Extracted from *Economic Trends* No. 199, May 1970

New method for seasonal adjustment of unemployment series

R. L. BROWN, *Chief Statistician*, A. H. COWLEY, *Assistant Statistician*
and J. DURBIN,* *Consultant, Central Statistical Office.*

Introduction

An investigation of the method of seasonal adjustment introduced by the Department of Employment and Productivity (DEP) in 1965(¹) was initiated during the second half of 1968 because it was felt that the method might be over-adjusting the recent observations. 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 normal 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. The work was done in the Research Division of the Central Statistical Office in collaboration with the DEP and the Treasury.

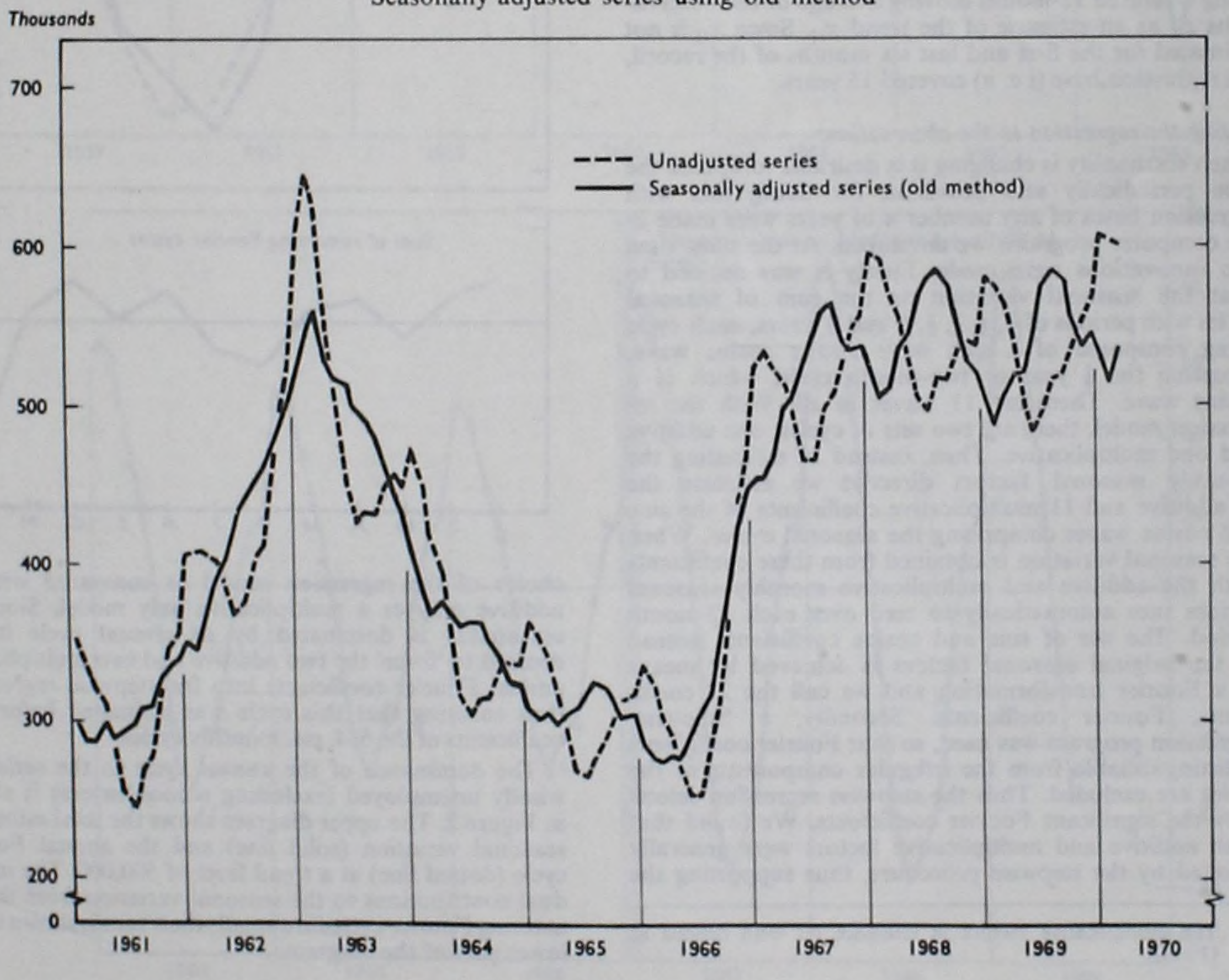
The investigation led to the development of a new method of adjustment adopted officially(²) in April 1970. A description of the new method is given in this note. A technical account of the entire investigation will be published in Studies in Official Statistics, Research Series. Although the main features of the new method are illustrated in this note, the reader is referred to the technical account for the detailed results that led to them.

The figures for 1969 confirmed that the old method was over-adjusting the series (see Figure 1). Our examination led us to the conclusion that an unusually large and abrupt change in seasonality had occurred during 1967, and we were able to use the observations for 1968 and 1969 to show that the main change was a reduction in the amplitude of the seasonal variation.

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

(¹) *Ministry of Labour Gazette*, September 1965, vol. 73, pp. 382-3.
(²) *Employment & Productivity Gazette*, April 1970, vol. 78, pp.285-7.

Figure 1 **Wholly unemployed (excluding school-leavers)**
Seasonally adjusted series using old method



Development of the regression model

The DEP model

A regression model was introduced by DEP in September 1965 because the additive model used previously was not thought to be working satisfactorily. The regression model contains monthly additive (α_j) and multiplicative (β_j) seasonal factors, where the subscript j refers to successive months. Using the subscript i for successive years of the base, the regression takes the form

$$E(z_{ij}-x_{ij}) = \alpha_j+\beta_j x_{ij} \qquad i = 1, 2, \ldots n \quad j = 1, 2, \ldots 12 \tag{1}$$

with $\sum_1^{12}\alpha_j = \sum_1^{12}\beta_j = 0$, where $E(z_{ij}-x_{ij})$ means the expectation of the deviation of the observation z_{ij} from the trend x_{ij}, n is the number of years in the base and the constraints $\Sigma\alpha_j = \Sigma\beta_j = 0$ ensure that the adjusted series sums each year to the sum of the original series when the trend level is constant. The seasonally adjusted series $\tilde{z}_{ij}$ is found from

$$\tilde{z}_{ij} = \frac{z_{ij}-a_j}{1+b_j} \tag{2}$$

where a_j, b_j are estimates of α_j, β_j respectively([3]).

The regression model was fitted by DEP to the 16 years of observations from June 1949 to May 1965, using a centred 12-month moving average of the observations z_{ij} as an estimate of the trend x_{ij}. Since x_{ij} is not estimated for the first and last six months of the record, the regression base (i.e. n) covered 15 years.

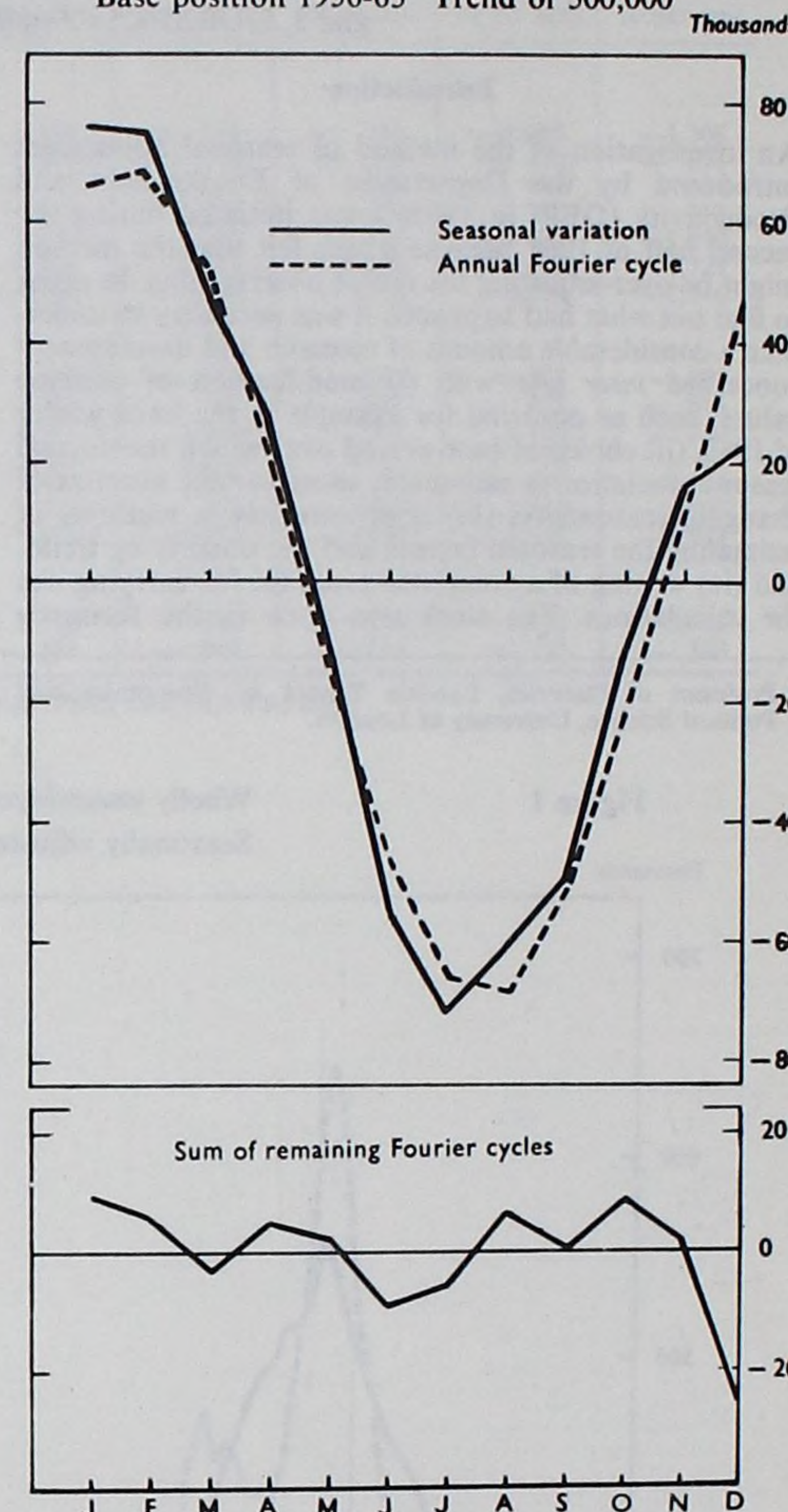
Fitting the regression to the observations

When seasonality is changing it is desirable to update the base periodically and provision for doing this with regression bases of any number n of years were made in the computer programs we developed. At the same time two innovations were made. Firstly it was decided to treat the seasonal variation as the sum of seasonal cycles with periods of 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$ and $\frac{1}{6}$ years, each cycle being composed of a sine wave and a cosine wave, excepting the $\frac{1}{6}$ year or two-month cycle, which is a cosine wave. There are 11 waves in all. With the regression model, there are two sets of cycles, one additive and one multiplicative. Then, instead of estimating the monthly seasonal factors directly, we estimate the 11 additive and 11 multiplicative coefficients of the sine and cosine waves comprising the seasonal cycles. When the seasonal variation is obtained from these coefficients both the additive and multiplicative monthly seasonal factors sum automatically to zero over each 12-month period. The use of sine and cosine coefficients instead of the original seasonal factors is achieved by means of a Fourier transformation and we call the 22 coefficients, Fourier coefficients. Secondly, a 'stepwise' regression program was used, so that Fourier coefficients indistinguishable from the irregular components of the series are excluded. Thus the stepwise regression selects only the significant Fourier coefficients. We found that both additive and multiplicative factors were generally selected by the stepwise procedure, thus supporting the

([3]) The multiplicative factors in reference ([1]) were defined as $(1+\beta j)$.

Figure 2 **Fourier cycles**

Base position 1956-65 Trend of 500,000



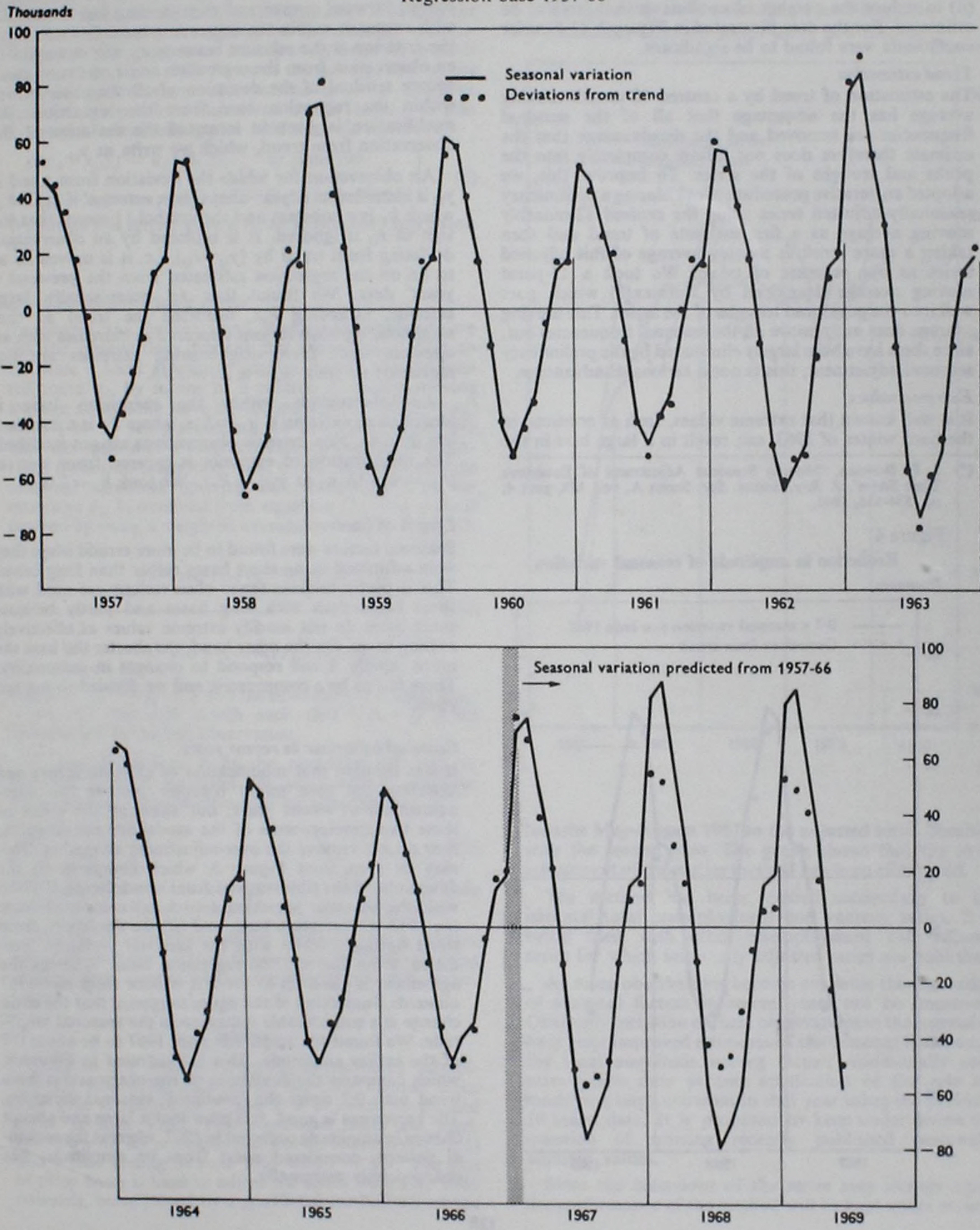
choice of the regression model as compared with an additive only or a multiplicative only model. Since the seasonality is dominated by its annual cycle it was decided to 'force' the two additive and two multiplicative annual Fourier coefficients into the stepwise regression, thus ensuring that this cycle was estimated before the coefficients of the 6, 4, etc. monthly cycles.

The dominance of the annual cycle in the series for wholly unemployed (excluding school-leavers) is shown in Figure 2. The upper diagram shows the total estimated seasonal variation (solid line) and the annual Fourier cycle (dotted line) at a trend level of 500,000. The individual contributions to the seasonal variation from the remaining Fourier cycles are small; their sum is shown in the lower part of the diagram.

Figure 3

Seasonal variation and deviations from trend

Regression base 1957-66



The effects of these two innovations are (i) to give a smoother seasonal variation, and hence less oversmoothing of the seasonally adjusted series, than would be obtained by estimating the 24 seasonal factors directly and (ii) to reduce the number of coefficients that have to be estimated. For the data illustrated in Figure 2, 11 Fourier coefficients were found to be significant.

Trend estimation

The estimation of trend by a centred 12-month moving average has the advantage that all of the seasonal frequencies are removed and the disadvantage that the estimate therefore does not follow completely into the peaks and troughs of the series. To improve this, we adopted an iterative procedure of calculating a preliminary seasonally adjusted series using the centred 12-monthly moving average as a first estimate of trend and then taking a more sensitive moving average of this adjusted series as our estimate of trend. We took a 13-point moving average developed by Burman([4]) which goes well into the peaks and troughs of the series. This moving average does not remove all the seasonal frequencies but, since these have been largely eliminated by the preliminary seasonal adjustment, this is not a serious disadvantage.

Extreme values

It is well known that extreme values, such as occurred in the hard winter of 1963, can result in a large bias in the estimated seasonal factors. After some experimentation we decided to modify extreme values in two steps, looking first for large extremes in the year following the regression base and modifying them before the base is moved forward a year and then dealing less drastically with extremes within the regression base. In both cases the criterion is the relation between r_{ij}, the deviation of an observation from the regression and s, the root mean square residual of the deviation of all the observations within the regression base from the regression: the modification is given in terms of the deviation of the observation from trend, which we write as y_{ij}.

An observation for which the deviation from trend is y_{ij} is identified as a 'year-ahead large extreme' if $|r_{ij}|>k_1 s$ where k_1 is a constant and the symbol | | means that the sign of r_{ij} is ignored. It is replaced by an observation deviating from trend by $(y_{ij}-r_{ij})$, i.e. it is moved so as to lie on the regression estimated from the previous n years' data. We found that an exceptionally large extreme, exceeding $k_o s$, distorted the trend so that neighbouring observations appeared as extremes with an opposite sign: these neighbouring extremes are not modified: we took $k_o = 6$, $k_1 = 3.75$.

An observation within the regression base is identified as extreme if $|r_{ij}|>k_2 s$, where k_2 is a constant less than k_1. Non-extreme observations are not modified. The modification of extremes is tapered from zero at $|r_{ij}| = k_2 s$ to r_{ij} at $|r_{ij}| = k_1 s$. We took $k_2 = 2.00$.

Length of base

Seasonal factors were found to be more erratic when they were estimated using short bases rather than long bases. This is partly because fewer observations are used with short bases than with long bases and partly because short bases do not modify extreme values as effectively as long bases. On the other hand, the shorter the base the more rapidly it will respond to changes in seasonality. There has to be a compromise and we decided to use ten years.

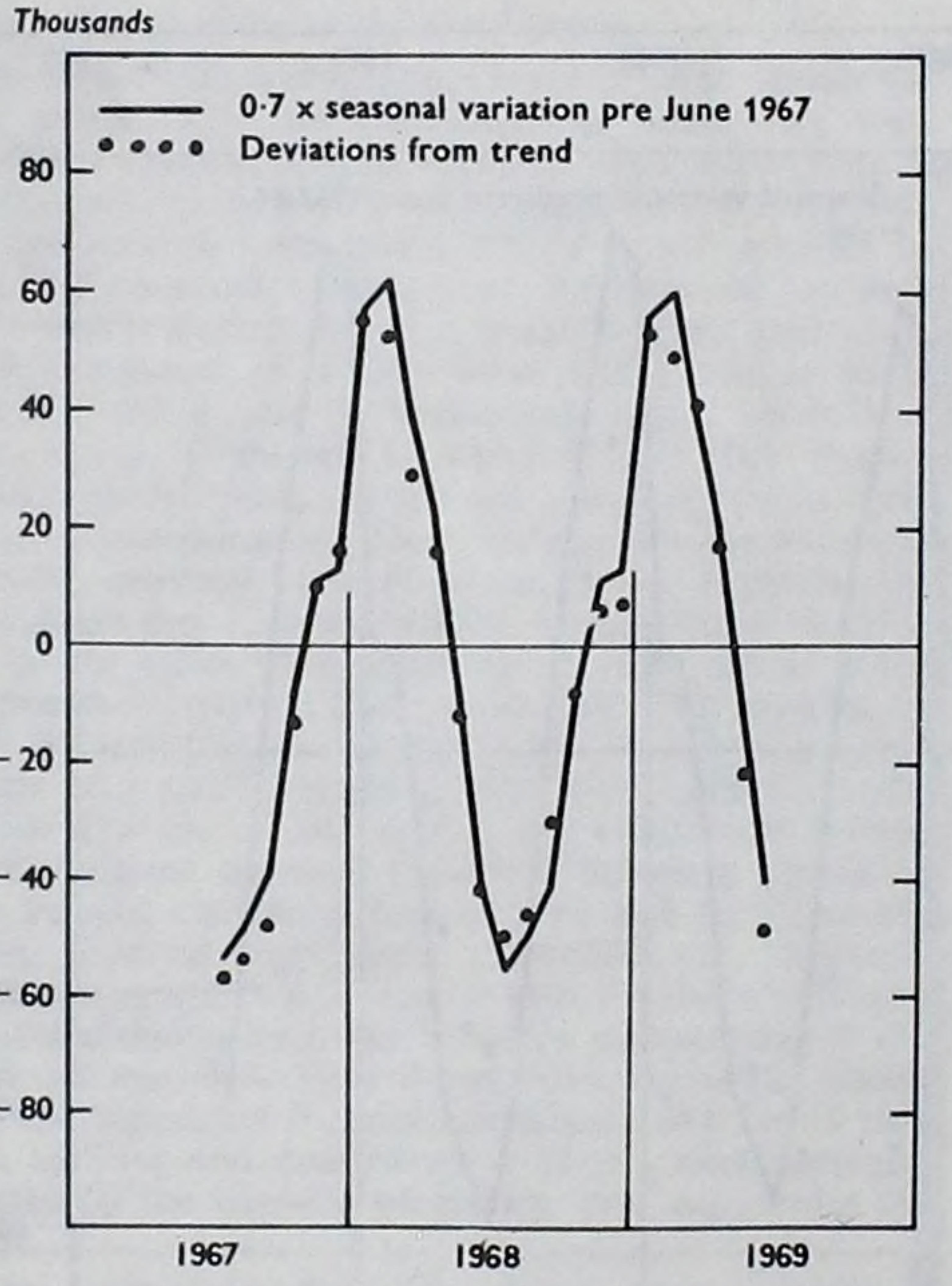
Seasonal behaviour in recent years

It was thought that modification of extreme values and updating the base might together remove the over-adjustment in recent years, but although the effect of these two developments of the model are considerable, they do not remove the over-adjustment altogether. This may be seen from Figure 3, which compares (i) the deviations of the observations from trend during 1957-66 with the seasonal variation $(a_j+b_j x_{ij})$ computed from the 1957-66 regression base, and (ii) the deviations from trend during 1967-69 with the seasonal variation 'predicted' from the 1957-66 regression base. Whereas the agreement is good up to 1966, it is poor from mid-1967 onwards. Inspection of the figure suggested that the main change is a considerable reduction in the seasonal amplitude. We found the amplitude after 1967 to be about 0·7 of the earlier amplitude. This is illustrated in Figure 4, which compares the deviations of the observations from trend with 0·7 times the 'predicted' seasonal variation. The agreement is good. It is clear that a large and abrupt change in amplitude occurred in 1967, whereas the seasonal pattern, considered apart from its amplitude, has changed only marginally.

([4]) J. P. Burman, 'Moving Seasonal Adjustment of Economic Time Series', *J. Roy. Statist. Soc.* Series A, vol. 128, part 4, pp. 534-558, 1965.

Figure 4

Reduction in amplitude of seasonal variation



Allowances for changes in amplitude

Improving the fit of the regression in recent years

We found that this large change in amplitude caused the fit of the regression to become increasingly unsatisfactory as the regression base was updated to include observations since 1967. So as to use information on the seasonal pattern from observations prior to 1967, we scaled their deviations from trend by the factor f, so that the model for estimating post-1967 seasonal factors is:

$$\left.\begin{aligned} E(z_{ij}-x_{ij})\,f &= \alpha_j+\beta_j\,x_{ij}, &&\text{to June 1967}\\ E(z_{ij}-x_{ij}) &= \alpha_j+\beta_j\,x_{ij}, &&\text{from July 1967}\end{aligned}\right\} \quad (3)$$

We found this device gave a more satisfactory regression fit.

Local amplitude scaling factor

To correct any further small changes in amplitude we use the observations for the last three years' data to calculate a local amplitude scaling factor. We estimate the trend x_{ij} by means of a centred 12-month moving average so that we have deviations from trend $(z_{ij}-x_{ij})$ from $\frac{1}{2}$ to $2\frac{1}{2}$ years preceding the last observation. We take the ratio d of a weighted average over these deviations from trend, ignoring sign, to the weighted average seasonal variation, ignoring sign, calculated from the estimates a_j, b_j obtained from equation (3). This ratio is centred by using a weighted average over 25 months, viz,

$$d=\frac{\sum_{\rho=-12}^{12} w_\rho\left|z_{12(u-1)+v+\rho}-x_{12(u-1)+v+\rho}\right|}{\sum_{\rho=-12}^{12} w_\rho\left|a_{v+\rho}+b_{v+\rho}\,x_{(12\,u-1)+v+\rho}\right|} \quad (4)$$

where $w_{-12}=w_{12}=1$ and the remaining w's equal 2; u, v are the year and month such that $(12u+v)$ is six months before the last observation.

It was decided to re-fit the regression (3) and to estimate d from equation (4) at quarterly intervals. Then the next three observations are adjusted by

$$\tilde{z}_{ij}=\frac{z_{ij}-da_j}{1+db_j} \quad (5)$$

The method responds to changes in the seasonal amplitude more quickly than to changes in the seasonal pattern.

The local amplitude scaling factor serves a further valuable purpose in that it provides an early warning of any future large change in seasonality that may occur.

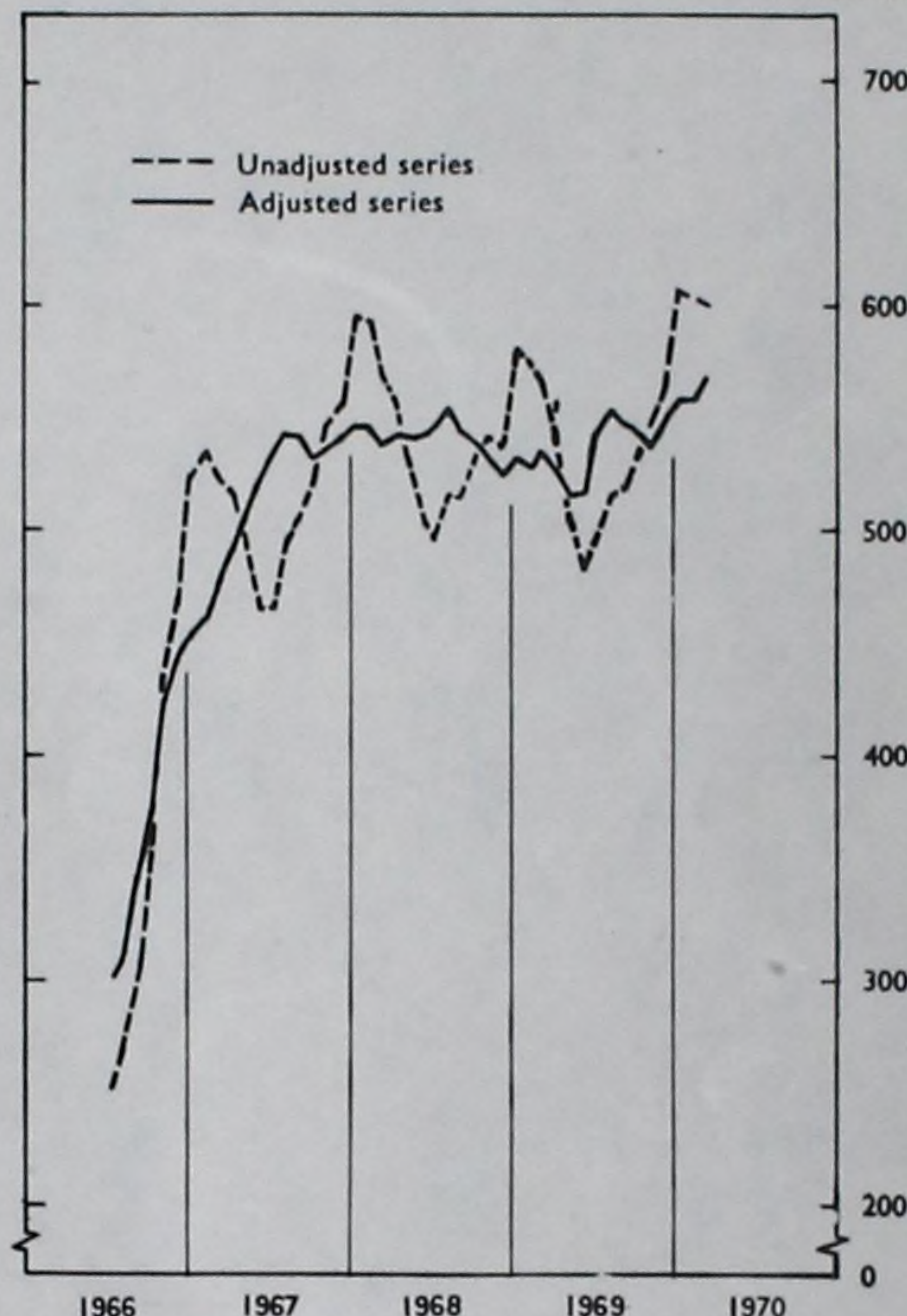
Results

The series for wholly unemployed (excluding school-leavers) seasonally adjusted by the new method since July 1966 is plotted in Figure 5: it is joined to the old series at June 1966. Two regression bases are used, 1957-66 with no preamplitude scaling factor (i.e. $f=1$) and a base ending in June 1969 with $f=0{\cdot}7$. The first of these bases is used to adjust the series from July 1966 onwards, being joined by a graduated method across the

Figure 5

Wholly unemployed (excluding school-leavers)

Seasonally adjusted series using new method



months May-August 1967 to the adjusted series obtained with the second base. The graph shows that the over-adjustment of the earlier method has been eliminated.

The method has been applied successfully to the national total unemployment and vacancy series. It is being tried with other unemployment and vacancy series for which seasonally adjusted series are published.

As more observations become available the estimation of seasonal factors in recent years can be improved. Obviously inclusion of these observations in the regression base gives improved estimates of the seasonal factors and the local amplitude scaling factor. Additionally each extra year's data permits application of the rule for modifying large extremes in that year using the previous 10 years' data. It is proposed to keep under review the question of revising recently published seasonally adjusted values.

Since the behaviour of the series may change again the performance of the method will be kept under review.

Produced in England for Her Majesty's Stationery Office by Commercial Colour Press, London.

Dd.503384 K.10 2.72

HER MAJESTY'S STATIONERY OFFICE

Government Bookshops

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
258 Broad Street, Birmingham B1 2HE
80 Chichester Street, Belfast BT1 4JY

Government publications are also available through booksellers

SBN 11 630107 4