

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER No. 151

Note: Cowles Foundation Discussion Papers are preliminary materials circulated to stimulate discussion and critical comment. Requests for single copies of a Paper will be filled by the Cowles Foundation within the limits of the supply. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has access to such unpublished material) should be cleared with the author to protect the tentative character of these papers.

SEASONAL ADJUSTMENT OF ECONOMIC TIME SERIES AND MULTIPLE REGRESSION

Michael C. Lovell

January 14, 1963

SEASONAL ADJUSTMENT OF ECONOMIC TIME SERIES AND MULTIPLE REGRESSION

Michael C. Lovell*

After demonstrating that any nontrivial technique for seasonally adjusting time series inevitably leads to certain distortions of the data, an effort is made to provide explicit motivation for the process of seasonal adjustment for purposes of appraising current economic conditions. Inherent advantages in terms of certain consistency requirements of a least squares procedure for seasonal adjustment are pointed out. Problems encountered by the econometrician when seasonally adjusted time series are to be employed in regression analysis are also explored. The dummy variable technique for dealing with seasonal fluctuations is generalized to encompass a flexible pattern of seasonal movement. It is argued that when seasonally adjusted data rather than the dummy variable procedure are employed, there is an inherent tendency to overstate the significance of regression coefficients; a correction procedure is suggested. Consideration is given to certain special problems created by autocorrelated residuals when seasonally adjusted data are utilized in regression analysis.

* I am indebted to Mr. Leopold A. Sveikauskas and to my friends at the Cowles Foundation for Research in Economics at Yale University for helpful comments and constructive criticism. This research has been supported by a National Science Foundation Research Grant.

SEASONAL ADJUSTMENT OF ECONOMIC TIME SERIES AND MULTIPLE REGRESSION

Michael C. Lovell*

After demonstrating that any nontrivial technique for seasonally adjusting time series inevitably leads to certain distortions of the data, an effort is made to provide explicit motivation for the process of seasonal adjustment for purposes of appraising current economic conditions. Inherent advantages in terms of certain consistency requirements of a least squares procedure for seasonal adjustment are pointed out. Problems encountered by the econometrician when seasonally adjusted time series are to be employed in regression analysis are also explored. The dummy variable technique for dealing with seasonal fluctuations is generalized to encompass a flexible pattern of seasonal movement. It is argued that when seasonally adjusted data rather than the dummy variable procedure are employed, there is an inherent tendency to overstate the significance of regression coefficients; a correction procedure is suggested. Consideration is given to certain special problems created by autocorrelated residuals when seasonally adjusted data are utilized in regression analysis.

* I am indebted to Mr. Leopold A. Sveikauskas and to my friends at the Cowles Foundation for Research in Economics at Yale University for helpful comments and constructive criticism. This research has been supported by a National Science Foundation Research Grant.

1. Introduction:

Seasonal fluctuations in economic time series present problems for both the business analyst and the econometrician. The business analyst interested in interpreting current economic movements is disconcerted by the frequent revisions to which seasonally adjusted time series are subjected as additional observations accumulate.* An equally difficult

* Seasonal patterns obtained in application of ratio-to-moving average procedures often are subject to considerable revision when additional observations become available. Mort S. Raff and Robert Stein [1960] report that "... changes of up to 7 percent in several of the monthly factors (were required) when 1958 data were added to the 1947-57 unemployment series. The unreasonableness of such large changes was recognized by an interagency committee, which decided to continue using the old seasonal factors until stronger evidence became available about changes in seasonal pattern." The difficulties have continued. New correction factors published in the February 1962 Monthly Labor Review reveal further revisions of fair magnitude of correction factors calculated with data through September 1961, although the differences were probably no larger than should be expected in any case from sampling error.

problem, emphasized by Professor Samuelson in a letter to the editor of the Sunday New York Times, November 12, 1961, concerns the failure of official seasonally adjusted time series to satisfy certain obvious consistency requirements. Samuelson pointed out that seasonally adjusted unemployment does not equal the excess of the adjusted labor force figure over the seasonally adjusted volume of employment. When one examines the movement of "residual adjusted unemployment," defined as the excess of adjusted labor force over adjusted employment, the experience of the early months of the

Kennedy administration appears much more satisfactory than it does in terms of the official figures.*

* John A. Brittain [1959] presents a detailed comparison of the residual unemployment concept with the official series.

In this paper we consider first the logical implications of certain consistency requirements that might appropriately be applied in appraising alternative procedures for decomposing time series. Then, in part 3, motivation for the process of time series decomposition is provided by examining the objectives of seasonal adjustment from the viewpoint of the business analyst interested in appraising the direction of current economic forces. Part 4 explores certain inherent advantages, in terms of the consistency criteria, of a technique for seasonal adjustment that may be executed by standard least squares procedures without special programming.

Current econometric studies reveal a lack of concensus concerning the most appropriate technique for handling seasonal movements when quarterly or monthly data are to be utilized for parameter estimation and hypothesis testing. Time series adjusted by the ratio to moving average procedure have been employed in studies by Barger and Klein [1954], by Colin Clark [1949], by Duessenberry, Eckstein and Frome [1960], and in a number of other recent econometric investigations. On the other hand, the Klein, Ball, Hazlewood and Vandome econometric study of the United Kingdom [1961] constitutes but one example of the introduction of seasonal dummy variables in an attempt

to net out the effects of seasonality present when working with unadjusted time series. Lawrence Klein warns in his Textbook of Econometrics [1953, p. 321]:

One must be careful in using data already adjusted for trend or seasonal variation not to attribute too many degrees of freedom to the sample data.

This caveat is generally ignored in reporting econometric results obtained with seasonally adjusted time series. Because Klein does not specify precisely the number of degrees of freedom that are sacrificed in the process of seasonal adjustment, the question of how one should proceed when data is available only in adjusted form remains unanswered.*

* The National Income Division publishes only seasonally adjusted quarterly data for certain components of the national income accounts. In a study of the behavior of prices, inventory, and production in the shoe, leather, hide sequence Kalman Cohen [1960, pp. 89-92] attempted to reintroduce estimated seasonal movements into time series available to him only in adjusted form.

Problems encountered when seasonally adjusted data are utilized in regression analysis are explored in part 5 of this paper. The possibility of bias when the data have been subjected to prior adjustment is appraised within the framework of Theil's procedure for treating specification error. Then an extension of a basic theorem of Ragnar Frisch and Waugh [1933] is presented. They had established that the least squares removal of linear trend from individual time series in advance of regression yields parameter

estimates identical to those that would have been obtained with unadjusted data if time had been included as an additional explanatory variable.*

* Tintner [1952, p. 304] generalized the argument to encompass polynomial trends.

The Frisch-Waugh argument will be extended in this paper to encompass seasonal as well as trend adjustment. Including seasonal dummy variables within a regression equation will be shown to be equivalent to working with time series from which the seasonal pattern has been removed by the least squares procedure. The two approaches are equivalent even when a changing seasonal pattern is considered or when dummy variables representing the role of such erratic factors as strikes are introduced. The analysis suggests a procedure for estimating the number of degrees of freedom sacrificed when seasonally adjusted data are utilized in regression analysis. A correction factor is advanced as an appropriate means for offsetting the inherent tendency to overstate the significance of regression coefficients when calculations are based upon adjusted time series. Consideration is also given to certain special problems created by autocorrelated residuals when seasonally adjusted data are utilized in regression analysis.

2. Criteria for Appraising Techniques of Seasonal Adjustment:

Seasonal adjustment hopefully transforms raw data into a new time series which in some sense will prove more revealing than the original observations. Choice among alternative techniques for seasonal adjustment is customarily made on the basis of a subjective appraisal of the quality of the results obtained in the experimental application of various procedures to actual or artificially constructed time series. An alternative approach is to analyze the implications of a set of explicitly stated requirements that any ideal technique of seasonal adjustment might reasonably be expected to satisfy. An analogous approach was followed by Irving Fisher in his defense of the ideal index number.

Let us first enumerate a set of properties that might generally be accepted as a reasonable, if incomplete list of criteria for appraising adjusted time series. Then we shall examine their implications:

Property I: The Adjustment Procedure Should Preserve Sums

$$(2.1) \quad x_t^a + y_t^a = (x_t + y_t)^a \quad \text{for all } t ,$$

where x_t and y_t are the original observations on any pair of time series and x_t^a and y_t^a are the adjusted observations.

Such definitional requirements as that arising with labor force and unemployment data will automatically be satisfied if an adjustment procedure meeting this additivity requirement is utilized. In addition, no discrepancies in aggregation will be generated if component series rather than totals are

subjected to deseasonalization with a procedure satisfying this requirement.*

* The most recent practice in the seasonal adjustment of employment data involves separate adjustment of four distinct groups in advance of aggregation. An interesting comparison of the effects of the direct adjustment of aggregates with the sum of adjusted components is presented by D. J. Daly [1960].

Property II: The Adjustment Procedure Should Preserve Products

$$(2.2) \quad x_t^a y_t^a = (x_t y_t)^a \quad \text{for any two time series } x_t \text{ and } y_t .$$

If the adjustment procedure preserves products, it is immaterial whether one deflates raw value series by an unadjusted price index and then subjects the deflated value series to seasonal adjustment or, alternatively, adjusts the quantity and price series in advance of deflation.* The ratio of seasonally

* I am indebted to Thomas M. Stanback Jr. for first bringing this problem to my attention.

adjusted unemployment to seasonally adjusted labor force would be identical to the time series obtained by deseasonalizing the unadjusted unemployment ratio.

Property III: The Adjusted Data Should Be Uncorrelated With The Correction Terms

$$(2.3) \quad \sum_t (x_t - x_t^a) x_t^a = 0 .$$

The seasonal movement of the time series cannot be properly defined by the correction terms $x_t - x_t^a$ if this independence condition is not satisfied, for if this were the correct definition, the fact that the correlation is not zero implies that some seasonality remains in the data.

Property IV: The Adjustment Procedure Should Be Idempotent

$$(2.4) \quad (x_t^a)^a = x_t^a$$

Subjecting to seasonal adjustment any time series that has already been deseasonalized should have no effect on the series. It is interesting to note that the Bureau of Labor Statistics' technique of subjecting unemployment data to repeated processing by the ratio to moving average procedure achieves this desirable property by iteration.*

* For a description see Raff and Stein [1960].

Property V: The Adjustment Procedure Should Be Symmetric

$$(2.5) \quad \frac{\partial x_t^a}{\partial x_{t'}^a} = \frac{\partial x_{t'}^a}{\partial x_t^a} \quad \text{for all } t \text{ and } t' .$$

An increase in one of the observations of an unadjusted time series may be expected to affect a number of components of the time series obtained by seasonally adjusting the perturbed series; it is reasonable to require that such effects be symmetric.

The following propositions concerning the implications of these various properties are established in the appendix:

Theorem 2.1: Properties I and III imply IV and V; Properties I, IV and V imply III

If the adjustment procedure preserves sums and yields adjustment terms that are uncorrelated with the adjusted data, then the adjustment procedure is also idempotent and symmetric. Any adjustment procedure that preserves sums, is idempotent, and symmetric necessarily yields seasonal adjustment factors that are uncorrelated with the adjusted data. Later, in part 4, we shall explain how standard least squares regression techniques may be applied in order to execute on an electronic computer without special programming any seasonal adjustment technique that possesses these desirable properties.

Theorem 2.2: If an adjustment procedure possesses both Properties I and II, then it is trivial in the sense that for each t either:

$$(2.6) \quad x_t^a = x_t \text{ or } x_t^a = 0 .$$

There exists no non-trivial seasonal adjustment procedure that preserves both sums and products. An adjustment procedure that preserves the definitional relationship between employment, unemployment, and the size of the labor force, cannot be expected, in general, also to preserve the relationship

between revenue and price times quantity or, for that matter, to yield an adjusted unemployment rate equal to the ratio of adjusted unemployment to adjusted labor force.

Theorem 2.2 is disturbing. Consideration of quite simple criteria rules out the possibility of a generally acceptable "ideal" technique for adjusting economic time series. At the same time, it should be recognized that if we are given two particular time series that are to be compared, it may be possible to construct a non-trivial adjustment procedure that preserves both sums and products for this particular application. One might hope to construct for a given class of economic time series an adjustment procedure that would come "reasonably" close to satisfying such criteria as those we have enumerated. The difficulties encountered in the seasonal adjustment of unemployment data suggest that such a search might well prove unsuccessful. Theorem 2.2 helps to explain why seasonal adjustment procedures that might have appeared to be reasonably satisfactory when tested upon historical data none the less prove unsatisfactory when new observations accumulate.

Since the process of seasonal adjustment has an inherent tendency to distort basic relationships between time series, one may well ask why one should deseasonalize data rather than leave it in unadjusted form. In the next section of this paper an attempt is made to provide explicit motivation for the seasonal adjustment of economic time series. In addition, various seasonal adjustment procedures will be appraised in terms of the criteria discussed in this section. We shall see that the choice as to the criteria which are to be considered as of more fundamental importance is not entirely a matter of

taste, but is dictated in part by the assumptions one chooses to make about the structure by which the time series is generated.

3. Why Deseasonalize?

Seasonal adjustment of economic time series is usually undertaken in the hope that the corrected series will offer a closer understanding of the direction of current economic forces. Donald Daly [1959] and Julius Shiskin [1957] have emphasized that seasonal correction of data on such current operating variables as sales, inventories and orders will provide a more useful guide for the business firm in making operating decisions than the common business practice of comparing present levels with the corresponding period of the previous year. Seasonally corrected data is usually regarded by the business analyst as providing a more accurate indication of current business movements than that which may be obtained from the inspection of unadjusted time series.

When justification is offered for the popular ratio to moving average procedure for seasonal adjustment, it is usually in terms of the assumption that the time series of interest is the product of a trend and cyclical forces, a moving seasonal, and an erratic disturbance;* the problem of

* Kendall [1946, p. 370] defines trend as "a smooth broad motion of the system over a long term of years." Seasonality is defined as "a fluctuation imposed on the series by a cyclic phenomenon external to the main body of causal influences at work upon it."

seasonal adjustment is that of filtering out the seasonal factor without seriously distorting the other elements generating the observed data.*

* Most writers substitute illustrative examples of the effects of eliminating seasonality for an explicit statement of the process by which the data is generated. In contrast, Hamman [1960] provides a precise statement of the problem. Bongard [1960] considers explicitly the problem of finding that centered 19 term moving average that will (a) eliminate a constant seasonal pattern, (b) faithfully reproduce a third degree curve and (c) minimize the variance of the residuals.

In contrast, the business analyst who utilizes adjusted time series is hardly likely to regard the data as having been generated in this way. He is generally all too aware of the multitude of interacting economic forces determining the movements of the time series in question. Since the trend, seasonal, irregular trichotomy underlying the traditional theory of time series decomposition is alien to any attempt at a causal explanation of the operations of the economy, explanation is required for the fact that economists and business analysts frequently find it useful to employ seasonally adjusted time series.**

** One econometrician, Stevan Valavanis [1959, p. 177], has argued that seasonal adjustment of data will help the forecaster only to the extent that businessmen themselves consciously take seasonal forces into account in planning their activities. This position is probably extreme, for systematic errors made by a firm in anticipating demand due to failure to properly allow for seasonal fluctuations may have a pronounced impact upon its inventory and hence effect its production planning and employment scheduling.

In order to appreciate how seasonal adjustment may assist the business analyst, let us suppose that the variable of interest, y_t , is generated by a linear model of a form familiar to both the economic theorist and the econometrician.

$$(3.1) \quad y_t = \alpha + \sum_i \beta_i x_{it} + s_t + \epsilon_t, \quad \mathcal{E}(\epsilon_t) = 0.$$

Here x_{it} is the level of the i^{th} explanatory variable at time t , ϵ_t is an unobserved random error term representing the effect of omitted variables or observational error, and the β_i and α are unknown parameters. Generally, the systematic seasonal disturbances s_t are unobserved. Although as a special case they may be assumed to change from one season to the next but to take on the same value in any given season each year,* the argument

* So that $s_{t+12} = s_t$ with monthly data. As we shall see, this is precisely the circumstance in which the practice of including one dummy variable for each season in regression analysis is appropriate. Section 5 of this paper generalizes this dummy variable regression approach in order to take a moving seasonal pattern into account.

that follows can be generalized in order to take a much more complicated type of seasonal pattern into account. The change in y_t is:

$$(3.2) \quad \Delta y_t = \sum_i \beta_i \Delta x_{it} + \Delta s_t + \Delta \epsilon_t.$$

For purposes of appraising current economic conditions, particular interest centers upon the extent to which changes in y_t are the consequences of

atypical movements of the x_t , apart from the regular movement to be expected on the basis of Δs_t . If observations on the x_{it} have been compiled, estimates $\hat{\beta}_i$ of β_i might be employed to calculate $\sum_i \hat{\beta}_i x_{it}$ as an approximation of the movements in y that are not the consequence of the seasonal forces. An alternative route leading to a related concept is provided by employing an appropriate seasonal adjustment technique. Suppose that we apply a seasonal adjustment procedure that preserves sums and, in addition, is capable of filtering out the s_t seasonal disturbances (i.e., $s_t^a = 0$ for all t). Such an adjustment procedure yields

$$(3.3) \quad y_t^a = \alpha + \sum_i \beta_i x_{it}^a + \epsilon_t^a .$$

For the change in the seasonally adjusted variable we have:

$$(3.4) \quad \Delta y_t^a = \sum_i \beta_i \Delta x_{it}^a + \Delta \epsilon_t^a$$

When a time series is generated by a linear model of form (3.1), examination of changes in the data adjusted by an appropriate procedure preserving sums yields an estimate of the extent to which the movements in the series are generated by atypical changes in nonseasonal economic forces $\sum_i \beta_i x_{it}^a$.

It must be emphasized that if all this does constitute a reasonable formulation of the objectives of seasonal adjustment, then it is clear that the choice between an adjustment procedure that preserves sums and one that preserves products must hinge in part upon the precise nature of the assumptions one chooses to make about the structure generating the data. If, for example, the underlying model were multiplicative (i.e., the

logarithms of the variables behave in accordance with equation (3.1)), then one might elect to employ a product preserving seasonal adjustment procedure. Since assumptions concerning the structure by which the data are generated are implicit in any data adjustment procedure, the task of data adjustment requires the services of an expert with specialized knowledge of the nature of underlying economic processes; seasonal adjustment is not a purely mechanical operation. At the same time, a statistician who subjects economic time series to seasonal adjustment might claim that his procedure is neutral with regard to controversy concerning precisely which economic variables actually generate the y_t in equation (3.1).

Quite simple sum preserving procedures may be employed to filter out the seasonal disturbance term in equation (3.1), provided it takes on the same value in any given season of each year. One needs only to equalize monthly means by subtracting an appropriate constant.* Alternatively, a centered

* If x_{my} is the unadjusted observation of the m^{th} month of the y^{th} year, calculate $x_{my}^a = x_{my} + \left[\frac{\sum_{ym} x_{my}}{(my)} \right] - \frac{\sum x_{my}}{y}$.

twelve month moving average may help to filter out a constant seasonal pattern; although this second procedure preserves sums it involves the sacrifice of the first and last six months of the series.** Furthermore, the twelve

** One procedure for extending the series is to assume that the adjustment factor for the last July is equal to the discrepancy between the actual observation and the moving average for the July of the preceding year, and similarly for August and the other remaining months.

month moving average procedure, while symmetric and sum preserving, is not idempotent. Therefore, by Theorem 2.2, the seasonal correction terms $x_t^a - x_t$ will be correlated with the adjusted series, suggesting that if they are indeed a correct description of the pattern of seasonal movements, not all seasonality has been eliminated from the data. A more complicated alternative is to work in terms of the deviations of the original series from the moving average; the deviations are classified by month and in turn subjected to a moving average in order to obtain the seasonal correction factors which are subtracted from the original observations. While this procedure preserves sums, it is not idempotent. The closely related ratio-to-moving-average procedure usually employed in preparing officially adjusted data does not preserve sums and is inappropriate when the data is assumed to be generated by a model of the form (3.1). In the next section we shall discuss a procedure for constructing filters capable of deleting quite complicated seasonal movements.

4. Least Squares Time Series Decompositions

Although seasonal adjustment of economic time series is usually attempted through the application of some variant of the ratio-to-moving average approach, least squares procedures for seasonal adjustment have been considered by Horst Mendershausen [1939], Dudley Cowden [1942], A. Hald [1948], and H. Eisenpress [1956]. The least squares approach utilizes the residuals obtained by regressing the time series upon an appropriate set of explanatory variables. Certain advantages of the least squares approach are revealed when it is

evaluated in terms of the criteria described in Section 2.

Theorem 4.1: Any adjustment procedure that possesses properties I, IV and V (equivalently I and III) can be executed by regressing the unadjusted time series upon an appropriate set of explanatory variables; conversely, the residuals obtained through the regression of the data upon an appropriate set of explanatory variables constitute an adjusted time series satisfying requirements I, II, and III.

The type of seasonal disturbance filtered out by this approach is determined by the specification of the explanatory variables. This proposition, which follows at once from the fact that the observed residuals obtained from regression analysis are necessarily uncorrelated with the explanatory variables, may be illustrated by considering once more the special case of a stable seasonal pattern that does not change from one year to the next. This type of seasonal variation may be filtered out by regressing x_{ym} , the observation for season m of year y , upon a set of dummy variables:

$$(4.1) \quad x_{ym} = \sum_{i=1}^k b_i s_{ymi} + e_{ym},$$

where k is the number of seasons in a year

$$s_{ymi} = 1 \quad \text{if } i = m$$

0 otherwise, and

$$x_{ym}^a = e_{ym} + \bar{x} \quad \text{is the adjusted observation.}$$

It is readily seen that the x_{ym}^a are identical to the set of seasonally adjusted data that would be obtained by equalizing monthly means, the procedure described at the end of Section 3.*

* After all, the data may be sorted by months and k separate regressions run without affecting the b_i . But each b_i is equal to the seasonal mean, for the second moment is at a minimum about the mean.

An alternative regression approach may prove particularly useful when the seasonal pattern can be assumed to display a certain amount of continuity from one season to the next. Fourier analysis can then be employed in order to economize on the use of explanatory variables. Specifically, the following regression is performed:

$$(4.2) \quad \begin{aligned} x_t &= a_0 + \sum_{i=1}^n a_i \sin(\gamma_i t + b_i) + e_t \\ &= a_0 + \sum_{i=1}^n \left\{ a_i \cos b_i [\sin(\gamma_i t)] + a_i \sin b_i [\cos \gamma_i t] \right\} + e_t, \end{aligned}$$

where $\gamma_i = \frac{2\pi i}{k}$, $k =$ number of seasons in a year,

$$2n < k,$$

$a_i \cos b_i$ and $a_i \sin b_i$ are the regression coefficients;

the expressions in brackets are the explanatory variables.

Once more $e_t + \bar{x} = x_t^a$ (the residuals plus the mean of the data) constitutes the adjusted series. This procedure may be particularly appropriate when weekly or daily data is available.

It is possible to utilize regression analysis to delete trend and seasonality simultaneously. A. Hald [1948], who has discussed this problem at length, recommends the following regression:

$$(4.3) \quad x_{ym} = \sum_{i=1}^k b_i s_{ymi} + \sum_{i=1}^n c_i t^i + e_{ym},$$

$$\text{where } t = y + \frac{m}{12}.$$

He presented for purposes of illustration an application to interwar percentage unemployment in Denmark.

Approaches (4.1), (4.2), and (4.3) all assume a stable seasonal pattern. There are compelling reasons, summarized by Morton S. Raff and Stein [1960, p. 825], for expecting a flexible pattern of seasonal disturbances:

"One basic problem is that the seasonal pattern for the current year will never be exactly like the average for any set of past years. For example, although outdoor activity expands and contracts each year in line with changes in weather conditions, outdoor employment in particular survey weeks may deviate from previous patterns for that time of year because of unusual weather. Although Easter occurs every spring, its timing in relation to the March and April surveys varies from year to year. Although new automobiles are introduced each year, the start of the plant shutdowns for retooling may vary from one year to the next by as much as 2 or 3 months."

Dudley Cowden [1942] has suggested that moving seasonal adjustment may be accomplished by fitting polynomial trends in separate regressions to each season. If x_{ym} is the observation for season m of year y we regress

$$(4.4) \quad x_{ym} = \sum_m s_{ym} (a_{0m} + a_{1m}t + \dots + a_{2m}t^d), + e_{ym}$$

$$\begin{aligned} \text{where } s_{ym} &= 1 && \text{in season } m \\ &= 0 && \text{otherwise,} \end{aligned}$$

and have as our deseasonalized series

$$x_{ym}^a = e_{ym} + \bar{x}.$$

Clearly, the researcher must be careful that he selects $\delta < \gamma$ if he is to avoid a singular moment matrix. Observe too, that it is possible to economize on degrees of freedom by introducing such restrictions as

$$a_{km} = a_k.$$

All of these approaches involve the introduction of proxies for the basic variables actually generating seasonal variation. Simon Kuznets [1933, pp. 9-10] has provided a classic statement of the causes of seasonal variation:

Nature and human institutions conspire to produce seasonal variations in industrial and trade activity. Climatic seasons bring changes in the length of the day and night, in temperature and precipitation... But all seasonal change is not induced by climate. Conventions also exercise a pervasive and often a conspicuous influence upon business and its periodical records. In monthly time series...the most pervasive of all seasonal variations is due primarily to the calendar, which makes February nearly 10 percent shorter than January... Appreciable in their influence upon many business processes are such conventional seasonal factors as religious observances, folk customs, fashions and business practices.

With the least squares regression procedure, if not with the popular ratio-to-moving-average technique, one is not compelled to utilize proxies for the underlying factors generating seasonal variation. In least squares seasonal adjustment of air conditioner sales, one might include the "comfort index" of temperature conditions as one explanatory variable; in a study of

fuel oil sales an index of "degree days" might be introduced. John A. Frechtling [1960, p. 14] has suggested that automobile sales and production series might be adjusted by regressing the raw time series upon such variables as the number of days elapsed since the last model change. Department store sales might be adjusted by number of trading days in the month. Dummy variables might also be included to indicate the occurrence of Easter and such irregular events as strikes or Presidential heart attacks. Of course, the length of the time series imposes a restriction on the variety of forces that may be introduced as explicit generators of seasonal variation.

Seasonal influences may be readily distinguished from trend when the application of the least squares procedure takes the form suggested by A. Hald; the first term on the right of the equality in (4.3) constitutes the estimated seasonal disturbance and the second term is trend. We shall see that the distinction between seasonality and trend becomes blurred when an attempt is made to introduce a flexible seasonal pattern, as in equation (4.4), or to include explicitly such variables as measures of climatic conditions that contribute directly to seasonality and/or trend.

In his analysis of seasonal movements in German data on unemployment Horst Mendershausen [1939] performed separate regressions for each month for data which had already been corrected for trend by taking the ratio of the actual observation to a twelve month moving average; deviations from trend were explained by variations in weather conditions and such economic variables as past levels of employment. With such an approach, only the deviations from trend will satisfy Properties I, III, IV and V; consequently,

the seasonally adjusted series obtained by combining the trend with the seasonally adjusted deviations from trend will no longer possess these properties.

When the least squares procedure is to be utilized in order to adjust data for trend as well as seasonal movements, it may be possible to follow a two stage procedure by partitioning the regressors into a set of trend and a set of seasonality determinants. First the raw data might be regressed on the trend variables; then the residuals of the trend analysis would be regressed on the seasonal factors. Such a two stage procedure involves certain ambiguities, for except in the special case in which the two sets of regressors are uncorrelated with each other the results are sensitive to whether we decide to delete trend or seasonality first.*

* The popular ratio-to-moving-average procedure for seasonal adjustment involves a three stage cycle: first trend is removed by taking ratios to a moving average; then a moving average of these ratios yields seasonal factors; in order to obtain the adjusted series the original observations are divided by these seasonal factors. The outcome of this first iteration may again be subjected to the same treatment in order to obtain a more refined set of seasonal patterns; improvement is hoped for on the grounds that the trend removal step will not be distorted by the presence of seasonality.

An alternative procedure for segregating seasonality from trend is to require that seasonal adjustment shall not distort annual totals. In any calendar year the sum of the seasonally adjusted data over all months must equal the sum over the same period of the unadjusted observations. Many applications of the ratio-to-moving average procedure achieve this by

blowing up the preliminary seasonally adjusted data by an appropriate conversion ratio. With least squares adjustment, the same effect might be achieved by placing certain restrictions on the regression coefficients or by adjusting the residuals obtained after completing the regression. But it must be observed that any such procedure is sensitive to an arbitrary convention as there is little basis for making the comparison of annual totals of adjusted with unadjusted data over the calendar year rather than over a fiscal year or indeed any twelve month period. If we avoid this ambiguity by requiring that totals over all consecutive twelve month periods shall be undistorted by the process of seasonal adjustment, we are forced to exclude from consideration any adjustment procedure that allows for a flexible seasonal pattern.* In this

* Suppose that the requirement $\sum_{\tau=k}^{k+12} x_{\tau}^a = \sum_{\tau=k}^{k+12} x_{\tau}$ is satisfied for all k ; then

$$\sum_{\tau=k'}^{k'+12} x_{\tau} - \sum_{\tau=k'+1}^{k'+13} x_{\tau} = \sum_{\tau=k'}^{k'+12} x_{\tau}^a - \sum_{\tau=k'+1}^{k'+13} x_{\tau}^a, \text{ or}$$

$$x_{k'} - x_{k'+13} = x_{k'}^a - x_{k'+13}^a. \text{ Therefore,}$$

$$x_{k'} - x_{k'}^a = x_{k'+13} - x_{k'+13}^a,$$

proving that the January correction factor has not changed. Iteration in this way proves that each season's correction factors must be a constant.

sense, then, the distinction between trend and seasonality is necessarily befuddled once the possibility of a moving seasonal pattern is introduced.

A final word must be added concerning computational expense. Any procedure for seasonal adjustment that preserves sums can be accomplished by matrix premultiplication; for t observations, t^2 multiplications, $t(t-1)$ additions, and t^2 storage locations are required. The least squares procedure, on the other hand, requires $2\mathcal{K}t$ multiplications, $(2\mathcal{K}-1)t$ additions, and $2t\mathcal{K}$ storage locations, where \mathcal{K} is an index of the complexity of the family of seasonal adjustment procedures under consideration.*

* If k is the number of seasons in a year, then $\mathcal{K} = k$ for the fixed seasonal pattern of (4.1); for the procedure described by (4.2), $\mathcal{K} = 2n$; for (4.3) $\mathcal{K} = k+n$ and for (4.4) $\mathcal{K} = md$. \mathcal{K} is the number of explanatory variables required in the regression, and we assume in terms of the notation of the appendix that the matrix $[(D'D)^{-1}D']$ is stored in the computer.

Of course, any measure of economy with regard to computations is sensitive to the state of technology.** Developing the Census Method II application

** After all, graphical procedures were once customarily employed for performing multiple regression; for a description see Geoffrey Shepherd [1947, pp. 122-142]. A sequence of simple regressions were performed by plotting the data and fitting the regression line by inspection. In order to regress $y = ax+bz$, y was first regressed on x and the residuals from that regression plotted against z in order to obtain a rough approximation of b that could be improved by iteration. Lester V. Manderscheid of Michigan State University has pointed out to me that the argument of the next section of this paper suggests that it would be more expeditious to regress z on x and plot y upon the residuals of that regression in order to obtain an approximation of b that would be precise if the draftsman succeeded in minimizing the sum of squares in each of the two simple regressions.

of the ratio to moving average procedure cost roughly \$250,000; the cost of adjusting a single time series is negligible.

In conclusion, it should be observed that although the least squares adjustment procedure preserves sums, it distorts products. But we have seen that no procedure can satisfy both the additivity and the multiplicative requirement. Furthermore, if one wishes to satisfy the multiplicative at the expense of the additivity requirement, a logarithmic transformation of the data in advance of regression analysis will accomplish this objective. Of course, if different D matrices are utilized in adjusting different series one cannot expect requirement (2.1) to be satisfied, although the magnitude of the discrepancy may not be substantial. It is also worth emphasizing that the least squares procedure has considerable flexibility. The user can delete or retain trend; he can obtain a constant or moving seasonal; economic and climatic variables may be included. The procedure makes explicit the element of judgment inherent in any process of time series decomposition.

5. Regression Analysis with Seasonally Adjusted Data

Seasonally adjusted time series rather than unadjusted data are frequently employed in regression analysis, sometimes as a matter of deliberate choice but on occasion also when unadjusted time series are not available. Suppose that the time series has been generated by a linear model of the type considered in section 3; in matrix notation we have

$$(5.1) \quad Y = XB + S + \epsilon, \quad \sum (\epsilon) = 0,$$

where Y is a t component column vector, X a txk matrix of explanatory variables, and ϵ a t component column vector of disturbances; $S = D\alpha$, the txl vector of seasonal disturbances, belongs to a particular family of possible seasonal movements characterized by the txd matrix D ; α is an unknown dxl vector. In order to determine the conditions in which the application of least squares to seasonally adjusted time series Y_a and X_a will yield unbiased estimates of the parameters of the underlying economic model, let us consider the following regression:

$$(5.2) \quad Y_a = X_a b + \epsilon .$$

Application of least squares to calculate the $kx1$ vector b of regression coefficients yields

$$(5.3) \quad b = (X_a' X_a)^{-1} X_a' Y_a$$

Following Theil's procedure for the analysis of specification error [1958, p. 213], we observe:

$$(5.4) \quad \begin{aligned} b &= (X_a' X_a)^{-1} X_a' (Y_a + X_a \beta - X_a \beta) \\ &= \beta + (X_a' X_a)^{-1} X_a' (Y_a - X_a \beta) . \end{aligned}$$

Unless the expected value of the second term on the right of the last equality vanishes, the application of least squares to seasonally adjusted data will yield biased estimates of the parameter vector β . Suppose that the technique of seasonal adjustment preserves sums; then there exists a txt

matrix A such that $Y_a = AY = AX\beta + AD\alpha + A\epsilon$. If, in addition, the adjustment procedure is capable of annihilating S in the sense that $AD = 0$, we have $Y_a - X_a\beta = A\epsilon$. When the adjustment procedure has these two properties, equation (5.4) reduces to

$$(5.5) \quad b = \beta + (X_a' X_a)^{-1} X_a' A\epsilon .$$

Clearly, b will be an unbiased estimator of β if the X are fixed from sample to sample.

There exist a whole family of techniques of seasonal adjustment that will both preserve sums and annihilate a given class D of seasonal movements. For example, when the method of least squares is utilized to adjust data, one computes $Y - Da_y^0$ and $X - Da_x^0$, where $a_y^0 = (D'D)^{-1}D'Y$ and $a_x^0 = (D'D)^{-1}D'X$; hence, $A = [I - D(D'D)^{-1}D']$ is the adjustment matrix. Now we have seen that application of least squares to data adjusted by any member of this family of seasonal correction techniques will yield an unbiased estimate of the vector β of parameters of (5.1). The following generalization of the Frisch-Waugh Theorem assists in specifying conditions under which the least squares method of seasonal adjustment will be the preferred procedure.

Theorem 5.1:

Consider the following alternative regression equations, where the subscript α indicates the data have been adjusted by the least squares procedure:

1. $Y = X b_1 + D a_1 + e_1$

2. $Y_\alpha = X_\alpha b_2 + e_2$

3. $Y = X b_3 + e_3$

4. $Y = X_\alpha b_4 + e_4$

5. $Y_\alpha = X b_5 + e_5$

6. $Y = X_\alpha b_6 + D a_6 + e_6$

7. $Y_\alpha = X b_7 + D a_7 + e_7$

Then if the rank of $[X;D]$ is $k+d < t$,* the regression coefficients

* If the matrix D of dummy variables is constructed carelessly, this rank condition may be violated. Caution is required when a column of X consists only of ones in order that the corresponding element of the vector b of regression coefficients constitutes the intercept; for example, when a constant seasonal is being introduced with quarterly data, D should have only three columns. For a discussion of dummy variables see Suits [1957].

obtained by the application of least squares satisfy:

$$\begin{aligned} b_1 &= b_2 = b_4 = b_6 = b_7 \\ &= b_3 + (X'X)^{-1}X' [(X-X_\alpha) b_1 - (Y - Y_\alpha)] \\ &= b_5 - (X'X)^{-1}X' (X_\alpha - X) b_1 \end{aligned}$$

$$a_1 = a_2 - a_x^0 b_1$$

$$a_6 = a_y^0$$

$$a_7 = - a_x^0 b_1 .$$

If e_i denotes the observed residuals of the i th regression, then

$$e_1 = e_2 = e_4 - (Y - Y_\alpha) = e_6 = e_7 .$$

The variance of the residuals satisfy:

$$s_1^2 = \frac{e_1' e_1}{n} = s_2^2 \leq \frac{e_4' e_4}{n} = s_1^2 + \frac{1}{n} (Y - Y_\alpha)' (Y - Y_\alpha)$$

$$s_1^2 \leq \frac{e_5' e_5}{n} .$$

The multiple correlation coefficients satisfy:

$$R_5 \leq R_7 = R_2 \leq R_1 = R_6$$

$$R_4 \leq R_6$$

$$R_3 \leq R_1 .$$

The proof of the theorem, which is related to suggestions of Wold [1953], Rao [1952, pp. 118-121] and discussions of residual analysis by Freund, Vail, Clunies-Ross [1961] and by Goldberger and Jochems [1961a] is relegated to the appendix.

Five alternative least squares regression procedures yield identical estimates of the parameters of (5.1). The identity $b_1 = b_2$

reveals that inclusion of the matrix of seasonal dummy variables in the regression analysis is equivalent to working with least squares adjusted time series.* The identity $b_2 = b_4$ reveals that it is immaterial whether

* Provided, of course, that the same matrix D is utilized in constructing the adjustment matrix A as is utilized in the dummy variable regression. It should be noted that in most applications of the dummy variable procedure D takes a particularly simple form implying an unchanging seasonal pattern. If in fact the D that is utilized in the dummy variable procedure is not capable of annihilating S , the erroneous results obtained may be interpreted as the effect of omitting variables that are correlated with other explanatory variables, a difficulty discussed by Wold [1953]. Lawrence Klein [1953, pp. 313-17] discusses the use of dummy variables in the case of an unchanging seasonal pattern and also explains how one may proceed when certain components of β are presumed to be subject to seasonality, a complication that cannot be readily taken into account by any additive procedure of seasonal adjustment.

the dependent variable is adjusted or not, provided the explanatory variables have been seasonally corrected.

The same residuals will be observed regardless of whether regression procedure 1, 2, 6, or 7 is applied. While the same unexplained variance will be obtained with any one of these five alternative regression procedures, the magnitude of the variance of the dependent variable depends upon whether one works with adjusted or unadjusted data. As a consequence, deseasonalizing the dependent variable in advance of running the regression, as in equation (2) or (7), worsens appearances by reducing the size of the multiple correlation coefficient.

With problems of hypothesis testing and confidence interval construction the choice between alternative regression procedures becomes crucial. Although procedures 1, 2, 6, and 7 yield the same observed residuals and have the same inverse relevant to the estimation of the standard errors of the components of β *, considerable care must be exercised

* It is easily verified that $(X_a'X_a)^{-1}$, the inverse of the moment matrix of the second regression, constitutes the upper left hand block of the inverse of the first regression's moment matrix. Let M be the moment matrix of the first regression, and compute $M^{-1} = (M^{-1}P^{-1})P = (PM)^{-1}P$,

where $P = \begin{bmatrix} I & -X'D(D'D)^{-1} \\ 0 & (D'D)^{-1} \end{bmatrix}$.

in making appropriate adjustments for the loss of degrees of freedom that occurs when deseasonalized data is employed. Utilizing deseasonalized data with regression 2 suggests that there are $n-k$ degrees of freedom; procedures 1, 6, and 7, on the other hand, imply that there are only $n-k-d$ degrees of freedom. Clearly, at least one of the approaches is misleading! The root of the problem may be perceived by premultiplying (5.1) by A , the adjustment matrix, in order to obtain an expression for the adjusted dependent variable:

$$(5.6) \quad Y_a = X_a \beta + A \epsilon,$$

where we suppose $\mathcal{E}(\epsilon) = 0$ and $\mathcal{E}(\epsilon\epsilon') = \sigma^2 I$; later problems created by autocorrelated error terms will be considered.

Now the disturbance terms in this last equation, $A \epsilon$, are necessarily free of any seasonal pattern; because of this artificial restriction, they cannot be truly independent.*

* If A is capable of filtering out some form of seasonal movement, then for some $S = D\alpha \neq 0$ we must have $AS = 0$. The rank of A is then $t-d$, where $d \geq 1$ is the dimension of the types of seasonal movement filtered out by A . Hence the residuals of (5.9) cannot span a space of dimension greater than $t-d$.

Fortunately, the loss of degrees of freedom may be appropriately taken into account, at least when the data have been adjusted by the least squares procedure described in part 4 above. While the point estimates of the regression coefficients are unaffected, their estimated standard errors are sensitive to the way in which the degrees of freedom are tabulated. If the calculations have been performed without considering the effects of seasonal adjustment, $n-k$ rather than $n-k-d$ degrees of freedom will have been employed in deriving the standard errors of the regression coefficients; they should be multiplied by the following correction factor:

$$(5.10) \quad \sqrt{\frac{n-k}{n-k-d}}$$

The same correction factor should also be employed if $n-k$ degrees of freedom were mistakenly assumed in calculating the unbiased estimate of σ_{ϵ} .

Corrected R^2 also requires adjustment.*

* In adjusting R^2 for loss of degrees of freedom the customary formulas lead to a corrected coefficient of determination equal to the ratio of the unbiased estimate of the explained variance to an unbiased estimate of the unexplained variance. The corresponding formula when the data has been seasonally adjusted is

$$R_c^2 = \left[1 - (1-R^2) \frac{n-d}{n-k-d} \right].$$

Seasonally adjusted data released by government agencies have been subjected to a ratio-to-moving average rather than the least squares adjustment procedure. It is common practice in reporting regressions to neglect the loss of degrees of freedom and the possibility of bias occasioned by the employment of adjusted data. While factor (5.10) can hardly be regarded as precise when the data have been subjected to prior adjustment by the ratio to moving average procedure, it is probably more appropriate to employ this factor as an approximation rather than to neglect the problem entirely. Since the ratio to moving average procedure allows for a shifting seasonal pattern, one might well choose $d = 3m - 1$, where m is the number of seasons.**

** Even this is a conservative figure, for it does not allow much flexibility to the moving seasonal. See equation (4.4). When an intercept is not included in the regression, we should let $d = 3m$.

For monthly data, neglect of the effects of seasonal adjustment may overstate the actual number of degrees of freedom by 35; for quarterly data, 11 too

many degrees of freedom may have been attributed to the data. The effect is to overstate considerably the significance of regression coefficients. In reporting regressions obtained in a study of inventory investment, I failed to consider the costs in terms of degrees of freedom imposed by the fact that only seasonally adjusted data were available. There were 29 quarterly observations and six coefficients in certain regression equations.* Formula (5.10) yields a correction factor of $\sqrt{\frac{29-6-11}{29-6}} \doteq .72$.

* Lovell [1961, p. 300].

Now if we had observed $b_1/s_{b_1} = 2.9$, the coefficient would have appeared to be significant at the 0.01 level with $n-k = 23$ degrees of freedom. In contrast, application of the correction factor reduces the ratio to 2.09, which is not significant at even the 0.05 level with $n-k-d = 12$ degrees of freedom. While this correction factor is not precise with data corrected by the ratio-to-moving average procedure, it seems eminently more reasonable than the current practice of ignoring the costs in terms of degrees of freedom when reporting regression results obtained with seasonally adjusted data.

A convenient assumption frequently invoked in regression analysis is that the disturbances are distributed with constant variance and are

free of auto correlation; more precisely, it is required that

$E(\epsilon \epsilon') = I\sigma^2$. * At least if the X are fixed from sample to sample,

* While it may be reasonable in many applications to assume that ϵ is distributed independently of D , a possible exception may arise with a "4-8-4" sampling pattern sometimes utilized in data collection; households that enter the sample in January, say, are also interviewed in February, March and April, and again in the corresponding four months of the following year.

this assures that the application of the method of least squares to regression 1 will yield best linear unbiased estimates of the parameters of $Y_t = X\beta + D\alpha + \epsilon$. If, in addition, the residuals are normally distributed the application of least squares yields maximum likelihood estimates of β . For any sample, however, precisely the same estimates of β will be obtained with regression 2 as with 1, provided the data have been adjusted with the least squares procedure. Consequently, best linear unbiased estimates of β are obtained when the regression is performed upon data adjusted with the least squares technique, provided the independence condition on the ϵ_t is satisfied. This is a distinctive property of the least squares adjustment procedure. If the explanatory variables are adjusted with any alternative to the least squares technique, best linear unbiased estimates will not be obtained.

Further complications arise with regard to the tasks of estimation and hypothesis testing when the disturbances of (5.1) are autocorrelated.*

* L. Hurwicz has argued [1950] that the shortening of the observation period contributes to more severe autocorrelation problems.

Aiken's generalized least squares analysis suggests that best linear unbiased estimators could be obtained by the application of least squares to the transformed equation

$$(5.11) \quad HY = HX\beta + HD\alpha + H\epsilon ,$$

where the transformation matrix H has the properties:

$$H \Omega H' = I \quad \text{and} \quad H'H = \Omega^{-1}$$

By the generalized Frisch-Waugh theorem, an identical estimate of β will be obtained by the application of least squares to

$$(5.12) \quad \tilde{A}Y = \tilde{A}X\beta + \tilde{A}\epsilon ,$$

$$\text{where } \tilde{A} = [I - HD(D'H'HD)^{-1}D'H']H .$$

This adjustment matrix annihilates D , and (5.12) is the same expression as would be obtained by adjusting (5.11) by the least squares adjustment matrix that filters out HD . This argument suggests that when the residuals of (5.1) are autocorrelated, time series adjusted with \tilde{A} rather than A are most appropriate for estimating β . In particular, the frequent practice of applying an estimated H directly to seasonally

adjusted data is not appropriate; regressing $HAY = HAX\beta + HA\epsilon$ will not in general yield the same estimated β as (5.11). Observe, however, that at least in principle data already adjusted by A may be transformed into the \tilde{A} adjusted form, and this despite the fact that the singularity of A means that adjusted data cannot be transformed back into the original form.* A practical problem is created by the fact that the

* Thrall and Thornheim [1957, p. 99] demonstrate, Theorem 3.7G, that if \tilde{A} and A are two txt matrices and if $AD = 0$ if and only if $\tilde{A}D = 0$, then there exists a non-singular matrix P such that $PA = \tilde{A}$. If A were not singular, $AD = 0$ implies $D = 0$.

variance covariance matrix Ω is not known and must be estimated in some way from the data. Since precisely the same residuals are observed with regressions performed on data adjusted by the least squares matrix A as are provided by unadjusted data with dummy variables explicitly included, the same information is available for estimating Ω . All this means that no new difficulties are created with regard to the problem of auto-correlation when the least squares seasonal adjustment procedure is applied.

The contrast between regressions 1 and 3 also deserves mention, for it reveals conditions under which best linear unbiased estimates of β may be obtained with unadjusted data without dummy variables. Although both the dependent and independent variables may be subject to considerable seasonal movement, it is possible that the seasonality in the dependent variable is entirely the consequence of seasonal influences acting indirectly

through the explanatory variables. In this case, the model is correctly formulated as:

$$(5.1^*) \quad Y = X\beta + \epsilon ,$$

and the regression is appropriately performed with unadjusted data. But if the seasonality in the dependent variable really is generated via the explanatory variable, if $(X-X_a) \beta = Y-Y_a$, the discrepancy between the estimates obtained with this procedure and alternatives 1 and 2 arises solely from sampling error and does not lead to bias. Degrees of freedom equal to the number of columns of D are lost when the regression is performed with the adjusted data.* It would be better, in these circumstances,

* An interesting application is provided by the conjecture of Jack Johnston [1961] that if manufacturers attempt to smooth the impact of seasonal variation in sales upon production, the effect can be adequately taken into account by including constant seasonal dummy variables in the regression (in this case D has a particularly simple form); the magnitude of the dummy variables coefficient indicates the extent to which planned inventories fluctuate in the attempt to stabilize production. The conjecture that manufacturers do smooth production in this way can be subjected to empirical test by determining whether the addition of seasonal dummy variables leads to a significant reduction in the unexplained variance.

to work with unadjusted rather than adjusted time series.

In conclusion, it must be emphasized once more that the argument concerning problems of seasonality encountered by the econometrician, as indeed most of the argument of this paper, has been developed only for the linear single equation models. Remember, however, that the analysis

may be generalized to multiplicative models linear in the logs, the case in which least squares seasonal adjustment preserves products rather than sums. It must also be conceded that nothing in this paper has ruled out the possibility that there exists some alternative class of models for which the ratio to moving average procedure is preferable; but surely it behooves anyone who would use data adjusted by this more popular technique to demonstrate that this is an appropriate approach to the problem of estimating the parameters of his model. While problems of seasonality have not been discussed within a simultaneous equation framework, it is obvious that much of the analysis admits an obvious extension to these more complicated problems; in particular, it is clear that the two-stage least squares procedure may be applied either with data subjected to prior adjustment by the least squares procedure or with seasonal dummy variables included explicitly as exogenous terms in the regression equations.

APPENDIX

PROOF OF THEOREM 2.1:

To prove that I, IV, and V imply III observe that I implies that the process of seasonal adjustment constitutes a linear transformation; if $X = \text{col}(x_t)$ and $X_a = \text{col}(x_t^a)$, then there exists a $n \times n$ matrix A such that $X_a = AX$.

Now properties IV and V imply that $A'A = AA = A$; hence $(I-A)'A = 0$. Therefore, $\sum_t (x_t - x_t^a)x_t^a = (X - X_a)'X_a = X'(I-A)'AX = 0$, as required.

To prove that I and III imply IV and V, first note that

$$\sum_t (x_t - x_t^a)x_t^a = X'(I-A)'AX = 0 \text{ for all } X \text{ implies that } Z =$$

$$(I-A')A = A-A'A = -Z', \text{ a skew symmetric matrix; hence, } A'A = \frac{1}{2}(A+A')$$

$$\text{and } Z = \frac{1}{2}(A-A'). \text{ We shall use an indirect proof to establish that}$$

$A' = A$. By a well known theorem [Thrall and Thornheim, 1957, p. 154]

there exists a nonsingular real matrix P such that all above

diagonal elements of $P'ZP$ are either zero or unity; since such

a congruent transformation constitutes a redefinition of variables,

we may assume that Z is already in this form. Hence $A' \neq A$ implies

that for some $k < i$,

$$2 z_{ki} = e_{ki} - a_{ik} = 2.$$

$$\text{Consider any diagonal element of } A'A = \begin{bmatrix} \sum_k a_{ki} a_{kj} \\ k \end{bmatrix}$$

$$= \frac{1}{2} \begin{bmatrix} a_{ij} + a_{ji} \end{bmatrix}, \text{ and we have } \sum_k a_{ki}^2 = a_{ii}, \text{ or } a_{ii}^2 + \sum_{k \neq i} a_{ki}^2 = a_{ii}$$

Since A is real, as X_a is to be real for all X , this last equality implies that

$$0 \leq a_{ii} - a_{ii}^2 = \sum_{k \neq i} a_{ki}^2 \leq \frac{1}{4};$$

Therefore, $-.5 \leq a_{ki} \leq .5$ contradicting the conjecture that

$$a_{ki} - a_{ik} = 2; \text{ hence } A' = A$$

To prove $A^2 = A$, remember that $Z = \frac{1}{2}(A - A') = 0$ implies

$$A - A'A = 0; \text{ hence } A' = A \text{ implies } A - A^2 = 0, \text{ as required.}$$

Note that the following matrix is idempotent but not symmetric:

$$\begin{bmatrix} .5 & 1 \\ .25 & .5 \end{bmatrix}$$

PROOF OF THEOREM 2.2:

Property I implies that the adjustment process constitutes a linear transformation and may therefore be represented by the expression:

$$x_t^a = \sum_{\tau} a_{t\tau} x_{\tau}$$

If, in addition, the procedure satisfies the multiplicative requirement we must have

$$(2.4) \quad x_t^a y_t^a = \left(\sum_{\tau} a_{t\tau} x_{\tau} \right) \left(\sum_{\tau} a_{t\tau} y_{\tau} \right) = (x_t y_t)^a = \sum_{\tau} a_{t\tau} x_{\tau} y_{\tau},$$

an equality that can hold for every two arbitrary time series only if

$$a_{t\tau} = 0 \text{ whenever } t \neq \tau, \text{ and } a_{tt} = 1 \text{ or zero.}$$

PROOF OF THEOREM 4.1

By properties I and IV there exists a matrix A such that for all X , $X_a = AX$ and, in addition $A(X - X_a) = 0$. Hence $X - X_a = D\alpha$, where D is a $t \times d$ matrix formed from the column vectors constituting a basis of the column kernel of A ; thus $AD = 0$ and $d = \text{rank of } D = t - \text{rank of } A$.

Premultiplication by D' yields $D'(X - X_a) = D'D\alpha$ where obviously

$$|D'D| \neq 0.$$

Now $(D'D)^{-1} D'(X - X_a) = \alpha = (D'D)^{-1} D'X$ for

$$D'X_a = D'AX = 0 \text{ as } AD = 0 \text{ and } A' = A \text{ implies } D'A = 0.$$

Thus α is the vector of least squares coefficients obtained by regressing X on D and $X_a = X - D\alpha$ are the residuals.

To establish the converse, observe that if

$$X^a = X - D\alpha, \text{ where } \alpha \text{ is obtained by regressing } X \text{ on } D, \text{ then}$$

$$\alpha = (D'D)^{-1} D'X. \text{ Hence } X_a = AX, \text{ where } A = I - D(D'D)^{-1}D'$$

Clearly $A' = A$. Also $AD = 0$; hence $AA = A$.

To show that the residuals plus the mean of the observations have the

desired properties, let $S = \text{col} \langle 1, 1, \dots, 1 \rangle$, a $t \times 1$ vector; thus $S'X^a = 0$ for all X ; hence $S'A = 0$. So if $A^* = A + \frac{1}{n} SS'$, then $A^* = A^{*'} and $A^*A^* = AA + \frac{1}{n^2} SS'SS' = AA + \frac{1}{n} SS' = A^*$.$

Now $\frac{1}{n} S'X = \bar{X}$, hence $A^*X = \left[A + S\left(\frac{1}{n} S'\right) \right] X = AX + S\bar{X}$.

PROOF OF THEOREM 5.1

The seven alternative regressions listed in Theorem 5.1, reduce to

$$(\alpha) \quad Y = Xb_{\alpha} + Da_{\alpha} + e_{\alpha} ,$$

$$(\beta) \quad Y_a = X_a b_{\beta} + Da_{\beta} + e_{\beta} , \text{ or}$$

$$(\gamma) \quad Y_a = X_a b_{\gamma} + e_{\gamma} ,$$

where we suppose that we are presented with seasonally adjusted observations $Y_a = Y - Da_y$ and $X_a = X - Da_x$, upon appropriate specification of a_y and a_x . Since the application of least squares to (α) selects the unique values \hat{b}_{α} and \hat{a}_{α} minimizing

$e_{\alpha}' e_{\alpha}$, we can achieve this same minimum sum of squares by letting

$$b_{\beta} = \hat{b}_{\alpha} \text{ and } a_{\beta} = a_{\alpha} - a_y + a_x b_{\beta} , \text{ for then } e_{\alpha} - e_{\beta} = Da_y$$

$$- Da_x b_{\beta} - X_a (b_{\alpha} - b_{\beta}) - D(a_{\alpha} - a_{\beta}) = 0 . \text{ These must be the coefficients}$$

minimizing the sum of squares of regression (β) , for if some alternative set achieved a smaller sum of squares, we could do as well with

regression (α) , in contradiction of the fact that \hat{b}_{α} and \hat{a}_{α}

minimize the sum of squares of (α) . By having $a_y = 0$ and $a_x = a_x^0$,

this yields $b_1 = b_6$ and $a_6 = a_1 + a_x^0 b_{\beta}$; with $a_y = a_y^0$ and

$a_x = 0$ we have $b_1 = b_7$ and $a_7 = a_1 - a_y^0$.

In order to determine circumstances in which the application of least squares to (γ) necessarily yield precisely the same vector of regression

coefficients, consider the normal equations of (β) :

$$\begin{bmatrix} X'_a X_a & X'_a D \\ D' X_a & D' D \end{bmatrix} \begin{bmatrix} b_\beta \\ a_\beta \end{bmatrix} = \begin{bmatrix} X'_a \\ D' \end{bmatrix} Y_a$$

premultiplying by $\begin{bmatrix} (X'_a X_a)^{-1} & 0 \end{bmatrix}$ yields:

$$b_\beta + (X'_a X_a)^{-1} X'_a D a_\beta = (X'_a X_a)^{-1} X'_a Y_a = b_\gamma .$$

Now $b_\beta = b_\gamma$ if $X'_a D = 0$, which requires from Theorem 4.1 that

$a_x = a_x^0$ (that the explanatory variables be adjusted by the least squares procedure); hence $b_2 = b_1$; furthermore, $X'_a Y_a = X' A A Y = X' A Y = X'_a Y$, implying that $b_4 = b_1$.

Premultiplying the normal equations of regression (4) by $(X'X)^{-1}$ yields

$$(X'X)^{-1} X'_\alpha X_\alpha b_4 = (X'X)^{-1} X'_\alpha X_\alpha b_4 = (X'X)^{-1} X'_\alpha Y_\alpha = (X'X)^{-1} X'_\alpha Y_\alpha \text{ or}$$

$$(X'X)^{-1} X'(Y_\alpha - X + X) b_4 = (X'X)^{-1} X'(Y_\alpha - Y + Y) .$$

Since $b_3 = (X'X)^{-1} X'Y$, we have

$$b_4 + (X'X)^{-1} X'(X_\alpha - X) b_4 = b_3 + (X'X)^{-1} X'(Y_\alpha - Y) , \text{ or}$$

$$b_4 = b_3 + (X'X)^{-1} X' [(X - X_\alpha) b_4 - (Y - Y_\alpha)] .$$

Since $b_5 = (X'X)^{-1} X'Y_\alpha$ we also have

$$b_4 + (X'X)^{-1} X'(X_\alpha - X) b_4 = (X'X)^{-1} X'Y_\alpha = b_5 .$$

Subtracting regression 2 from 4, and remembering $b_2 = b_4$, yields

$$Y - Y_a = e_4 - e_2, \text{ or } e_4 - (Y - Y_a) = e_2$$

The restrictions on the residuals follow at once, provided one

observes that $Y - Y_a = D a_y$, and consequently, by

regression 1, is distributed independently of $e_1 = e_4$.

The restrictions on the coefficients of multiple correlation are

easily derived once it is observed that $Y_a' Y_a \leq Y'Y$.

REFERENCES

- Barger, Harold and L. R. Klein [1954], "A Quarterly Model for the United States Economy," Journal of the American Statistical Association, Vol. 49, pp. 413-437.
- Bongard, J. [1960], "Some Remarks on Moving Averages," Seasonal Adjustment on Electronic Computers, Proceedings of Conference of November, 1960 Paris, Organization for Economic Co-operation and Development, 1961.
- Brittain, John A. [1959], "A Bias in the Seasonally Adjusted Unemployment Series and a Suggested Alternative," Review of Economics and Statistics, Vol. 41.
- _____, [1962], "A Regression Model for Estimation of the Seasonal Component in Unemployment and other Volatile Time Series," Review of Economics and Statistics, Vol. 44.
- Clark, Colin [1949], "A System of Equations Explaining the United States Trade Cycle, 1921 to 1941," Econometrica, Vol. 17, pp. 93-124.
- Cohen, Kalman [1960], Computer models of the Shoe, Leather, Hide Sequence, Englewood Cliffs, N.J.: Prentice-Hall.
- Cowden, Dudley J. [1942], "Moving Seasonal Indexes," Journal of the American Statistical Association, Vol. 37.
- Daly, Donald [1959], "Seasonal Variations and Business Expectations," Journal of Business, Vol. 32.
- _____, [1960], "The Direct Adjustment of Aggregates Compared with the Sum of Adjusted Components" Seasonal Adjustment on Electronic Computers, Proceedings of Conference of November, 1960, Paris, Organization for Economic Co-operation and Development, 1961.
- David, F. N. and J. Neyman [1938], "Extension of the Markoff theorem on Least Squares," Statistical Research Memoirs, Vol. II, J. Neyman and E. S. Pearson, eds., University of London.
- Duesenberry, James S., Otto Eckstein, and Gary Fromm [1960] "A Simulation of the United States Economy in Recession," Econometrica, Vol. 28, pp. 749-809.
- Eisenpress, Harry [1956], "Regression Techniques Applied to Seasonal Corrections and Adjustments for Calendar Shifts," Journal of the American Statistical Association, Vol. 51, pp. 615-621.
- Frechtling, John A. [1960], "Problems of Adjusting Automotive Series," Proceedings, American Statistical Association, pp. 13-21.

- Freund, Rudolf J., Richard W. Vail, and C. W. Clunies-Ross [1961], "Residual Analysis," Journal of the American Statistical Association, Vol. 56, pp. 98-104.
- Frisch, Ragnar and F. V. Waugh [1933], "Partial Time Regression as Compared with Individual Trends," Econometrica, Vol. 1.
- Goldberger, Arthur S. and D. B. Jochems [1961a], "Note on Stepwise Least Squares," Journal of the American Statistical Association, Vol. 56, pp. 105-110.
- Goldberger, Arthur S. [1961b], "Stepwise Least Squares: Residual Analysis and Specification Error," Journal of the American Statistical Association, Vol. 56, pp. 998-1001.
- Hald, A. [1948], The Decomposition of a Series of Observations Composed of a Trend, a Periodic Movement, and a Stochastic Variable. Copenhagen: G.E.C. Gads Forlag.
- Hannan, E. J. [1960], "The Estimation of Seasonal Variation," The Australian Journal of Statistics, Vol. 2.
- Hurwicz, L. [1950], "Variable Parameters in Stochastic Process: Trend and Seasonality," Statistical Inference in Dynamic Economic Models, T. C. Koopmans, ed., New York: John Wiley.
- Johnston, J. [1961], "An Econometric Study of the Production Decision," Quarterly Journal of Economics, Vol. 75.
- Kendall, Maurice G. [1946], The Advanced Theory of Statistics, London: Charles Griffin & Co.
- Klein, Lawrence [1953], Econometrics, New York: Row, Peterson and Co.
- Klein, L. R., R. J. Ball, A. Hazlewood, and P. Vandome [1961], An Econometric Model of the United Kingdom, Oxford: Basil Blackwell.
- Kuznets, Simon [1933], Seasonal Variations in Industry and Trade, New York: National Bureau of Economic Research.
- Lovell, Michael C. [1961], "Manufacturers' Inventories, Sales Expectations, and the Acceleration Principle," Econometrica, Vol. 29, pp. 293-314.
- Mendershausen, Horst [1939], "Eliminating Changing Seasonals by Multiple Regression Analysis," Review of Economics and Statistics, Vol. 21.

- Raff, Morton S. and Robert L. Stein [1960], "New Seasonal Adjustment Factors for Labor Free Components," Monthly Labor Review, Vol. 83.
- Rao, C. R. [1952], Advanced Statistical Methods in Biometric Research, New York: John Wiley.
- Shepherd, Geoffrey S. [1947], Agricultural Price Analysis, Ames: Iowa State College Press, 1947.
- Shiskin, Julius [1957], Electronic Computers and Business Indicators, Occasional Paper 57, New York: National Bureau of Economic Research, Inc.
- Shiskin, Julius and Harry Eisenpress [1957], "Seasonal Adjustment by Electronic Computer Methods," Journal of the American Statistical Association, Vol. 52, pp. 415-449.
- Suits, Daniel B. [1957], "Use of Dummy Variables in Regression Equations," Journal of the American Statistical Association, Vol. 52.
- Theil, H. [1958], Economic Forecasts and Policy, Amsterdam: North Holland Publishing Company.
- Thrall, Robert M. and Leonard Tornheim [1957], Vector Spaces and Matrices, New York: John Wiley & Sons.
- Tincher, Gerhard [1952], Econometrics, New York: John Wiley.
- Valavanis, Stevan [1959], Econometrics, New York: McGraw Hill.
- Wold, Herman [1953], Demand Analysis: A Study in Econometrics, New York: John Wiley.