Studies in Official Statistics
Research Series no. 4

CSO

Seasonal adjustment of unemployment series

R. L. Brown, A. H. Cowley and J. Durbin

NATIONAL INSTITUTE OF
12 OCT 1971
ECONOMIC & SOCIAL RESEARCH

Her Majesty's Stationery Office Price 50p net

STUDIES IN OFFICIAL STATISTICS

This collection forms a useful set of reference papers on a variety of statistical sources and methods.

The list below gives a selection of the titles still in print and available from Her Majesty's Stationery Office.

No. 6 Method of Construction and Calculation of the Index of Retail Prices (February 1967)

No. 13 National Accounts Statistics: Sources and Methods (September 1968)

No. 14 Agricultural and Food Statistics: A Guide to Official Sources (May 1969)

No. 15 New Contributions to Economic Statistics (Fifth Series) (Reprinted from *Economic Trends*, May 1967—August 1969) (January 1970)

No. 16 Input—Output Tables for the United Kingdom 1963 (March 1970)

No. 17 The Index of Industrial Production and other Output Measures (November 1970)

No. 18 Qualified Manpower in Great Britain—The 1966 Census of Population (June 1971)

The present paper is No. 4 in the separate RESEARCH SERIES of the STUDIES. The earlier titles published are:

No. 1 Duration of unemployment on the register of wholly unemployed (R. F. Fowler) (August 1968)

No. 2 The numbers of coins in circulation in the United Kingdom (R. G. de Glanville) (January 1970)

No. 3 Some problems of index number construction (R. F. Fowler) (March 1970)

Published by
HER MAJESTY'S STATIONERY OFFICE

and obtainable from the Government Bookshops at the addresses listed on back page of cover (post orders to P.O. Box 569, London SE1) or through booksellers

STUDIES IN OFFICIAL STATISTICS
RESEARCH SERIES No. 4

Seasonal adjustment of unemployment series

R. L. Brown *Chief Statistician,*
A. H. Cowley *Assistant Statistician*
and **J. Durbin*** (*Consultant*)

CENTRAL STATISTICAL OFFICE

* *Professor of Statistics, London School of Economics and Political Science, University of London*

LONDON
HER MAJESTY'S STATIONERY OFFICE 1971

Preface

In 1968 it was found that the method of seasonal adjustment then being used for the unemployment series was over-adjusting so that the series appeared to be rising or falling when in fact the underlying trend was substantially flat. Since it is often easier to interpret the movement of economic indicators when they have been seasonally adjusted, the Research and Special Studies Division of the Central Statistical Office was invited by the Department of Employment to investigate what had happened.

The unemployment series has both a constant and a trend dependent seasonal component and a regression method was used to estimate a mixture of additive and multiplicative monthly seasonal factors. The detailed investigation covered the problems of estimating the trend and the significant seasonal factors, together with modification of extreme values such as occurred for example in the hard winter of 1963. The research showed that it was possible to make estimates of the seasonality in recent years because although the size of the seasonal movement had fallen (leading to the overadjustment mentioned above) the relationship between individual months was largely unchanged during the historical period examined.

The method developed in this work was adopted officially in April 1970 and has since been applied to component series differentiated by sex, by industry and by region. It has also been used for the national series for vacancies and it may prove to have wider applications.

Contents

1. Introduction

1.1 Reason for undertaking the research

The interpretation of the movement of economic time-series is often helped if the series is adjusted so as to eliminate the seasonal variation. This gives a seasonally adjusted series consisting of the trend (sometimes called the trend-cycle) and an irregular component. The latter has two parts: (i) chance or random variations arising from numerous causes that are not individually identifiable, (ii) extreme movements usually having identifiable causes, e.g. severe winter weather or strikes, and usually distinguishable from the smaller random movements. Thus the estimation of the seasonal variation requires the estimation of the trend and the identification and modification of extreme values. Because of the random component it is usual to estimate an average seasonal variation over a number of years and to use this average estimator for adjusting current observations.

It sometimes happens that current seasonality differs markedly from the average estimated seasonal variation based on past experience. This has occurred in recent years with the series for wholly unemployed (excluding school leavers), where the average seasonality based on experience up to 1965 has resulted in over-adjustment of the series since 1966. This is evident from Figure 1, which shows the original unadjusted and the seasonally adjusted series. The research reported in this paper was undertaken to find out what had happened and to devise a new method for seasonal adjustment. A brief description of the new method has already been published (Brown, Cowley, Durbin, 1970)*. This paper gives a more complete account. The new method was adopted officially in April 1970 (*Employment and Productivity Gazette*, 1970).

The previous method used by the Department of Employment (called here the 1965-70 method) was introduced in 1965 (*Ministry of Labour Gazette*, 1965) since an earlier method (*Ministry of Labour Gazette*, 1960) was not proving satisfactory. These two methods are described in the remainder of this section.

Most of the work discussed in later sections was done with the series for wholly unemployed (excluding school leavers), which for simplicity we shall call the total series. Some results are given for the national series for males and females separately.

1.2 The 1965-70 method

Prior to 1965 a simple additive method was used by the Ministry of Labour (now the Department of Employment). The method may be regarded as derived from the following model:

$$z_{ij}-\xi_{ij} = \alpha_j+\varepsilon_{ij} \qquad \left.\begin{array}{l} i = 1, 2\ldots, n \\ j = 1, 2\ldots, 12 \end{array}\right\} \qquad \ldots(1)$$

where ξ_{ij} is the trend and α_j are the monthly seasonal factors constrained to sum to zero, ε_{ij} are the irregular or error components. However this simple model did not give satisfactory adjustments and another model was introduced officially in September 1965 (*Ministry of Labour Gazette*, 1965). This was based on the regression model

$$z_{ij}-\xi_{ij} = \alpha_j+\beta_j\xi_{ij}+\varepsilon_{ij} \qquad \ldots(2)$$

where β_j are multiplicative seasonal factors and α_j are, as before, additive factors. The factors were constrained by:

$$\sum_1^{12} \alpha_j = \sum_1^{12} \beta_j = 0 \qquad \ldots(3)$$

so that, at a constant trend level, the annual totals of the original and adjusted series are approximately equal. In formulating the regression model a constant corresponding to the general mean has been excluded. This is because the sum of the series at constant trend level and the sum of the trends are equal if the series is infinite in length and are nearly equal for a finite series. The seasonally adjusted series $\tilde{z}_{ij}$ was calculated from

$$\tilde{z}_{ij} = \frac{z_{ij}-a_j}{1+b_j} \qquad \ldots(4)$$

where a_j, b_j are estimates of α_j, β_j subject to the constraints (3). The trend was estimated by a 12-month centred moving average of z_{ij}. Observations over the 16 years from June 1949 to May 1965 were used. Since the trend was not calculated for the first and last 6 months of the record, the regression was fitted to deviations from trend for the 15 years from December 1949 to November 1964. We shall call the period over which regressions are fitted the regression base.

If the series has an additive seasonality, the β_j's are all zero. For a multiplicative seasonality the α_j's are all zero. The advantage of the regression model is that it enables one to deal with the intermediate case in which both additive and multiplicative effects are present.

Seasonal factors for males, females and the total series are given in Table 1. This table suggests that the seasonality of the series for females is more additive than that for males. To make this more precise, it is desirable to relate the seasonal factors to the trend levels characteristic of each series.

* Detailed references are given at the end of the paper.

The fitted regression may be written

$$z_{ij}-x_{ij} = a_j+b_jx_{ij}+r_{ij} \qquad \ldots(5)$$

where x_{ij} is the trend estimated by a centred 12-month moving average of the z_{ij}'s, a_j and b_j are the estimated seasonal factors, and the r_{ij}'s are the residuals from the regression. Then the amount of the seasonal adjustment is

$$z_{ij}-\tilde{z}_{ij} = z_{ij}-x_{ij} - \frac{r_{ij}}{1+b_j}$$

$$= (a_j+b_jx_{ij}) + \left(\frac{b_jr_{ij}}{1+b_j}\right) \qquad \ldots(6)$$

We shall call $(a_j+b_jx_{ij})$ the seasonal variation. Since $b_j r_{ij}$ is generally small, the seasonal variation approximately equals the seasonal adjustment and it may therefore be used to exhibit the effects of the seasonal adjustment procedure.

If x_1 is a low and x_2 a high trend level, put

$$\left.\begin{aligned} S_{1,j} &= a_j+b_jx_1 \\ S_{2,j} &= a_j+b_jx_2 \end{aligned}\right\} \qquad \ldots(7)$$

We now define the amplitude of the seasonal variation as half the distance from the peak to the trough, i.e.

$$A_k = \tfrac{1}{2}\left\{ \max_j |\text{ positive } S_{k,j}| + \max_j |\text{ negative } S_{k,j}| \right\}, \quad k = 1, 2 \qquad \ldots(8)$$

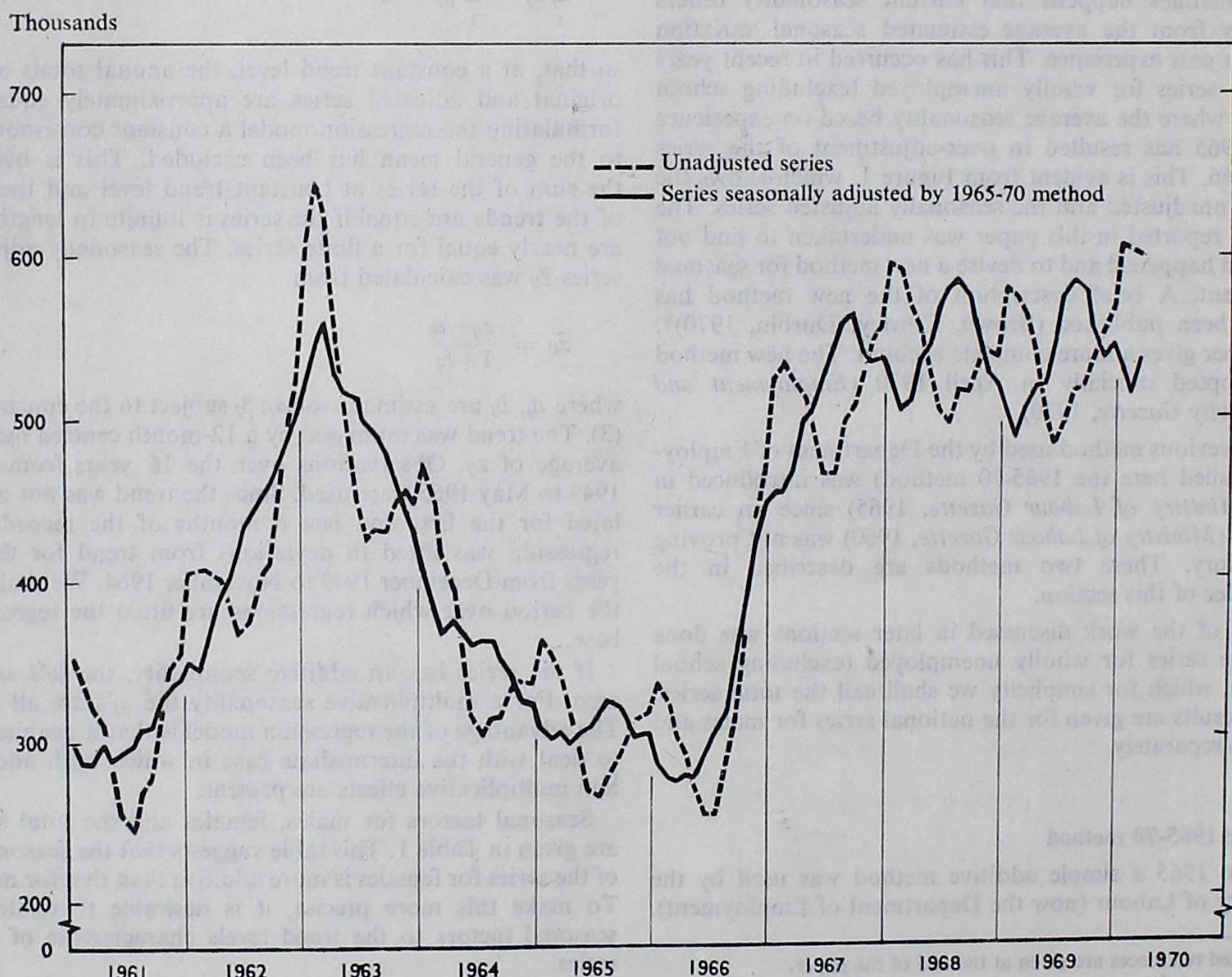
For the purpose of exhibiting the dependence of the amplitude on trend we shall use the ratio

$$M = \left(\frac{x_2+x_1}{x_2-x_1}\right) \left(\frac{A_2-A_1}{A_2+A_1}\right) \qquad \ldots(9)$$

If the peaks and troughs of S_1, S_2 occur in the same months, M is unity for a multiplicative model and zero for an additive model. Thus we call M the degree of multiplicativity. M summarises the information about the nature of the seasonality of the series when fitted by model 5.

The values of x_1, x_2, $S_{1,j}$, $S_{2,j}$, A_1, A_2 and M for the three series of Table 1 are given in Table 2. It is not easy to estimate the standard error of M, but it is high and the values are reported only to one decimal place. The table shows that the series for males is multiplicative or nearly so, whereas that for females is additive or nearly so. This is in line with the finding that the multiplicativity of the total series is 0·8. The series for the six broad industry groups also have a range of multiplicativities from nearly additive to nearly multiplicative. Thus, the regression model appears to be well suited for seasonally adjusting unemployment series.

Figure 1 **Wholly unemployed (excluding school-leavers)**



1.3 Aims of the research

It was felt early in 1968 that the 1965-70 method might not be giving satisfactory results and it was decided to initiate an investigation of the recent behaviour of the series. The situation at that time can be seen by blanking off the observations since mid-1968 in Figure 1. A preliminary examination of the problem convinced us that the regression model (2) provided an appropriate basis for studying the seasonal behaviour of the series. The remainder of our work therefore assumed the basic validity of this model.

In order to find out what had happened it was necessary to undertake a considerable amount of research and development concerned *inter alia* with (i) modification of extreme values, such as occurred for example in the hard winter of 1963, (ii) choice of base period over which the seasonal variation is estimated, so as to take account of changing seasonality, (iii) improvements in methods of estimating the seasonal factors and the underlying trend, and (iv) writing of a computer program for carrying out the calculations.

This work led to the adoption of the new method of seasonal adjustment in April 1970. Not all stages of the work that led to the various features of the new method will be described in this paper. We shall confine ourselves on the whole only to those parts of the work directly related to the final form of the method.

The first question discussed is the estimation of trend. This is dealt with in Section 2 where the two procedures used in this paper are described and the effect of replacing the trend ξ_{ij} in (2) by a moving average is discussed. The third section is concerned with the fitting of the regression; it describes the use of a Fourier transformation so that instead of estimating seasonal constants for each month separately the seasonality is considered in terms of an annual, a six month, etc. cycle. This transformation makes the use of stepwise fitting for the regression possible so that only those cycles that are significantly non-zero are estimated. Section 4 discusses the effect of extreme values on regression estimates and describes the two-stage method of modifying such values. It is shown in Section 5 that the unsatisfactory results of the 1965-70 method in recent years are largely due to a major change in seasonal amplitude during 1967. A method of prior amplitude scaling by which this change can be taken account of is introduced together with a system of local amplitude scaling factors for monitoring the behaviour of the series and correcting for small changes in amplitude. Section 6 combines the various pieces of development work discussed in the preceding sections into the new seasonal adjustment method. For clarity a large part of this section is written in tabular form with a close correspondence to the steps of the computer program. Some aspects of the use of the method with particular reference to future behaviour of the series are discussed in Section 7, the final section.

Seasonal factors for the unemployment series for males, females and total: the 1965-70 method (1)

Table 1

more digits are used to find $\tilde{z}_{ij}$

Month	Series: Males		Females		Total	
	a per thousand	100b	a per thousand	100b	a per thousand	100b
January	11	14	18	− 6	21	11
February	− 3	21	16	− 2	0	18
March	− 9	16	8	3	− 3	13
April	1	4	3	2	2	4
May	0	− 4	−13	14	− 3	− 2
June	− 2	−12	−16	4	−11	− 9
July	− 5	−14	−17	− 2	−19	−11
August	− 3	−11	−12	− 4	−16	− 9
September	− 1	−11	− 4	− 5	0	−11
October	3	− 6	4	− 2	11	− 5
November	3	0	5	2	11	0
December	5	2	8	− 4	7	2

(1) The series considered are:

(i) Males registered as wholly unemployed (excluding school leavers)

(ii) Females registered as wholly unemployed (excluding school leavers)

(iii) Total registered as wholly unemployed (excluding school leavers)

Table 2 **Seasonal variations using the 1965-70 method**

		Series					
		Males		Females		Total	
Trend level	thousands	225	450	45	90	250	500
Seasonal variation	thousands						
January		43	75	15	13	48	74
February		43	90	15	14	45	89
March		27	62	9	11	28	60
April		10	19	4	5	12	22
May		−9	−18	−7	−1	−7	−12
June		−29	−56	−14	−12	−34	−58
July		−36	−67	−18	−19	−47	−74
August		−28	−53	−14	−16	−38	−61
September		−25	−49	−6	−8	−26	−53
October		−9	−22	3	2	−3	−17
November		4	5	6	7	11	12
December		10	15	6	4	12	17
Amplitudes	thousands	40	78	17	16	47	82
Degree of multiplicativity, *M*		1·0		0·0		0·8	

2. Trend estimation

2.1 Moving averages and frequency response functions

The trend ξ_{ij} in (2) is not itself observed and must therefore be estimated from the data. We decided to use a moving average estimator for this purpose as is almost invariably done in the seasonal adjustment of economic time series. It is convenient to discuss the properties of the trend estimator in terms of its effects at (i) the seasonal frequencies, which have periods of 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$, and $\frac{1}{6}$ years corresponding respectively to frequencies of 30, 60, 90, 120, 150 and 180 degrees, and (ii) the lower frequencies, particularly those near 6 to 7·5 degrees which characterise the underlying cycle, and those nearer zero degrees which may be regarded for the present purpose as characterising the long term trend. We shall use the terms filter and moving average estimator indifferently. The purpose of the trend estimator is to pass as much as possible of the low frequencies and to filter out as much as possible of the remaining frequencies, particularly the seasonal frequencies. If the filter is chosen so that it reproduces exactly or approximately a local polynomial of cubic or higher order in time, then frequencies lower than the annual will be passed almost unchanged. Unfortunately, however, although such a filter does not normally pass the six month and higher frequencies it does pass some of the annual seasonal frequency, a point of particular importance with unemployment series for which seasonality is dominated by the 12-month cycle. On the other hand, the centred 12-month moving average, which does not pass any of the six seasonal frequencies, only reproduces a linear trend exactly and therefore gives a trend estimate that is biased at the peaks and troughs of the series.

The effectiveness of a filter in meeting these conflicting requirements depends on the number of observations used. Since a symmetrical filter through an odd number $(2m+1)$ of points gives a trend estimate at its mid-point for which the phase at any frequency is unchanged, it is desirable to use such filters. Then the filter provides a trend estimate short of the ends of the series by m months. Although trend can be estimated for these m months, this requires either a notional continuation of the series or the use of asymmetric filters. Both these methods introduce amplitude and phase changes in the low and seasonal frequencies and we felt that these were undesirable features. We therefore confined our attention to symmetrical filters.

We chose for our work two short filters both through 13 points. Besides the centred 12-month moving average we have used a filter developed by Burman (1965) which reproduces locally a cubic polynomial in time and removes most of the six monthly and higher seasonal frequencies.

The properties of these two filters can be expressed simply in terms of their frequency response functions T, which show the effect of the filter on particular frequencies ψ of the series being filtered. T is calculated as follows.

Suppose the moving average has weights w_p $(p = -m, -m+1, \ldots, m)$, then the frequency response function for the frequency ψ in the series being filtered is given by,

$$T(\psi) = w_o + 2 \sum_{p=1}^{m} w_p \cos p\psi \quad \ldots \quad (10)$$

Figure 2

Frequency response functions of moving averages

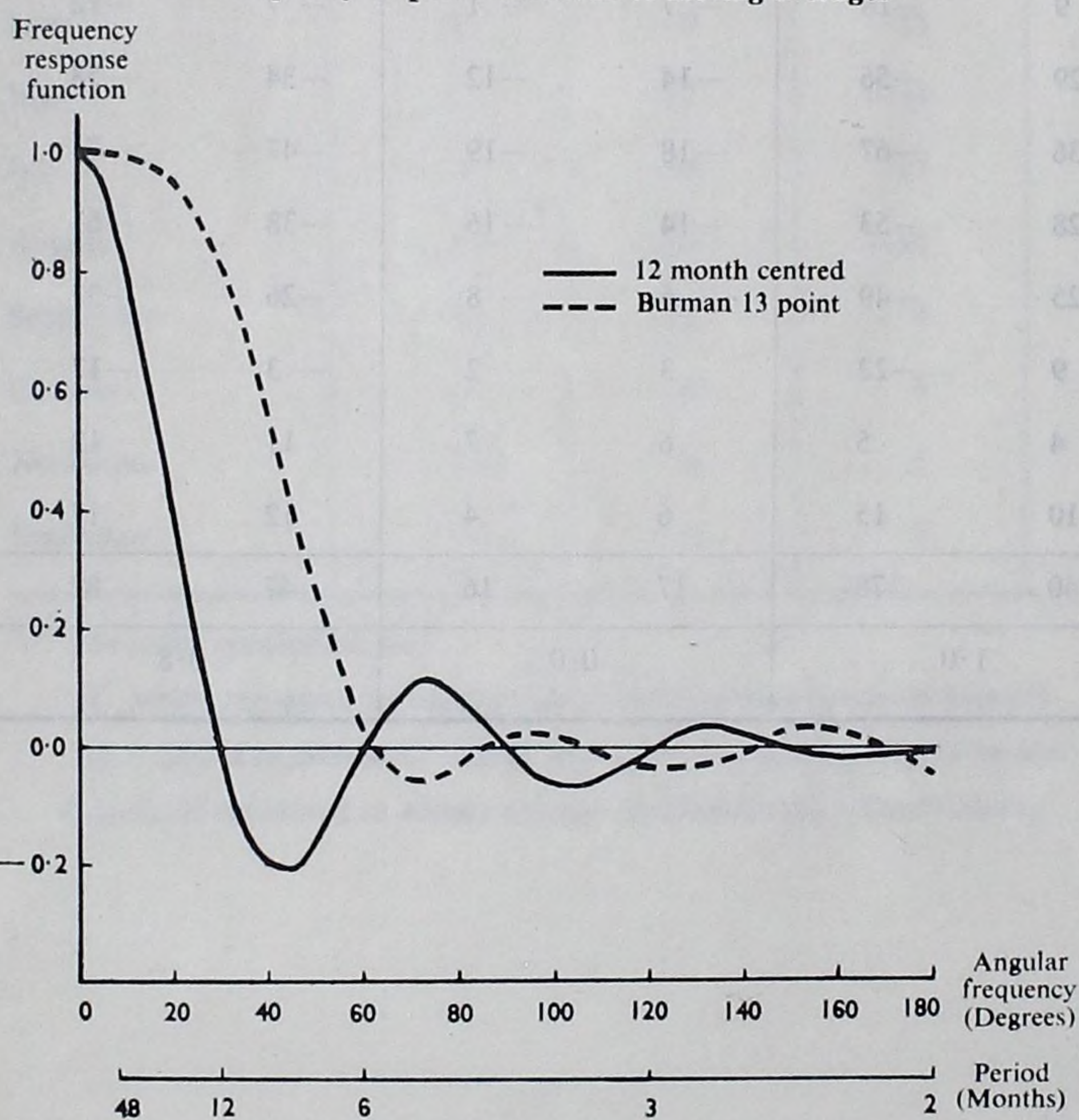
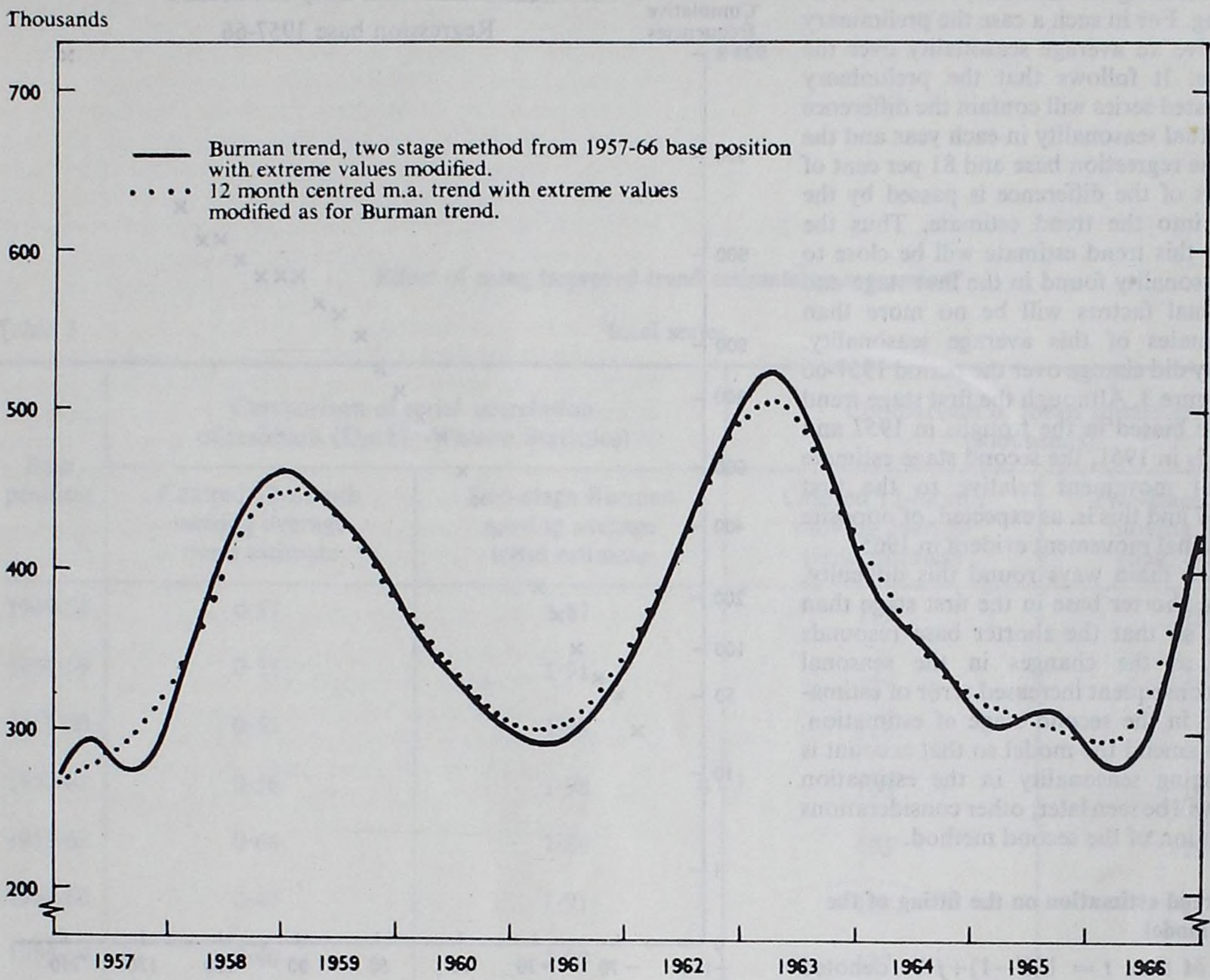


Figure 3 **Improved trend estimation**



The weights for the 12-month centred filter are

$$\frac{1}{24}(1, 2, 2, 2, 2, 2, 2)$$

and for the Burman 13-point filter are

(−0·0331, −0·0208, 0·0152, 0·0755, 0·1462, 0·2039, 0·2262)

The frequency response functions are plotted in Figure 2. The centred 12-month moving average gives $T(\psi)$ equal to zero at the seasonal frequencies and about 0·9 at a frequency of 7°, corresponding to a period of 4·3 years, which is close to the period of the underlying cycle in the series. This means that 0·1 of the underlying cycle is removed as well as the seasonal frequencies. This is due to the fact that the 12-month centred moving average does not penetrate into the peaks and troughs of the series.

With the Burman filter, $T(\psi)$ is almost equal to unity at 7°, but this improvement is at the expense of the annual seasonal cycle (30°) at which $T(\psi)$ is now 0·81. Thus the Burman filter passes about four-fifths of the annual cycle as well as the trend. This filter was designed to be comparable to the Spencer 15-point moving average.

With an additive model, the amplitude at seasonal frequencies can be restored by dividing the Fourier transform of the deviations from trend by $(1-T)$, but with the regression model there is no obvious way of making this restoration.

2.2 Two-stage method of estimating trend

The usual way of overcoming the deficiencies of filters is to adopt a two-stage method of estimating the trend. Preliminary seasonal factors are estimated using a centred 12-month moving average and a preliminary seasonal adjustment is made. The adjusted series is then filtered, using the Burman 13-point filter to give an improved trend estimate: when residual seasonality in the preliminary adjusted series is small, the fact that the Burman filter passes 81 per cent of the residual annual cycle is not important. The seasonal factors are then recalculated using this improved trend estimate. The improvement in the trend estimation is shown in Figure 3, which shows how the two-stage procedure gives a trend estimate higher at the peaks of the series in 1958-59 and 1963 and lower at the trough of the series in 1961.

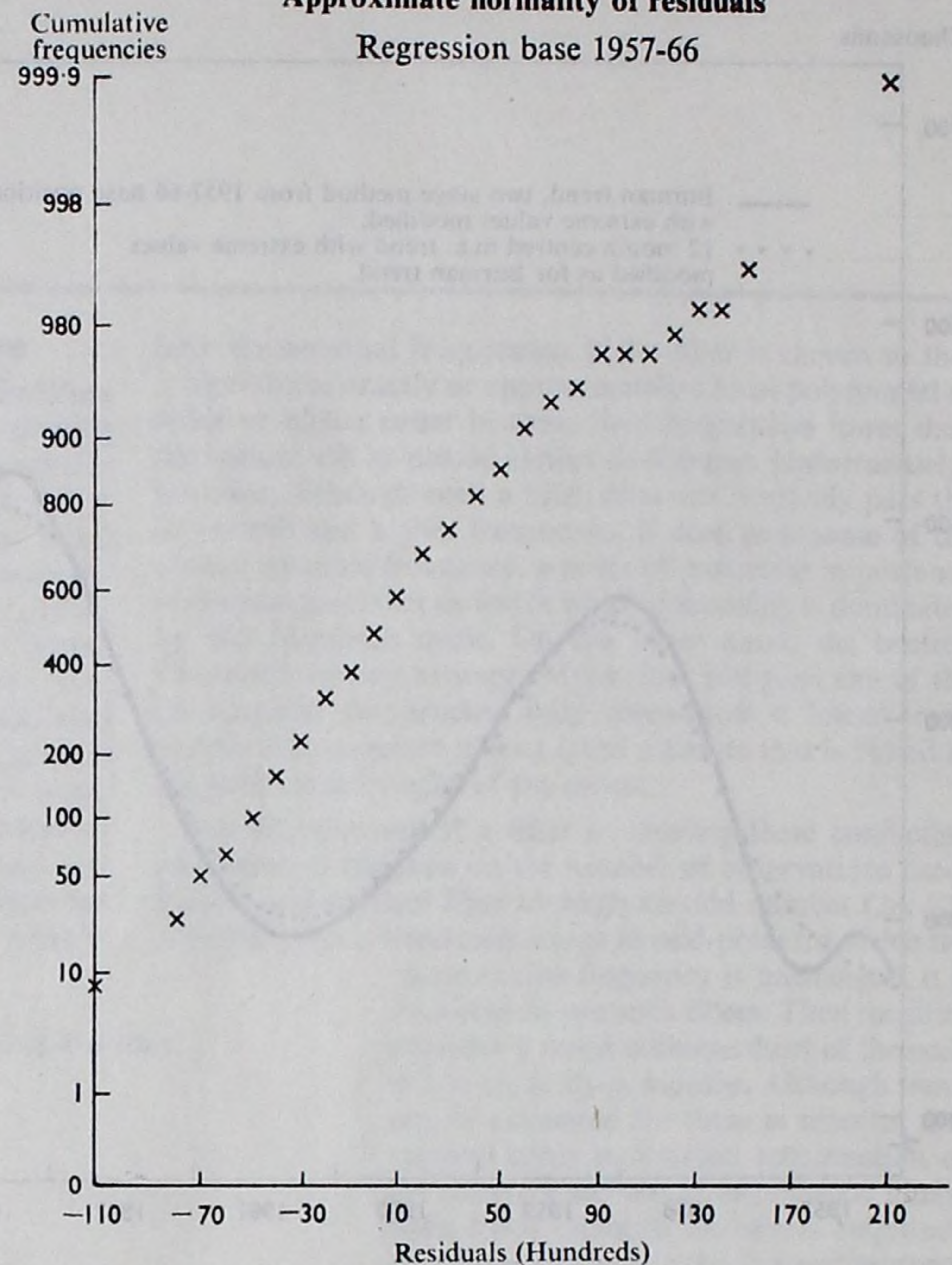
This method has been adopted in the present work, notwithstanding its deficiency when seasonality is changing. For in such a case the preliminary estimate will give an average seasonality over the regression base. It follows that the preliminary seasonally adjusted series will contain the difference between the actual seasonality in each year and the average over the regression base and 81 per cent of the annual part of the difference is passed by the Burman filter into the trend estimate. Thus the deviation from this trend estimate will be close to the average seasonality found in the first stage and the final seasonal factors will be no more than improved estimates of this average seasonality. That seasonality did change over the period 1957-66 is evident in Figure 3. Although the first stage trend estimate will be biased in the troughs in 1957 and 1965, just as it is in 1961, the second stage estimate has a seasonal movement relative to the first estimate in 1957 and this is, as expected, of opposite sign to the seasonal movement evident in 1965.

There are two main ways round this difficulty. One is to use a shorter base in the first stage than in the second, so that the shorter base responds more quickly to the changes in the seasonal variation; the consequent increased error of estimation is reduced in the second stage of estimation. The other is to amend the model so that account is taken of changing seasonality in the estimation procedure. As will be seen later, other considerations led to the adoption of the second method.

Figure 4

Approximate normality of residuals

Regression base 1957-66



2.3 Effect of trend estimation on the fitting of the regression model

Let the trend at time $t = 12(i-1)+j$ be denoted indifferently by ξ_{ij} or ξ_t and let the corresponding estimate be denoted similarly by x_{ij} or x_t. We must now consider the implications for the fitting of the regression model of the replacement of ξ_{ij} by x_{ij} in (2). In order to simplify the discussion we shall consider the first order of approximation only. Let us take the first stage of fitting based on the centred 12-month estimate,

$$x_t = \frac{1}{24}(z_{t-6}+2\sum_{p=-5}^{5} z_{t+p}+z_{t+6}) \qquad \ldots (11)$$

The regression model (2) may be written in the form

$$z_t = \alpha_j+(1+\beta_j)\,\xi_t+\varepsilon_t \qquad \ldots (12)$$

To the first order of approximation let us suppose that the trend estimation filter (11) annihilates the seasonal component $\alpha_j+\beta_j\,\xi_t$ completely and reproduces the trend ξ_t exactly. Then we have

$$x_t = \xi_t+\frac{1}{24}(\varepsilon_{t-6}+2\sum_{p=-5}^{5}\varepsilon_{t+p}+\varepsilon_{t+6}) \qquad \ldots (13)$$

so that, on substitution in (12) we obtain

$$z_t = \alpha_j+(1+\beta_j)x_t+\eta_t \qquad \ldots (14)$$

where

$$\eta_t = \varepsilon_t-\frac{1}{24}(\varepsilon_{t-6}+2\sum_{p=-5}^{5}\varepsilon_{t+p}+\varepsilon_{t+6}) \qquad \ldots (15)$$

to the first approximation on regarding

$\frac{1}{24}(\varepsilon_{t-6}+2\sum_{p=-5}^{5}\varepsilon_{t+p}+\varepsilon_{t+6})\beta_j$ as negligible.

We therefore finish up with a new regression model (14) which has the same form as (12) except that the original disturbances ε_t are replaced by the filtered disturbances η_t given by (15). From the graph of the frequency response function of the centred 12-month moving average given in Figure 2 it follows that the properties of the new disturbances η_t are very close to those of ε_t in the neighbourhood of the seasonal frequencies. Since the properties in this neighbourhood are the most relevant for the estimation of α_j and β_j we infer that the fitting of the regression model is little affected by the substitution of x_t for ξ_t.

Effect of using improved trend estimate on regression*

Table 3 total series

Base position	Comparison of serial correlation of residuals (Durbin-Watson Statistics): Centred 12-month moving average trend estimate	Two-stage Burman moving average trend estimate	Comparison of mean square errors (thousands)²: Centred 12-month moving average trend estimate	Two-stage Burman moving average trend estimate
1949-58	0·57	1·87	119	16
1950-59	0·51	1·91	125	17
1951-60	0·53	1·98	118	16
1952-61	0·56	1·98	106	16
1953-62	0·64	1·84	95	19
1954-63	0·47	1·91	124	17
1955-64	0·40	2·00	127	16
1956-65	0·47	1·99	139	16
1957-66	0·37	1·93	169	14
†1958-67	0·45	2·17	122	14
†1959-68	0·44	2·03	102	14
†1960-69	0·44	1·85	92	18

*Modified for extreme values

†Model fitted with prior amplitude scaling: see Section 5·2

Of the assumptions made above the most critical is that the estimator (11) reproduces ξ_t exactly. This is strictly true only where the trend is linear. When the trend is non-linear a bias is introduced into the regression model arising from the discrepancy between

$$\xi_t \text{ and } \frac{1}{24}(\xi_{t-6}+2\sum_{p=-5}^{5}\xi_{t+p}+\xi_{t+6})$$

It is to eliminate this bias that the Burman estimator is used to estimate ξ_t in the second stage of fitting. Let the Burman trend estimator be denoted by

$$x_t = \delta+\sum_{p=-6}^{6}\delta_p z_{t+p} \qquad (\delta_p = \delta_{-p})$$

Making approximations similar to those above it can be shown that to the first order of approximation we obtain

$$\eta_t = \varepsilon_t - \sum_{p=-6}^{6} w_{jp}\, \varepsilon_{t+p} \qquad \ldots (16)$$

where w_{jp} depends to a small extent on j and $t = 12(i-1)+j$. It was found empirically that the disturbances η_t were distributed approximately normally (see Figure 4) and were not serially correlated (see Table 3). Thus the main effect of the trend estimation is to replace the original disturbance ε_t by a new disturbance η_t, the relation between these two being indicated by the transfer function given in Figure 2.

The question now arises whether it is legitimate to fit the derived regression model (14) by least squares bearing in mind the possibility that the original disturbances (12) may be serially correlated and taking into account the above mentioned consequences of trend estimation. Now Grenander and Rosenblatt (1957), chapter 7, have shown that for estimating coefficients of periodic functions such as α_j and β_j of slowly changing regressors such as x_t, least squares is an asymptotically efficient procedure under wide variations in the serial properties of the disturbances. We infer that, at least to a reasonably satisfactory order of approximation, it is justifiable to fit model (14) by least squares after both first and second-stage estimation of trend.

We calculated the Durbin-Watson statistic d for assessing the serial correlation of the residuals r_{ij} from the regression in the original data space (see Table 3). It will be noted for different base positions that with the centred 12-month moving average, d lies between 0·37 and 0·64. In sharp contrast, the two-stage method gives d between 1·91 and 2·18, showing no serial correlation. Hence in the two-stage procedure, the efficiency of fitting by least squares and the significance tests in the stepwise procedure are not seriously in question due to autocorrelation of the errors.

A further measure of the improvement resulting from the use of the two-stage method in preference to the 12-month centred moving average as the trend estimate is the reduction in mean square error. This is shown in Table 3 where it can be seen that the model fitted using the two-stage method gives a squared error about one fifth of that using the 12-month centred moving average.

3. Fitting the regression model

3.1 Fourier transformation of the model

In the 1965-70 method, the model (2), subject to the constraints (3) was fitted by least squares, the trend ξ_t of course being replaced by an estimate x_t; since there are 24 seasonal factors and 2 constraints, 22 coefficients are obtained. For the further development of the method of fitting it is convenient to write the fitted regression in vector notation as:

$$\mathbf{y}_i = \mathbf{a}+\mathbf{x}_i\mathbf{b}+\mathbf{r}_i \qquad \ldots(17)$$

where

$$\mathbf{y}_i = \mathbf{z}_i-\mathbf{x}_i$$

and $\mathbf{y}_i$ is the column vector $[y_{i1}, y_{i2}, \ldots, y_{i,12}]'$

$\mathbf{a}$ is the column vector $[a_1, a_2, \ldots, a_{12}]'$

$\mathbf{x}_i$ is the diagonal matrix
$$\begin{bmatrix} x_{i1} & & & \bigcirc \\ & x_{i2} & & \\ & & \ddots & \\ \bigcirc & & & x_{i,12} \end{bmatrix}$$

$\mathbf{b}$ is the column vector $[b_1, b_2, \ldots, b_{12}]'$

and $\mathbf{r}_i$ the column vector $[r_{i1}, r_{i2}, \ldots, r_{i,12}]'$

Since seasonal variation is a periodic function of period 1 year it may be represented as a sum of sine and cosine waves of periods 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$ and $\frac{1}{6}$ years, i.e. 12, 6, 4, 3, 2·4 and 2 months. Rewriting the model (17) in terms of these cycles, we can estimate the amplitude and phase of each of the sine and cosine waves and thence derive the seasonal factors for each month separately. The rewriting is done by the following Fourier transformation.

We take the 12 ×12 Fourier matrix **F** with elements f_{kp} given by, for $p = 1, 2, \ldots 12$,

$$\left.\begin{aligned} f_{1p} &= 1/\sqrt{12} \\ f_{kp} &= 1/\sqrt{6}\cos\frac{kp\pi}{12} && k = 2, 4, \ldots, 10 \\ &= 1/\sqrt{6}\sin\frac{(k\text{-}l)p\pi}{12} && k = 3, 5, \ldots, 11 \\ f_{12p} &= 1/\sqrt{12}\cos\frac{12p\pi}{12} \end{aligned}\right\}\ldots(18)$$

Put $\mathbf{v}_i = \mathbf{F}\mathbf{y}_i$, $\mathbf{c}_1 = \mathbf{F}\mathbf{a}$, $\mathbf{c}_2 = \mathbf{F}\mathbf{b}$, so that the transform of (17) becomes

$$\mathbf{v}_i = \mathbf{c}_1+\mathbf{F}\mathbf{x}_i\,\mathbf{b}+\mathbf{F}\mathbf{r}_i$$

$$\text{or} \quad \mathbf{v}_i = \mathbf{c}_1+(\mathbf{F}\mathbf{x}_i\,\mathbf{F}')\,\mathbf{c}_2+\mathbf{g}_i \qquad \ldots(19)$$

where $\mathbf{g}_i = \mathbf{F}\mathbf{r}_i$.

Now let **I** be the 12×12 diagonal matrix
$$\begin{bmatrix} 1 & & & & \bigcirc \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ \bigcirc & & & & 1 \end{bmatrix}$$

then $\mathbf{v}_i = (I, \mathbf{F}\mathbf{x}_i\mathbf{F}')\mathbf{c}+\mathbf{g}_i \qquad \ldots(20)$

where $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2)$ are the Fourier coefficients.

Since the Fourier transformation is orthogonal, the application of least squares to model (20) gives results identical to those given by the application of least squares to model (17).

3.2 Stepwise regression program

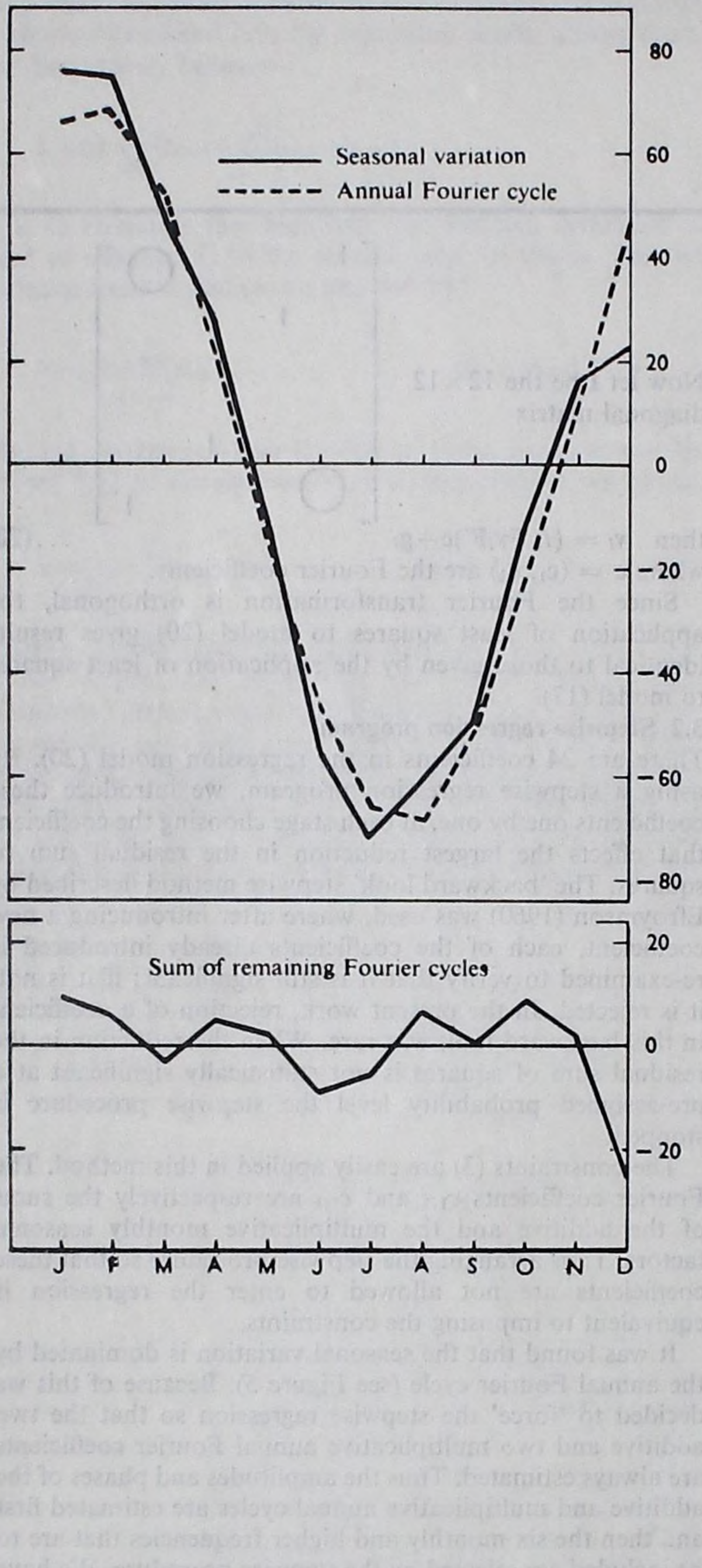
There are 24 coefficients in the regression model (20). By using a stepwise regression program, we introduce these coefficients one by one, at each stage choosing the coefficient that effects the largest reduction in the residual sum of squares. The 'backward look' stepwise method described by Efroymson (1960) was used, where after introducing a new coefficient, each of the coefficients already introduced is re-examined to verify that it is still significant; if it is not, it is rejected. In the present work, rejection of a coefficient in this backward look was rare. When the reduction in the residual sum of squares is not statistically significant at a pre-assigned probability level the stepwise procedure is stopped.

The constraints (3) are easily applied in this method. The Fourier coefficients $c_{1,1}$ and $c_{2,1}$ are respectively the sums of the additive and the multiplicative monthly seasonal factors. Thus arranging the stepwise procedure so that these coefficients are not allowed to enter the regression is equivalent to imposing the constraints.

It was found that the seasonal variation is dominated by the annual Fourier cycle (see Figure 5). Because of this we decided to 'force' the stepwise regression so that the two additive and two multiplicative annual Fourier coefficients are always estimated. Thus the amplitudes and phases of the additive and multiplicative annual cycles are estimated first and then the six monthly and higher frequencies that are to be included are selected by the stepwise procedure. We have used a 10 per cent level throughout, which has seemed generally satisfactory. Including the four coefficients of the annual cycles, about half the 22 coefficients are introduced. We should perhaps remark that the use of significance tests for the purpose of deciding which variables to include seems to us justifiable more on the pragmatic grounds that it seems to give sensible results in practice rather than in terms of the formal theory of testing hypotheses.

Figure 5 **Fourier cycles**

Base position 1956-65. Trend of 500,000 Thousands



The advantage of the stepwise method is that the excluded non-significant cycles are classed correctly with the irregular components and thus the estimated seasonal variation is smoother than would be obtained if all 22 Fourier coefficients were estimated: it follows that the seasonally adjusted series is not 'over-smoothed' in the sense that the irregular component is separated better from the seasonal component.

A disadvantage of the stepwise procedure when applied to the regression model is that a slight element of instability may occur as the regression base is updated due to a shift in the selection from additive to multiplicative components at the same frequency or *vice versa.* However this effect is reduced by the forced inclusion of both additive and multiplicative annual components. The small amount of instability remaining is a small price to pay for the positive advantages of the stepwise regression procedure.

3.3 Length of base

To respond to changing seasonality, the program is organised so as to update the regression base periodically. An important question is the number of observations or base length over which the regression model is fitted. To get some idea of what would be suitable, provision was made in the early programs for calculation of a mean square prediction error. In these programs the regression base was updated at annual intervals, the last month being a December. For a base ending in year n, the deviations of the observations from the regression were calculated for the 12 months from July of year $(n+1)$ to June of year $(n+2)$ and the mean square prediction error was found as 1/12th the sum of the squares of these deviations.

Working with a trend estimated by a 12-month centred moving average it was found that the mean square prediction errors fell rapidly as the number of years in the regression base was increased to 10 years and thereafter fell slowly. In the light of this result, and bearing in mind the importance of responding quickly to changing seasonality, we adopted a base length of 10 years for our further work.

4. Extreme values

4.1 Introduction

A basic part of our work was the development of a suitable procedure for dealing with extreme observations such as occurred in the winter of 1963 when unemployment rose drastically because of the unusually severe weather. Observations as extreme as this occur only very infrequently and it is clear that if they are included without modification in the estimation formulae substantial biases can result.

There are several papers in the literature dealing with the study of extreme values in regression and other linear models, notably Anscombe (1960), but we did not find these very helpful since they seemed to be concerned mainly with the detection of extremes arising from contamination of the data, e.g. from errors of observations or the inclusion of observations from another population and these considerations did not apply in our case. We were therefore forced to develop our own *ad hoc* analysis.

In order to demonstrate the effect that an extreme value can have on a regression estimate it is instructive to examine the simple regression case.

Consider $E(y) = \alpha+\beta x \qquad \ldots(21)$

fitted to n observations (x_i, y_i). Writing

$$n\bar{x} = \sum_{i=1}^{n} x_i,$$

$$CS(x, y) = \sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y}) \text{ etc.,}$$

the least squares estimates of α, β are

$$\left.\begin{aligned} a &= \bar{y}-b\bar{x} \\ b &= \frac{CS(xy)}{CS(xx)} \\ \text{and } R &= CS(yy)-b^2CS(xx) \\ r_i &= y_i-a-bx_i \end{aligned}\right\} \quad \ldots(22)$$

where R is the residual sum of squares and r_i the residual of the ith observation from the regression.

If the mth observation (x_m, y_m) is extreme in y_m then it seems natural to look at the change in the estimates when this observation is omitted.

In fact we have

$$\left.\begin{aligned} b' &= b-\gamma\delta r_m \\ R' &= R-\gamma r_m^2 \end{aligned}\right\} \quad \ldots(23)$$

$$\left.\begin{aligned} \text{where } \delta &= \frac{(x_m-\bar{x})}{CS(xx)} \\ \frac{1}{\gamma} &= \frac{n-1}{n}-\frac{(x_m-\bar{x})^2}{CS(xx)} \end{aligned}\right\} \quad \ldots(24)$$

For example consider the regression for Februarys, 1951-67, for the total series taking x as a 12-month centred moving average. Working in hundreds $\delta = 1{\cdot}57\times10^{-4}$, $\gamma = 1{\cdot}50$, $r_m = 400$, and $b = 0{\cdot}23$. We know that February 1963 is an extreme value and if we omit it from the regression we find $b' = 0{\cdot}14$ compared with an original value of $0{\cdot}23$. Though this is a simplified situation (only 17 observations) it demonstrates that extreme values can have a considerable effect.

Omitting an extreme value is clearly a drastic step. Moreover the use of moving averages and Fourier transforms in seasonal adjustment procedures makes it difficult if not impossible to omit an observation. An obvious alternative is to make some modification Δ, say, to the observation y_m. It can be shown that

$$\left.\begin{aligned} b' &= b+\delta\Delta \\ R' &= R+2r_m\Delta+\frac{\Delta^2}{\gamma} \\ r'_m &= r_m+\frac{\Delta}{\gamma} \end{aligned}\right\} \quad \ldots(25)$$

If the observation is displaced so that it lies on the original regression i.e. $\Delta = -r_m$ then

$$\left.\begin{aligned} b' &= b-\delta r_m \\ R' &= R-(2-\frac{1}{\gamma})r_m^2 \\ r'_m &= r_m(1-\frac{1}{\gamma}) \end{aligned}\right\} \quad \ldots(26)$$

In the example discussed earlier the effect of moving the observation for February 1963 to lie on the original regression is to give a b' of $0{\cdot}17$ which is reasonably close to the estimate got by omitting the observation entirely. However it is not difficult to show that this device is unsatisfactory for a very large extreme.

It is interesting to note that if we choose Δ so that R' in equation (25) is minimised we find $\Delta = -\gamma r_m$ and

$$\left.\begin{aligned} b' &= b-\gamma\delta r_m \\ R' &= R-\gamma r_m^2 \\ r'_m &= 0 \end{aligned}\right\} \quad \ldots(27)$$

the same as omitting the observation y_m. This also follows if we displace the observation so that it lies on the re-calculated regression, i.e. we require r'_m to be zero.

4.2 Modification of extreme values

An obvious way of identifying observations that are extreme is to look for residuals outside the range $\pm k_2 s$ when s is the standard error from the regression and k_2 is a constant. When we tried displacing such observations to be on the regression as is suggested by the previous section an observation which was extreme and therefore put onto the regression line for one position of the base sometimes was not found to be extreme for another position of the base. This difficulty is avoided by using a tapering method in which extremes outside the range $\pm k_2 s$, where $k_2 > k_1$, are moved on to the regression line and those in the ranges $k_2 s$ to $k_1 s$ and $-k_2 s$ to $-k_1 s$ are moved on a linearly graduated scale such that residuals at $\pm k_1 s$ are not displaced.

Experimental runs suggested that this method did not deal adequately with the very large extreme observations recorded for the hard winter of 1963. As the base was updated to include 1963 observations it seemed possible that the estimated regression coefficients were being biased so much that merely putting the extreme observations onto the regression was inadequate. Assuming that there was no appreciable change in seasonality in 1963 an idea of the extremity of this winter can be obtained by looking at the deviations of the 1963 observations from the regression base 1953-62. We calculate the trend of the observations for 1963 by seasonally adjusting the series to June 1964 with the seasonal factors for the 1953-62 base and then taking a Burman moving average. Thence we derive the 'predicted' residuals of the 1963 observations from the 1953-62 regression. Table 4 shows a 'predicted' residual of about 5s for February and 8s for March. The large negative residuals for April and May are discussed later. It was decided to calculate residuals of observations a year ahead of each position of the regression base and to displace those observations having residuals outside the range $\pm k_2 s$ to lie on the predicting regression. As was noted in Section 4.1 such a procedure with a simple regression is equivalent to omitting the observation.

Year-ahead residuals—1963

Table 4 hundreds

January	− 44
February	229
March	346
April	−126
May	−157
June	− 57
July	− 60
August	65
September	48
October	− 22
November	7
December	74

(Standard error 1953-62 Base position = 44)

With these considerations in mind a two-step procedure for modifying extreme values was adopted. The process is started by modifying internal extremes in the first base position. Then the base is updated annually and for all subsequent base-positions the two-step procedure is applied. First, year-ahead large extremes are displaced to lie on the 'predicting' regression. This displacement is then maintained as the regression base is updated to include the year in which these large extremes were detected, i.e. the original observation is replaced for all subsequent work by the modified value. However, the modified large extremes together with all other observations within a regression base are examined and modified as necessary by the tapering procedure. Occasionally an extreme residual outside the range $\pm k_2 s$ is found within the regression base, in which case it is displaced to lie on the regression estimated for this base position.

This two-step procedure was evolved after a number of trials, including an examination of the Winsorisation technique (Tukey, 1962).

The choice of control limits, i.e. of k_1 and k_2, was empirical. Various experiments and scrutiny of residuals led us to take $k_1 = 2{\cdot}00$, $k_2 = 3{\cdot}75$. It is worth noting that with these values, observations having residuals just exceeding 2s are displaced so that their residuals are just less than 2s: thus if such residuals do not correspond to extreme observations but merely arise from normal sampling the price paid for modifying these fluctuations is not large.

Because trend is a moving average of either the original or a preliminarily seasonally adjusted series, any extreme values affect the trend estimate. This gives rise to two problems; first a value may not appear extreme until its more extreme neighbours have been modified and secondly the modifications must take account of the effect of these extreme neighbours. The first of these problems is discussed in Section 4.3 in the light of some difficulties encountered with the component unemployment series. The second requires some algebra and since this is identical for internal and year-ahead extremes no distinction is made.

We define residuals r_{ij}

$$r_{ij} = z_{ij} - a_j - x_{ij}(1+b_j) \quad \ldots(28)$$

If s is the standard error from the regression and k_1 and k_2 are two control limits then z_{ij} is said to be a large extreme if

$$|r_{ij}| > k_2 s \quad \ldots(29)$$

and a small extreme if

$$k_1 s \leqslant |r_{ij}| \leqslant k_2 s \quad \ldots(30)$$

If $(r_{ij} - \Delta r_{ij})$ is the new residual obtained by modifying z_{ij}, then for large extremes

$$\Delta r_{ij} = r_{ij} \quad \ldots(31)$$

and for small extremes

$$\Delta r_{ij} = r_{ij} - \frac{|k_2 s - r_{ij}|}{(k_2 - k_1)} k_1 \text{ (sign } r_{ij}) \quad \ldots(32)$$

We need now to find the modified observation $(z_{ij} - \Delta z_{ij})$ having a residual $(r_{ij} - \Delta r_{ij})$. Changing z_{ij} to $(z_{ij} - \Delta z_{ij})$ also changes x_{ij} to $(x_{ij} - \Delta x_{ij})$. Using (28), we therefore have

$$r_{ij}-\Delta r_{ij} = (z_{ij}-\Delta z_{ij}) - a_j-(x_{ij}-\Delta x_{ij})(1+b_j) \quad \ldots(33)$$

and neglecting $b_j\,\Delta x_{ij}$, which is generally small,

$$\Delta r_{ij} = \Delta z_{ij}-\Delta x_{ij} \quad \ldots(34)$$

Translating this to the notation used in Section 3,

$$\Delta r_t = \Delta z_t-\Delta x_t \quad \ldots(35)$$

where $t = 12\,(i-1)+j$.

We can write the centred 12-month moving average as

$$x_t = \sum_{-6}^{+6} w_p\, z_{t+p} \quad \ldots(36)$$

$$\text{whence } \Delta x_t = \sum_{-6}^{+6} w_p\, \Delta z_{t+p} \quad \ldots(37)$$

where Δz_{t+p} is zero unless the observation is extreme. To emphasise that we are considering z_t to be extreme, let Σ' denote summation over p excluding $p = 0$, whence

$$\Delta x_t = w_o\, \Delta z_t+\Sigma'\, w_p\, \Delta z_{t+p} \quad \ldots(38)$$

Substituting in (35) and re-arranging,

$$\Delta z_t = \left(\frac{1}{1-w_o}\right) \left\{ \Delta r_t+\Sigma' w_p\, \Delta z_{t+p} \right\} \quad \ldots(39)$$

Substituting (39) in the right hand of (37) we get Δx_t and thence Δz_t from (35). Neglecting terms $0(w_p^2)$ we obtain

$$\Delta z_t = \Delta r_t+\frac{\sum_{-6}^{+6} w_p\, \Delta r_{t+p}}{1-w_o} \quad \ldots(40)$$

which equation takes into account the effect on z_t of any extremes lying within 6 observations either side. When $\Delta r_{t+p} = 0$ for all $p \neq 0$,

$$\Delta z_t = \Delta r_t+\frac{w_o\, \Delta r_t}{1-w_o} = \frac{\Delta r_t}{1-w_o} \quad \ldots(41)$$

If x is a Burman moving average of a preliminarily seasonally adjusted series

$$x_t = \sum_{-6}^{+6} w'_p\, \frac{z_{t+p}-a_h}{1+b_h} \quad \ldots(42)$$

where $h = (t+p-1) \bmod 12+1$

$$\text{whence } \Delta x_t = \sum_{-6}^{+6} w'_p\, \frac{\Delta z_{t+p}}{1+b_h} \quad \ldots(43)$$

Thus in a similar manner to (41) we derive

$$\Delta z_t = \left\{ \Delta r_t+\Sigma'\, w'_p\frac{\Delta z_{t+p}}{1+b_h}\right\}\left\{ 1-\frac{w'_o}{1+b'_h}\right\} \quad \ldots(44)$$

where $h' = (t-1) \bmod 12+1$

and substituting into (43) and neglecting terms $0(w'_p{}^2)$ we obtain

$$\Delta z_t = \Delta r_t+ \sum_{-6}^{+6} \frac{w'_p\, \Delta r_{t+p}}{1+b_h-w'_o} \quad \ldots(45)$$

Formula (40) is applied when trend is estimated by a 12-month centred moving average and gives the modification to be made to an observation at time t due to an extreme at time t and to neighbouring extremes in the interval $(t-6)$ to $(t+6)$. It is based on the modifications of the corresponding residuals as found from formulae (31, 32). Formula (45) is applied at the second stage of trend estimation using the Burman moving average of a preliminary seasonally adjusted series and modification of residuals from formulae (31, 32).

4.3 Experience with the method

As far as the total series is concerned the method has worked well so far. Table 5 gives both year-ahead and small internal modifications for various positions of the base. The modifications being made to 1963 by the year-ahead rule now appear adequate inasmuch as the fit of the model is substantially unchanged by the inclusion of the modified values for 1963 in the estimation period. This point can be seen clearly in Figure 7 of Section 5 where regression lines for January and July both inclusive and exclusive of 1963 modified extreme observations are shown.

Figure 4, which was a normal plot of residuals with year-ahead extremes modified is evidence for the choice of inner control limit. The residuals in the range $\pm 2s$ (i.e. ± 88 in Figure 4) lie practically on a straight line and are consistent with the assumption that they are distributed normally with the same mean and variance whereas those few outside the control limit appear quite different.

Some difficulties have been experienced, however, with component series when 1963 appears even more extreme in relation to the rest of the series than is the case in the total series. The effect on the trend estimation is so serious as to make neighbouring values appear to be large extremes in the opposite direction. This effect is visible in the moderately large negative residuals for April and May shown in Table 4 though in the total series these do not exceed the upper control limit. It has therefore been found necessary to introduce a further control limit k_3 where $k_3 > k_2$. If any residual is found outside the range $\pm\, k_3 s$ then only neighbouring observations with residuals of the same sign are modified ($k_2 s$ is still the level above which residuals are identified as extremes). In practice the choice of k_3 equal to 6 has been satisfactory for most of the components where this difficulty has been encountered but one or two series have needed special attention.

Attention has already been drawn to the necessary assumption of constant seasonality for the year-ahead procedure. If a major alteration takes place then this can be found from the local amplitude scaling factors described in Section 5. However it might be that only one particular month is affected; this month appears as a year-ahead extreme which is then modified permanently and of course as the base is updated the same month in the following year will appear as a year-ahead extreme. We propose that if the same month is identified two years running the year-ahead modification should not be made and the model will then adapt to the new situation. Although the performance of the method should always be kept under scrutiny, in practice it is reasonably automatic. Obviously any empirical method should be used with care: e.g. control limits suitable for the total series may prove unsatisfactory for component series etc. Another case calling for caution is when several months in the year-ahead are identified as large extremes for which, unlike the hard winter in 1963, there is no assignable cause.

Modification of extreme values

Table 5 thousands

		Original value	Year-ahead extremes	Modification of extremes inside base					
				1950-59	1952-61	1954-63	1956-65	1958-67†	1960-69†
Feb	1954	362		355	358	359			
Nov	1956	251		262	259	260			
Jan	1957	336			332		327		
Feb		355		345	342	349	330		
Oct		264		282	276	274	270		
Feb	1961	339			340				
Feb	1962	408				412			
Jan	1963	605				592	588	590	555
Feb		647	610				605	597	557
Mar		628	578						538
Apr		553							528
Aug	1966	274						284	281
Nov		436	416						
Dec		465	442						
May	1969	506	525						
June		481	501						

†Model applied with prior amplitude correction: see Section 5.2.

5. Fitting the model to recent observations

5.1 Recent behaviour of the series

It was hoped that improvements effected in the method of fitting the model and updating the base (Section 3), use of the two-stage method for trend estimation (Section 2) and modification of extreme values (Section 4), taken together, might remove the over-adjustment of the series evident in recent years from Figure 1. However Figure 6 shows that the total series, adjusted with a regression base 1957-66, still shows a contra-seasonal movement. The figure suggests that a large change in the seasonal variation has taken place since 1967.

This change is also apparent from the fit of the regression as the base is updated to include data for 1967 and 1968. For the base 1959-68 the seasonal variation at high trend (500,000) is estimated as 52,000 and at low trend (250,000) as 50,000. This compares with a high trend figure from the 1956-65 base of 74,000 and a low trend figure of 42,000. These estimation periods have seven years in common and it is clearly unacceptable to have such differing estimates. The figures from the 1959-68 base would appear to have an increased additivity and thus prompt further consideration of the type of model fitted up to 1967. Figure 7 shows the regression lines for January and July (continuous lines) as estimated from the regression 1957-66. It will be clear that there is a definite multiplicative element present. It is also clear that this relationship had not altered greatly up to 1966; the broken lines on Figure 7 show the regressions obtained from the base 1950-59 and they are close to the continuous lines. Also shown on the figures are the actual deviations from trend from 1957-69. Up to January 1967 the lines are a good fit but the observations after that deviate markedly from the lines.

The nature of the change is evident from Figure 8 which compares (i) the deviations of the observations, corrected for extreme values, from trend during 1957-66 with the seasonal variation $(a_j+b_jx_{ij})$ computed from the 1957-66 base using the two stage trend and (ii) the deviations from trend during 1967-69 with the seasonal variation 'predicted' from the 1957-66 regression base. Inspection of the figure suggested that the main change is a considerable reduction in the amplitude of the seasonal cycle.

Thus our problem now is to consider how to fit a regression to data which manifestly show a marked change in the dependent variable in 1967. It would not be unreasonable to say that one should start a new adjustment procedure from mid-1967 onwards, notwithstanding that this gives only a couple of years' data for estimating the seasonality. However as has already been stated it would seem that the pattern (or shape) of the seasonal component is substantially unaltered and a method which would compensate for the change in the amplitude after 1967 while estimating the pattern over a 10 year period seems preferable. That this may be possible is evident from Figure 8b which shows that the post-1967 amplitude is about 0·7 of the pre-1967 amplitude, the changes in seasonal pattern being marginal.

5.2 Prior amplitude scaling

The manner in which we decided to do this was to fit the model in the following way;

$$\left.\begin{array}{ll} \text{up to June 1967,} & (z_{ij}-x_{ij}) = \dfrac{1}{f}(\alpha_j+\beta_j x_{ij})+\varepsilon_j \\ \text{from July 1967,} & (z_{ij}-x_{ij}) = (\alpha_j+\beta_j x_{ij})+\varepsilon_{ij} \end{array}\right\} \quad \ldots(46)$$

where f is called the prior amplitude scaling factor. This formulation assumes that the variances of the irregular components before and after mid-1967 are the same. We fitted (46) by minimising with respect to the seasonal factors the sum of squares

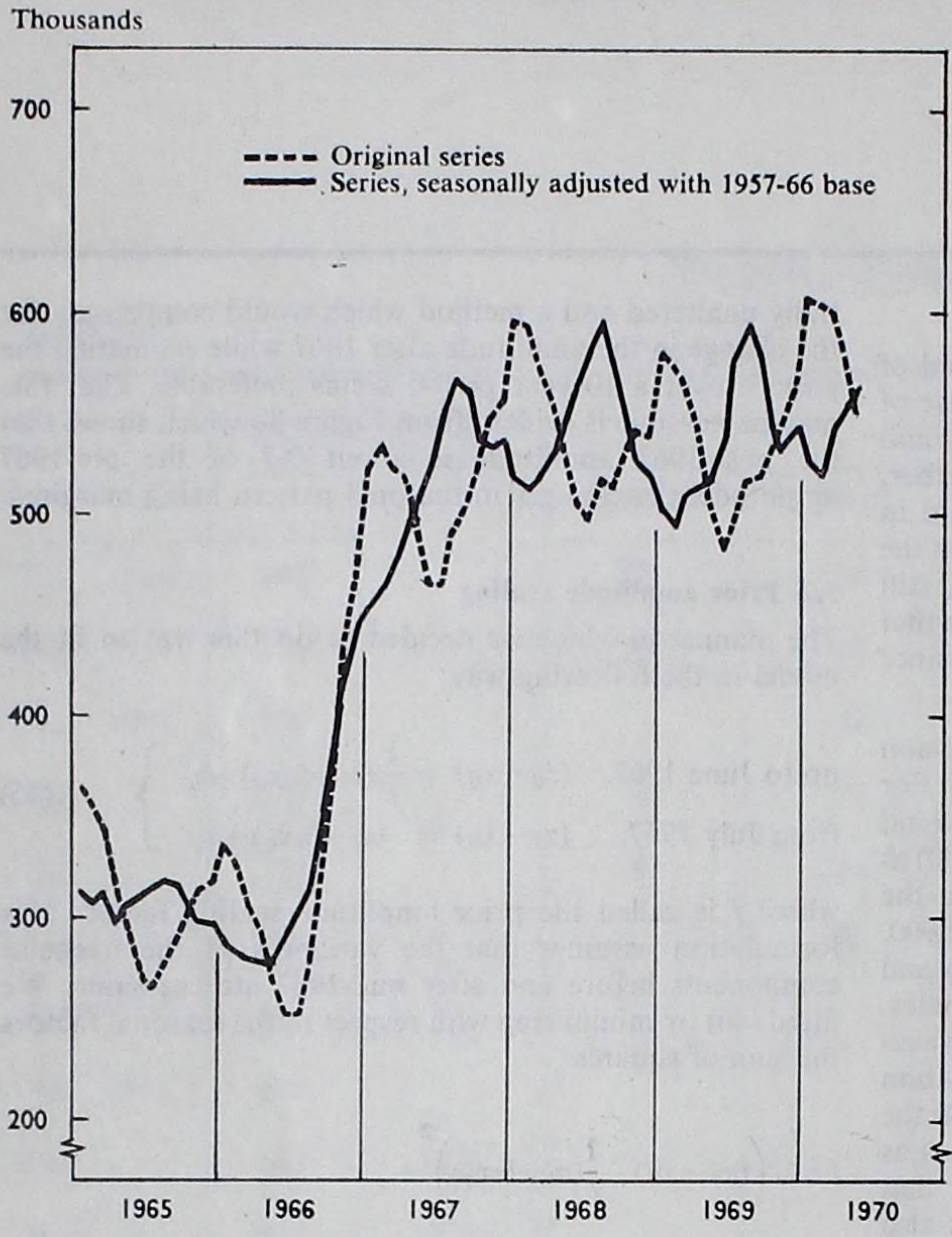
$$f^2\sum_{1}^{t_o}\left((z_{ij}-x_{ij})-\frac{1}{f}(a_j+b_jx_{ij})\right)^2+\sum_{t_o+1}^{120}\left((z_{ij}-x_{ij})-(a_j+b_jx_{ij})\right)^2$$

where t_o is June 1967. It will be noted that the first term in the minimisation expression is multiplied by the factor f^2 and since the prior amplitude scaling factor is less than one for all the series examined the least squares estimates are weighted in favour of observations after mid-1967; the greater the change in amplitude after 1967 the more emphasis that is given to recent observations. It seemed reasonable to us to allow the prior amplitude scaling factor to be used as a weighting factor on the grounds that the more severe the change in amplitude the more likely there are to be alterations in the pattern thus making observations prior to the amplitude discontinuity of less value. However we feel that in the further development of the program the weighting ought to be treated independently. It is evident from the mean square residuals given in Table 3 that the weights we have chosen and the assumption of constant irregular components are reasonable.

The seasonal variation at high trend estimated by the revised model (46) using the 1959-68 base is 50,000. In order to compare it with pre-June 1967 figures it must be divided

Figure 6

Series, seasonally adjusted, using 1957-66 regression base



by 0·7 giving 71,000 which is now of the same order as the 74,000 from the 1956-65 base. Similarly the low amplitude figure of 46,000 is comparable with the 1956-65 42,000. The regression fit for 1959-68 using the revised model also is slightly better than when the prior correction is not made.

5.3 Local amplitude scaling factors

In view of the sharp change in amplitude in 1967 we wanted to include in our adjustment procedure a means of responding rapidly to future changes in amplitude, assuming the seasonal pattern to be held constant. With this in view we decided to make local comparisons between the area of the actual seasonal wave and the area of the wave using the averaged constants. Thus we define the ratio

$$d_{ij} = \frac{\sum_{p=-12}^{12} w_p \left| z_{t+p} - x_{t+p} \right|}{\sum_{p=-12}^{12} w_p \left| a_{j+p} + b_{j+p}\, x_{t+p} \right|} \qquad \ldots(47)$$

$$t = 12(i-1)+j \qquad a_o = a_{12},\ a_{-1} = a_{11} \text{ etc.}$$

$$w_{-12} = w_{12} = 1 \qquad w_{-11} = w_{-10} \ldots = w_{11} = 2$$

We are thus comparing a 25-term centred average of deviations from trend to the same average of the seasonal variation using the estimated seasonal constants, in both cases independent of signs. We call this ratio the local amplitude scaling factor.

The values z and x used in (47) are freed from large 'year-ahead' extreme values when the observation lies within the regression base and for the following 12 months. When factors are calculated using observations more than 13 months ahead of the regression base, these observations may include some extreme values.

This concept of estimating the average seasonal component over a longish period and then examining local behaviour in relation to this average was discussed in a little known paper by Wald (1936) summarised and discussed by Godfrey and Karreman (1967). However Wald's method differs in several respects from the method used here.

For the trend estimate we decided to use the 12-month centred moving average. This is because when the seasonal component is changing the two-stage trend estimate will contain a seasonal element which will bias the local amplitude scaling factors in the region where they are of most value. The price paid for use of the 12-month centred moving average is a possible bias in peaks and troughs but experience so far has not found this to be severe. The estimation period for the local amplitude scaling factors was taken as two years; we felt that it should be sufficiently long as to be little affected by any unusually erratic behaviour of the series while responding quickly to genuine changes in the size of the seasonal component. A 3-year period could be used with erratic series.

The manner in which the local amplitude scaling factors are used in the final seasonal adjustment is discussed in Section 6.1; for the present we shall confine ourselves to considering their use for analysing the behaviour of the series.

Tables 6a, b, c give the factors for each June from 1960 to 1969 computed from different regression bases. Taking first the results for the total series given in Table 6a, for any year up to 1967 the d_{ij}'s predicted for each year from the preceding base and those found when the year in question is at the middle of the base do not differ by more than 10 per cent. The years from 1964 onwards cannot be at the middle of a base, so the table is completed by giving the local amplitude scaling factors for these years using the 1958-67 base without prior amplitude scaling. This small difference in the d_{ij}'s is to be expected in view of the relative stability and independence of the base of estimated seasonal constants up to 1966. However the d_{ij}'s predicted from the 1957-66 and 1958-67 bases for 1968 drop sharply and the value in 1968 of 0·6 is reasonably in line with the adopted prior amplitude scaling factor of 0·7. For the total series the 0·7 factor was put into practice before the development of the local scaling factors. For component series we propose the use of the predicted local amplitude scaling factor as the prior amplitude scaling factor (see Section 6.2).

Figure 7 **The fit of the regression**

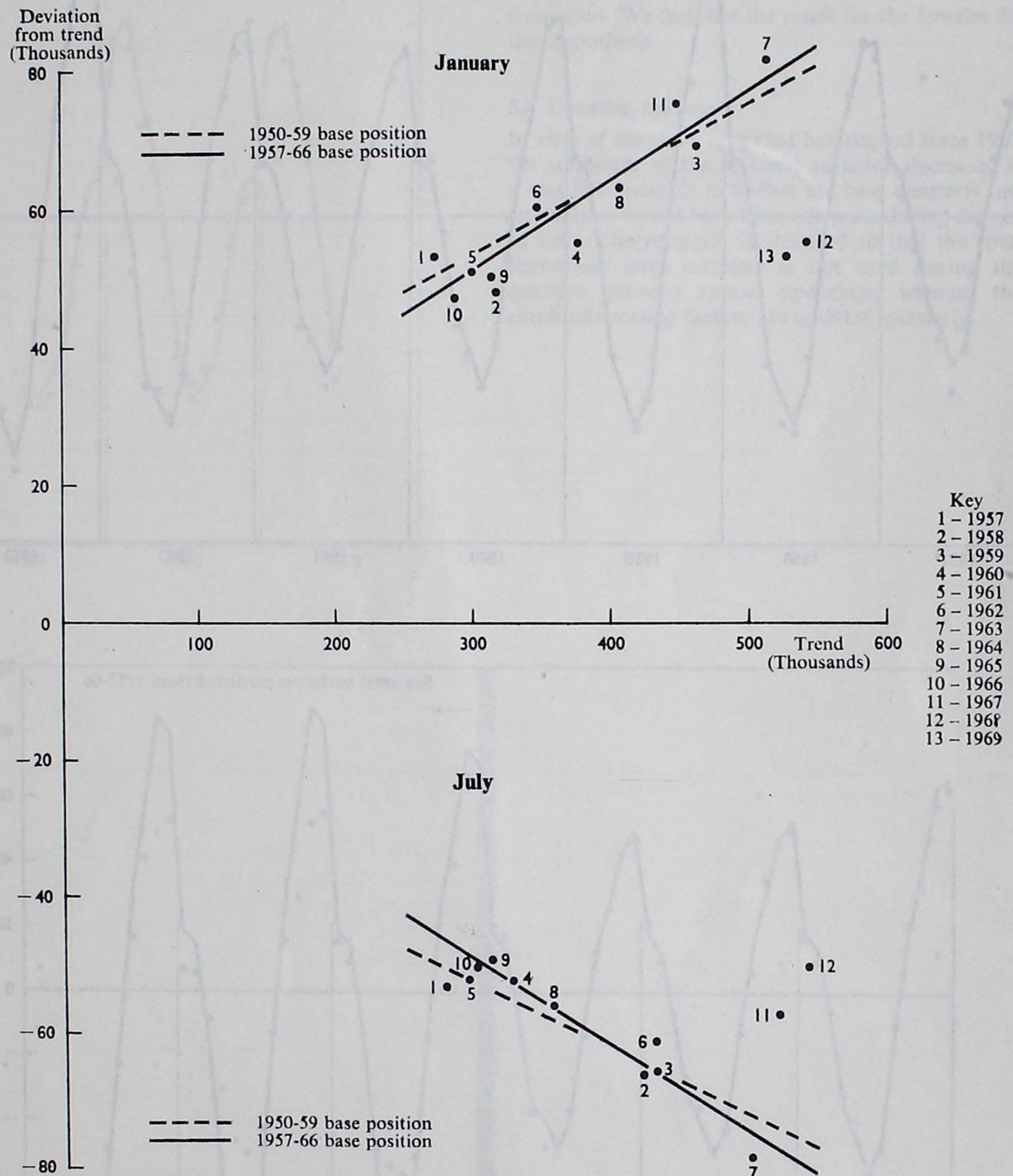
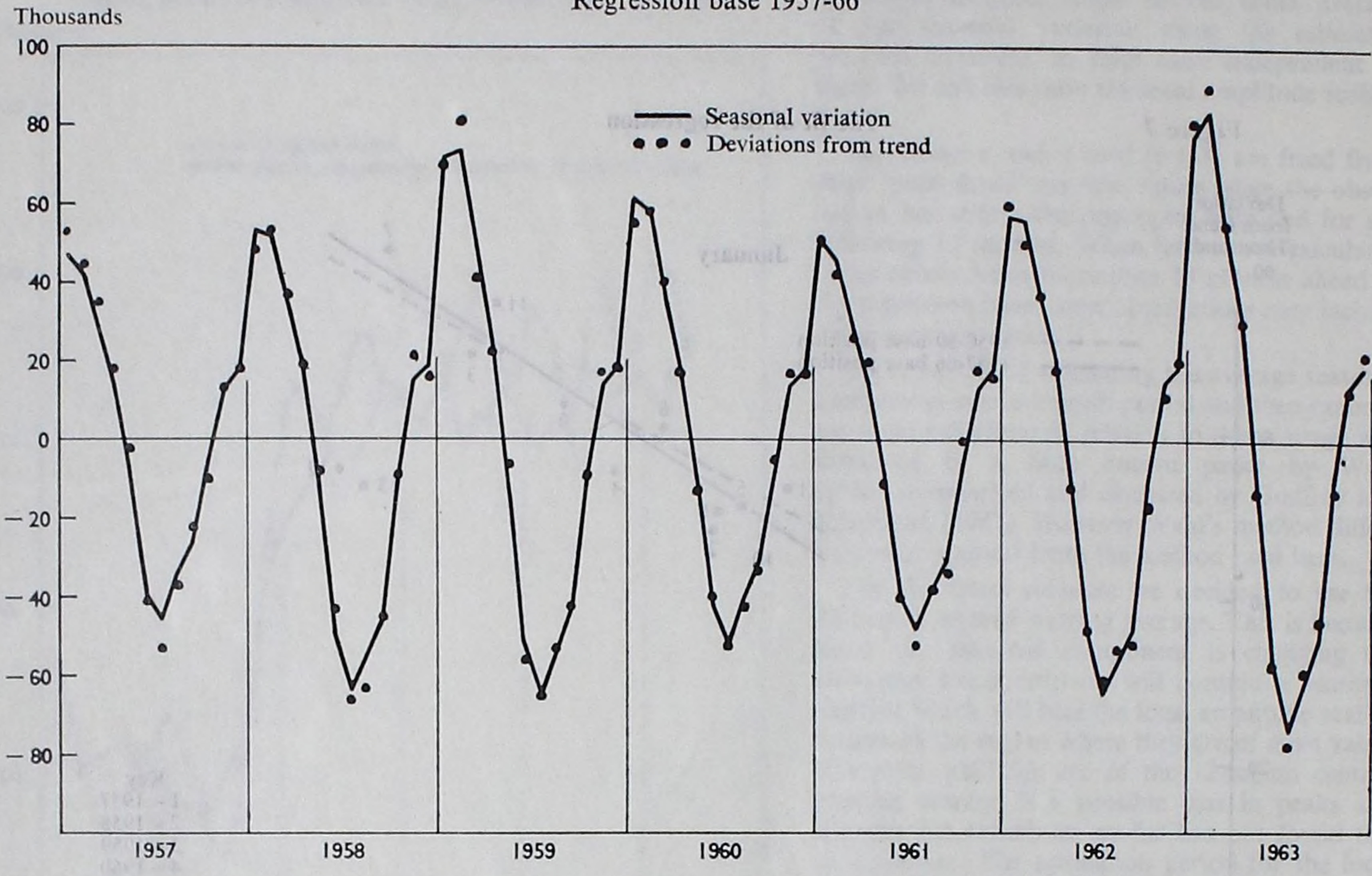


Figure 8 a

Seasonal variation and deviations from trend

Regression base 1957-66



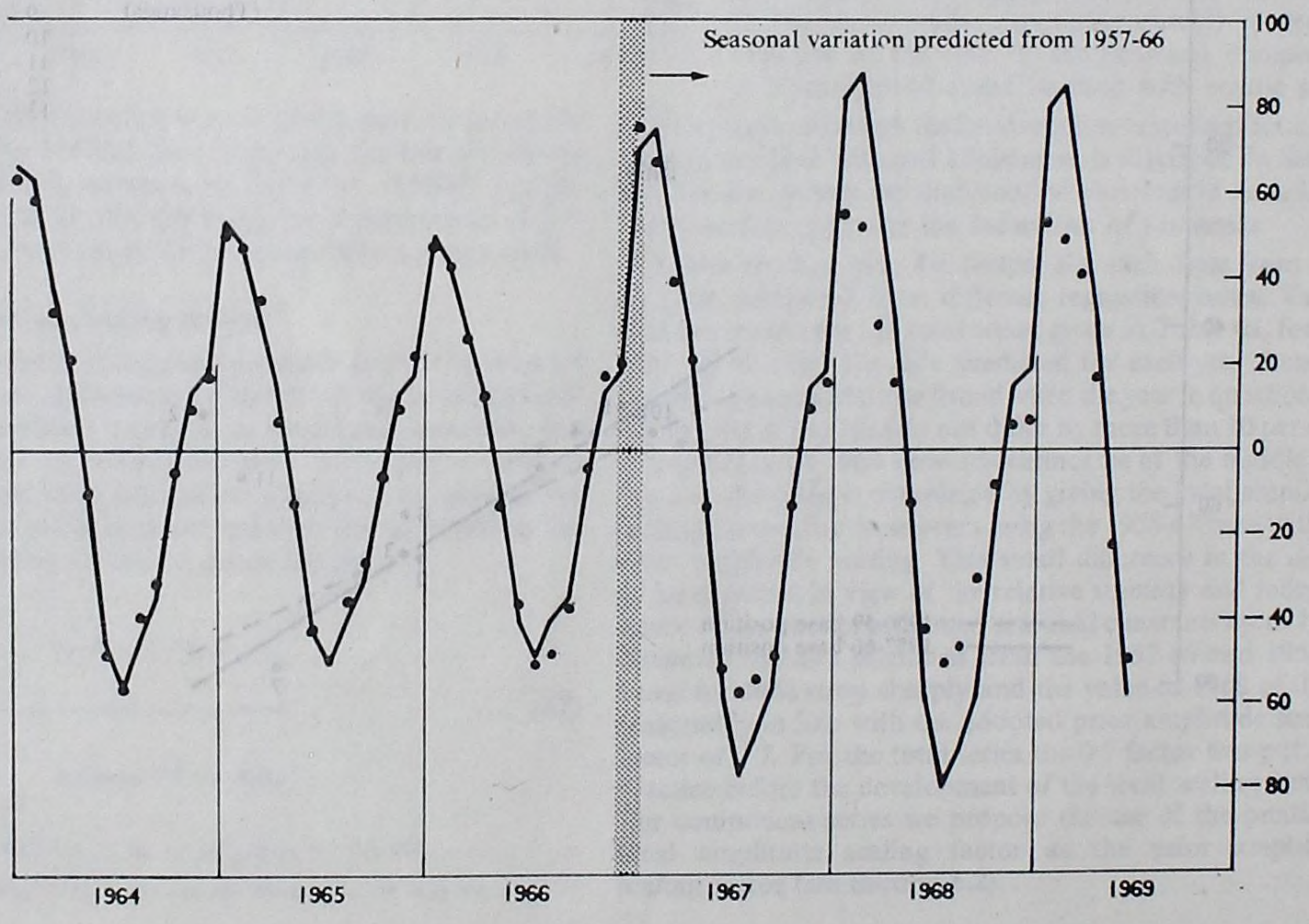
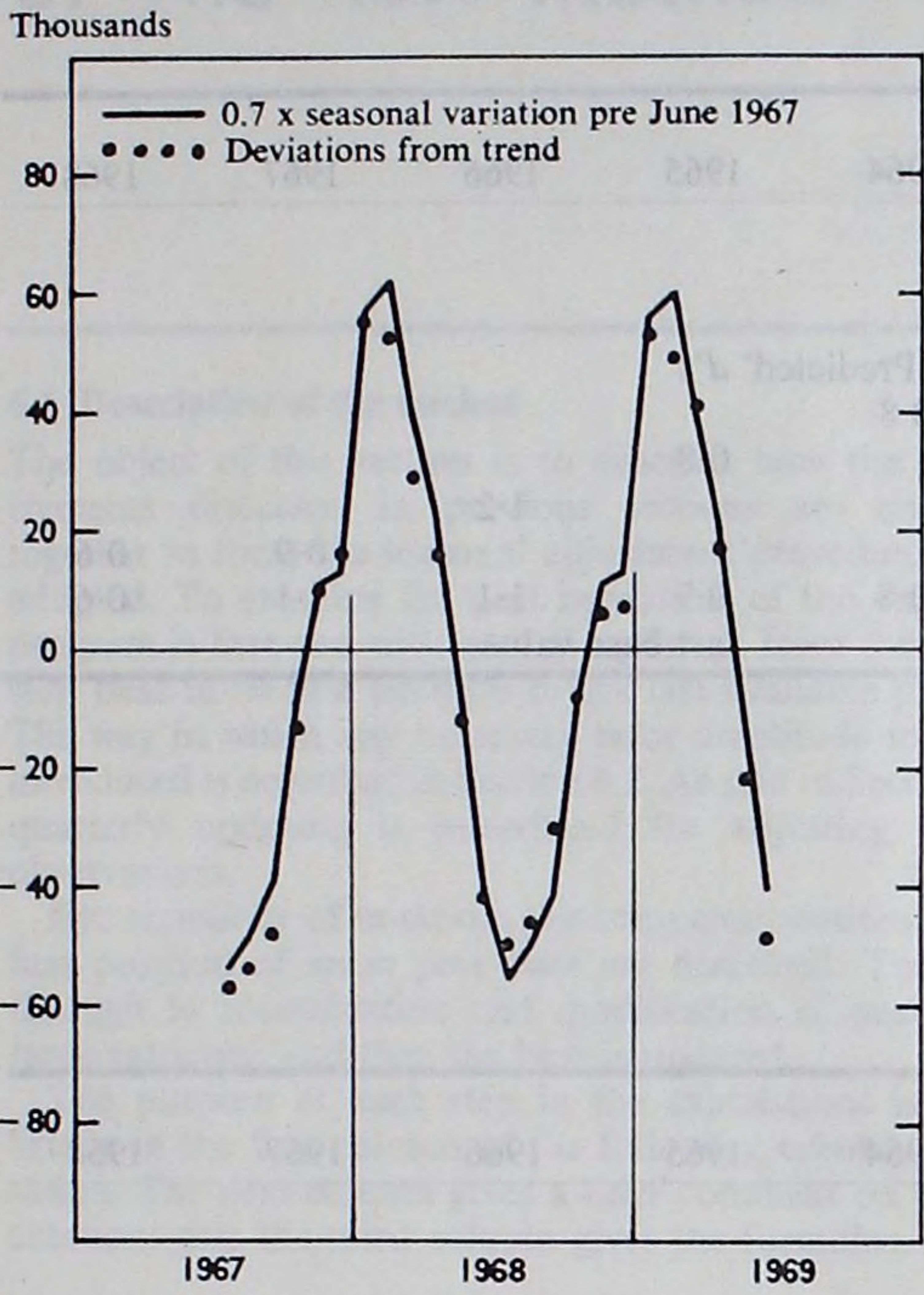


Figure 8 b

Reduction in amplitude of seasonal variation



Tables 6b and 6c give the same June scaling factors for the males and females components of the total series. The predicted fall from 1967 for the females is of particular interest since this series is predominantly additive, having for the base 1957-66 a multiplicativity M of 0·4. Now one possible hypothesis for the change in the behaviour of the total series was that, rather than a fall in amplitudes post-1966, there was a switch from a mixed to an additive model from 1964. We feel that the result for the females disproves this hypothesis.

5.4 Updating the base

In view of the short time that has elapsed since 1967, when the amplitude of the seasonal variation decreased sharply, it was felt desirable to update the base quarterly instead of annually as would have been adequate during the period up to 1967. The program is arranged so that the routine for year-ahead large extremes is not used during the three quarters between annual updatings, whereas the local amplitude scaling factors are updated quarterly.

Local amplitude scaling factors at June in each year

Table 6a

Total series

Base position	1960	1961	1962	1963	1964	1965	1966	1967	1968
1950-59	1·0								
1951-60		1·0							
1952-61			1·2						
1953-62				1·1	'Predicted' *d*'s				
1954-63					0·8				
1955-64	1·0					0·8			
1956-65		1·0					1·2		
1957-66	mid-base values		1·1					0·9	0·6
†1958-67				1·0	0·8	0·9	1·1	1·1	0·6
						last base values			

Table 6b

Males

Base position	1960	1961	1962	1963	1964	1965	1966	1967	1968
1950-59	1·0								
1951-60		1·0							
1952-61			1·3						
1953-62				1·2	'Predicted' *d*'s				
1954-63					0·8				
1955-64	0·9					0·8			
1956-65		1·0					1·1		
1957-66	mid-base values		1·1					0·8	0·5
†1958-67				1·0	0·8	0·9	1·1	0·9	0·6
						last base values			

Table 6c

Females

Base position	1960	1961	1962	1963	1964	1965	1966	1967	1968
1950-59	1·0								
1951-60		0·9							
1952-61			1·0						
1953-62				0·9	'Predicted' *d*'s				
1954-63					0·8				
1955-64	0·9					0·8			
1956-65		1·0					1·1		
1957-66	mid-base values		1·1					0·9	0·5
†1958-67				1·0	0·9	0·9	1·1	1·0	0·7
						last base values			

†Model applies without prior amplitude correction.

6. The new method

6.1 Description of the method

The object of this section is to describe how the various elements discussed in previous sections are combined together to form the seasonal adjustment procedure finally adopted. To examine the past behaviour of the series, the program is first run with annual updating from the regression base in its first position to the last available position. The way in which any necessary prior amplitude scaling is introduced is described in Section 6.2. As said in Section 5.4, quarterly updating is introduced for adjusting current observations.

For simplicity of notation, the computer routines for the first position of an m year base are described. This leads through to identification and modification of year-ahead large extremes, and then the base is updated.

The purpose of each step in the calculations is stated briefly in the first column of the following tabular presentation. The next column gives a brief comment on the calculations and the third column gives the formulae. Again, for simplicity, the formulae do not include the pre-1967 amplitude scaling.

It is convenient to change the notation from writing i, j as subscripts to writing them on the line; thus z_{ij} is now written as $z(ij)$. This enables us to use other subscripts and superscripts as follows:

Subscript 1 when (directly or indirectly) a 12-month centred moving average is used,

Subscript 2 when (directly or indirectly) a Burman moving average of a preliminary seasonally adjusted series is used,

Subscript e when a trend free from extremes is used or implied,

Subscript g when a series free from extremes is used or implied,

Superscript s to denote a seasonally adjusted series.

Step	*Comment*	*Formulae*
Original series	Observations for $n+1$ years.	(1) $\{z(ij)\}$ $i = 1$, $j = 7, \ldots, 12$ $i = 2, \ldots n$, $j = 1, \ldots, 12$ $i = n+1$, $j = 1, \ldots, 6$ also given by (2) $\{z(t)\}$ $t = 12(i-1)+j$
First trend estimate	A 12-month centred moving average of the original series is calculated for years 2 to $m+1$. The subscript 1 is used whenever an equation involves (directly or indirectly) the 12-month centred moving average.	(3) $x_1(t) = \sum_{p=-6}^{+6} w_{1p} z(t+p)$ $i = 2, \ldots, m+1$, $j = 1, \ldots, 12$
First fit of model	The model is fitted to years 2 to $m+1$ using the first trend estimate. The fitting is done in the manner described in Section 3 i.e. a Fourier transformation is made and the model is fitted by a step-wise procedure.	(4) $z(ij) - x_1(ij) = a_1(j) + b_1(j)x_1(ij) + r_1(ij)$ $i = 2, \ldots, m+1$, $j = 1, \ldots, 12$
First modification of extremes	Internal extremes in years 2 to $m+1$ are modified using the residuals from the first fit in order to provide a better trend estimate. The subscript e is used whenever trend free from extremes is used or implied. A revised series is produced for years 1 to $m+2$.	(5) $\{z_e(ij)\}$ $i = 1$, $j = 7, \ldots, 12$ $i = 2, \ldots m+1$, $j = 1, \ldots, 12$ $i = m+2$, $j = 1, \ldots, 6$

Step	Comment	Formulae
Preliminary seasonally adjusted series	In order to use the Burman moving average as a better trend estimate the revised series is seasonally adjusted. The adjusted series is distinguished by a superscript *s*.	(6) $z_e^s(ij) = \dfrac{z_e(ij)-a_1(j)}{1+b_1(j)}$ $i = 1 \quad j = 7, \ldots, 12$ $i = 2, \ldots m+1 \quad j = 1, \ldots, 12$ $i = m+2 \quad j = 1, \ldots, 6$
Second trend estimate	A Burman 13-point moving average of the preliminary seasonally adjusted series is calculated for years 2 to $m+1$. The subscript 2 is used whenever an equation involves (directly or indirectly) the Burman moving average.	(7) $x_{2e}(t) = \sum_{p=-6}^{+6} w_{2p}\, z_e^s(t+p)$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$
Second fit of the model	The model is refitted to years 2 to $m+1$ using the second trend estimate and the original series.	(8) $z(ij)-x_{2e}(ij) = a_{2e}(j)+b_{2e}(j)x_{2e}(ij)+r_{2e}(ij)$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$
Second and final modification of internal extremes	Internal extremes in years 2 to $m+1$ are finally modified using the residuals from the second fit of the model. These modifications are then used to obtain final seasonal factors. The subscript *g* is used to show that a series modified for extremes is implied.	(9) $\{z_g(ij)\}$ $i = 1 \quad j = 7, \ldots, 12$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$ $i = m+2 \quad j = 1, \ldots, 6$
REFIT OF MODEL USING SERIES MODIFIED FOR EXTREME VALUES		
Third trend estimate	A 12-month centred moving average of the revised series is calculated for years 2 to $m+1$.	(10) $x_{1g}(t) = \sum_{p=-6}^{+6} w_{1p}\, z_g(t+p)$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$
Third fit of the model	The model is fitted to years 2 to $m+1$ using the modified series.	(11) $z_g(ij)-x_{1g}(ij)=a_{1g}(j)+b_{1g}(j)\, x_g(ij)+r_{1g}(ij)$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$
Preliminary seasonally adjusted series	A further preliminary seasonally adjusted series using the constants from the third fit is computed.	(12) $z_g^s(ij) = \dfrac{z_g(ij)-a_{1g}(j)}{1+b_{1g}(j)}$ $i = 1 \quad j = 7, \ldots, 12$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$ $i = m+2 \quad j = 1, \ldots, 6$
Fourth trend estimate	The final trend estimate is computed as a Burman moving average of the preliminary seasonally adjusted series calculated at the previous step.	(13) $x_{2g}(t) = \sum_{p=-6}^{+6} w_{2p}\, z_g^s(t+p)$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$
Fourth and final fit of model	The model is now fitted using the modified series and the fourth trend estimate. This fit provides the final seasonal constants for the base position.	(14) $z_g(ij)-x_{2g}(ij)=a_{2g}(j)+b_{2g}(j)\, x_{2g}(ij)+r_{2g}(ij)$ $i = 2, \ldots, m+1 \quad j = 1, \ldots, 12$

Step	Comments	Formulae
Modification of year-ahead extremes	In order to extend the fourth trend estimate a preliminary seasonally adjusted series beyond the base is calculated. A Burman moving average of this adjusted series is then computed. The predicted residuals for year $m+2$ are then obtainable.	(15) $z_g^s(ij) = \frac{z_g(ij)-a_{2g}(j)}{1+b_{2g}(j)}$ $i = m+1 \quad j = 7, \ldots, 12$ $i = m+2 \quad j = 1, \ldots, 12$ $i = m+3 \quad j = 1, \ldots, 6$ (16) $x_{2g}(t) = \sum_{p=-6}^{+6} w_{2p}\, z_g^s(t+p)$ $i = m+2 \quad j = 1, \ldots, 12$
	Large extremes can now be modified in year $m+2$ and a revised series obtained. The procedure is then re-applied from steps 15 to 17 using the revised series and after any further modifications the resulting revised series is used for all further estimation.	(17) $r'_{2g}(ij) = z(ij)-a_{2g}(j)-(1+b_{2g}(j))\, x_{2g}(ij)$ $i = m+2 \quad j = 1, \ldots, 12$
Local amplitude scaling factors	First the final 12-month moving average (third trend estimate) is extended 18 months beyond the base. Local scaling factors are then calculated for the second half of the base plus six months.	(18) $x_{1g}(t) = \sum_{p=-6}^{+6} w_{1p}\, z_g(t+p)$ $i = m+1 \quad j = 7, \ldots, 12$ $i = m+2 \quad j = 1, \ldots, 12$ $i = m+3 \quad j = 1, \ldots, 6$
		(19) $d(t)= \frac{\sum_{p=-12}^{+12} v_p\lvert z_g(t+p)-x_{1g}(t+p)\rvert}{\sum_{p=-12}^{+12} v_p\lvert a_{2g}(j+p)+x_{1g}(t+p)a_{2g}(j+p)\rvert}$ $i = \left[\frac{m+1}{2}\right] \quad j = 12$ $i = \left[\frac{m+1}{2}\right]+1, \ldots, m+1 \quad j = 1, \ldots, 12$ $i = m+2 \quad j = 1, \ldots, 6$ $a(-1) = a(11)$ etc. and $v_{-12} = v_{12} = 1 \; v_{-11} = v_{-10} \ldots = v_{11} = 2$
Seasonally adjusted series	Using the local amplitude scaling factors a seasonally adjusted series is then calculated for the second half of the base *plus* six months. To preserve the constraints $\Sigma a_j = \Sigma b_j = 0$ only one scaling factor is used for each year (months 7 to 6).	(20) $z^s(ij) = \frac{z(ij)-d(i-1,\, 12)\, a_{2g}(j)}{1+d(i-1,\, 12)\, b_{2g}(j)}$ $i = \left[\frac{m+1}{2}\right]+1, \ldots, m+2 \quad j = 1, \ldots, 6$
		(21) $z^s(ij) = \frac{z(ij)-d(i\; 12)\, a_{2g}(j)}{1+d(i\; 12)\, b_{2g}(j)}$ $i = \left[\frac{m+1}{2}\right]+1, \ldots, m+1 \quad j = 7, \ldots, 12$

The whole procedure is now re-applied to obtain estimates of seasonal constants for the regression based on years 2 to $m+2$ and so on to the regression ending in year n. Because the local scaling factor is calculated as a function of 25-month centred moving averages which themselves use a trend estimate, the last scaling factor is centred 18-months short of the end of the series. This final scaling factor $\{d(n-1, 12)\}$ is then used for the last two years of the series. The modified seasonal constants obtained with this final scaling factor are also used to seasonally adjust the next quarter's figures as they become available, i.e. months 7, 8, 9 of year $n+1$. At this point the regression base and the final scaling factor are both advanced one quarter to provide estimates for months 10, 11, 12 of year $n+1$ and quarterly updating continues until another complete year's data is available. When a complete year's data is available in order to modify any year-ahead extremes in year $n+1$ the program is rerun starting two base positions before the end of the series. (In fact to remove any bias due to year-ahead extremes in the local amplitude scaling factor the program has to run through the last three positions; this is discussed further in the sub-section on revisions in Section 7.)

So far we have discussed the method without reference to pre-1967 amplitude scaling (Section 5.2). As a full description would require the re-writing of the first part of this subsection with an additional equation at each step we shall confine ourselves to general differences. If the amplitude after month r of years is reduced by a factor f then the model is fitted as:

$$\left.\begin{array}{l} f\{z(ij)-x(ij)\} = a(j)+b(j)x(ij)+r(ij) \\ \quad i=2,\ldots s-1 \quad j=1,\ldots 12 \quad i=s \quad j=1,\ldots r \\ \text{and } z(ij)-x(ij) = a(j)+b(j)x(ij)+r(ij) \\ \quad i=s \quad j=r+1,\ldots 12 \quad i=s+1,\ldots m+1 \quad j=1,\ldots 12 \end{array}\right\} \quad (22)$$

Equations (22) with appropriate subscripts then replace equations (4), (8), (11) and (14).

Seasonal adjustment is then given by

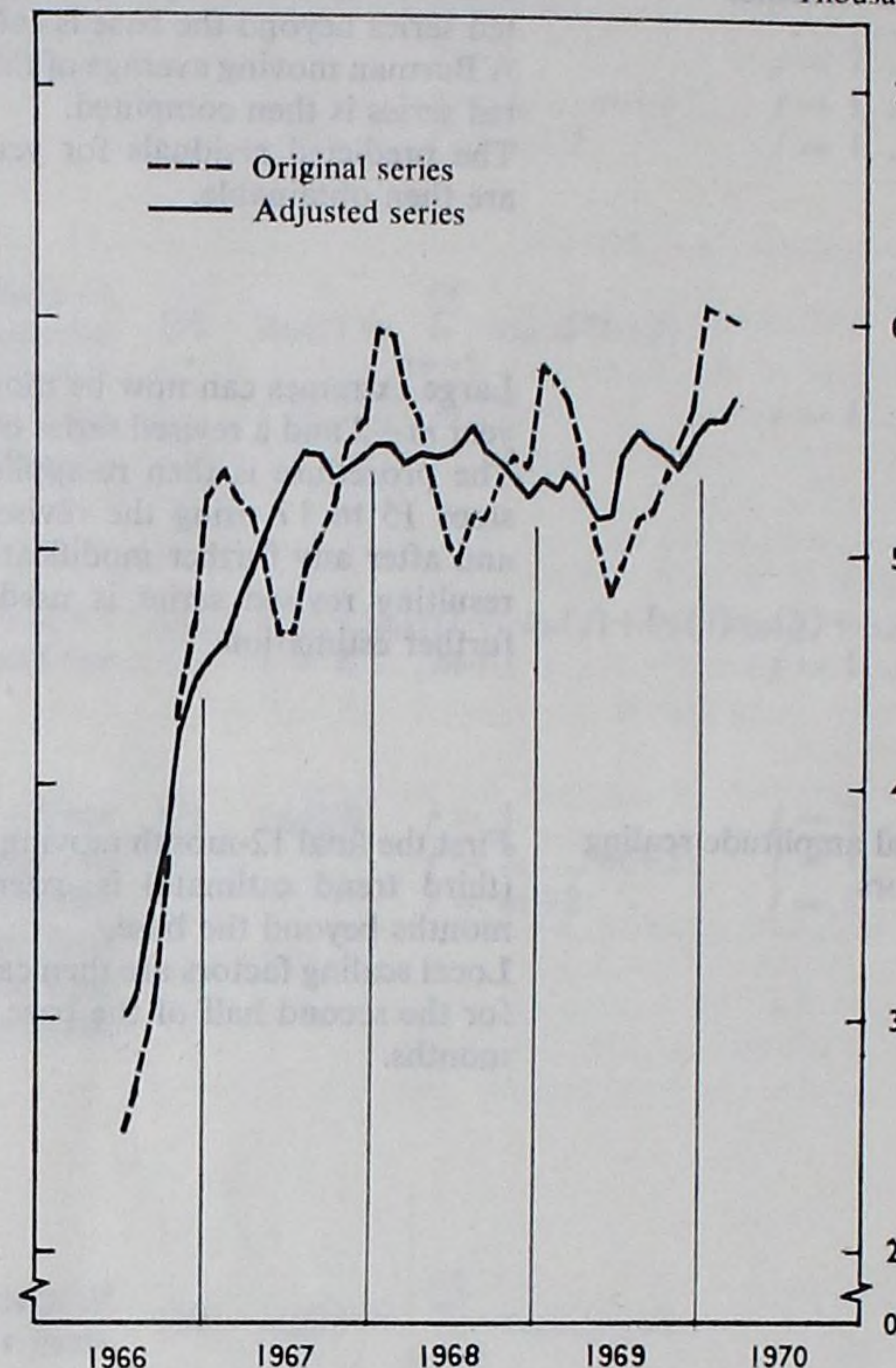
$$\left.\begin{array}{lll} z^s(ij) = \dfrac{z(ij)-(a(j)/f)}{1+(b(j)/f)} & i=2,\ldots s-1 & j=1,\ldots 12 \\ & i=s & j=1,\ldots r \\ \text{and } z^s(ij) = \dfrac{z(ij)-a(j)}{1+b(j)} & i=s & j=r+1,\ldots 12 \\ & i=s+1,\ldots m+1 & j=1,\ldots 12 \end{array}\right\} \quad (23)$$

Equations (23) replace equations (6), (12), (15) and (20).

Similar alterations are necessary in the equations for modification of extreme values i.e. Δr_{ij} becomes $\Delta r_{ij}/f$ whenever $(i, j) < (r, s)$. The mean square residual about the regression is calculated, as indicated in Section 5.2, on the assumption that the variance of the irregular components did not decrease in 1967 with the seasonal amplitude.

Figure 9

Seasonally adjusted series using new method



6.2 Setting up the model

In order to apply the method to component series it is necessary to estimate f and the point at which it should be applied (r, s). The easiest way to investigate this is to run the program with annual updating and with no preliminary scaling factor and to use the local scaling factors as a monitoring device. This was discussed in Section 5.2 with reference to the total series. In this way a preliminary estimate of f is made (and of (r, s)) and the program is rerun. If the first estimate was accurate the local scaling factors should approach unity otherwise a further estimate is made. Clearly, although the local scaling factors provide a useful tool they should not be used in isolation and the various statistics discussed above all provide useful indications to the stability achieved.

6.3 The adjusted series

For the total series the prior amplitude scaling factor was applied before July 1967. It seemed desirable to use a regression which does not include observations with the later amplitude to estimate the seasonal coefficients to be used up to June 1967. The obvious base to use is that ending in December 1966 and this base was in fact used up to April 1967. After June 1967 it is desirable to include as much of the new data in the estimation period as possible, i.e. to use the adjusted series from the latest possible base position. When the method was introduced the latest base available ended in June 1969 and this was used to adjust from October 1967 onwards. Between April and October a graduated system was used in order to provide a smooth transition across the discontinuity. The constants used were those from the final base position and the adjustment is therefore given by:

$$z^s(67, j) = \frac{z(67, j) - \dfrac{5d(67, 6)a(j)}{f(10-j)}}{1 + \dfrac{5d(67, 6)b(j)}{f(10-j)}} \qquad (24)$$

$$j = 5, \ldots 9$$

The change between the two sets of seasonal constants used is then made between April and May where the seasonal amplitude is small. The total unemployment series adjusted using this method is shown in Figure 9. The graph shows that the over-adjustment of the earlier method has been eliminated.

7. Concluding remarks

7.1 Review of performance of the method

It will be noted that the prior amplitude scaling was applied to the seasonal variation, i.e. to both additive and multiplicative factors and not just to one of them. The reasons were (i) the seasonalities of the total series and the series for males with the 1957-66 regression base were strongly multiplicative, as is evident for the total series from Figure 7, and (ii) the series for females with the 1957-66 base is strongly additive and yet shows a large reduction in amplitude after 1967, as may be seen from the local amplitude scaling factors in Table 6c. Thus the seasonal amplitude of the total series for a trend level falling to 250,000 would be, on present evidence, about 33,000. Only a change in trend level can show whether or not future seasonality will have such a dependence on the trend level. It follows that the performance of the model must be reviewed from time to time.

Another reason for reviewing the performance of the model is that there may be another sharp change in the seasonal amplitude. Should this happen, either a further prior amplitude scaling may be desirable or, if the series reverts to its pre-1967 behaviour, it may be desirable to apply the prior amplitude scaling to the relevant portion of the series after 1967.

However, neither of the eventualities mentioned in the previous paragraphs is likely to cause immediate concern, since small changes in seasonal amplitude are taken into account by the local amplitude scaling factors and these factors provide an early warning of large changes. What the local amplitude scaling factor cannot do well is take account of sudden changes in seasonal pattern, considered apart from the amplitude; Changes in seasonal pattern are taken into account progressively as the base is updated.

In this connection it is to be noted that the calculation of the local factors using the centred 12-month moving average may be biased during any future rapid changes in trend level.

7.2 Updating and revisions

In the previous section the manner in which current estimates are updated quarterly was described. However the local amplitude scaling factor used at the end of the series is centred eighteen months out of date and the estimation period may contain as-yet-unmodified year-ahead extremes. As more data become available it is possible to provide revised estimates free from the influence of possible year-ahead extreme values and also using correctly centred local amplitude scaling factors. If we consider the situation when we have an additional 12 months data for a series previously ending in June of year $(n+1)$ then large extremes in that year may be modified using the regression base ending in December of year n. If any large extremes are found then by re-running the base ending in December of year n with such extremes modified in the input data, genuinely unbiased estimates of the seasonal constants are obtained. Furthermore the local amplitude scaling factor centred on December of year $(n—1)$, as it only uses data to June of year $(n+1)$, is also unbiased by extreme values so that the twelve months, July of year $(n—1)$ to June of year n, seasonally adjusted using this scaling factor applied to the constants from the regression base ending in December of year n, may be considered as final estimates. It is inherent in the logical use of the new method that a seasonally adjusted current observation may be revised three times, i.e. it only reaches its final value when three years' further observations become available.

7.3 The historical series

Throughout the research attention has been concentrated on the problem of adjusting current observations as well as possible and little work has been done on the generation of a definitive historical series in which, for each year for which the definitive adjustment is effected, use is made of several years of subsequent observations.

The adjustments of the total series up to 1966 effected by the 1965-70 method described in Section 1 appear to be reasonably close to those given by the new method for the same period and have been accepted as an adequate series. The most noticeable discrepancies are at the peaks of the series for which the new method would generally give slightly higher results than the old. However, with some of the component series it was necessary to use the new method back to 1964.

The main use of an historical seasonally adjusted series is in econometric analyses using a long run of data and it is not felt that such discrepancies as there may be between the new and old methods prior to 1966 will have any serious consequences.

7.4 Use of the new method

Since the initial work on the total series was done the method has been applied to all the component series (unemployment classified by sex, by industry and by region) and in all except one case a recent reduction in the seasonal amplitude was found. The method was applied precisely as described in Section 6.2 with the local scaling factors being used to estimate the preliminary scaling factor; the program was then re-run successively, adjusting the preliminary amplitude scaling factor until the local scaling factors were close to unity.

This procedure was not followed for the total series, for which a preliminary scaling factor of 0·7 was chosen by another method: this reduction in the seasonal variation is shown in figure 9b. Since then, it has been found that the seasonally adjusted total series did not agree with the three series formed by adding the components, although the three series agreed between themselves.

We feel that these discrepancies are caused by the different procedures used to obtain the preliminary scaling factors. Because the factor for the total series was chosen rather too large, not all the post-1967 amplitude reduction is accounted for, and as a result the regression takes more time to adapt than it need. Also the use of a larger factor for the total series means that the regression is not so heavily weighted with respect to recent observations. If a preliminary factor is calculated by the method of Section 6.2 a value of 0·55 as compared with previous value of 0·7 is found. Re-running the program for the total series produces a series very close to the aggregates and as a result the Department of Employment intend to adopt this value for future work. This shows that the procedure described in Section 6.2 is a powerful one. However, we feel that this experience underlines the necessity for a further development of the program in which the weighting effect is separated from the preliminary scaling factor.

The method has also been applied to the vacancy series and again a recent reduction in seasonal amplitude was found.

Acknowledgements

The work on this project was done in the Research and Special Studies Division of the Central Statistical Office (CSO) and was discussed at intervals by an interdepartmental group comprising A. R. Thatcher, R. Ash and D. E. Bradbury of the Department of Employment; A. D. Roy and H. P. Evans of the Treasury; R. E. Beales and W. D. Stedman Jones of the CSO, and the authors. Professor C. A. Moser, Director of the CSO, was chairman. The authors are indebted to the members of the group for their advice and help and to the Department of Employment for providing some of the results reported in the paper.

We are grateful to Mr. J. P. Burman of the Economic Intelligence Department, Bank of England for several helpful discussions at various stages of the work and particularly for some valuable comments on the first draft of this paper.

We also wish to record our thanks to Mrs. V. Isham (Assistant Statistician, CSO) who worked on the project during the first half of the research and to Miss S. Brown and R. R. Heard of the Statistics Research Division, London School of Economics, who helped us during the early stages of the development of the computer programs.

References

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, Vol. 2, pp 123-147.

Burman, J. P. (1965). Moving Seasonal Adjustment of Economic Time Series. *Journal of the Royal Statistical Society, Series A*, Vol. 128, Part 4, pp 534-558.

Employment and Productivity Gazette (1970), Vol. 78, pp 285-287.

Economic Trends (1970), No. 199, pp xvi-xx.

Efroymson, M. A. (1960). Multiple Regression Analysis (Stepwise). *Mathematical Methods for Digital Computers.* Ed. by Ralston, A. and Wilf, H. S. Wiley, New York.

Godfrey, M. D. and Karreman, H. (1967). A Spectral Analysis of Seasonal Adjustment. *Essays in Mathematical Economics in Honour of Oskar Morgenstern.* Ed. by Shulik, M. Princeton University Press.

Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, chapter 7. Wiley, New York.

Ministry of Labour Gazette (1960), Vol. 68, pp 89-92.
Ministry of Labour Gazette (1965), Vol. 73, pp 382-383.

Tukey, J. W. (1962). The Future of Data Analysis. *The Annals of Mathematical Statistics*, Vol. 33, pp 1-68.

Wald, A. (1936). Berechnung und Ausschaltung von Saisonschwankungen. *Beiträge zur Konjunkturforschung, No. 9.* Verlag von Julius Springer. Wien.

Printed in England for Her Majesty's Stationery Office by H. O. Lloyd & Co. Ltd., London N1 8HX
Dd. 500982 K.10

© *Crown copyright* 1971
Published by
HER MAJESTY'S STATIONERY OFFICE
To be purchased from
49 High Holborn, London WC1V 6HB
13a Castle Street, Edinburgh EH2 3AR
109 St Mary Street, Cardiff CF1 1JW
Brazennose Street, Manchester M60 8AS
50 Fairfax Street, Bristol BS1 3DE
259 Broad Street, Birmingham B1 2HE
80 Chichester Street, Belfast BT1 4JY
or through any bookseller

SBN 11 630094 9

34,079