 3). Substituting $d = a\sqrt{T-k}$ and $c = 2a/\sqrt{T-k}$ we obtain the equation

$$Q(3a) + \exp(-4a^2)(1 - Q(a)) = \frac{1}{2}\alpha,$$

to be solved for a .

It has been assumed that the probability that W_r crosses both lines is negligible, which will be justifiable for values of α normally used for significance testing, say 0.1 or less. Useful pairs of values of a and α are

$$\alpha = 0.01, \quad a = 1.143,$$

$$\alpha = 0.05, \quad a = 0.948,$$

$$\alpha = 0.10, \quad a = 0.850.$$

From the standpoint of data analysis, the function of these lines is to provide a yardstick against which to assess the observed behaviour of the sample path, though of course they can be used to provide a formal test of significance by rejecting if the sample path travels outside the region between the lines.

2.4. The Cusum of Squares Test

This test uses the squared recursive residuals, w_r^2 , and is based on the plot of the quantities

$$s_r = \left(\sum_{j=k+1}^r w_j^2 \right) / \left(\sum_{j=k+1}^T w_j^2 \right) = S_r/S_T, \quad r = k+1, \dots, T.$$

The test provides a useful complement to the cusum test, particularly when the departure from constancy of the β 's is haphazard rather than systematic. On H_0 , s_r may be shown to have a beta distribution with mean $(r-k)/(T-k)$. This suggests drawing a pair of lines $s_r = \pm c_0 + (r-k)/(T-k)$ on the diagram parallel to the mean-value line such that the probability that the sample path crosses one or both lines is α , the required significance level.

To find the significance values, c_0 , it is convenient to consider first the case when $T-k$ is even. Then the joint distribution of the $\frac{1}{2}(T-k)-1$ statistics $s_{k+2}, s_{k+4}, \dots, s_{T-2}$ is the same as that of an ordered sample of $\frac{1}{2}(T-k)-1$ independent observations from the uniform (0, 1) distribution. This may be shown by writing

$$n = \frac{1}{2}(T-k)-1 \quad \text{and} \quad z_j = (w_{k+2j}^2 + w_{k+2j-1}^2)/2\sigma^2, \quad j = 1, \dots, n+1.$$

Then the z_j 's are independent, exponentially distributed random variables with mean one. If Z is the sum of the z_j 's we have

$$s_{k+2j} = (z_1 + \dots + z_j)/Z, \quad j = 1, \dots, n.$$

The required result follows by transforming the variables z_1, \dots, z_{n+1} to give the joint distribution of $s_{k+2}, \dots, s_{T-2}, Z$ and then integrating out Z .

The distribution of an ordered sample of independent observations from the uniform (0, 1) distribution plays an important part in the theory of non-parametric statistics, and the distribution of each of the statistics c^+ and c^- , defined by

$$c^+ = \max_{j=1, \dots, m-1} (s_{k+2j} - j/m), \quad c^- = \max_{j=1, \dots, m-1} (j/m - s_{k+2j})$$

where $m = \frac{1}{2}(T-k)$, can be recognized as being equivalent to that of Pyke's (1959) modified Kolmogorov-Smirnov statistic, C_n^+ , with $n = m-1$. The statistics c^+ and c^- are the maximum positive and negative deviations respectively of the set of statistics $(s_{k+2}, \dots, s_{T-2})$ from their mean-value line.

A table of significance values of the quantity C_n^+ for $n = m-1$ is given by Durbin (1969) (Table 1, p. 4). The procedure suggested for the cusum of squares test is to take these values as approximations to the significance values of

$$c_1^+ = \max_{i=1, \dots, T-k-1} \left(s_{k+i} - \frac{i}{T-k} \right) \quad \text{and} \quad c_1^- = \max_{i=1, \dots, T-k-1} \left(\frac{i}{T-k} - s_{k+i} \right),$$

which are the maximum positive and negative deviations of the whole set of s_i 's from the mean-value line. For the value of the significance level, α , normally chosen, say 0·1 or less, the probability of crossing both lines is negligible, so that given a significance level α , to find the value of c_0 we may take the value obtained by entering the table at $n = \frac{1}{2}(T-k)-1$ and $\frac{1}{2}\alpha$. If $T-k$ is odd the procedure suggested is to interpolate linearly between the values for $n = \frac{1}{2}(T-k)-\frac{3}{2}$ and $n = \frac{1}{2}(T-k)-\frac{1}{2}$. Monte Carlo runs made by M. C. Hutchison have shown that this test gives significant results more often than the exact test would give, but that the discrepancy is very small when $(T-k)$ exceeds 30.

It may sometimes be appropriate to consider a one-sided test. For example, if it is assumed that $\beta_t = \beta^*$ for $t \leq r$ and $\beta_t = \beta^{**} \neq \beta^*$ for $t > r$ while $\sigma_t^2 = \sigma^2$ for all t , then $E(w_t^2) = \sigma^2$ for $t \leq r$ and $E(w_t^2) > \sigma^2$ for $t > r$. One would therefore expect the departure from the null hypothesis to be indicated by a tendency for the sample path s_r to lie below the mean value line, and would therefore use a one-sided test. For this purpose, one would take the significance value of c_0 corresponding to significance level α , not $\frac{1}{2}\alpha$. However, whether the two- or one-sided situations are envisaged we ourselves prefer to regard the lines constructed in this way as yardsticks against which to assess the observed sample path rather than providing formal tests of significance.

If the two plots described above do indicate departures from constancy it may be useful to examine plots of the components of b_r against time to try to identify the source. Further, to help locate the point of change it is often informative to look at the set of plots which are obtained by running the analysis backwards through time as well as forwards.

2.5. Moving Regressions

Another useful way of investigating the time-variation of β_t is to fit the regression on a short segment of n successive observations and to move this segment along the series. The graphs of the resulting coefficients against time provide further evidence of departures from constancy. In addition, the estimated residual variance may be computed and plotted to investigate the constancy of σ^2 .

The quantities required for each new segment are computed by first adding a new observation to the segment just dealt with using formulae (3)–(5) and then allowing for the effect of dropping an observation from the beginning by means of the following

analogues of (3)–(5):

$$\begin{aligned} (\tilde{\mathbf{X}}'_n \tilde{\mathbf{X}}_n)^{-1} &= (\mathbf{X}'_{n+1} \mathbf{X}_{n+1})^{-1} + (\mathbf{X}'_{n+1} \mathbf{X}'_{n+1})^{-1} \mathbf{x}_1 \mathbf{x}'_1 (\mathbf{X}'_{n+1})^{-1} / (1 - \mathbf{x}'_1 (\mathbf{X}'_{n+1} \mathbf{X}_{n+1})^{-1} \mathbf{x}_1), \\ \tilde{\mathbf{b}}_n &= \mathbf{b}_{n+1} - (\tilde{\mathbf{X}}'_n \tilde{\mathbf{X}}_n)^{-1} \mathbf{x}_1 (y_1 - \mathbf{x}'_1 \mathbf{b}_{n+1}), \\ \tilde{S}_n &= S_{n+1} - (y_1 - \mathbf{x}'_1 \tilde{\mathbf{b}}_n)^2 / (1 + \mathbf{x}'_1 (\tilde{\mathbf{X}}'_n \tilde{\mathbf{X}}_n)^{-1} \mathbf{x}_1), \end{aligned}$$

where $\tilde{\mathbf{X}}_n$, $\tilde{\mathbf{b}}_n$, \tilde{S}_n are the values of the regressor matrix, the coefficient vector and the residual sum of squares based on observations from $t = 2$ to $n+1$. For simplicity of notation we have given the formulae for the first update only but the further formulae are, of course, similar. Proofs are similar to those of (3)–(5).

A significance test for constancy based on this approach, called by us the homogeneity test, is derived from the results of regressions based on non-overlapping time segments, using the analysis of variance. The time segments used by the program, for a moving regression of length n , are $(1, n)$, $((n+1), 2n)$, ..., $((p-2)n+1, (p-1)n)$, $((p-1)n+1, T)$, where p is the integral part of T/n , and the variance ratio considered, called by us the homogeneity statistic, is

$$\frac{(T-kp) S(1, T) - \{S(1, n) + S(n+1, 2n) + \dots + S(pn-2n+1, pn-n) + S(pn-n+1, T)\}}{(kp-k) \{S(1, n) + S(n+1, 2n) + \dots + S(pn-n+1, T)\}},$$

where $S(r, s)$ is the residual sum of squares from the regression calculated from observations from $t = r$ to s inclusive. This is equivalent to the usual “between groups over within groups” ratio of mean squares and under H_0 is distributed as $F(kp-k, T-kp)$.

The TIMVAR program also calculates the quantity M_1 , the mean square prediction error one period ahead. This is defined as

$$M_1 = \sum_{m=n+1}^T \{y_m - \mathbf{x}'_m \mathbf{b}(m-n, m-1)\}^2 / (T-n),$$

where the vector $\mathbf{b}(m-n, m-1)$ is the estimate of the vector of regression coefficients from the time segment $(m-n, m-1)$. When calculated for moving regressions of several different lengths its minimum value gives a useful criterion for the length of record to use when predicting one period ahead. Also calculated are M_2 , defined by

$$M_2 = \sum_{m=1}^{T-n} \{y_m - \mathbf{x}'_m \mathbf{b}(m+1, m+n)\}^2 / (T-n),$$

which is equivalent to M_1 calculated from a moving regression passed in the reverse direction, and M the sum of M_1 and M_2 . Of these M_1 will normally be the most useful. Finally, M_3 is calculated. This is $\sum_{m=n_1+n}^T (y_m - \mathbf{x}'_m \mathbf{b}(m-n, m-1))^2 / (T-n_1)$ where n_1 is the maximum length of regression considered. This fulfils the same function as M_1 , except that the different regression lengths are now compared over the same part of the record, namely that from n_1 to T .

2.6. Time-trending Regressions

This technique introduces time variation into the regression model explicitly by allowing the regression coefficients to become polynomials in time. To determine whether this extended model will produce a significantly better fit than one based on constancy, and further to determine what degree of polynomial should be employed, the program calculates the sum of squares removed by each of the following nested

hypotheses:

$$(0): \quad y_t = \mathbf{x}'_t \boldsymbol{\beta}_{(0)} + \varepsilon_t$$

$$(1): \quad y_t = \mathbf{x}'_t (\boldsymbol{\beta}_{(0)} + \boldsymbol{\beta}_{(1)} t) + \varepsilon_t$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$(e): \quad y_t = \mathbf{x}'_t (\boldsymbol{\beta}_{(0)} + \boldsymbol{\beta}_{(1)} t + \dots + \boldsymbol{\beta}_{(e)} t^e) + \varepsilon_t.$$

The $\boldsymbol{\beta}$'s are all vectors of length k , and e is a positive integer specified by the user.

Comparison of the mean-square increase in the explained variation with an estimate of the error variance gives a test for determining whether each model gives a significantly better fit than the one before. This estimate of the error variance may be derived from either the residual sum of squares from the model in the next higher degree in t or from the residual sum of square of the full model (e) and so two F-ratios are calculated, one for each estimate.

2.7. Quandt's Log-likelihood Ratio Technique

This technique, described in two papers by Quandt (1958, 1960), is appropriate when it is believed that the regression relationship may have changed abruptly at an unknown time point $t = r$ from one constant relationship specified by $\boldsymbol{\beta}^{(1)}$, σ_1^2 , to another constant relationship specified by $\boldsymbol{\beta}^{(2)}$, σ_2^2 . For each r from $r = k+1$ to $r = T-k-1$ the program computes and plots

$$\lambda_r = \log_{10} \left(\frac{\text{max likelihood of the observations given } H_0}{\text{max likelihood of the observations given } H_1} \right),$$

where H_1 is the hypothesis that the observations in the time segments $(1, \dots, r)$ and $(r+1, \dots, T)$ come from two different regressions. This is the standard likelihood ratio statistic for deciding between the two hypotheses H_0 and H_1 , and it is easy to show that

$$\lambda_r = \frac{1}{2}r \log \hat{\sigma}_1^2 + \frac{1}{2}(T-r) \log \hat{\sigma}_2^2 - \frac{1}{2}T \log \hat{\sigma}^2,$$

where $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$ and $\hat{\sigma}^2$ are the ratios of the residual sums of squares to number of observations when the regression is fitted to the first r observations, the remaining $T-r$ observations and the whole set of T observations, respectively. The estimate of the point at which the switch from one relationship to another has occurred is then the value of r at which λ_r attains its minimum. Unfortunately, no test has yet been devised for $\min \lambda_r$ since its distribution on H_0 is unknown. However, the behaviour of the graph of λ_r against r sheds light on the stability of the regression and in particular indicates whether changes have occurred as an abrupt transition or gradually.

3. EXAMPLES OF THE APPLICATION OF TIMVAR

In this section we present three examples which illustrate the use of TIMVAR techniques. The first and third examples reveal evidence of change while the second does not. The graphs have been chosen to illustrate different kinds of TIMVAR output but in each case the graphs shown are only a small fraction of the total output available.

Example 1. This example was made available by the Statistics and Business Research Department of the Post Office. As part of a wider study of posts and telecommunications described in Turner (1973), a regression model was developed to

explain growth in the number of local telephone calls (i.e. the differences between the numbers of calls in consecutive years) in terms of a linear model involving a constant and four independent variables. These four variables, which were used in first difference forms, were a measure of economic activity, the number of residential telephones, the "real" price of local calls and the "real" price of residential telephones. (The deflator used to arrive at the "real" prices was the retail price index.) The data ran from 1950/51 to 1971/72, but it had been felt that the estimates of the number of local calls and hence of local call growth were subject to some uncertainty after 1964/65; however, the 1971/72 figure was thought to be reliable. In order to use this model as part of a larger model, reliable estimates for the growth of local calls were required for the whole period. Thus it was important that the stability of the relationship over time should be investigated.

Figs 1–4 give respectively the plots of the least-squares residuals, the cusum of least-squares residuals, the cusum of recursive residuals and the cusum of squares of

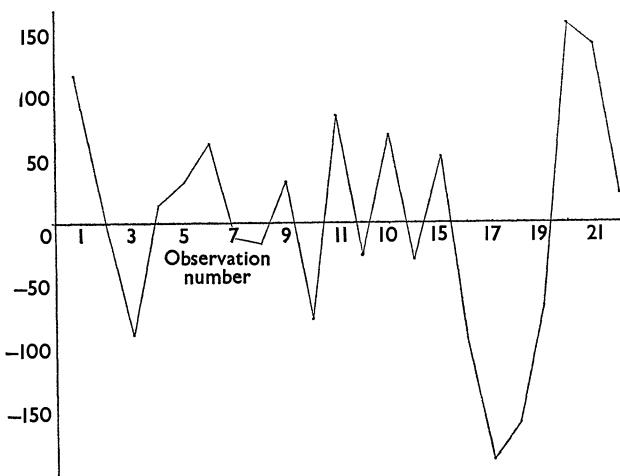


FIG. 1. *Example 1:* Ordinary least-square residuals.

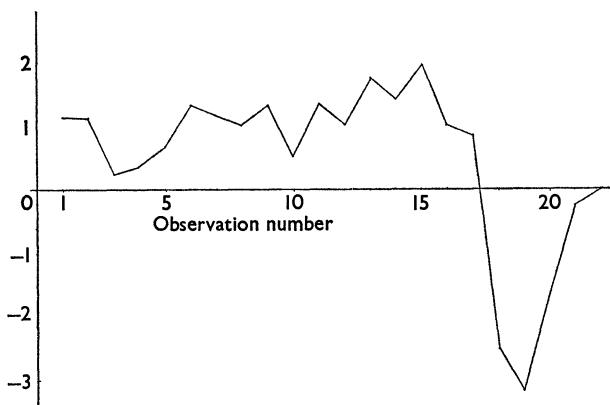


FIG. 2. *Example 1:* Cusum of ordinary least-squares residuals.

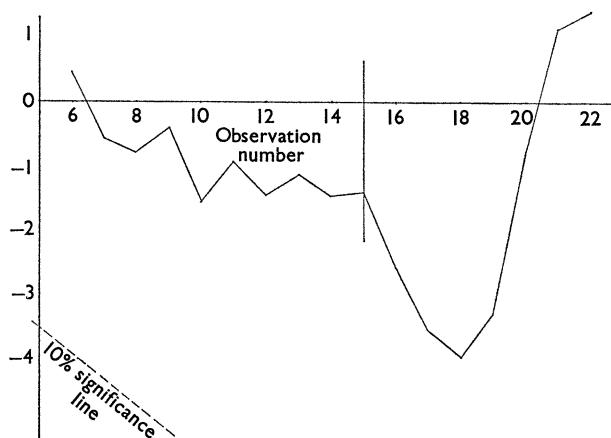


FIG. 3. Example 1: Cusum of recursive residuals, forward recursion.

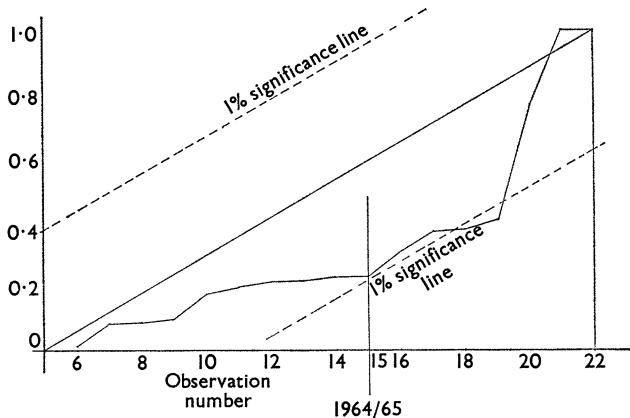


FIG. 4. Example 1: Cusum of squares of recursive residuals, forward recursion.

least-squares residuals. These plots provide increasing indication of evidence of instability after 1964/65, 1 per cent significance being attained in Fig. 4. The fact that significance is achieved by the cusum of squares plot but not by the cusum of recursive residuals suggests that instability may be due to a shift in residual variance than to shifts in values of regression coefficients. However, examination of the plots of coefficients and residual variance estimated from moving regressions showed that the instability was due to local changes in the regression coefficients and not to changes in variance. The model shows no sign of instability in the years up to 1964/65 and it was further found that a forecast of the 1971/72 figure from the model fitted to the data up to 1964/65 was very close to the actual 1971/72 estimate, which had been accepted as reliable. In the circumstances it was decided to ignore the suspect estimates between 1964/65 and 1971/72, to use the model fitted from the remaining data to provide the explanatory equation required and to replace the discarded data by forecasts derived from it.

Example 2. The second example, using data made available by Dr M. S. Khan of the International Monetary Fund, is based on a study of the demand for money

function for the United States, 1901–65 in Khan (1974). In this paper, Khan considers several possible specifications for the function and uses TIMVAR tests to investigate their stability over time. He argues that the question of stability of the function over time is of crucial importance for the effectiveness of monetary policy. The particular model considered here expresses the “narrow” real *per capita* stock of money M_t in terms of the long-term interest rate R_t and the permanent real *per capita* income Y_t in an equation of the form

$$\Delta \log M_t = \alpha + \beta \log \Delta R_t + \gamma \log \Delta Y_t + w_t,$$

where Δ is the first difference operator and w_t is an error term with the property $w_t \sim NID(0, \sigma_w^2)$. This is the one of eight specifications tested in the paper using annual data from 1901 to 1965.

None of the TIMVAR results were significant at 5 per cent. Figs 5 and 6 show the cusum and cusum of squares graphs from the forward recursion. The results are clearly consistent with the hypothesis of stability over time. In his paper Khan goes on to draw conclusions from the results for this and the other model specifications.

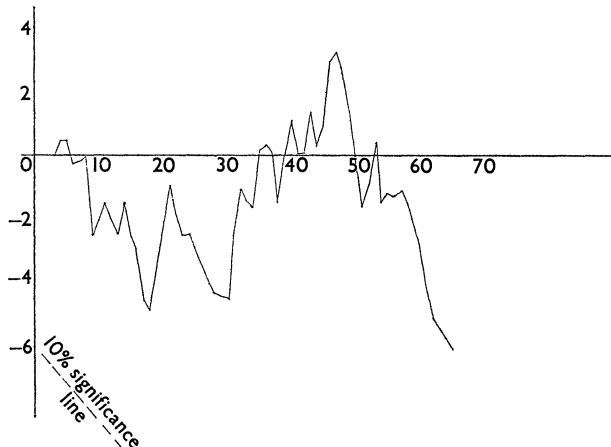


FIG. 5. Example 2: Cusum of recursive residuals, forward recursion.

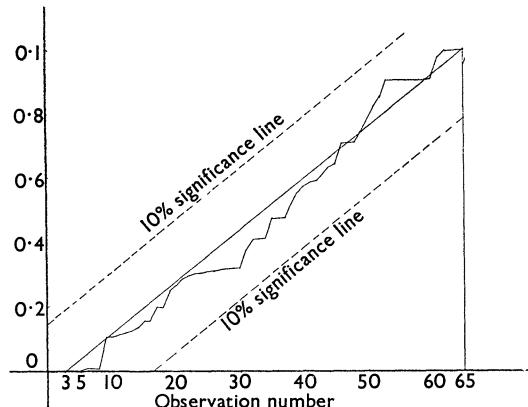


FIG. 6. Example 2: Cusum of squares of recursive residuals, forward recursion.

Example 3. This example uses data provided by the Civil Service Department and is concerned with the staff requirement S_t of an organization expressed as a function of workloads of 9 different categories. The example is studied in detail in Cameron and Nash (1974) and uses quarterly data from the first period of 1960 to the third quarter of 1970 (43 observations). Cameron and Nash found that the 9 workload categories were highly intercorrelated so they employed factor analysis to reduce them to three uncorrelated factors F_1, F_2, F_3 . They then fitted the regression model:

$$S_t = \beta_0 + \sum_{j=1}^3 \beta_j F_{jt} + e_t,$$

where e_t is a disturbance term.

Forward and backward cusum and cusum squared recursive residuals plots showed strong evidence of instability. Fig. 7 gives the graph of Quandt's log likelihood ratio and this indicates clearly that an abrupt change took place just after the 27th quarter.

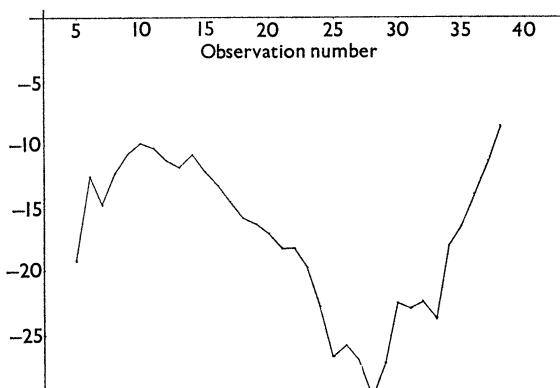


FIG. 7. *Example 3:* Quandt's log-likelihood ratio.

In fact two separate bodies were amalgamated during this quarter to form the organization under study so there was an obvious administrative explanation in this case for the results observed. However, the example serves to illustrate the way in which the Quandt log-likelihood ratio can serve to pinpoint a change in the relation. Fig. 8 gives the graph of estimates of β_3 calculated from segments of length six quarters moved along the series. These graphs show the consequences of the abrupt change in the relation just after the 27th quarter. As a result of these and other tests Cameron and Nash decided to fit the model for forecasting purposes from the last 14 observations only.

4. THE TIMVAR PROGRAM

The TIMVAR program was written to calculate all the results necessary for the tests and techniques described in the paper and to produce a considerable amount of graphical output to aid the interpretation of the results. Any or all of the following results are available:

- (a) The standard regression. The analysis of variance over the whole time span and the DW statistic. Tables and plots of the residuals and regression estimates.

- (b) The results of the time-trending regressions (see Section 2.6).
- (c) The results of the recursive regressions, both forward and backward, giving tables and plots of the successive regression estimates, the recursive residuals and their cusums, and the test statistics (see Sections 2.2, 2.3, 2.4).
- (d) The values of Quandt's log-likelihood ratio (see Section 2.7).
- (e) The results of the moving regressions for each specified length giving tables and plots of the successive regression estimates, mean square errors, the quantities M_1 , M_2 , M and M_3 , and the statistic for the homogeneity test (see Section 2.5).

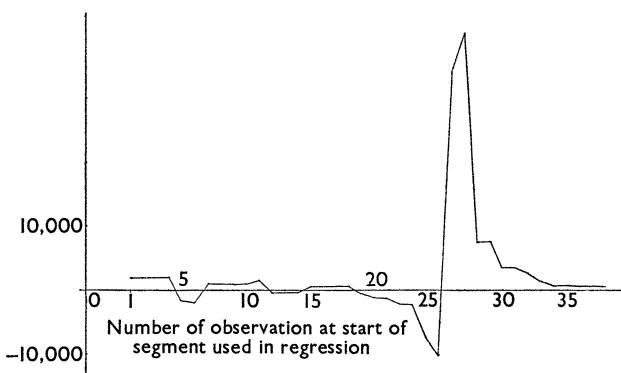


FIG. 8. Example 3: Estimate of coefficient of third independent variable derived from moving regression of length 6.

The program makes full use of the formulae described in Sections 2.2 and 2.5 during the calculations of the recursive and moving regressions. Because of the large number of successive matrix operations performed during these calculations there is a danger that some of the matrices may become ill-conditioned. Any such tendency is usually reduced by subtracting the means from each of the variables and this is done automatically by the program if the model contains a constant. The value of the constant term is then recovered by another mechanism. In the case where the model contains a constant and one of the other regressors is constant at the beginning or end of the record for a number of observations greater than k , this regressor, if supplied last, is dealt with by the program during the recursive regressions in the manner described in Section 2.2. If it is constant at the start only it can be dealt with in a similar fashion during the moving regressions. The extension to the case where two or more regressors are constant at the beginning or end of the record has not been programmed.

REFERENCES

- ANScombe, F. J. (1961). Examination of residuals. *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, 1, 1–36.
 ANSCOMBE, F. J. and TUKEY, J. W. (1963). The examination and analysis of residuals. *Technometrics*, 5, 141–160.
 BARNARD, G. A. (1959). Control charts and stochastic processes. *J. R. Statist. Soc. B*, 21, 239–271.
 BARTLETT, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. *Ann. Math. Statist.*, 22, 107–111.

- BROWN, R. L. and DURBIN, J. (1968). Methods of investigating whether a regression relationship is constant over time. *Selected Statistical Papers, European Meeting*, Mathematical Centre Tracts No. 26, Amsterdam.
- CAMERON, M. H. and NASH, J. E. (1974). On forecasting the manpower requirements of an organization with homogeneous workloads. *J. R. Statist. Soc. A*, **137**, 200–218.
- CHERNOFF, H. and ZACKS, S. (1964). Estimating the current mean of a normal distribution which is subjected to changes in time. *Ann. Math. Statist.*, **35**, 999–1018.
- DURBIN, J. (1969). Tests for serial correlation in regression analysis based on the periodogram of least squares residuals. *Biometrika*, **56**, 1–15.
- (1970). An alternative to the bound test for testing serial correlation in least-squares regression. *Econometrica*, **38**, 422–429.
- (1971). Boundary-crossing probabilities for the Brownian motion and Poisson processes and techniques for computing the power of the Kolmogorov–Smirnov test. *J. Appl. Prob.*, **8**, 431–453.
- EVANS, J. M. (1973). User's Guide to TIMVAR. Research Exercise Note 10/73 of the Central Statistical Office.
- HEYADAT, A. and ROBSON, D. S. (1970). Independent stepwise residuals for testing homoscedasticity. *J. Amer. Statist. Ass.*, **65**, 1573–1581.
- HINKLEY, D. V. (1972). Time-ordered classification. *Biometrika*, **59**, 509–523.
- KENDALL, M. G. and STUART, A. (1969). *The Advanced Theory of Statistics*, Vol. I, 3rd ed. London: Griffin.
- KHAN, M. S. (1974). The stability of the demand for money function in the United States, 1901–65. *J. Polit. Econ.*, **82**, 1205–1219.
- MEHR, C. B. and McFADDEN, J. A. (1965). Certain properties of Gaussian processes and their first passage times. *J. R. Statist. Soc. B*, **27**, 505–522.
- PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika*, **41**, 100–114.
- PLACKETT, R. L. (1950). Some theorems in least squares. *Biometrika*, **37**, 149–157.
- PYKE, R. (1959). The supremum and the infimum of the Poisson process. *Ann. Math. Statist.*, **30**, 569–576.
- QUANDT, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *J. Amer. Statist. Ass.*, **53**, 873–880.
- (1960). Tests of the hypothesis that a linear regression system obeys two separate regimes. *J. Amer. Statist. Ass.*, **55**, 324–330.
- THEIL, H. (1965). The analysis of disturbances in regression analysis. *J. Amer. Statist. Ass.*, **60**, 1067–1079.
- (1968). A simplification of the BLUS procedure for analysing regression disturbances. *J. Amer. Statist. Ass.*, **63**, 242–251.
- TUKEY, J. W. (1962). The future of data analysis. *Ann. Math. Statist.*, **33**, 1–67.
- TURNER, W. M. (1973). A critical reappraisal of the interaction of posts and telecommunications. Report No. 29, Statistics and Business Research Department, Post Office.
- WOODWARD, R. H. and GOLDSMITH, P. L. (1964). *Cumulative Sum Techniques*. Monograph No. 3, ICI Series on Mathematical and Statistical Techniques for Industry, Edinburgh: Oliver & Boyd.

DISCUSSION OF THE PAPER BY DR BROWN, PROFESSOR DURBIN AND MR EVANS

Professor D. R. Cox (Imperial College): This is an important and interesting paper; it deals with a common practical problem, gives valuable new methods and their theory and concludes with cogent illustrations.

My comments concern two theoretical aspects of the paper. First there is the efficiency of the procedures in idealized situations which, notwithstanding the remarks in Section 1 of the paper, seems of some interest in understanding the applicability of the methods. I shall deal only with the very simplest situation, in particular where the fitted model contains just a constant term, so that the recursive residuals are defined by the standard Helmert transformation and are

$$\frac{y_2 - y_1}{\sqrt{2}}, \frac{2y_3 - y_2 - y_1}{\sqrt{6}}, \dots$$

Suppose further that $E(y_i) = \mu$ ($i = 1, \dots, m$), $E(y_{m+j}) = \mu + \delta$ ($j = 1, \dots, n$).

The simplest analogue of the procedures of the paper is to use the cumulative sum of the last n recursive residuals as a test statistic for $\delta = 0$, the change-point being regarded as known and the amount of data fixed, both assumptions in contrast with the situations contemplated by the authors. Both for the cumulative sum statistic and for the efficient test, the ratio of the expectation of the test statistic to its standard deviation can be found. The values are respectively

$$\frac{\delta}{\sigma \sqrt{n}} \sum_{s=1}^n \{(m+s-1)(m+s)\}^{-\frac{1}{2}}$$

and

$$\frac{\delta}{\sigma} \left(\frac{mn}{m+n} \right)^{\frac{1}{2}}.$$

Thus the ratio can be found, simple results emerging in the limit $m, n \rightarrow \infty, n/m = k$. The relative efficiency is somewhat conventionally defined as the square of the ratio and asymptotically this is

$$(1+k) \{ \log(1+k) \}^2 k^{-2}.$$

This is close to 1 except for large k , some representative values being

k	1	2	5	10
ARE	0.961	0.905	0.770	0.632

For small k , the asymptotic relative efficiency is $1 - k^2/12 + O(k^3)$. Very similar results hold for small m, n . Thus except in the untypical case when the discontinuity emerges relatively close to the start of the data, the cumulative sum method is very efficient.

A similar calculation can be made for the departure $E(y_{m+j}) = \mu + j\beta$ ($j = 1, \dots, n$) comparing the efficient test both with the cumulative sum of the last n recursive residuals and with their regression coefficient on time. There is no difficulty in principle in making similar calculations for more general models.

A second theoretical point concerns the effect of serial correlation. It would, of course, be possible to define recursive residuals relative to an assumed or estimated covariance matrix; a simpler possibility is to keep to the definitions of the paper and to examine the effect of serial correlation in the data. While general formulae can be written down, I have investigated again only the case where the fitted model contains just an unknown mean. It can then be shown that for large m

$$\text{var}(w_m) = \sigma^2 \{1 + O(m^{-1})\}, \quad \text{corr}(w_m, w_{m+h}) \sim \rho_h,$$

where $\{\rho_h\}$ is the autocorrelation function of the data. This suggests that the standard deviation of sums of w 's is inflated over its value in the independent case by the factor

$$\left(1 + 2 \sum_{h=1}^{\infty} \rho_h\right)^{\frac{1}{2}};$$

in particular, for a first-order autoregressive process this factor is $\{(1+\rho_1)/(1-\rho_1)\}^{\frac{1}{2}}$. Provided this holds also for more general models, it would be reasonable to inflate the limits of the paper by a rough estimate of this factor.

I propose a most cordial vote of thanks to the authors.

Mr P. R. FISK (University of Edinburgh): I should like to start by paying homage to Dr R. L. Brown who, as Professor Durbin has said, was the first Director of the Research and Special Studies Unit in the Central Statistical Office. The CSO was very fortunate to have him in that position because during his period of office he demonstrated what the unit was capable of doing. His untimely death was regretted by all who knew him. It is noteworthy that this evening's paper is the first to be read before the Society from that unit. It is my earnest hope that we will receive more contributions, either of read or published papers, on methodology from this or other sources in the Government statistical service.

A basic feature of the procedures given in the paper is the set of recursive residuals described in equation (2). Various properties of these have been mentioned by the authors. One notable feature mentioned in the paper is that under the assumptions made about the terms in equation (1) the recursive residuals have zero correlation between any pair. It may prove of some interest to give a little attention to the nature of the transformation involved. I always tend to operate here in terms of matrix transformations, which is implicit in the paper but not spelt out.

As Professor Durbin said, the transformation is clearly linear from a T -dimensional space to a $(T-k)$ -dimensional space. It is possible to write the transformation matrix in such a way that it is orthogonal to the regressor matrix \mathbf{X}_T . This is sufficient to show that the recursive residuals may be described in terms of the same transformation matrix applied to the vector of dependent variables, or to the vector of least-squares residuals, or to the vector of errors in the equation provided the assumptions underlying the model in equation (1) are correct. This variety of ways in which the recursive residuals may be described makes it conceptually possible to describe the nature of the distribution of the recursive residuals, and so of statistics based on those residuals, under alternative assumptions such as heteroscedasticity of the errors.

The transformation matrix mentioned above has some nice properties. One that is implied in the paper is that when post-multiplied by its transpose we get an identity matrix. When pre-multiplied by its transpose we get the idempotent matrix used in the definition of least-squares residuals. The structure of the transformation matrix makes the interpretation of the statistic s_r in the paper as a residual sum of squares obvious at a glance.

Interest in the transformation matrix does not stop there. This matrix, of order $(T-k) \times T$, may be partitioned by the first k columns. The remaining $(T-k)$ columns form a square matrix, \mathbf{D} say, which is lower triangular and satisfies the equation

$$\mathbf{D}[\mathbf{I} + \mathbf{Z}\mathbf{Z}'] \mathbf{D}' = \mathbf{I},$$

where \mathbf{I} is an identity matrix of order $(T-k)$ and $\mathbf{Z} = \mathbf{X}_{T-k}' \mathbf{X}_k^{-1}$. Here I have partitioned \mathbf{X}_T' as $(\mathbf{X}_k' : \mathbf{X}_{T-k}')$ in which \mathbf{X}_k must be non-singular for recursive residuals to be derivable at all. It is evident from the properties of the transformation matrix mentioned above that recursive residuals are members of the Theil system of residual transformations. The distinction is that whereas Theil, and those who follow his particular approach, have used the spectral resolution of the matrix $(\mathbf{I} + \mathbf{Z}\mathbf{Z}')$, the authors of the present paper have used the Choleski factorization. I have been told, although I have seen no published demonstration, that the Householder transformation recommended by Golub yields transformed residuals which are also members of the Theil system of transformations. This leads me to wonder whether when one attempts any linear transformation of the least-squares residuals to a set of uncorrelated random variables we are perhaps producing just another member of the Theil system.

One noticeable feature of the recursive residual transformation is that it is not unique. The rows of \mathbf{X}_T can be arranged in any order we please, so long as the first k rows given by \mathbf{X}_k forms a non-singular matrix. The authors have a natural order in their examples which is induced by time, but there is no reason why they should be inhibited from trying some other order. Thus, before the analysis was conducted there was a suspicion in Example 1 that 1971/72 was more like the earlier periods than the period after 1964/65. That particular observation could have been inserted between those for 1964/65 and 1965/66 for the purpose of the test applied.

I have not looked very closely at the moving regressions described in the paper. I confess that I find the procedures appealing, possibly because I have used similar techniques without any attempt at a theoretical justification. I was examining the errors in preliminary estimates of economic time series, defined as the difference between the first published figure of the value of the series and the figure published three years hence. I was interested in detecting any marked changes in the values of variances and first-order autoregression coefficients over the length of the observed record. I think this kind of non-stationarity

may be expected with such error series and might be found with other types of economic time series also. Simple procedures, such as moving regressions, are of benefit in searching for marked changes, although I confess that I had some difficulty in interpreting the meaning of the charts that I constructed. This was mainly because I could recognize that the patterns observed could have been induced by changes in the series other than the one of prime interest. I experienced a similar feeling when looking at the authors' examples.

The authors are to be congratulated on their very interesting paper.

I am very pleased to second the vote of thanks.

The vote of thanks was passed by acclamation.

Sir MAURICE KENDALL (World Fertility Survey): I fully endorse what previous speakers have said about the merits of this paper. It is the most important contribution to regression problems that we have had for some time. I should like to ask three questions about the paper itself and make three suggestions for further work.

The procedure suggested for computing successive residuals relies on the result given by our Chairman and Professor Bartlett enabling the covariance matrix of regressors to be up-dated when a new observation becomes available. If this were not so the arithmetic would be very tedious. However, the up-dating does require matrix multiplication and if it is done over a fairly long sequence there are dangers of cumulative rounding-off errors. My inclination would be, having arrived at the end of a series, to recalculate the covariance matrix and to check whether the iterative process has arrived at the correct result. The point applies equally and possibly more strongly to moving regressions.

A point which was not entirely clear to me was how the authors distinguish between departures from the null hypothesis concerning the regression coefficients and those concerning the magnitude of the residual errors. I should be glad to know how they decide between one or the other explanation of a significant result.

A third point on which I should value the authors' comments concerns the application of their technique to data which do not have a temporal order. So far as I can see, one could use their method for data arising in any order, but if the order were to be determined by reference to the values of the regressor variables, which of them should one choose?

I think further examination is required of Quandt's method of testing where a regression changes routine. Maximizing likelihood has the disadvantage that maxima are flat, by which I mean that much the same maximum value is reached for a fairly wide range of the variables. In certain cases, as for example when a metal changes molecular shape, it is important to narrow down the point of change with great accuracy and some further research on this subject is desirable.

Finally, two directions in which an extension of the work in this paper would be very valuable. One would be an application to autoregressive series. The other would be an extension to the case where all variables are subject to error. Both cases are more likely to arise in practice than the one of pure regression on non-autocorrelated regressors.

Professor M. B. PRIESTLEY (University of Manchester Institute of Science and Technology): The problem discussed by the authors is indeed an interesting one, and they have presented us with a rich variety of techniques for its solution. Essentially, what they seem to be saying is that in the real world one would expect relationships between variables to be "dynamic" rather than "static"; hence simple approximations (such as "static" linear models) cannot be expected to remain valid over indefinitely long periods of time unless we are prepared to modify continually the values of the parameters so as to allow these models to adapt themselves to "local" conditions. The proposed techniques for monitoring and testing changes in the parameter values should therefore prove extremely useful in many fields of application.

However, although the authors assume at the outset that the x variables are deterministic (so that, presumably, model (1) may be treated within the framework of classical

regression theory), this would seem to be a rather unnatural assumption—particularly in the context of the examples discussed in Section 3. Indeed, the independent variables in Examples 1 and 2 seem just as “stochastic” as the dependent variable! Moreover, although the *dynamic* element is incorporated via the possible time dependence of the parameter, it would seem more natural to express this notion in the more conventional manner by introducing lagged terms into the regression relationships. (These would allow, for example, for the effects of “inertia” in the interrelationships between the variables.) If we now combine both these ideas we are led to a more general form of equation (1), namely,

$$y_t = \sum_{j=1}^k \sum_{s=0}^{\infty} \beta_{s,t}^{(j)} x_{t-s}^{(j)} + u_t. \quad (*)$$

Equations of the form (*) are familiar in the study of relationships between time series, and arise, for example, in the context of linear stochastic control systems. Model (1) reduces to a special case of (*) in which only $\beta_{0,t}^{(j)}$ is non-zero, all the remaining β 's being zero.

Of course, if only finitely many of the $\beta_{s,t}^{(j)}$ are non-zero, (*) may still be expressed in the form (1) where now the $x_{t-s}^{(j)}$ are simply regarded as additional variables. However, this would, in general, lead to a model with an extremely large number of parameters, and if the $x_t^{(j)}$ are regarded as stationary processes, the serial correlation within each process and the cross-correlation between $x_t^{(j)}, x_s^{(j)}$, would no doubt invalidate the distribution theory of the various test statistics. More importantly, if the original relationship between y_t and the $\{x_t^{(j)}\}$ involved lagged values of $\{y_t\}$, the formulation (*) would require an infinite set of the parameters $\{\beta_{s,t}^{(j)}\}$. For these reasons it is usually more convenient to transform (*) into its equivalent frequency domain representation. If we assume, for the moment, that the $\beta_{s,t}^{(j)}$ are time invariant, i.e. $\beta_{s,t}^{(j)} = \beta_s^{(j)}$, all t , then the inter-relationships between y_t and $\{x_t^{(j)}\}$ are completely characterized by the *sequence of transfer functions*,

$$\beta^{(j)}(\omega) = \sum_{s=0}^{\infty} \beta_s^{(j)} e^{-iws}, \quad j = 1, 2, \dots$$

If each $\beta^{(j)}(\omega)$ is a sufficiently “smooth” function of ω it is then possible to estimate all these functions “non-parametrically”, i.e. without making any specific assumptions about the form of the $\{\beta_s^{(j)}\}$, and this, in turn, provides estimates of the complete set of parameters $\{\beta^{(j)}\}$. It is well known, for example, that if there is no cross-relation between the $\{x_t^{(j)}\}$ processes, then the least-squares estimate of $\beta^{(j)}(\omega)$ is (asymptotically) given by

$$\hat{\beta}^{(j)}(\omega) = \hat{f}_{y,j}(\omega)/\hat{f}_{x,j}(\omega),$$

where $\hat{f}_{y,j}(\omega)$ is the estimated cross-spectral density function between y_t and $\{x_t^{(j)}\}$ and $\hat{f}_{x,j}(\omega)$ is the estimated spectral density function of $x_t^{(j)}$. When the $\beta_s^{(j)}$ are time-dependent y_t is, in general, no longer a stationary process, but the essential point is that *the same basic ideas may still be used to obtain a frequency domain description of (*) even in this more general case*. The transfer functions themselves now become time-dependent, i.e. in place of the $\{\beta^{(j)}(\omega)\}$ we now have the “generalized transfer functions”

$$\beta_t^{(j)}(\omega) = \sum_{s=0}^{\infty} \beta_{s,t}^{(j)} e^{-iws},$$

but, assuming that the $\beta_{s,t}^{(j)}$ do not vary “too rapidly” over time, the $\beta_t^{(j)}(\omega)$ may still be estimated “non-parametrically” by introducing the notion of *evolutionary (i.e. time-dependent) spectra and cross-spectra*. This approach has been studied by myself and my colleagues, Dr Subba Rao and Dr Tong, and the main ideas were reported in Priestley (1965) and Priestley and Tong (1973). Moreover, although models of the form (*) are more complicated than (1) in that they involve “lagged” regressor variables, *it is still possible to construct tests for the constancy over time of the complete sequence of transfer functions, $\{\beta_t^{(j)}(\omega)\}$, using a MANOVA approach*. Such tests have been applied and tested on “real data”, and were described in recent published papers by Subba Rao and Tong (1972, 1973). In a very loose sense the ideas underlying this approach are not unlike those

discussed in Section 2.5 on "moving regressions", but its advantage is that it allows one to examine possible variations over time of the *complete form* of each of the transfer functions, $\beta_t^{(j)}(\omega)$. (The authors' model (1) assumes, in effect, that each $\beta_t^{(j)}(\omega)$ is a constant, independent of ω .)

The more general approach described above does, of course, require the availability of fairly long series of observations on y_t and the $\{x_t^{(j)}\}$, but this requirement is surely inherent in the basic nature of the problem, irrespective of the approach adopted for its solution. There is, after all, a limit to the amount of information which one can extract from a finite amount of data; once the parameters are allowed to become time-dependent the accuracy of the estimators is related in quite a fundamental way to the maximum rate at which the parameters can be allowed to change—cf. the "Uncertainty Principle" (Priestley, 1965). It may well prove to be false economy to try to extract additional information from the data by over-simplifying the model.

It is well known (for example, Rosenbrock, 1965), that the recursive relations for updating regression coefficients—due to Plackett and Bartlett and mentioned in Section 2.2—are very closely related to the recursive relations which arise in Kalman filtering theory, and it may be interesting, therefore, to investigate whether the (now massive) literature on Kalman filtering could be exploited to provide further results on the properties of the recursive residuals, $\{w_t\}$. It may be noted also that both Harrison and Stevens (1971) and Bohlin (1968) have considered models of the form (*) in which the coefficients $\beta_{s,t}^{(j)}$ are themselves *stochastic processes*, and Kalman filtering then emerges almost inevitably as the appropriate technique for updating (or more precisely in this case, *predicting*) the coefficients in such models. Models involving "stochastic parameters", although not conceived in a true Bayesian spirit, may perhaps be regarded as a first step in this direction.

The points mentioned in the preceding paragraph illustrate the close relationship which exists between certain branches of stochastic control theory and problems in time series analysis. Unfortunately, much of the control theory literature tends to be written in a language and style which, at first sight, may seem unfamiliar to statisticians, and this may account for the fact that many statisticians are unaware of the full potentialities of this work. It would be to the mutual advantage of both control theorists and statisticians to foster closer collaboration between workers in these two fields. On the other hand, it would be wholly regrettable if, by adopting a parochial and introverted approach to their work, statisticians failed to appreciate the relevance and importance of much of the recent control theory research.

Dr PETER C. YOUNG (Centre for Resource and Environmental Studies, Australian National University, Canberra†): It has been suggested (Kailath, 1974), although without specific reference, that lying somewhere in the Collected Works of C. F. Gauss there is a recursive formulation of the least-squares equations. In recent years, however, there can be no doubt that the development of the recursive least-squares algorithm is due, in large part, to the work of our Chairman tonight, R. L. Plackett, who published an important paper on the subject in 1950. Indeed Professor Plackett's paper was to me, as a very young research worker in the early nineteen sixties, a great revelation and it had an important influence on my future work; an influence for which I am extremely grateful and for which I am, after more than a decade, now able to thank him personally.

But like most works of innovative quality Professor Plackett's paper was, in many ways, ahead of its time and its significance was rather lost on the pre-computer statistical audience of the day. Indeed it was left to a control theorist, Rudolf Kalman, to continue the saga of recursive least squares in 1960 when he published his influential paper on state variable estimation theory. In effect, Kalman utilized the principle of orthogonal projection to evolve a more general form of the recursive least-squares equations for the case where

† Formerly at Control Division, Department of Engineering, University of Cambridge.

the unknown *parameters* are no longer considered constant coefficients (as assumed implicitly in Professor Plackett's formulation) but are treated as time variable *states* of a dynamic system described by a set of linear stochastic state equations of a Gauss-Markov type.

Since 1960 many papers have appeared in the control literature on both recursive least squares and state variable estimation (or filtering as it is referred to in the literature), certainly too many to discuss here: it will suffice merely to mention a recent survey paper on filtering theory by Kailath (1974), which lists 390 references from both the control and statistical literature and provides an excellent appreciation of the subject, albeit somewhat orientated towards the information and communication theory audience to whom it was principally directed. It is a little surprising, however, that the statistical literature has not been similarly influenced by the early work on recursive least-squares methods, particularly now that the availability of electronic computers makes the recursive formulation so attractive in practical terms. With this in mind, it is especially welcome to hear a paper read at the Royal Statistical Society which recognizes the importance of the recursive least-squares formulation in analysing data with general non-stationary statistical properties and, in particular, which discusses how the recursive residuals can be used to detect the possibility of temporal change in the coefficients of a regression relationship. The paper tonight will, I am sure, prove of practical use in day-to-day statistical analysis for I know from experience with real data from a variety of different sources that the possibility of parameteric non-stationarity is ever present and classical methods of block data analysis simply do not have the flexibility to handle such problems.

One minor criticism of the paper, which I may perhaps be allowed to voice, is its notable lack of reference to parallel developments in the control literature; developments which have considerable bearing on the type of analysis suggested by the authors and which should, I believe, be brought to the attention of the audience. I have attempted, at a previous meeting of the Society (Young, 1971), to correct the apparent lack of contact between the disciplines but, apart from some exceptions, I have clearly failed to get my message across and, with the Chairman's indulgence, I will try again.

Much of the paper tonight is concerned with the statistical properties of a normalized function w_r of the recursive residuals $y_r - \mathbf{x}'_r \mathbf{b}_{r-1}$. In the control literature these recursive residuals have been termed the *innovations process*; a term that appears to have been first introduced in this connection by Wiener and Masani in the mid-fifties (see Kailath, 1974). It was not until 1968, however, that Kailath showed that this process (or its continuous-time equivalent), is, under assumptions similar to those of regression analysis, zero mean, gaussian and serially uncorrelated, (although it should be emphasized that these properties are to a certain extent implicit in the orthogonal projection arguments of Kalman).

Bearing the "white noise" properties in mind, it is not surprising that statistical tests on the recursive residuals have been the basis of many methods of verifying the efficacy of recursive estimation schemes. In our own work on autoregressive-moving-average time-series model estimation (see, for example, Young *et al.*, 1971), for instance, it is normal to compute the sample autocorrelation function of the residuals and assess whether the statistical assumptions are satisfied; in particular whether the recursive residuals are serially independent. And in recent years, there has been considerable research into the statistical properties of the innovations process when, for example, a sub-optimal Kalman filter-estimation algorithm is applied to a stochastic system (as is often the case, since the optimal algorithm requires *a priori* information on the nature of the dynamic system and the covariance properties of the stochastic disturbances; information which, if it is available at all, can be subject to considerable uncertainty). Such research has produced algorithms for both (a) explicitly estimating either the statistical properties of the stochastic disturbances, or the gain matrix of the Kalman filter, from the sample correlations of the innovations process (for example, Mehra, 1970; Carew and Bélanger, 1973; Neethling and Young, 1974) or (b) implicitly adapting the Kalman filter to ensure a satisfactory innovations process (Neethling, 1974).

The authors, with their experience of statistical hypotheses testing, suggest various checks that can be applied to the normalized innovations process which appear to be of great value in assessing the existence of non-stationarity. A somewhat similar approach to detecting the possibility of parametric change using a serial correlation test has been suggested by Hancock (1971). But hypothesis tests are only diagnostic tools and they do little to tell us how, having discovered the existence of parametric non-stationarity, we are able to modify our analysis to account for its presence. In this sense, I think the authors have erred in not recommending that the user refers more often to the recursive estimation of the parameters in the regression relationship, as well as to the recursive residuals. Certainly our own work over the past ten years shows that such information can be particularly useful in practical terms providing, as it does, information not only on the existence of time variability but also, and perhaps more important, on the physical cause of such non-stationarity; information which could well be used, often in an iterative manner, to "identify" better the process under investigation, in some cases removing the cause of non-stationarity completely and so yielding a model of greater practical utility.

As an example of this latter approach consider the rainfall (u_k)-runoff (y_k) data shown in Fig. 1 for a stretch of the River Ouse near Bedford, in the year 1972. (This example has arisen in connection with a systems analysis study of water quality in the Great Ouse River system; a study being carried out by Paul Whitehead and myself in collaboration with the Great Ouse River Division of the Anglian Water Authority and the Department of the Environment; see Whitehead and Young, 1975.) The recursive estimates of the moving average parameters b_1 and b_2 in an autoregressive-moving average model relating y_k and u_k are shown in Fig. 2. These estimates were, in fact, obtained from a special recursive instrumental variable (I.V.) algorithm which may be interesting to Professor Durbin whose early paper on the instrumental variable method (Durbin, 1954) proved useful in our initial development of this approach to time-series analysis. The algorithm is special in the sense that it can be given the ability to estimate parametric non-stationarity if the user has reason to believe that significant changes may occur over the observation interval. In this present case, it is clear that the parameters show a marked tendency to both long- and short-term variation. Reference to the physical nature of the system indicates that such non-stationarity is probably due to the effects of evaporation etc. (in the long term) and soil moisture deficit (in the short term) and suggest that the system might be "purged" of its non-stationary behaviour by pre-processing or filtering the data with these factors in mind in order to yield an *effective* rainfall input, as shown in Fig. 1(b), such that the resulting estimates are, at least approximately, time-invariant as shown in Fig. 2(b). In this figure the estimates do not indicate strict time invariance because of the nature of estimation algorithm: in effect, the ability to track parameter variation is only obtained at some cost in estimation efficiency unless the *exact* nature of the time variation is known *a priori*. In this case such information is not available and the algorithm is instructed to expect only random variation in the b parameters between samples, with the result that the estimates have a fairly high residual variance. But, at the same time, it is now clear that the *apparent* estimated parameter variations are probably due in large part to the residual random noise effects that have not been sufficiently "smoothed out" by the modified algorithm which, in effect, has difficulty in differentiating between random noise effects and any random changes in parameters of the model (see, for example, Young, 1974). As a result, further analysis can now proceed *under the assumption that the parameters are now sensibly constant over the observation interval*. In this case, such analysis was carried out using an iterative version of the I.V. algorithm which is able to refine the estimates by making multiple passes through the data, each time updating the instrumental variables and so improving the statistical efficiency.

The success of the stationary time-series model obtained from the iterative I.V. algorithm is demonstrated in Fig. 3, which shows the forecasted river flow compared with the observed flow for the year 1972, and Fig. 4, which gives similar results for the following year on a *different* reach of the river but using the model as fitted to the 1972 data.

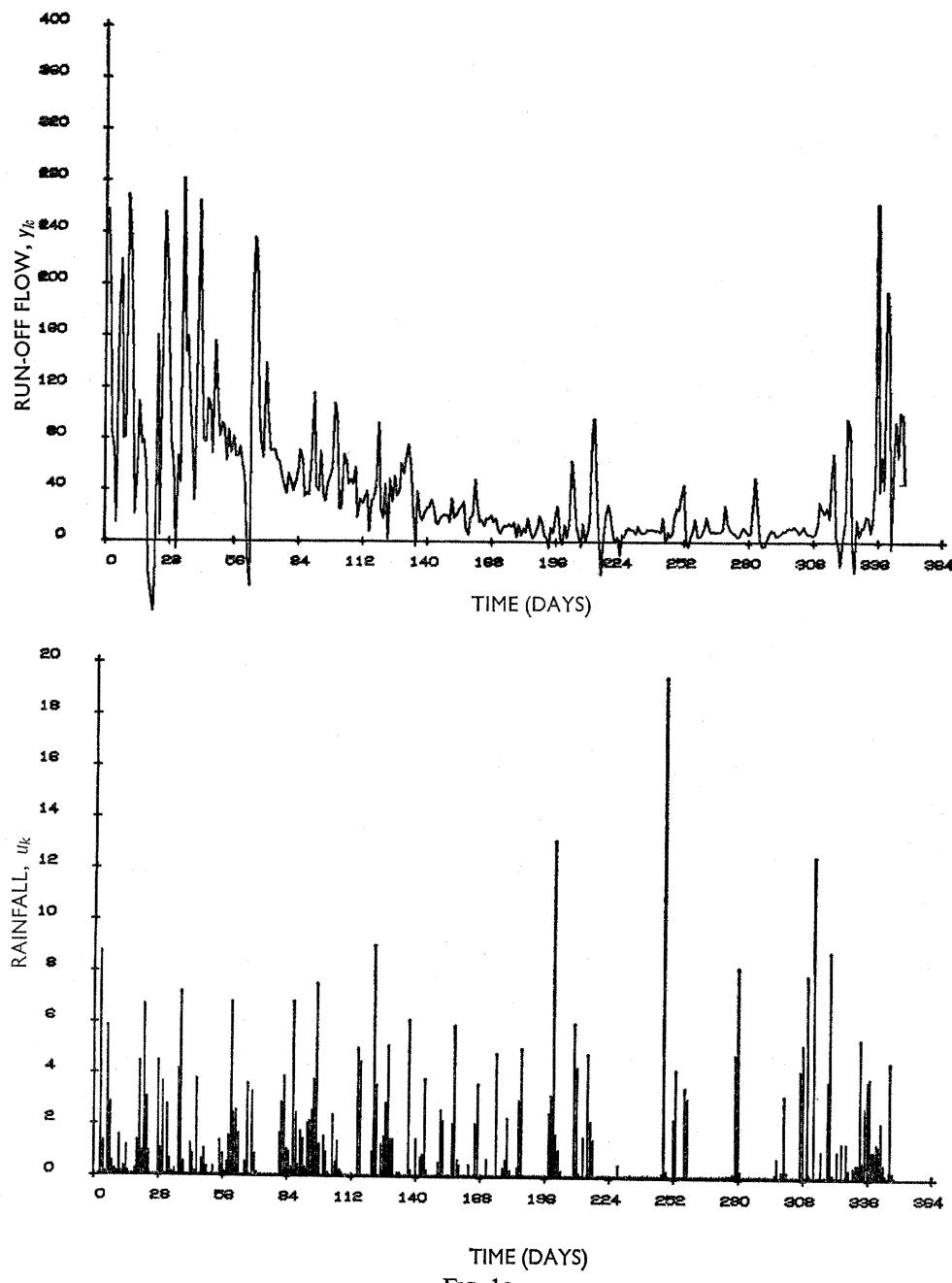


FIG. 1a

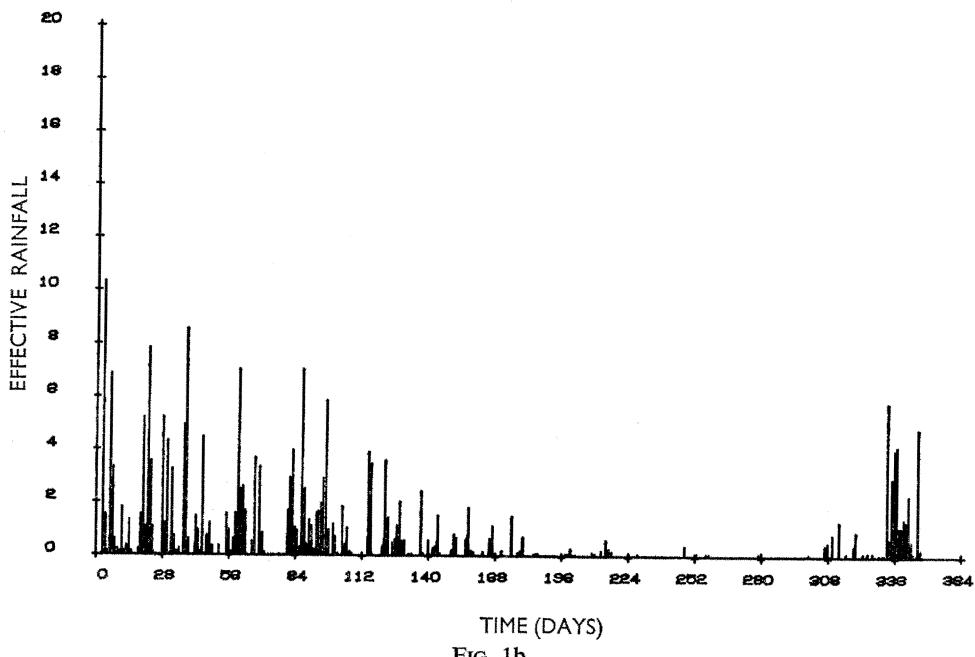


FIG. 1b

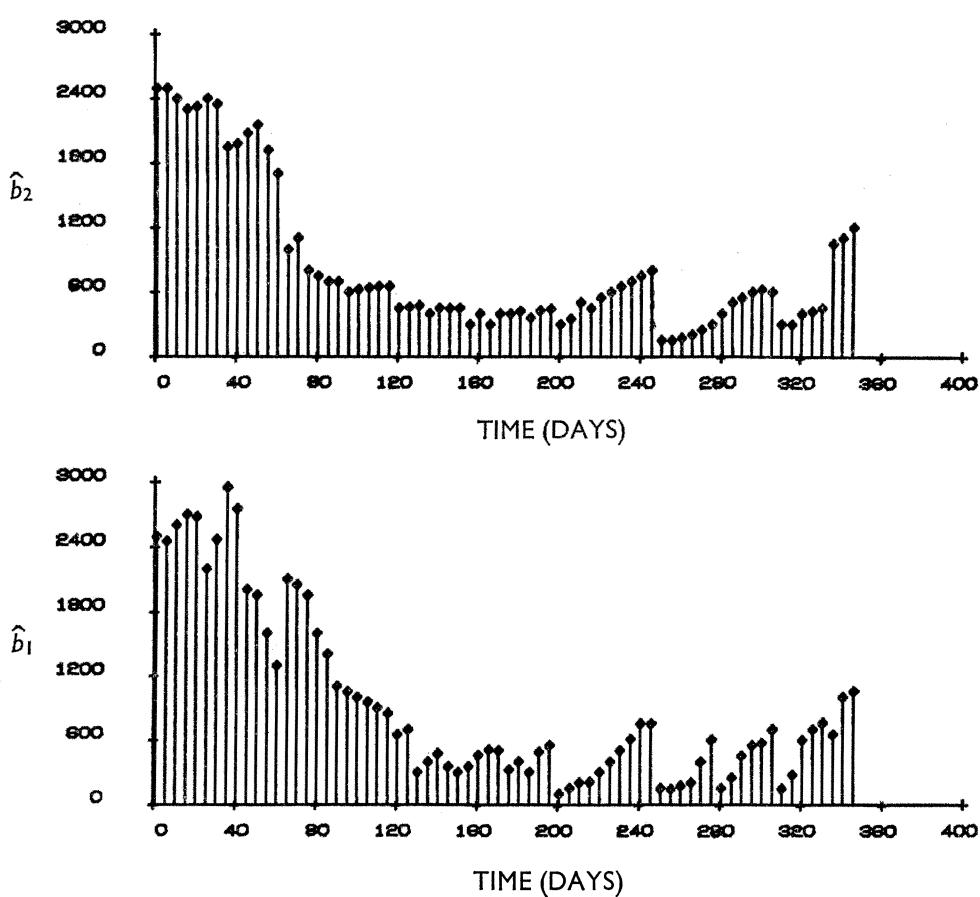


FIG. 2a

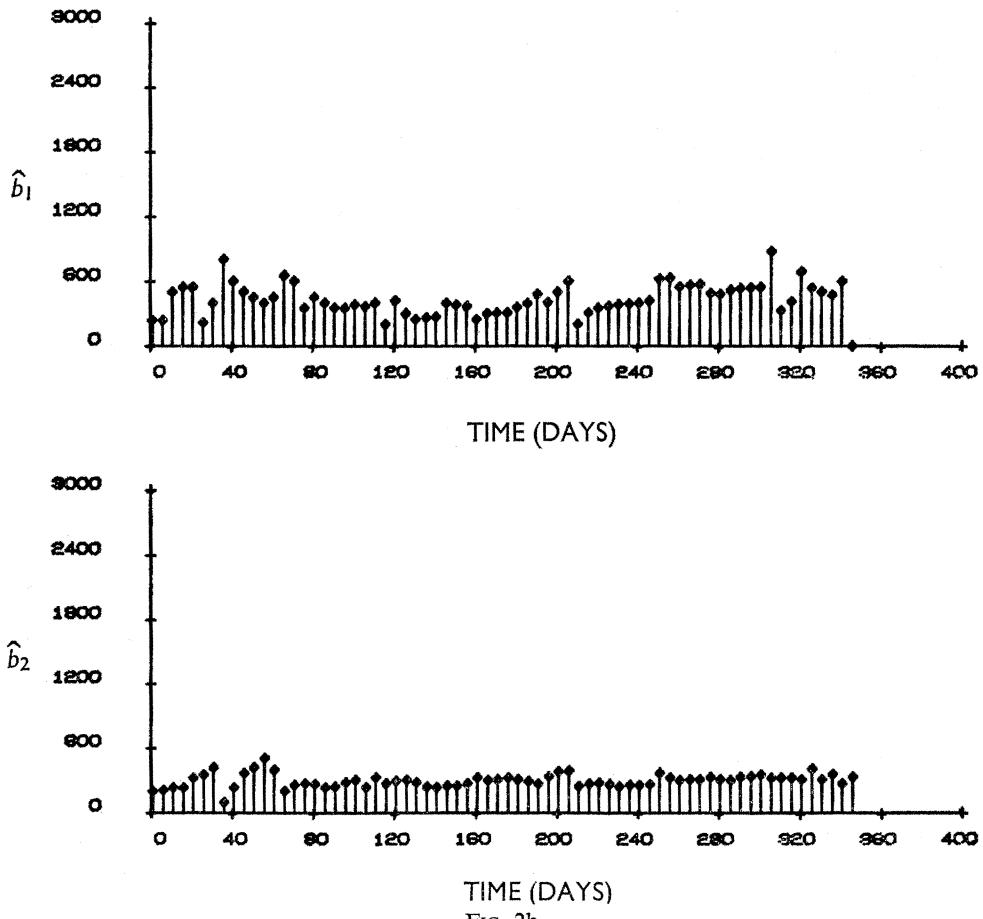


FIG. 2b

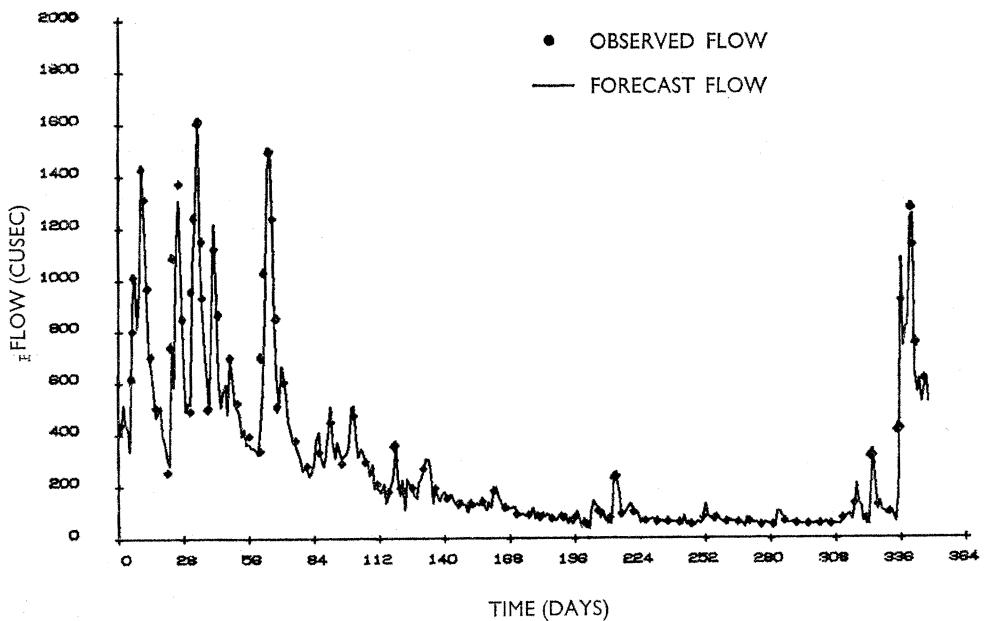


FIG. 3

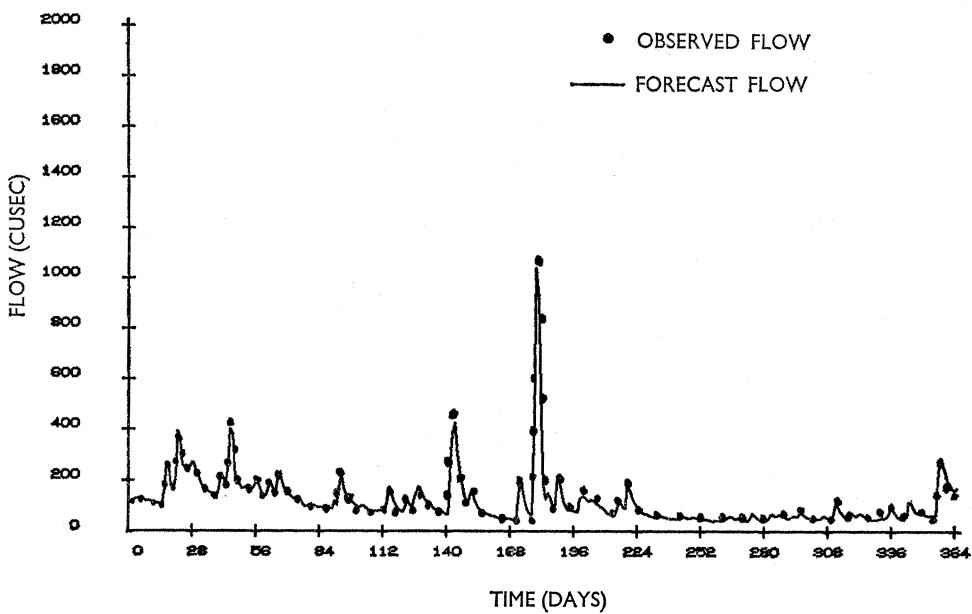


FIG. 4

Before leaving the topic of time-series analysis, I think it is a little misleading to a general audience when the authors, in the first sentence of their paper, refer to time-series data, so tending to give the impression that the techniques they describe are particularly useful for time-series analysis. And yet it is well known that, except in special cases, regression analysis yields asymptotically biased estimates when applied to more general time-series problems such as the example I have just outlined, where there are clearly problems of errors in variables. (This is probably no problem in the examples quoted in the paper since the authors are not, apparently, interested in the structural parameters but only in the forecasting ability of the models.)

I merely wish to note here that recursive techniques of time-series analysis that are not prone to such disadvantages are, as we have seen, available and have been applied successfully to real data from a variety of systems, in addition to the water resources problem discussed above. Descriptions of these IV-AML (Instrumental Variable-Approximate Maximum Likelihood) methods of time-series analysis have appeared both in the control literature (Young and Hastings-James, 1970; Young, 1972) and, more recently, in the mathematical literature (Young, 1974), while a report describing the techniques in detail is available from the author (Young *et al.*, 1971). In addition, a paper emphasizing their importance in general dynamic systems analysis will be submitted to this Society in the near future and will deal not only with single input-single output systems of the conventional time-series type (see, for example, Box and Jenkins, 1970) but also with multi-input, multi-output systems of the kind met so often in practice (Young and Whitehead, 1975).

Mr G. PHILLIPS (University of Kent): I found the paper extremely interesting, and it clearly indicates the usefulness of the recursive residuals in testing for specification errors in linear regression models. Further evidence for this is provided in two forthcoming papers, the first of which (Phillips and Harvey, 1974) discusses tests for serial correlation using recursive residuals, and the second (Harvey and Phillips, 1974) which discusses tests for heteroscedasticity.

However I find some difficulty in deciding the likely usefulness of the proposed tests in the absence of any investigations of their power under a range of alternative hypotheses. I noticed that the authors referred to the fact that they have not investigated the problem of serial correlation on their tests. There is some evidence in a paper by R. A. Johnson and M. Bagshaw, in *Technometrics* (1974), 16, 103–112, which suggests that the cusum tests are not robust to departures from independence.

A further point is that the recursive residuals are independent only when disturbances are normal. When there is a departure from normality perhaps recursive residuals may be no more effective than the ordinary least-squares residuals.

Finally, in the first example of an application of TIMVAR, the problem as posed appeared to be one of errors of measurement on the dependent variable. I was not clear how the analysis came to its conclusion that the instability was due to local changes in regression coefficients although I agree with the course of action taken.

Dr T. W. ANDERSON (Stanford University and London School of Economics): This paper was interesting to me on many counts because its contributions touch several of my own areas of research. Let me comment on different aspects. I thought it would be amusing to suggest to Professor Durbin that the recursive residuals could be used to construct tests of serial correlation which would be alternative to the well-known Durbin-Watson procedure, but that idea has already come up earlier in the discussion.

The generalization of the Helmert transformation is particularly useful in its natural time sequence for indefinitely long series. Before reading the paper I had used the transformation to prove the following theorem. (My reading of the discussions in the *JRSS* indicate that not infrequently the opportunity is used for the discussant to display his own results, and I seldom get the chance.) In the model of the paper, with H_0 true, $\mathbf{b}_T \rightarrow \boldsymbol{\beta}$ as $T \rightarrow \infty$ with probability 1 if and only if $(\mathbf{X}'_T \mathbf{X}_T)^{-1} \rightarrow 0$.

Another aspect of the paper that struck a familiar note was the use of the continuous Gaussian process to approximate the probability that the cusums lie between two lines for $r = k+1, \dots, T$. At one time I worked very hard to obtain the probability that the Brownian motion process remain between two specified lines. (Apparently, these RSS discussions permit a discussant to refer to his earlier work as well.) Using Corollary 4.2 of “A modification of the sequential probability ratio test to reduce the sample size” (*Annals of Mathematical Statistics* (1960), 31, 165–197) I find, for example, that when $a = 0.850$, the probability of going out of the interior region is 0.0987, which the authors use for 10 per cent significance. This is an error of only 1.3 per cent. Since the error will decrease with significance level, my calculations justify the authors’ assumption that the probability that W_n crosses both lines is negligible. It might be noted that the continuous time computation exaggerates the probability of cusums calculated discretely crossing a line.

It was instructive to me to write (4) as

$$\mathbf{b}_r - \mathbf{b}_{r-1} = w_r \times (\mathbf{X}'_r \mathbf{X}_r)^{-1} \mathbf{x}_r \sqrt{1 + \mathbf{x}'_r (\mathbf{X}'_{r-1} \mathbf{X}_{r-1})^{-1} \mathbf{x}_r}.$$

This shows that the recursive residuals are based on changes in the estimate of the regression parameter vector and w_r^2 is a quadratic form in the difference.

Dr A. F. M. SMITH (University College London): The topics dealt with in this paper are of considerable practical importance and provide a number of challenging problems for the theoretical statistician. The authors have adopted an unashamedly exploratory, data-analytic approach, and are careful to point out that their proposals “should be regarded as yardsticks for the interpretation of data rather than leading to hard and fast decisions”.

I wonder if this is really satisfactory? Data exploration is certainly a necessary preliminary, but it seems to me that sooner or later one must provide a more formal,

theoretical framework within which to assess the outcomes of the significance tests, or the departures occurring in the cusum plots. Other contributors have already touched on the problems of assessing power against particular alternatives, distinguishing changes in regression coefficients from changes in variances, and taking into consideration the effects of autocorrelated errors. The latter point, in particular, seems crucial given the type of problem under consideration. Have the authors any idea of how procedures based on recursive residuals are affected by the introduction of serial correlation?

Turning more specifically to the cusum techniques, is it not possible to improve the procedure by adopting a moving V-mask? Considering Fig. 5, for example, one notices that the plot has returned to the origin by about observation 35. Could this information not be utilized by resetting the mask (or some modification thereof) in order to begin again at this point?

As they stand, the procedures put forward by the authors seem to require a great deal of informal use of personal judgment (as evidenced in Section 2.4, for example, where we encounter such phrases as "we ourselves prefer", "it may be useful to examine" and "it is often informative to"). The Bayesian approach offers a more formal framework for the inclusion of personal judgments, and I should like to bring to the authors' attention two such Bayesian offerings.

The first of these is due to Harrison and Stevens (1971), and was developed in a time-series context. Basically, they take as their model a weighted average of *submodels*, the latter being selected to cover the range of potential behaviour of the process in question. The weights are the continuously updated probabilities of the submodels. Changes in the underlying relationship are reflected in changes in the weights, and the submodels can be chosen to include a whole range of possible departures. The method has been applied, in particular, to forecasting demand in a mail-order context where fashions are prone to sudden change (see Green and Harrison, 1973).

The second approach is one I have been working on myself in the general context of change-point inference. The simplest version concerns a series of n observations which may all follow the same distribution, with density $f_1(\cdot | \theta_1)$, or may have changed at some (unknown) point to a different distribution, with density $f_2(\cdot | \theta_2)$. A change at time r would generate a likelihood

$$l(r, \theta_1, \theta_2 | x_1, x_2, \dots, x_n) = \prod_{i=1}^r f_1(x_i | \theta_1) \prod_{i=r+1}^n f_2(x_i | \theta_2).$$

If $p_0(r)$ represents the prior probability of a change at time r , then $p_n(r)$, the posterior probability given the data, is defined by

$$p_n(r)/p_0(r) \propto \int \int l(r, \theta_1, \theta_2 | x_1, x_2, \dots, x_n) p(\theta_1, \theta_2) d\theta_1 d\theta_2,$$

where $p(\theta_1, \theta_2)$ denotes the joint prior density for θ_1 and θ_2 . The $p_n(r)$ provide a starting point for inferences about r , θ_1 and θ_2 .

In the particular case of detecting a possible change in the regression coefficients corresponding to a particular set of regressor variables, if \mathbf{X}_r , $\tilde{\mathbf{X}}_{n-r}$ denote the portions of the design matrix corresponding to the first and r the last $n-r$ observations, respectively, and SS_r and \widetilde{SS}_{n-r} the corresponding residual sums of squares, the assignment of standard vague priors for the regression coefficients and the variance (assumed constant throughout) leads to the result

$$p_n(r)/p_0(r) \propto |\mathbf{X}'_r \mathbf{X}_r|^{-\frac{1}{2}} \times |\tilde{\mathbf{X}}'_{n-r} \tilde{\mathbf{X}}_{n-r}|^{-\frac{1}{2}} \times (SS_r + \widetilde{SS}_{n-r})^{-d},$$

where d is a function of n and the number of regression coefficients. In so far as there is any connection here with the authors' analysis, the Bayesian approach would seem to favour techniques based on the squares of the recursive residuals (cf. the authors' equation (5)).

More complicated situations involving a number of possible change points can be dealt with straightforwardly by performing a similar prior to posterior analysis for appropriate k -tuples (r_1, r_2, \dots, r_k) , where $1 \leq r_1 < r_2 < \dots < r_k \leq n$.

The authors are to be congratulated on stimulating this evening's discussion. By their own admission, a number of problems remain unsolved, but, as was said of St Denis when he walked some considerable distance with his head in his hand: "*La distance n'y fait rien; il n'y a que le premier pas qui coûte*".

The following contributions were received in writing, after the meeting.

Mr M. R. B. CLARKE (University of London): I have a brief comment to make about the updating formulae for moving regressions in Section 2.5. Many people have pointed out possible dangers arising from numerical ill-conditioning in forming sums of squares and products in order to solve the normal equations. More recently updating formulae such as those quoted at the beginning of Section 2.5 have come in for severe criticism, notably by Chambers, when used for numerical rather than theoretical computations. Broadly speaking such problems can be avoided by using one of the orthogonal decomposition methods such as Householder or Gram-Schmidt. These decompose the data matrix into the form

$$[\mathbf{X} | \mathbf{y}] = [\mathbf{Q}] \begin{bmatrix} \mathbf{U} & \mathbf{v} \\ \mathbf{0} & \mathbf{w} \end{bmatrix},$$

where \mathbf{Q} is $(n \times n)$ orthogonal and \mathbf{U} is the $(k \times k)$ upper triangular square root of $\mathbf{X}'\mathbf{X}$. \mathbf{Q} need not be known explicitly as all the information required is in \mathbf{U} and \mathbf{v} , the coefficients being the roots of $\mathbf{U}\beta = \mathbf{v}$ and the regression sum of squares $\mathbf{v}'\mathbf{v}$.

If we now add another observation (x, y) we have

$$\begin{bmatrix} \mathbf{X} & \mathbf{y} \\ \mathbf{x} & y \end{bmatrix} = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0}' & 1 \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{v} \\ \mathbf{0} & \mathbf{w} \\ \mathbf{x} & y \end{bmatrix}$$

and since $\begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0}' & 1 \end{bmatrix}$ is orthogonal we need only write down some simple equations to determine the k -plane rotations that annihilate the \mathbf{x} part of (x, y) . These updating formulae are numerically stable being linear in the data and very nearly as economical, once the original decomposition is completed, as those quoted in the paper.

Professor A. S. C. EHRENBURG (London Business School): I welcome a paper addressing itself to practical problems in regression analysis, but I have difficulties with it. For example, I do not understand the data-analysis "yardsticks" which the authors proffer in place of classical tests of significance. In what probabilistic or other units are the yardsticks calibrated?

My main difficulty, however, arises from the assumption at the beginning of Section 2 that the regressors are non-stochastic. This is obviously not true in most practical cases, and certainly not in the authors' own examples in Section 3. The failure of this assumption leads not merely to minor technical difficulties or small biases, but radically affects the authors' null hypothesis and indeed the applicability of regression analysis as a whole. This can be readily demonstrated in terms of the "moving regression" approach developed in Section 2.5.

Here regressions are fitted to non-overlapping sets of n successive observations. The authors' null-hypothesis is that the regression for the first n readings is the same as that for another n , say the last n . This is illustrated in Fig. A for two variables, y and x . (For simplicity I consider only two variables here, but the argument generalizes to more than two.)

If x is non-stochastic (as assumed in the paper) this null-hypothesis says that the expected values of y for each x in the first set of n readings lie on a straight line and that the expected values of y for each x in the second set of data lie on the same straight line. This is of course perfectly feasible and would therefore be worth testing against any data.

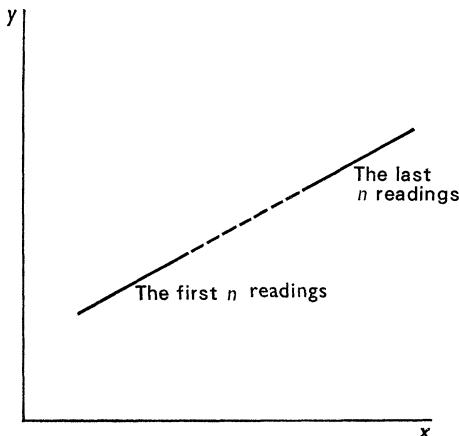


FIG. A. The null-hypothesis with a non-stochastic x -variable.

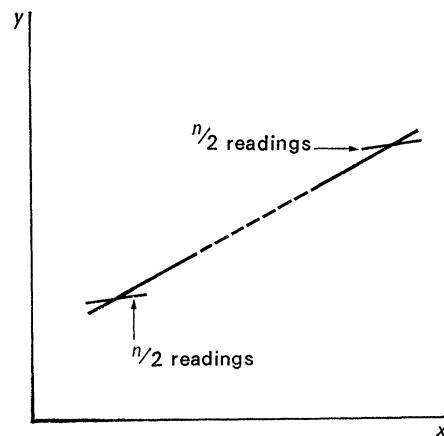


FIG. B. Regressions of y on x for the first and last n readings, and for the first and last $n/2$ readings.

But this null-hypothesis is inherently impossible if the x -variable is stochastic (other than in trivial cases). It is then well known that the regression of y on x for the first n readings will generally differ from the regression for a systematic subset of the readings, such as the first $n/2$. Similarly, the regression of y on x for the last n readings will be different from that for the last $n/2$ readings. This is illustrated in Fig. B.

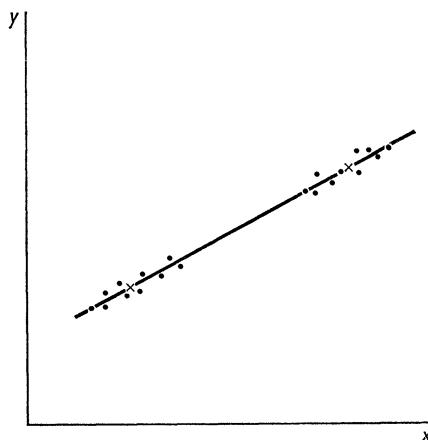


FIG. C. A line fitted to means of the first and last n readings.

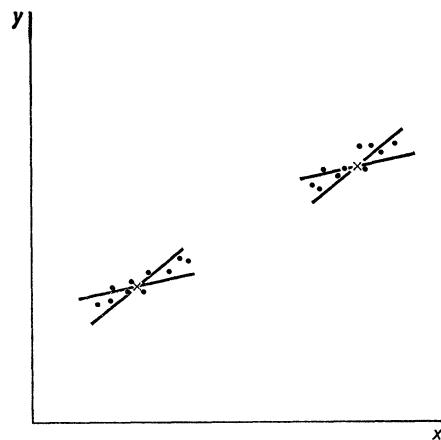


FIG. D. The regressions of y on x and of x on y for the first and last n readings.

It follows that if the null-hypothesis that the regressions of y on x are the same holds for the first n and last n readings, the corresponding hypothesis cannot hold for the first $n/2$ and last $n/2$ readings. But n is an arbitrary number. Hence the regressions of y on x with

stochastic x -variables cannot in general be the same for any two sets of n readings. The null-hypothesis is therefore untrue, and this can be seen without any data-analysis, or even without the data. The distinction between the non-stochastic situation considered in the paper and the realistic one where the regressors are in some sense stochastic is therefore crucial.

An alternative formulation which avoids the above dilemma is discussed more generally elsewhere (Ehrenberg, 1975). In terms of the approach of analysing separate sets of n readings as developed in Section 2.5 it is that for a straight line to hold for both the first and the last sets of n successive readings (and for any other sets of readings), the line must go through the mean values of each set of readings, as shown in Fig. C.. This is both a necessary and a sufficient condition.

Such a line is in general not a regression equation for any of the data. This is illustrated in Fig. D which indicates the regressions of y on x and of x on y for both sets of n readings.

Since the line in Fig. C is fitted without recourse to the residuals from the line (i.e. with no minimization procedures), the residuals are in practice relatively easy to examine for systematic deviations. The line is also not affected by unbiased errors (of measurement or the like) in either of the variables.

Dr A. C. HARVEY (University of Kent at Canterbury): Recursive residuals have two great attractions; one is their simplicity, and the other, which is perhaps more important, is their flexibility. For a given set of T observations, there are $T!/k!$ different sets of recursive residuals. Which set is actually computed depends on the result of two closely connected decisions. The first concerns which k observations should be used to form the "basis" (i.e. used to form the initial estimate of β); the second concerns how the remaining $T-k$ observations should be ordered. If the basis is chosen in a certain way it is possible to obtain a set of recursive residuals which follow a similar pattern to that produced by the O.L.S. residuals. Exact tests against such alternative hypothesis as serial correlation (Phillips and Harvey, 1974) and heteroscedasticity (Harvey and Phillips, 1974) may then be constructed. On the other hand, it is sometimes possible to choose the basis and the ordering in such a way that under certain misspecifications of the model, the recursive residuals have a distinctive pattern, very different to that produced by O.L.S. residuals. Tonight's paper has provided one example of this type of distinctive pattern in the context of structural change. Another example concerns the case when the functional form of one of the regressors is incorrectly specified. Mr P. Collier and myself have recently completed a paper (Harvey and Collier, 1975) in which we show that a test based on recursive residuals is relatively powerful compared with a number of other tests. The test is based on the statistic

$$\psi = \left\{ (T-k-1)^{-1} \sum_{j=k+1}^T (w_j - \bar{w})^2 \right\}^{-\frac{1}{2}} \left\{ (T-k)^{-\frac{1}{2}} \sum_{j=k+1}^T w_j \right\}, \quad (1)$$

where \bar{w} is the arithmetic mean of the recursive residuals. This statistic follows a t -distribution with $(T-k-1)$ degrees of freedom under the null hypothesis. Under the alternative hypothesis the recursive residuals tend to have the same sign and so ψ tends to be large in absolute value, thus leading to rejection of the null hypothesis.

Perhaps I can now turn to some specific points on tonight's paper. The first concerns the definition of the cusum quantity, W_r . It is suggested that this be obtained by using, as a deflator of the recursive residuals, the estimator $\hat{\sigma}$, defined as the square root of $\sum w_j^2/(T-k)$. It seems much more sensible to me, however, to estimate σ by the square root of $\sum (w_j - \bar{w})^2/(T-k-1)$. This does not affect the theory behind the cusum test, but it is likely to make the procedure more effective as the cusum will tend to be larger in absolute value under the alternative hypothesis. (Another advantage of defining the cusum in this way is that when the last cusum quantity has been obtained it only needs to be divided through by $(T-k)^\frac{1}{2}$ in order to yield the statistic ψ , which, as I have already said, has a t -distribution under the null hypothesis.)

My second point concerns the type of variation in the β 's which we are interested in detecting. The authors prefer to leave their specification of the alternative hypothesis in a rather vague form. However, if a more concrete alternative is proposed more powerful tests are available. For example, the test proposed by Farley and Hinich (1970) is likely to be much more powerful than the cusum test against certain special types of structural changes in the β 's. Now it seems to me perfectly acceptable to leave the alternative hypothesis vague, but this leaves us with a test which may be very weak when used in the presence of many types of structural change likely to occur in practice. Of course, I think the authors implicitly recognize this when they say that the significance lines drawn should be interpreted as yardsticks rather than as part of a formal test. However, if the lines yield an ineffective test they may well be ineffective as yardsticks also.

Finally may I suggest an alternative, or rather additional, way of calculating moving regressions, which are the subject of Section 2.5 of the paper. Suppose estimates of β based on r observations are calculated using an exponential weighting system. In this system the current set of observations would receive a weight of unity, the previous set of observations would receive a weight of q , the set before a weight of q^2 and so on. Of course, $0 < q \leq 1$. Successive estimates could be computed recursively since if we define

$$\mathbf{Q}_j = \sum_{i=1}^j \mathbf{x}_i \mathbf{x}'_i q^{j-i} \quad (2)$$

we have

$$\mathbf{Q}_r = q\mathbf{Q}_{r-1} + \mathbf{x}_r \mathbf{x}'_r \quad (3)$$

and so

$$\mathbf{Q}_r^{-1} = q^{-1} \mathbf{Q}_{r-1}^{-1} - \frac{q^{-2} \mathbf{Q}_{r-1}^{-1} \mathbf{x}_r \mathbf{x}'_r \mathbf{Q}_{r-1}^{-1}}{1 + q^{-1} \mathbf{x}'_r \mathbf{Q}_{r-1}^{-1} \mathbf{x}_r}. \quad (4)$$

I wonder if the authors have used this type of weighted recursion? It would be interesting to know if it is more sensitive to changes in the β 's than the moving average method.

Dr AGNES M. HERZBERG (Imperial College, London): Andrews (1972) proposed a simple way of plotting higher dimensional data in two dimensions, i.e. if the data are m -dimensional each point $\mathbf{x}' = (x_1, \dots, x_m)$ defines a function

$$f_{\mathbf{x}}(t) = 2^{-\frac{1}{2}}x_1 + x_2 \sin t + x_3 \cos t + x_4 \sin 2t + x_5 \cos 2t + \dots,$$

the function being plotted over the range $-\pi < t < \pi$.

The basic regression model considered in equation (1) is $y_t = \mathbf{x}'_t \beta_t + u_t$ ($t = 1, \dots, T$). In Section 2.5, the authors suggest a way of investigating the time-variation of β_t by fitting equation (1) on n successive observations and then on the next n , and so on. They suggest that the graphs of the resulting estimates of the coefficients in β_t against time provide evidence of departures from constancy.

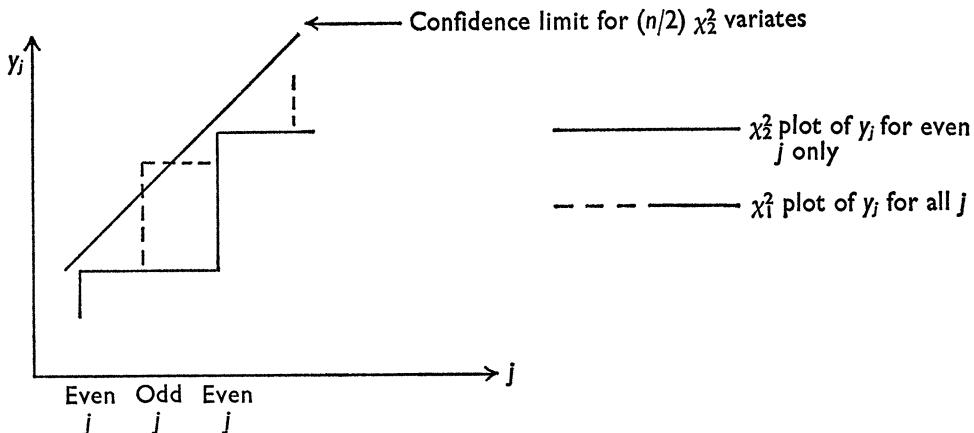
Let $\hat{\beta}_i = (\hat{\beta}_{1i}, \dots, \hat{\beta}_{ki})$ be the $k \times 1$ vector of estimates of β_i obtained from the i th set of n observations, i.e. $\hat{\beta}_1$ is estimated from the first n observations, $\hat{\beta}_2$ is estimated from the second observation to the $(n+1)$ st observation, et cetera. From each $\hat{\beta}_i$, form the function $f_{\hat{\beta}_i}(t)$ as above. Plot the resulting functions. The plots of the functions should show the gradual change of the whole vector of coefficients from constancy by the change of the clusters of the plots.

Mr M. C. HUTCHISON (Department of Health and Social Security): I would like to congratulate the authors in providing a most interesting paper which, in the form of TIMVAR is especially useful to statisticians working with regressive time series.

The cusum of squares test is of the form

$$y_j = \sum_{i=1}^j x_i / \sum_{i=1}^n x_i \quad (j = 1, \dots, n),$$

where $n = T - K$ and the x_i are distributed as χ_1^2 variates under H_0 . The approximate test given in the paper essentially considers only half of the y_j for even j or equivalently adds x_{j-1} to x_j for even j and treats $x_{j-i} + x_j$ as a χ_2^2 variate. Obviously this test will give more significant results than the exact test because the y_j for odd j can overstate the confidence limit calculated on even values only while the y_j for even j stay within it. The following diagram shows this clearly.



An exact test for y_j can be obtained. Considering the probabilities of sample paths of y_j ($j = 1, \dots, n$) crossing a linear boundary using Durbin (1971), formulae can be obtained for calculating the confidence limits of the joint distribution of the y_j for even values of n only. A program has been written to calculate these percentage points but unfortunately due to the lengthy summations involved in integration by parts combined with an iterative process, the running time is prohibitive. Consequently, percentage points for probabilities of 0.05 and 0.005 of crossing one boundary have been calculated only for even values of n up to 34. Work is in hand to obtain percentage points for further values of n and further probabilities. This will give a test for detecting changes in the mean of normal variates.

My other comments concern the use of TIMVAR in practice. I have used TIMVAR in the past on regressive time-series data and it has successfully picked a point of discontinuity in the regression coefficients which was believed to be highly likely *a priori*. Since then, other data have been brought to my notice for which TIMVAR does not seem readily applicable. First, there is the case of data for a small number of years (say, $m < k$) but with many observations within each year. If one can be certain that data within each year come from the same equation then I can find no reason why TIMVAR should not be used, which implies a random ordering of observations within each year to give a new "time dimension". Then one is concerned with discontinuities at m points only which are the changes of data from one year to the next. If one is unsure that the data within each year come from the same equation then presumably a simple test between corresponding coefficients calculated from regressions on data within years should be applied. Secondly, the "other dimension" need not be time. One may want to consider points of discontinuity over an ordered variable, say adjusted income (income minus needs) collected from a survey together with other variables. One might wish to relate expenditure on luxury goods to other factors as adjusted income increases. A point of discontinuity may be suspected for high adjusted income. In this case one may not have ready access to more data than just averages within adjusted income bands. These averages will almost surely not lie equidistant from each other. How robust is TIMVAR to changes in the "other dimension" from a discrete to

a continuous variable? One could perhaps use TIMVAR if the deviations of the data from discreteness were small in some sense.

I would be interested to hear the views of the authors on these points.

Professor MOHSIN S. KHAN (International Monetary Fund): I am very pleased to have the opportunity to discuss this interesting paper, although unfortunately I was unable to be present at its presentation. I believe that this paper will be of considerable importance in the field of applied economics where more and more researchers are becoming interested in regression relationships that contain parameters that vary over time. In discussing this paper I would like to make two points about the tests of stability that have been proposed, particularly the tests utilizing the cusums of residuals.

Since the authors caution the reader against the use of the tests of significance in any rigid way, it may be useful to report the results from some Monte Carlo experiments that I have conducted. In these experiments I have examined the cusums tests of the Brown-Durbin-Evans paper, for a random coefficients model. It is interesting that the cusums tests have reasonably high power even for sample sizes of 20 and 30 as compared to the maximum-likelihood ratio test and a test based on the estimated values of the variances that is due to Hildreth and Houck (1968).

My other point has to do with the question of the exclusion of autoregressive models from consideration. Since it is precisely these models, namely ones involving the use of lagged dependent variables as regressors, that are currently of greatest interest to economists, it would be extremely useful to have tests of stability that would be applicable to this class of models. In certain special cases it is possible to apply the methods contained in the paper, for example, if one has a regression equation of the form:

$$y_t = \alpha_0 + \alpha_1 x_t^e + u_t, \quad (1)$$

where the dependent variable, y , is related to the "expected" value of the independent variable, x . The error term, u_t , is assumed to have classical properties. The expected variable is generated by recursive mechanism such as

$$x_t^e = \beta x_t + (1 - \beta) x_{t-1}^e, \quad (2)$$

where β is the coefficient of expectations, $1 > \beta > 0$. Substituting (2) into (1) and eliminating x_t^e we obtain:

$$y_t = \alpha_0 \beta + \alpha_1 \beta x_t + (1 - \beta) y_{t-1} + u_t - (1 - \beta) u_{t-1}. \quad (3)$$

Obviously the cusums method cannot be applied to equation (3) because of the appearance of y_{t-1} as a regressor and the moving-average nature of the error process. However, it is possible to generate x_t^e from (2) for various values of β (as it is bounded) and substitute the generated series of x_t^e into equation (1). Such a procedure would fit into the framework of the tests described in the paper although, unfortunately, it is fairly time consuming. With other models involving lagged dependent variables even this does not appear to be possible.

As I said earlier, the cusums tests of Brown *et al.* have considerable potential use in economics. In addition to the applications described by the authors, the techniques have been used to evaluate the stability of Phillips' curve relationships in the U.K. and import functions for the U.S.

J. A. NELDER (Rothamsted Experimental Station): The updating procedure given by relation (3), and more particularly its analogue for deleting a point, is unstable numerically. Chambers (1971) describes better methods based on updating the components of, for example, a QR decomposition of the data matrix.

I would like to make a small protest about the misuse of the word "recursive" in the paper. The authors are following what appears to be a well-established practice in talking about "recursive residuals" *et cetera*, but nonetheless the underlying procedures are in no sense recursive. A *recursive* procedure is one that invokes itself in the course of execution;

the procedures in the paper are *sequential* or *updating* procedures in which new units are added to (or dropped from) an existing fit. This type of algorithm also needs to be distinguished from an *iterative* procedure, which is applied sequentially, but to the same set of data. "Sequential" seems as good an adjective as any.

RICHARD E. QUANDT (Princeton University, N. J.): Parameter variation in regression models has recently been considered in a variety of contexts (see *Annals of Economic and Social Measurement*, 1973, entire issue) and econometric model builders have increasingly been willing to entertain the notion that regression coefficients may have different values corresponding to *unknown* partitions of the sample (see Davis *et al.*, 1966; Fair and Jaffee, 1972). In such cases it is important to be able to test the hypothesis that no shift in parameter values has taken place. There are basically two types of mechanisms that may be thought to be responsible for shifts: (a) deterministic mechanisms according to which the parameter vector β has the value β_1 if some specified function $\phi(z, \pi) \leq 0$, where z is a vector of some observable exogenous variables and π a vector of unknown parameters, and β has value β_2 if $\phi(z, \pi) > 0$; (b) stochastic mechanisms among which we include random coefficients regression models as well as mixture models according to which, for each observation, $\beta = \beta_1$ with probability λ and β_2 with probability $1-\lambda$. Brown, Durbin and Evans address themselves to the first case with the further specialization that the function $\phi(z, \pi)$ is of the form $t + \pi$ where the only exogenous variable responsible for the shift is the time index t . Their procedures involving recursive residuals are particularly appealing because of (a) the particular suitability of the recursive residuals, as contrasted with, say, Theil's BLUS residual, for the test at hand; (b) the ease of computation of the tests; (c) the fact that the tests serve the combined purpose of testing hypotheses and performing data analysis in the Tukey sense; and (d) their adaptability to cases where the shift in parameter values occurs according to the values of some extraneous variable other than t . If, for example, it were posited that $\beta = \beta_1$ for values of a variable $z_t \leq z_0$ and $\beta = \beta_2$ otherwise, where z_t is observable and z_0 unknown, all that would be necessary for the Brown, Durbin and Evans procedures would be to sort the observations according to the values of z_t and then apply the tests. It is particularly laudable that their several procedures as well as the log-likelihood ratio technique are operational in the TIMVAR program.

Several questions and problems remain, however. Some of these are as follows. (1) What are the asymptotic properties of the various procedures? One of the questions here is the manner in which one proceeds to the limit. In a recent paper, Farley *et al.* (1973) suggest holding constant the period spanned by the observations and letting the intervals between observations converge to zero; only by such a device can one guarantee that the fraction of observations belonging to the two regression regimes remains unaltered as we pass to the limit. On this basis they indicate that both cusum tests have undesirable asymptotic properties, such as the power of the cusum of squares test not converging to 1 as n goes to infinity. (2) What are the powers of the various procedures in finite samples? Some preliminary Monte Carlo experiments by Farley *et al.* have compared the power of their own procedure with that of the Chow test (Chow, 1960) performed on the assumption that the shift occurs at the midpoint of the data series and the log-likelihood ratio test using empirically derived critical values. Their own procedure involves estimating $y = X\beta + \epsilon$ and $y = X\beta + H\delta + \epsilon$, where H has (tj) th element tx_{jt} , computing the sum of squares residuals S_0 and S_1 from the two regressions and rejecting $H_0: \delta = 0$ if $(S_0 - S_1)/S_1$ is significantly different from zero. Among the three procedures compared they find that there is no test most powerful uniformly in the value of the true shift point. Unpublished results by Goldfeld and Quandt suggest that the cusum test performs well in samples of size 30–60 observations and in some cases has power equal to 100 per cent. It seems possible to gain additional power by performing the test both "forward" and "backward" on the data series, although we have not determined how to use the thus gained information "rigorously". It would clearly be desirable to have more systematic information about the

several procedures proposed by the authors in contrast with the procedures proposed by others. (3) What are sensible procedures if the error terms are serially correlated? It seems fairly certain that the tests mentioned heretofore do not remain acceptable in the presence of serial correlation, yet it is precisely in time series models that serial correlation is most likely to occur. It would be interesting to see to what kind of modification the authors' methods would have to be subjected in order to cope with this problem. An approximate maximum-likelihood procedure for estimating the parameters of the two regression regimes which can then be used in likelihood ratio tests is as follows (see Goldfeld and Quandt, 1974).

Let

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_1 + u_{1t} \quad \text{if } t \leq t_0,$$

$$y_t = \mathbf{x}'_t \boldsymbol{\beta}_2 + u_{2t} \quad \text{if } t > t_0$$

with β_1, β_2, t_0 and the error variances unknown.

Define $D_t = 0$ if $t \leq t_0$ and $D_t = 1$ otherwise, and posit that the error terms are generated by

$$u_{1t} = \rho_1 \{(1 - D_{t-1}) u_{1,t-1} + D_{t-1} u_{2,t-1}\} + \varepsilon_{1t},$$

$$u_{2t} = \rho_2 \{(1 - D_{t-1}) u_{1,t-1} + D_{t-1} u_{2,t-1}\} + \varepsilon_{2t},$$

where ε_{1t} and ε_{2t} are jointly normal and

$$E(\varepsilon_{1t}) = E(\varepsilon_{2t}) = E(\varepsilon_{1t} \varepsilon_{1,t-1}) = E(\varepsilon_{2t} \varepsilon_{2,t-1}) = 0,$$

$$E(\varepsilon_{1t}^2) = \sigma_1^2, \quad E(\varepsilon_{2t}^2) = \sigma_2^2, \quad E(\varepsilon_{1t} \varepsilon_{2t}) = \sigma_{12}.$$

The two regression regimes may be combined in obvious fashion to yield

$$\begin{aligned} y_t = & (1 - D_t) [\mathbf{x}'_1 \boldsymbol{\beta}_1 + \rho_1 \{(1 - D_{t-1}) (y_{t-1} - \mathbf{x}'_{t-1} \boldsymbol{\beta}_1) + D_{t-1} (y_{t-1} - \mathbf{x}'_{t-1} \boldsymbol{\beta}_2)\}] \\ & + D_t [\mathbf{x}'_2 \boldsymbol{\beta}_2 + \rho_2 \{(1 - D_{t-1}) (y_{t-1} - \mathbf{x}'_{t-1} \boldsymbol{\beta}_1) + D_{t-1} (y_{t-1} - \mathbf{x}'_{t-1} \boldsymbol{\beta}_2)\}] \\ & + (1 - D_t) \varepsilon_{1t} + D_t \varepsilon_{2t}. \end{aligned}$$

From this the likelihood function (conditional on y_0) can be derived and is a function of all the parameters including the discontinuous D_t . It is possible to replace D_t in the likelihood function with a continuous approximation with the correct qualitative properties such as

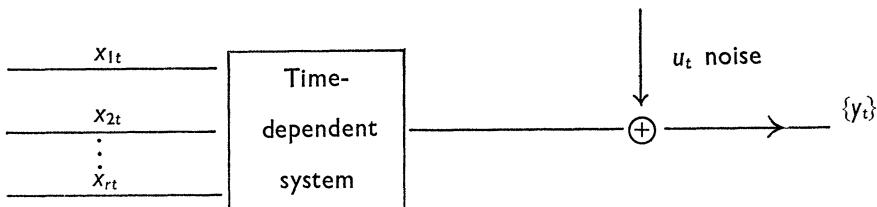
$$D_t = \int_{-\infty}^t \frac{1}{\sqrt{(2\pi) \sigma}} \exp \left\{ -\frac{1}{2} \left(\frac{\xi - t_0}{\sigma} \right)^2 \right\} d\xi;$$

the likelihood function then becomes a function of two new parameters t_0 and σ but all the parameters can now be estimated by fairly routine numerical optimization techniques. Preliminary indications seem to be that if we form the likelihood ratio λ by dividing the maximum of the likelihood function under H_0 by the maximum of the likelihood function suggested above, the quantity $-2 \log \lambda$ has approximately χ^2 distribution with appropriate degrees of freedom as suggested by asymptotic theory even in moderate-sized samples.

The preceding comments and suggestions are merely intended to give some indications of directions in which future research might go. It is clear that the tests proposed in the paper are already useful since several researchers have been using them and indeed the cusum test has been programmed by several econometricians.

Dr T. SUBBA RAO (University of Manchester Institute of Science and Technology): The problem considered by the authors is interesting, and a similar problem has been considered earlier by Subba Rao and Tong (1972, 1973) and Bohlin (1971). By first formulating the problem in the context of a control system, I will show how the problem considered by the authors can be deduced as a particular case of the one considered by Subba Rao and Tong (1972, 1973).

Let us assume we have a time-dependent system with k inputs $\{x_{rt}, r = 1, 2, \dots, k\}$ and a single output $\{y_t\}$ contaminated by the noise $\{u_t\}$. The system can be described schematically as follows:



Further let us assume that the stochastic processes $\{x_{rt}, r = 1, 2, \dots, k\}, \{u_t\}, \{y_t\}$ are all oscillatory processes (Priestley, 1965), with zero mean and spectral representations

$$\left. \begin{aligned} x_{rt} &= \int_{-\pi}^{\pi} e^{it\omega} A_{t,r}(\omega) dZ_{x,r}(\omega) \quad (r = 1, 2, \dots, k), \\ u_t &= \int_{-\pi}^{\pi} e^{it\omega} A_{t,u}(\omega) dZ_u(\omega) \end{aligned} \right\} \quad (1)$$

and

$$y_t = \int_{-\pi}^{\pi} e^{it\omega} A_{t,y}(\omega) dZ_y(\omega),$$

where $\{dZ_{x,r}(\omega)\}, \{dZ_u(\omega)\}$ and $\{dZ_y(\omega)\}$ are all orthonormal processes. Let the system be described by the linear relationship

$$y_t = \sum_{j=1}^k \sum_{l=0}^{\infty} h_{j,t}(l) x_{j,t-l} + u_t, \quad (2)$$

where it is assumed that $\{u_t\}$ and $\{x_{j,s}\}$ are independent. We note that by choosing the impulse response functions

$$h_{j,t}(l) = h_{j,t} \delta_t(l) \quad (j = 1, 2, \dots, k), \quad (3)$$

where

$$\delta_t(l) = \begin{cases} 1 & \text{if } l = 0, \\ 0 & \text{if } l \neq 0, \end{cases}$$

the model (1) considered by the authors can be obtained.

By substituting the spectral representations in (2), we can show that

$$dZ_{t,y}(\omega) = dZ'_{t,x}(\omega) \mathbf{H}_t(\omega) + dZ_{t,u}(\omega), \quad (4)$$

where

$$dZ_{t,y}(\omega) = A_{t,y}(\omega) dZ_y(\omega), \quad dZ_{t,u}(\omega) = A_{t,u}(\omega) dZ_u(\omega),$$

$$dZ'_{t,x}(\omega) = \{A_{t,1}(\omega) dZ_{x,1}(\omega), \dots, A_{t,k}(\omega) dZ_{x,k}(\omega)\},$$

$$\mathbf{H}_t(\omega) = \{H_{1,t}(\omega), H_{2,t}(\omega), \dots, H_k(\omega)\},$$

$$H_{j,t}(\omega) = \sum_{l=0}^{\infty} h_{j,t}(l) e^{-it\omega} \quad (j = 1, 2, \dots, k).$$

For the type of impulse response functions given by (3),

$$H_{j,t}(\omega) = h_{j,t} \quad (j = 1, 2, \dots, k). \quad (5)$$

Multiplying both sides of equation (4), by $dZ'^*_{t,x}(\omega)$ and taking expectations, we get

$$\mathbf{F}_{t,xx}(\omega) \mathbf{H}_t(\omega) = \mathbf{F}_{t,xy}(\omega),$$

whence

$$\mathbf{H}_t(\omega) = \mathbf{F}_{t,xx}^{-1}(\omega) \mathbf{F}_{t,xy}(\omega), \quad (6)$$

where

$$\mathbf{F}_{t,xx}(\omega) = E\{d\mathbf{Z}_{t,x}^*(\omega) d\mathbf{Z}_{t,x}'(\omega)\},$$

$$\mathbf{F}_{t,xy}(\omega) = E\{d\mathbf{Z}_{t,x}^*(\omega) d\mathbf{Z}_{t,y}'(\omega)\}.$$

For the type of impulse response functions given by (3) and (5), we have the estimate of $\mathbf{H}_t(\omega)$,

$$\hat{\mathbf{H}}_t = \hat{\mathbf{F}}_{t,xx}^{-1}(\omega) \hat{\mathbf{F}}_{t,xy}(\omega). \quad (7)$$

We note \mathbf{H}_t (or equivalently β_t) is a *gain* vector and hence testing the constancy of the coefficient vector \mathbf{H}_t is equivalent to testing the constancy of gain vector, and this problem has been considered by Subba Rao and Tong (1972, 1973). Briefly this approach is as follows: Estimate $\hat{\mathbf{H}}_t$ at several frequencies, covering the whole frequency range $(-\pi, \pi)$. Since $\hat{\mathbf{H}}_t$ is approximately a multivariate normal, we can perform single-factor multivariate analysis of variance test for the hypotheses

$$\mathbf{H}_{t_1} = \mathbf{H}_{t_2} = \dots = \mathbf{H}_{t_T}$$

on the lines suggested by Subba Rao and Tong (1972, 1973). From the spectral representation of $\{u_t\}$, we have

$$\text{var } u_t = \sigma_t^2 = \int_{-\pi}^{\pi} f_{t,u}(\omega) d\omega.$$

Testing the constancy of $f_{t,u}(\omega)$ for all ω is equivalent to testing the constancy of σ_t^2 . This can be performed following the two-factor analysis of variance technique suggested by Priestley and Subba Rao (1969).

Bohlin (1971) considered the following time domain approach. Consider the time series $\{y_t\}$ generated from the model

$$y_t + a_1(t) y_{t-1} + \dots + a_n(t) y_{t-n} = \lambda e_0(t) + k_t(t), \quad (8)$$

where

$$a_i(t) = a_i(t-1) + q_i e_i(t) \quad (i = 1, 2, \dots, n),$$

$$k(t) = q_{n+1} e_{n+1}(t),$$

where $\{e_i(t)\}$ is a sequence of i.i.d. random variables $N(0, 1)$. Bohlin (1971, equation 5) derived the updating equations for the parameter vector $\theta'(t) = \{a_1(t), a_2(t), \dots, a_n(t)\}$, $k(t)$ based on the sample (y_t, y_{t-1}, \dots) . Assuming $q_1 = q_2 = \dots = q_n = q$ he obtained the maximum-likelihood estimate of q and tested the null-hypothesis $q = 0$. If the null-hypothesis is accepted it implies that the coefficient vector $\theta(t)$ is time invariant.

By choosing $\lambda = 0$, $-y_{t-1} = x_{1t}$, $-y_{t-2} = x_{2t}$, ... *et cetera* we can reduce the model (8) to the one considered by the authors, and hence the problem has been solved by Bohlin (1971) in a more general set-up.

Dr H. TONG (University of Manchester Institute of Science and Technology): The study of the dependency of regression relationships on time is, of course, a very important one and the authors are to be congratulated for bringing forward a very timely paper.

Although the authors have confined themselves to the case of *non-stochastic* regressors, many of the problems studied and some of the results obtained have their counterparts in the case of *stochastic* regressors, as could arise in, for example, a time-dependent system with a stochastic input subject to an additive stochastic noise disturbance. This type of problem has been systematically studied. See, for example, Subba Rao and Tong (1972, 1973), Priestley and Tong (1973) and Tong (1974).

If we pose the problem in the above general form, with stochastic regressors, then tests for constancy of relationships over time have been proposed by Subba Rao and myself (1972, 1973). I am glad to report that these tests have been successfully applied to real data. (See Subba Rao and Tong, 1973, 1974. The latter paper is currently available in the form of a technical report issued from the Department of Mathematics, UMIST.)

Coming back to the problem considered in the paper, I would make the following comments.

(i) In practice, it would seem that restricting the regression model (1) to a *fixed* number, k , of regressors over *all* time is sometimes unrealistic. Just as in the case of stationary autoregressive model building, the determination of the number of regressors is an important problem. In the former case, Dr H. Akaike of the Institute of Statistical Mathematics, Japan, has recently obtained important results and practical procedures (see, for example, Akaike, 1973).

(ii) The idea of using W_r ($r = k+1, \dots, T$) to study the adequacy of the proposed model has its counterparts in the case of stochastic regressors. For example, Mehra and Peschon (1971) have summarized the experiences of control engineers in "Fault detection and diagnosis in dynamic systems" and their approach is based on what is commonly known as the "innovation sequence" which in the control literature, is defined to be $Y_r - Y_{r/r-1}$, $Y_{r/r-1}$ being the minimum linear mean-square predictor of Y_r in terms of Y_{r-1}, Y_{r-2}, \dots .

(iii) In designing moving regressions, it would seem that the choice of the length n is quite important. Have the authors any systematic procedure for its selection? If so, is it in any way related to something like the "maximum width over which the short segment may be regarded to follow one and the same regression relationship"?

(iv) I would just mention that a similar technique as that based on Quandt's log-likelihood ratio has recently been proposed by Ozaki and me in a paper to be presented at the Eighth Hawaii International Conference in Systems Science, 1975, for the detection of abrupt changes over time in auto-regressive relationships. An alternative approach to this problem may be formulated in a Bayesian framework. Recently, some control engineers in Russia have studied this problem by the Bayesian method. See, for example, Telksnys (1973) and the references quoted there.

(v) In the case of stochastic regressors, Jones and Brelsford (1967) have considered Fourier series expansion of regression coefficients, using an approach due to Gladyshev (1961). Here, the problem of determining an *optimal* number of Fourier terms is important, and a Ph.D. student at Manchester (Mrs M. Green) has recently studied this problem. This approach is somewhat similar to the polynomial parameterization suggested in Section 2.6 of the paper, but must be applied with caution if prediction is the objective.

Dr W. G. GILCHRIST (Sheffield Polytechnic): Though every facet of daily life proclaims that the world is non-stationary, the literature of regression tends to ignore it. The methods and plots proposed by the authors provide useful tools that will help those who wish to check their assumptions. I would be interested to know what proportion of the data the authors have looked at satisfies the assumption of constant β .

The methods used by the authors apply least squares to either all the data up to a time r or to a moving sequence of n observations up to that time. Thus the estimates of β treat all the data used as being equally important. The deviations from the fitted model, as represented by the recursive residuals, are used to indicate possible changes in the β 's. An alternative approach is to seek to find "local" estimates of the values of β_t . This can be done, for example, by using discounted least squares, e.g. Gilchrist (1967). Moving forward through time the estimates that minimize the discounted least-squares criteria,

$$\sum a^r \hat{u}_{t-r}^2,$$

are given by

$$\mathbf{b}_r = \mathbf{b}_{r-1} + \mathbf{P}_r \mathbf{x}_r (\mathbf{y}_r - \mathbf{x}'_r \mathbf{b}_{r-1}),$$

where

$$\mathbf{P}_r = [\mathbf{x}'_r \mathbf{A}_r \mathbf{x}_r]^{-1}, \quad \mathbf{A}_r = \begin{pmatrix} a^{r-k-1} & & & \mathbf{0} \\ & \ddots & \ddots & \\ & & a & \\ \mathbf{0} & & & 1 \end{pmatrix}$$

and \mathbf{P}_r may be updated by

$$\mathbf{P}_r = \frac{\mathbf{P}_{r-1}}{a} - \frac{\mathbf{P}_{r-1} \mathbf{x}_r \mathbf{x}'_r \mathbf{P}_{r-1}}{a(a + \mathbf{x}'_r \mathbf{P}_{r-1} \mathbf{x}_r)}.$$

These correspond to equations (4) and (3) in the paper. The above estimate of β_r puts the greatest emphasis on data close to time r . The value of this approach is that we can plot components of \mathbf{b}_r against time and see how the regression coefficients actually vary over time. Clearly the ordinary least-squares estimate is a special case of this when $a = 1$. Where we wish to look at past values of \mathbf{b}_r , a criterion that discounts $\hat{u}_{r+\tau}^2$ by a factor $a^{|\tau|}$ ($\tau = \dots, -1, 0, 1, \dots$) provides the equivalent to the moving regression.

An implication of the possible variation in coefficients, that has been explored by Singleton (1971), is that, in selecting variables for regression, the best variables to select in one locality in time may not be the same as those at a different locality in time.

Professor DURBIN and Mr EVANS replied briefly at the meeting and subsequently in writing as follows:

We agree with Professors Cox and Quandt that it would be useful to compare the powers achieved by a variety of tests, including those we have suggested, against alternatives of interest. Professor Cox has made a useful start but takes the cusum of a fixed number of terms. Where, as in our case, the number of terms varies the mathematics becomes intractable, though of course one could use simulation.

As Mr Fisk indicates, there are many ways of transforming the least-squares residuals into an orthogonal set. This raises the interesting question: which of them is best for a given test statistic and a given alternative hypothesis?

Sir Maurice Kendall, Mr Clarke and Dr Nelder point to the possibility of undesirable build-up of rounding errors. In order to guard against this, all the recursive calculations in our program are performed in double precision and, if the model contains a constant, are carried out on deviations from means. In addition, various checks can be made on the output. First, the final estimates of the regression coefficients after all the recursive calculations have been performed may be compared with those obtained in one step from the entire data set. Secondly, the final cusum of squares may be compared with the theoretical value of unity. Thirdly, in the case of the moving regressions the coefficients obtained from the final segment of the data after all the recursions have been performed may be compared with the corresponding estimates derived from the backward recursive regression. These checks have been carried out as a matter of routine and have rarely shown any sizeable discrepancies. For the three examples in the paper the discrepancies were negligible.

Sir Maurice Kendall's question about how to distinguish between changes in regression coefficients and changes in residual variance lends support to the emphasis we have placed on the examination of the data from several different standpoints. The information on variance change obtained as part of the output of the moving regression technique discussed in Section 2.5 has to be balanced against the information on coefficient changes provided by the other tests. The result of this examination will often indicate fairly clearly which is the more likely explanation. This was the case with our Example 1 as we stated in the paper, notwithstanding Mr Phillips's remark about this example.

Sir Maurice's suggestions on extensions are well worth following up. As regards autoregressive series, Dr Young has referred to his use of recursive residuals for dynamic models and Dr Khan has pointed to the need for treatment of the general model including both exogenous variables and lagged dependent variables. At present we do not know

which of our tests, if any, are valid even asymptotically for models containing lagged dependent variables. That the situation requires care is clear from the study of similar problems for ordinary least-squares residuals (Durbin, 1970). Models containing errors in variables might be amenable to the Kalman technique.

Mr Fisk, Sir Maurice Kendall, Mr Phillips, Mr Harvey and Professor Quandt all point out that the observations can be ordered by criteria other than time. This extends the domain of application of the techniques to other tests of model specification. One could go further and transform the data first and then order by some appropriate quantity. For example, one could follow Duncan and Jones (1966) and transform first to the frequency domain and then use TIMVAR to investigate the stability of the regression relationships with respect to frequency.

We recognize the merit of the work on time-varying techniques of time-series analysis developed by Professor Priestley, Dr Subba Rao and Dr Tong and are grateful for the references to it. For the problem mentioned by Dr Tong of choosing the length n of the moving regression our approach is purely pragmatic. We plot the mean-square one-step-ahead prediction error, derived as indicated in Section 2.5, as a function of n and choose the largest n for which this value has attained its effective minimum. This technique has been used to determine the length of base over which to average in seasonal adjustment work (Durbin and Murphy, 1975).

As regards Professor Priestley's comments on the control theory literature and related remarks by Dr Young, we of course agree that it is important that communications between statisticians and control engineers working on related problems should be kept open. Both contributors have done valuable work in this respect. At the same time we do not think that statisticians are quite as parochial as their remarks might be understood to suggest. The original Kalman (1960) paper was an outstanding achievement which is surely well known to all time-series specialists. Papers relating Kalman's work to statistical problems were published in British statistical journals by Jones (1966) and Walker and Duncan (1967). This Society held an Ordinary Meeting devoted to control theory in 1969 at which papers were read by Wishart (1969), Whittle (1969) and Bather (1969). The principal speaker at the Society's Conference at Nottingham in 1972 was Professor K. J. Åström who is a leading control theorist. These are just a few examples of the influence of control theory on British statistics.

Having included Kalman's equations in lecture courses we were well aware of the relation between them and our relations (2)–(5) and should probably have referred to this. The reason we did not was that our relations owe nothing to Kalman historically. The definition (2) of recursive residuals was used by us in the form of a generalization of the Helmert transformation in lectures in the mid-1950's and is surely "well known" and much older. The remaining relations come from the papers of Plackett and Bartlett referred to.

Dr Young recommends greater emphasis on the recursive estimates of the parameters. These are in fact produced by the program and study of the resulting plots along the lines he suggests has been found useful in practice. Mention of this and other aspects of the work was omitted only in order to try to keep the paper short and simple. We certainly agree with him that if time-variation is found, it is important to investigate the physical nature of the system in order to seek transformations of the data which will yield a system which is time-invariant.

Mr Phillips questions the robustness of the tests against non-normality. We have not investigated this but would conjecture that the tests would be sufficiently robust for most practical work. As regards the effects of serial correlation, raised also by Dr Smith and Professor Quandt, these are likely to be substantial. It would be worth investigating whether simple correction factors along the lines suggested by Professor Cox can be developed.

We thank Professor Anderson for confirming our assertion that at the usual significance levels the probability that a sample path crosses both lines is negligible. In fact the results in his (1960) paper could be used to produce a complete set of percentage points for a .

Dr Smith's suggestion of using a V-mask is a very good one which can certainly be expected to lead to a gain in power against some types of alternative. His development of related Bayesian techniques is to be encouraged, though we believe that our own approach, in which one looks at the data from a variety of standpoints and uses different test statistics to measure different types of departure from the null-hypothesis, is more informative when there is no specific alternative in mind at the outset.

We agree with Professor Ehrenberg that in particular practical situations a model based on stochastic regressors might well be appropriate. But our approach in such a situation would be simply to perform a conditional analysis given the observed values of the regressors. This would immediately reduce to the model considered in the paper. Leaving aside sampling fluctuations, we cannot follow his argument that when the regressors are stochastic two half-samples would necessarily give different regression lines.

Mr Harvey makes a useful point about the choice of the estimator of σ and we agree that the use of his estimator should lead to an increase of power. As regards the use of exponential weights, suggested by him and also by Dr Gilchrist, we considered this possibility and obtained the relevant formulae at an early stage of the work, but gave up the idea and we have not, in fact, ever used the formulae in practice.

Dr Herzberg makes an imaginative suggestion but we have not been able to put it to the test and find it hard to evaluate the merit of the idea in the abstract. This is one of those cases where the proof of the pudding will be in the eating.

We wish to encourage Mr Hutchison to produce his table which will be a useful contribution. While the use of TIMVAR after randomization within years as he suggests will give a test which is valid in the sense of giving the right rejection probabilities on the null-hypothesis, we feel that power would inevitably be lost relative to the corresponding test based only on year-to-year changes.

We have some sympathy with Dr Nelder's remarks on the use of the word "recursive" but feel that the usage is too well-established to change now. Dr Young has referred to the use of the term "innovations process" in the control literature. In the time-series literature (Wiener and Masani, 1957; Cramér, 1961) the term "innovation" is normally used in connection with stochastic processes whose realizations are of infinite length. The innovation at time t is then defined as the difference between the observation at time t and its best linear predictor given all the observations up to and including time $t-1$. If one were to regard y_1, \dots, y_T as a sample from an underlying infinite population, one would take u_t in (1) as the innovation. Extending the definition to finite samples of data, however, one could take as the "sample innovation" at time t the difference between y_t and its best linear estimate based on y_1, \dots, y_{t-1} , i.e. $y_t - x'_t b_{t-1}$, $t = k+1, \dots, T$ in our situation. But it would still be necessary to introduce a new term, such as our term "recursive residual", to denote the standardized residual w_t .

We are grateful to Professor Quandt for the brief summary of recent work by him and his colleagues. It is a matter for regret that there are no distributional results available for the original Quandt log-likelihood ratio statistic considered in Section 2.6. The idea of replacing the discontinuous D_t by a continuous approximation is ingenious and we hope it will make these difficult problems more tractable.

REFERENCES IN THE DISCUSSION

- AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. In *International Symposium in Information Theory*, pp. 267–281 (B. N. Petrov and F. Csaki, eds). Budapest: Akadémiai Kiadó.
- Annals of Economic and Social Measurement* (1973), 2, No. 4 (whole issue).
- ANDREWS, D. F. (1972). Plots of high-dimensional data. *Biometrics*, 28, 125–136.
- BATHER, J. A. (1969). Diffusion models in stochastic control theory (with Discussion). *J. R. Statist. Soc. A*, 132, 335–352.
- BOHLIN, T. (1968). Information pattern for linear discrete-time models with stochastic coefficients. Report No. T.P. 18.192. I.B.M. Nordic Laboratory.

- BOHLIN, T. (1971). Analysis of EEG-signals with changing spectra. Report No. T.P. 18.212, I.B.M. Nordic Laboratory.
- BOX, G. E. P. and JENKINS, G. M. (1970). *Time Series Analysis, Forecasting and Control*. San Francisco: Holden-Day.
- CAREW, B. and BELANGER, P. R. (1973). Identification of optimum filter steady state gain for systems with unknown noise covariances. *I.E.E.E. Trans. on Aut. Cont.*, **AC-18**, 582–587.
- CHAMBERS, J. M. (1971). Regression updating. *J. Amer. Statist. Ass.*, **66**, 744–748.
- CHOW, G. C. (1960). Tests of the equality between two sets of coefficients in two linear regressions. *Econometrica*, **28**, 561–605.
- CRAMÉR, H. (1961). On some classes of nonstationary stochastic processes. *Proc. 4th Berk. Symp.*, **2**, 57–78.
- DAVIS, O. A., DEMPSTER, M. A. H. and WILDAVSKY, A. (1966). A theory of the budgetary process. *Amer. Polit. Sci. Rev.*, **60**, 529–547.
- DUNCAN, D. B. and JONES, R. H. (1966). Multiple regression with stationary errors. *J. Amer. Statist. Ass.*, **61**, 917–928.
- DURBIN, J. (1954). Errors in variables. *Rev. Int. Statist. Inst.*, **22**, 23–32.
- (1970). Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables. *Econometrica*, **38**, 410–421.
- DURBIN, J. and MURPHY, M. J. (1975). Seasonal adjustment based on a mixed additive-multiplicative model. *J. R. Statist. Soc. A*, **138**, Part 3 (to appear).
- EHRENBERG, A. S. C. (1975). *Data Reduction*. London and New York: Wiley.
- FAIR, R. C. and JAFFEE, D. M. (1972). Methods of estimation for markets in disequilibrium. *Econometrica*, **40**, 497–514.
- FARLEY, J. U. and HINICH, M. J. (1970). Testing for a shifting slope coefficient in a linear model. *J. Amer. Statist. Ass.*, **65**, 1320–1329.
- FARLEY, J. U., HINICH, M. J. and MCGUIRE, T. W. (1973). Testing for a shift in the slopes of a multivariate linear time series model. Working paper, Carnegie-Mellon University, June 1973.
- GILCHRIST, W. G. (1967). Methods of estimation using discounting. *J. R. Statist. Soc. B*, **29**, 355–369.
- GLADYSHEV, E. G. (1961). Periodically correlated random sequences. *Soviet Maths*, **2**, 385–388.
- GOLDFELD, S. M. and QUANDT, R. E. (1974). The estimation of structural shifts by switching regressions. *Ann. Econ. Soc. Measur.*, **2**, 475–485.
- GREEN, M. and HARRISON, P. J. (1973). Fashion forecasting for a mail-order company using a Bayesian approach. *Oper. Res. Quart.*, **24**, 193–205.
- HANCOCK, R. A. (1971). Problems relating to the identification of linear stochastic systems. Ph.D. Thesis, Department of Engineering, University of Cambridge.
- HARRISON, P. J. and STEVENS, C. F. (1971). A Bayesian approach to short-term forecasting. *Oper. Res. Quart.*, **22**, 341–362.
- HARVEY, A. C. and COLLIER, P. (1975). Testing for functional mis-specification in regression analysis. University of Kent Q.S.S. Discussion Paper No. 12.
- HARVEY, A. C. and PHILLIPS, G. D. A. (1974). A comparison of the power of some tests for heteroscedasticity in the general linear model. *J. Econometrics*, **2**, 307–316.
- HILDRETH, C. and HOUCK, J. P. (1968). Some estimators for a linear model with random coefficients. *J. Amer. Statist. Ass.*, **63**, 584–595.
- JONES, R. H. (1966). Exponential smoothing for multivariate time series. *J. R. Statist. Soc. B*, **28**, 241–251.
- JONES, R. H. and BRELSFORD, W. M. (1967). Time series with periodic structure. *Biometrika*, **54**, 403–408.
- KAILATH, T. (1968). An innovations approach to least squares estimation—Part I. Linear filtering in additive white noise. *I.E.E.E. Trans. on Aut. Cont.*, **AC-13**, 655–660.
- (1974). A view of three decades of linear filtering theory. *I.E.E.E. Trans. on Inf. Theory*, **IT-20**, 145–181.
- KALMAN, R. E. (1960). A new approach to linear filtering and prediction problems. *Trans. A.S.M.E., J. Basic Engng*, **82**, 35–45.
- MEHRA, R. K. (1970). On the identification of variances and adaptive Kalman filtering. *I.E.E.E. Trans. on Aut. Cont.*, **AC-15**, 175–184.
- MEHRA, R. K. and PESCHON, J. (1971). An innovation approach to fault detection and diagnosis in dynamic systems. *Automatica*, **7**, 637–640.

- NEETHLING, C. G. and YOUNG, P. C. (1974). Comments on "Identification optimal filter steady-state gain for systems with unknown noise covariances". *I.E.E.E. Trans. on Aut. Cont.*, AC-19, 623-624.
- PHILLIPS, G. D. A. and HARVEY, A. C. (1974). A simple test for serial correlation in regression analysis. *J. Amer. Statist. Ass.*, 69, 935-939.
- PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc. B*, 27, 204-237.
- PRIESTLEY, M. B. and SUBBA RAO, T. (1969). A test for non-stationarity of time series. *J. R. Statist. Soc. B*, 31, 140-149.
- PRIESTLEY, M. B. and TONG, H. (1973). On the analysis of bivariate non-stationary processes (with Discussion). *J. R. Statist. Soc. B*, 35, 153-188.
- ROSEN BROCK, H. (1965). Sur les relations entre les filtres linéaires discrets et quelques formules de Gauss. Conference on *Identification, Optimisation et Stabilité des Systèmes Automatiques*. Paris:
- SINGLETON, P. W. (1971). Multiple regression methods in forecasting. M.Phil. Thesis, University of London.
- SUBBA RAO, T. and TONG, H. (1972). A test for time dependence of linear open-loop systems. *J. R. Statist. Soc. B*, 34, 235-250.
- (1973). On some tests for the time dependence of a transfer function. *Biometrika*, 60, 589-597.
- (1974). Linear time-dependent systems. *I.E.E.E. Trans. on Aut. Cont.*, AC-19, 735-737.
- TELKSNYS, L. (1973). Determination of changes in the properties of random processes by the Bayes method. In *Identification and System Parameter Estimation* (P. Eykoff, ed.). Amsterdam: North-Holland.
- TONG, H. (1974). On time-dependent linear transformations of non-stationary stochastic processes. *J. Appl. Prob.*, 11, 53-62.
- WALKER, S. H. and DUNCAN, D. B. (1967). Estimation of the probability of an event as a function of several independent variables. *Biometrika*, 54, 167-179.
- WHITEHEAD, P. G. and YOUNG, P. C. (1975). A dynamic-stochastic model for water quality in part of the Bedford-Ouse river system. In *Proc. IFIP Working Conference on Modelling and Simulation of Water Resource systems*. Ghent, Belgium: (to appear).
- WHITTLE, P. (1969). A view of stochastic control theory (with Discussion). *J. R. Statist. Soc. A*, 132, 320-334.
- WIENER, N. and MASANI, P. (1957). The prediction theory of multivariate stochastic processes. *Acta Math.*, 98, 111-150.
- WISHART, D. M. G. (1969). A survey of control theory (with Discussion). *J. R. Statist. Soc. A*, 132, 293-319.
- YOUNG, P. C. (1969). Applying parameter estimation to dynamic systems. *Control*, 16, (10), 119-125; 16 (11), 118-124.
- (1971). Comments on "Dynamic equations for economic forecasting with G.D.P.—unemployment relation and growth of G.D.P. in the U.K. as an example". *J. R. Statist. Soc. A*, 134, 220-223.
- (1972). Comments on "On line identification of linear dynamic systems with application to Kalman filtering". *I.E.E.E. Trans. on Aut. Cont.*, AC-17, 269-270.
- (1974). Recursive approaches to time series analysis. *Bull. Inst. Maths. & Applic.*, 10, 209-224.
- YOUNG, P. C. and HASTINGS-JAMES, R. (1970). Identification and control of discrete dynamic systems subject to disturbances with rational spectral density. In *Proc. 9th I.E.E.E. Symp. on Adaptive Processes: Decision and Control*. New York: I.E.E.E.
- YOUNG, P. C., SHELSWELL, S. H. and NEETHLING, C. G. (1971). A recursive approach to time series analysis. Department of Engineering, Cambridge University, Report No. CUED/B-Control/TR16.
- YOUNG, P. C. and WHITEHEAD, P. G. (1975). A recursive approach to time series analysis for multivariate systems. In *Proc. IFIP Working Conference on Modelling and Simulation of Water Resource Systems*. Ghent, Belgium: (to appear).