Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://amstat.tandfonline.com/loi/uasa20

Decomposition of Seasonal Time Series: A Model for the Census X-11 Program

W. P. Cleveland & G. C. Tiao

To cite this article: W. P. Cleveland & G. C. Tiao (1976) Decomposition of Seasonal Time Series: A Model for the Census X-11 Program, Journal of the American Statistical Association, 71:355, 581-587

To link to this article: https://doi.org/10.1080/01621459.1976.10481532

Published online: 05 Apr 2012.

Submit your article to this journal

Article views: 87

View related articles

Citing articles: 111 View citing articles

Full Terms & Conditions of access and use can be found at
https://amstat.tandfonline.com/action/journalInformation?journalCode=uasa20

Decomposition of Seasonal Time Series: A Model for the Census X-11 Program

W. P. CLEVELAND and G. C. TIAO*

This paper shows that the linear filter version of the Census X-11 program for time-series decomposition can be approximately justified in terms of an additive model with stochastic trend, seasonal and noise components. Optimal estimates of the trend and seasonal components are obtained from the model and found to be in close agreement with the corresponding estimates for the Census procedure. This approach makes it possible to assess the appropriateness of the Census method. Two examples are given, one showing that the use of the X-11 procedure is largely appropriate and the other much less so.

1. INTRODUCTION

One of the goals people often have in approaching a seasonal time series is removal of the seasonal component. This may be undertaken for several reasons. It leaves a series with a simpler pattern to be studied for its implications. It is particularly helpful in revealing the latest trends, since averaging by eye is not as easy in this nonsymmetric situation. Another common reason for wanting a deseasonalized series is to make comparisons of series with different seasonal patterns.

One way of effecting the decomposition of a series into seasonal and trend components is to fit a model which has deterministic seasonal and nonseasonal parts to the data. Another is to apply moving-average filters to the data to separate out these components. The filters are often designed on the basis of allocating various parts of the power spectrum of the series to the seasonal and trend components.

The Bureau of the Census has developed a computer program for seasonal adjustment which has been widely used in government and industry. The basic feature of the program is that it uses a sequence of moving average filters to decompose a series into a seasonal component, a trend component, and a noise component. The strength and weakness of the program lies in its use of roughly the same filters for most series. Such a decomposition has the superficial advantage of uniform interpretation of the seasonal and trend components of most series. On the other hand, if one believes that observed phenomena are generated according to the physical circumstances of the problem, one could certainly be misled as to their nature by the results of the census program if no checks on its adequacy are made. Such checks would, however, be difficult to make unless one had some ideas as to the kinds of *underlying mechanisms* for which the census program would be appropriate.

This paper proposes a stochastic model for which the census procedure is nearly optimal. Specifically, we suppose that an observed time series y_t consists of three additive random components: a seasonal component s_t, a trend component p_t, and a noise component e_t, such that if particular autoregressive integrated moving average models are given for these components, the optimum estimators for s_t and p_t turn out to be very close to those obtained from the basic set of moving average filters employed in the census program. The results shed considerable light on both the merits and the demerits of the census program. In particular, one can see why the census method is generally satisfactory for a variety of series, why the residuals of the census decomposition will have smaller variance than the residuals of an autoregressive integrated moving average model fit to the same series, and why the census residuals may be correlated.

2. THE BUREAU OF THE CENSUS PROGRAM

The census procedure is summarized in Shiskin and Eisenpress [12] and described fully in Shiskin, Young, and Musgrave [13]. The program assumes the additive decomposition[1]

$$y_t = p_t + s_t + e_t \ , \tag{2.1}$$

where y_t is the observed series, p_t is the trend component, s_t is the seasonal component, and e_t is the noise. For each series the estimates of p_t and s_t, denoted, respectively, by $\hat{p}_t$ and $\hat{s}_t$, may be obtained. The basic device used to estimate the seasonal and trend components is the symmetric moving-average operator (filter) with weights summing to one. General properties of moving average operators are discussed in Kendall and Stuart [9, p. 366] and in Whittaker and Robinson [16, p. 288]. A symmetric moving average operator $S(\delta, k)$ generates a series

* W.P. Cleveland is Assistant Professor of Biostatistics, Department of Community Health Sciences, and Resident Statistician, Center for the Study of Aging and Human Development, Box 3003, Duke University Medical Center, Durham, NC 27710. G.C. Tiao is Professor of Statistics, University of Wisconsin, Madison, WI 53706. Research was supported by U.S. Army Grant DAHCO4-76-G-0010. The authors wish to thank the referees for many helpful comments which resulted in considerable improvement in the presentation of the paper.

[1] There is also a multiplicative version of the program. By taking logarithms, the multiplicative version is essentially the same as the version discussed.

© Journal of the American Statistical Association
September 1976, Volume 71, Number 355
Applications Section

x_t from a series y_t according to the relation

$$x_t = S(\delta, k)y_t = \sum_{j=-k}^{k} \delta_j y_{t-j} \tag{2.2}$$

where k is a non-negative integer, and δ_j are the weights (constants) used in averaging y_t such that $\delta_{-j} = \delta_j$.

For a series $y_1, \ldots, y_T$ of length T, x_t cannot be computed according to (2.2) for $t = 1, \ldots, k$ and $t = T - k + 1, \ldots, T$. The operator $S(\delta, k)$ must, therefore, be modified at both ends of the series according to some extrapolation principle, as is done in the census procedure. In the analysis here, we shall be concerned mainly with version (2.2). However, the results developed can be logically extended to cover the end effects of the series.

Wallis [15] describes the sequence of moving average operators used in the census program. Although it includes procedures for adjusting outliers and for making trading day adjustments, the basic linear filter feature of the program is of the form (2.2). Also, the census program has a built-in procedure to select the 9-term, 13-term, or 23-term Henderson trend filters. In what follows we have assumed the 13-term filter because it seems to be the one applicable to the majority of series, and also because it would be rather difficult to model the procedure for selecting this one from the other two alternatives. The model in Section 3 applies to this linear filter version of the program.

3. A MODEL FOR THE BUREAU OF THE CENSUS PROGRAM

Without a specific underlying model the Bureau of the Census program lends itself to all the criticisms of moving-averages in general. If the observed series were a polynomial in time plus a trigonometric series plus noise, one should do a regression. If the preceding with time changing coefficients were the model, one could parameterize the pattern of changes in the coefficients and still do a regression. Models which fit a polynomial over a finite number of points to estimate the center point are not consistent, in the sense that different polynomials represent the same points at different times.

For a discussion of the properties of moving average filters and criticisms of the Census procedure in the context of spectral theory, the reader is referred to [3–8, 10, 11].

The best justification of moving-average filters seems to be in terms of stochastic models. For the additive decomposition in (2.1), we employ the autoregressive integrated moving average (ARIMA) processes [1] for the components. Specifically, we suppose that

$$y_t = p_t + s_t + e_t \tag{3.1}$$

with

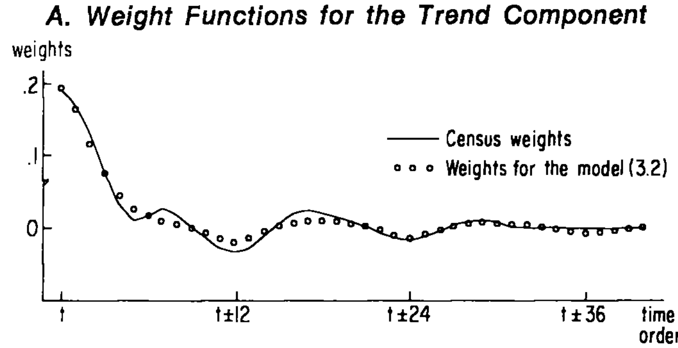
$$\phi_{r_1}(B)(p_t - \mu) = \psi_{q_1}(B)b_{1t}$$
$$\phi_{r_2}(B)s_t = \psi_{q_2}(B)b_{2t} ,$$

where μ is a constant, B is the backward shift operator such that $Bx_t = x_{t-1}$, and $\psi_{q_1}(B)$, $\psi_{q_2}(B)$, $\phi_{r_1}(B)$ and $\phi_{r_2}(B)$ are real polynomials in B of degrees q_1, q_2, r_1 and r_2, respectively. We shall require that the zeroes of $\psi_{q_1}(B)$ and $\psi_{q_2}(B)$ lie outside the unit circle and those of $\phi_{r_1}(B)$ and $\phi_{r_2}(B)$ lie on or outside the unit circle. In (3.1), e_t, b_{1t} and b_{2t} are three independent white noise processes, normally distributed with zero means and variances σ_e^2, $\sigma_{b_1}^2$ and $\sigma_{b_2}^2$, respectively. In this framework, the minimum mean square error estimators of p_t and s_t are, respectively, the conditional expectations $E(p_t|\mathbf{y})$ and $E(s_t|\mathbf{y})$ where $\mathbf{y} = (y_1, \ldots, y_T)'$ is the vector of observations. It is shown in the appendix that, for t not close to either end of the series, $E(p_t|\mathbf{y})$ and $E(s_t|\mathbf{y})$ are, to a close approximation, symmetric moving averages of the observations $\mathbf{y}$. Thus, if one can find a model for which the conditional expectations give the same weights as those of particular symmetric moving average filters, it may then be argued that this model represents an underlying stochastic mechanism for those filters.

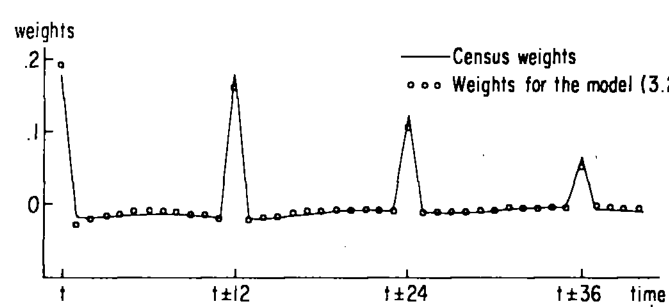
In searching for a model of the form (3.1) which would give weight functions similar to those of the census program, we kept several things in mind. When $\phi_{r_1}(B)$ contains the factor $(1 - B)$ to a degree at least one higher than $\phi_{r_2}(B)$, the weights in $E(p_t|\mathbf{y})$ and $E(s_t|\mathbf{y})$ sum to one and zero, respectively, as do the Census weights for $\hat{p}_t$ and $\hat{s}_t$. The use of a moving average filter for $\hat{p}_t$ rather than a polynomial suggests that one should allow for changing trend slopes. Similarly, the seasonal estimator should allow for changing amplitude and phase. Otherwise, a trigonometric series could be used.

The model stated in (3.2) is consistent with these principles and gives good correspondence to the census program using a minimal number of parameters. The weight functions for $E(p_t|\mathbf{y})$ and $\hat{p}_t$ are exceedingly close, as are the weight functions for $E(s_t|\mathbf{y})$ and $\hat{s}_t$. These are illustrated in Figures A and B.

A. Weight Functions for the Trend Component



B. Weight Functions for the Seasonal Component



$$y_t = p_t + s_t + e_t \ ,$$
$$(1 - B)^2 p_t = (1 - \psi_{11}B - \psi_{12}B^2)b_{1t} \ ,$$
$$(1 - B^{12})s_t = (1 - \psi_{21}B^{12} - \psi_{22}B^{24})b_{2t} \ , \quad (3.2)$$
$$\psi_{11} = -.49 \ , \ \psi_{12} = .49 \ , \ \psi_{21} = -.64 \ , \ \psi_{22} = -.83 \ ,$$
$$\sigma_{b_2}^2/\sigma_{b_1}^2 = 1.3 \quad \text{and} \quad \sigma_e^2/\sigma_{b_1}^2 = 14.4 \ .$$

Models differing from (3.2) in the order of differencing and in the numbers of autoregressive and moving average parameters in the two component models were tried. This model was found to be as good as models with more parameters and superior to others, particularly in matching $E(s_t|\mathbf{y})$ to the corresponding census weights.

A general purpose nonlinear least-squares program was used to estimate the parameter values for each model. For a given set of parameter values, the generating functions (A.20) and (A.21) in the appendix were expanded for the first 42 terms on either side of the center. The program used these as predictions of the corresponding weights in the linear filter version of the census program, and it adjusted the parameter values iteratively until the sum of squared deviations over the seasonal and trend weights was minimized.

While we used the sum of squares as one measure in choosing among candidate models, we also visually inspected the nature of the discrepancies in the weight functions, as displayed in Figures A and B, when evaluating the adequacy of a model. Occasionally tails of the census and model weights did not appear to be converging or only the trend weights and not the seasonal were being well approximated.

3.1 The Overall Model for y_t

From (3.2), we may write

$$\begin{aligned}(1 - B)(1 - B^{12})y_t &= ((1 - B^{12})/(1 - B))(1 - \psi_{11}B - \psi_{12}B^2)b_{1t} \\ &+ (1 - B)(1 - \psi_{21}B^{12} - \psi_{22}B^{24})b_{2t} \\ &+ (1 - B)(1 - B^{12})e_t \ , \quad (3.3)\end{aligned}$$

so that the autocorrelations of $w_t = (1 - B)(1 - B^{12})y_t$ are

k	ρ_k									
1–10	−.25	.13	.12	.11	.09	.08	.07	.05	.04	.03
11–20	.18	−.35	.16	.00	.00	.00	.00	.00	.00	.00
21–25	.00	.00	−.01	.03	−.01					

with $\rho_k = 0$ for $k > 25$. The explicit overall model for y_t is found to be,

$$\begin{aligned}(1 - B)(1 - B^{12})y_t &= (1 - .337B + .144B^2 + .141B^3 + .139B^4 \\ &+ .136B^5 + .131B^6 + .125B^7 + .117B^8 + .106B^9 \\ &+ .093B^{10} + .077B^{11} - .417B^{12} + .232B^{13} \\ &- .001B^{20} - .003B^{21} - .004B^{22} - .006B^{23} \\ &+ .035B^{24} - .021B^{25})c_t \ . \quad (3.4)\end{aligned}$$

where c_t is a white noise process, normally distributed with zero mean and variance σ_c^2.

It is important not to confuse the basic additive model (3.2) and the overall form (3.4). In particular,

1. the overall form is a logical consequence of the basic model,
2. the conditional expectations $E(p_t|\mathbf{y})$ and $E(s_t|\mathbf{y})$ are determined not by (3.4) but by (3.2),
3. different additive models can lead to the same overall model,
4. the overall form is, however, important in that it can be identified from the data and, hence, provides a means to partially assess the appropriateness of the census procedure for a given set of data.

The model in (3.2) and its consequent overall form (3.4) helps explain why the census program fits series as well as it often does. A nonstationary seasonal pattern is presumed, and the model for p_t allows for a trend of changing slope. The autoregressive part of the overall model for y_t is $(1 - B)(1 - B^{12})$. This suggests that a series obeying (3.4) or at least $(1 - B)(1 - B^{12})y_t = \theta(B)c_t$ for some polynomial $\theta(B)$ might be fairly accurately analyzed by the census program. The overall model is broadly similar to the multiplicative model discussed by Box and Jenkins [1, Ch. 9], though not the same.

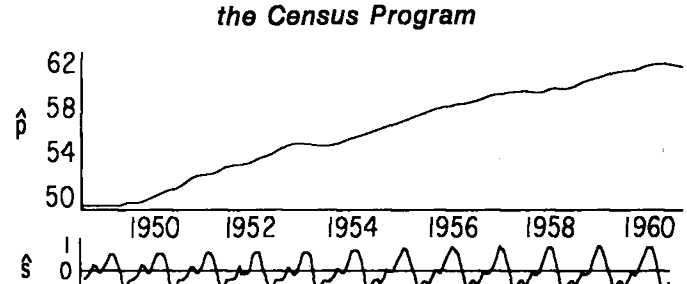
4. A COMPARISON OF THE CENSUS PROGRAM WITH OVERALL ARIMA SEASONAL MODELS FOR TWO SERIES

Two series are presented which have been fit with ARIMA models according to the procedures suggested in Box and Jenkins [1]. In doing so, the models sought were the ones that fit best in the sense of matching the correlation structure of the observed series and having uncorrelated residuals. The goal in each case was a model for forecasting rather than seasonal decomposition. It is, however, of interest to compare the results of this procedure with those of the census program.

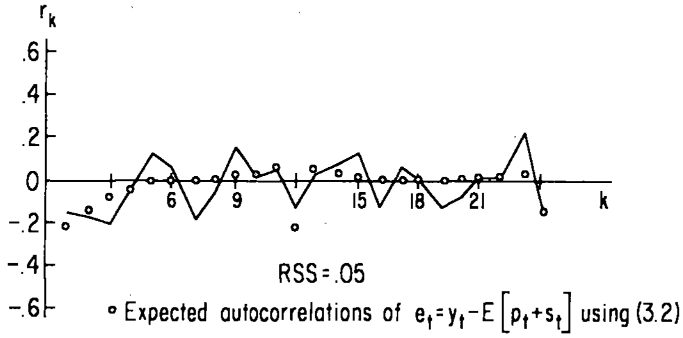
4.1 The Airline Data

These data may be found in [2, p. 429], or in [1, p. 304]. They represent logarithms of monthly passenger totals. Figures C and D show the census estimates of its components, $\hat{\mathbf{p}}$ and $\hat{\mathbf{s}}$, and the autocorrelation pattern of the residuals, $\mathbf{y} - \hat{\mathbf{p}} - \hat{\mathbf{s}}$. Also shown in Figure D are the expected residual autocorrelations, which will be explained later. The immediate impression of the series is that it exhibits a regular seasonal pattern with almost

C. Estimates of the Seasonal and Trend Components of the Airline Data from the Bureau of the Census Program



D. Residual Autocorrelations of the Airline Data from the Bureau of the Census Program



linear slope. The plots of the component estimates bear this out to a certain extent, but some evolution of the seasonal pattern and deviation from linearity are indicated. Some of the sample autocorrelations of the residuals appear to be rather large in magnitude, but the autocorrelation function as a whole exhibits no discernible pattern. The residual sum of squares, RSS, is .05.

The model for the airline data obtained by Box and Jenkins is

$$(1 - B)(1 - B^{12})y_t = (1 - \theta_1 B)(1 - \theta_{12}B^{12})c_t$$

$$\hat{\theta}_1 = .4 \ , \quad \hat{\theta}_{12} = .6 \ , \quad \hat{\sigma}_c^2 = .00134 \ , \quad \text{RSS} = .175 \ . \tag{4.1}$$

The autoregressive part of the model in (4.1) is the same as that of (3.4). Also, the estimates of θ_1 and θ_{12} in (4.1) are reasonably close to the coefficients associated with B and B^{12} on the right side of (3.4). One is, however, immediately struck by the fact that the RSS in (4.1), .175, is much larger than the corresponding number, .05, for the census program.

To see why, consider a simpler case where $y_t = z_t + e_t$ and the model for z_t is $(1 - B)z_t = (1 - \psi B)b_t$ where $|\psi| < 1$. This gives

$$\begin{aligned}(1 - B)y_t &= (1 - \psi B)b_t + (1 - B)e_t \\ &= (1 - \theta B)c_t \ , \quad \text{where } |\theta| < 1 \ . \end{aligned} \tag{4.2}$$

Thus,

$$\begin{aligned}c_t = {} & [(1 - \psi B)/(1 - \theta B)]b_t \\ & + [(1 - B)/(1 - \theta B)]e_t \ . \end{aligned} \tag{4.3}$$

By expanding the right side of (4.3) we find

$$\sigma_c^2 = \left\{1 + \frac{(\theta - \psi)^2}{1 - \theta^2}\right\}\sigma_b^2 + \frac{2}{1 + \theta}\sigma_e^2 \tag{4.4}$$

so that

$$\sigma_c^2 > (2/(1 + \theta))\sigma_e^2 > \sigma_e^2 \ .$$

From (A.16) in the appendix, the variance of the residuals $\hat{e}_t = y_t - E(z_t|\mathbf{y})$ is

$$\text{Var}\ (\hat{e}_t) = \sigma_e^2 \cdot \frac{\sigma_e^2(2/(1 + \theta))}{\sigma_c^2} \cdot \tag{4.5}$$

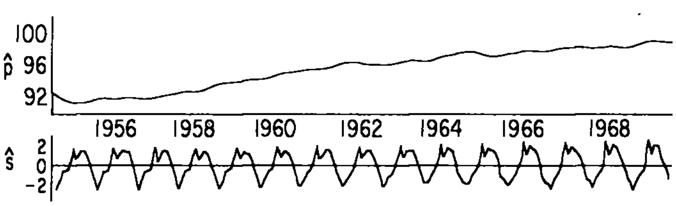
Hence, $\sigma_c^2 > \text{Var}\ (\hat{e}_t)$. Now the residuals from Box and Jenkins's method of model fitting are $\hat{c}_t$ which are different from the residuals $\hat{e}_t$. Thus the difference in RSS is not in this situation evidence of one method being better than the other. Two different kinds of residuals are involved. As we have said, the model for y_t in (4.1) has the same differencing structure as (3.4), indicating that the census program is a nearly appropriate decomposition procedure for the airline series.

Further evidence that the census procedure works reasonably well for the airline data is provided by the following analysis. The theoretical autocorrelations of the residuals $y_t - E(p_t + s_t|\mathbf{y})$ corresponding to the underlying model (3.2) may be computed using the general result (A.16) in the appendix. These correlations are given by the circles in Figure D. The pattern of the expected correlations is fairly close to that computed from the data using the census procedure, and the agreement in sign of the correlations is nearly perfect.

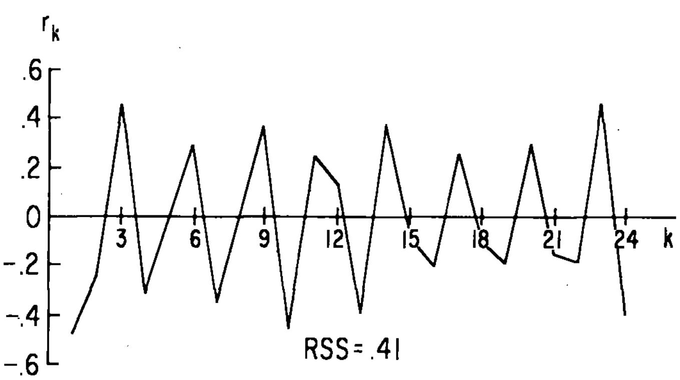
4.2 The Outward Telephone Data

The data in [14] are logarithms of monthly telephone disconnections, which the telephone companies call outward movement. Figures E and F show the census estimates of its components, and the residual autocorrelations. Note that the autocorrelations in this case exhibit a strong cyclical pattern. It seems clear that not all of the seasonality in the data has been accounted for by the program.

E. Estimates of the Seasonal and Trend Components of the Telephone Data from the Bureau of the Census Program



F. Residual Autocorrelations of the Telephone Data from the Bureau of the Census Program



The telephone data were analyzed by Thompson and Tiao and found to obey the model

$$\begin{aligned}(1 - \phi_3 B^3)(1 - \phi_{12}B^{12})y_t & \\ = (1 - \theta_9 B^9 - \theta_{12}B^{12} - \theta_{13}B^{13})c_t & \end{aligned}$$

$$\begin{gathered}\hat{\phi}_3 = .49 \ , \quad \hat{\phi}_{12} = 1.005 \ , \quad \hat{\theta}_9 = .23 \ , \\ \hat{\theta}_{12} = .334 \ , \quad \hat{\theta}_{13} = .17 \ , \quad \hat{\sigma}_c^2 = .0035 \ , \quad \text{RSS} = .59 \ . \end{gathered} \tag{4.6}$$

The autoregressive part of (4.6) is now quite different from that in the model (3.4) underlying the census program. Again the residual sum of squares, RSS, for this model is larger than that of the census program (.59 to .41). However, the census residuals exhibit a strong quarterly correlation pattern, while the residuals $\hat{c}_t$ do not. This is probably due to the differences between the two models (3.2) and (4.6), suggesting that using the census program on this series is not justified.

5. CONCLUSIONS

In the preceding sections, it was shown that the census procedure can be approximately justified in terms of an additive model which consists of a stochastic trend and seasonal components. The specific models for the components are given in (3.2), and the resulting overall model for the observations in (3.4).

If this model could be *assumed* for a given series, the census program could be used to estimate the unobserved trend and seasonal components during the observational period. In addition, the components could be forecast by calculating the conditional distributions of their future values given the observations. However, this result also raises a number of serious questions. First, should this model or minor deviations from it, corresponding to the use of the alternative 9-term or 23-term trend filters, be assumed for all series as implied by the census program? Second, is the census procedure robust to departures from this model? The answer to the first question is clearly no. Some clues to the answer for the second are provided by the analysis of the airline and telephone data.

The model (4.1) fitted by Box and Jenkins to the airline data has the same autoregressive part as the model (3.4), but the moving average parts are somewhat different. If we take the model fitted by Box and Jenkins to be correct, the appropriateness of the decomposition (3.2) is thrown in doubt. Yet the census program appears to perform quite well on this set of data, suggesting a certain degree of robustness. On the other hand, the census residuals from the telephone data show a strong quarterly correlation pattern. Since both the autoregressive and moving-average parts of (3.4) are quite different from those of the model (4.6) fitted to the telphone data, one suspects that this robustness does not extend to models which differ markedly from (3.4).

It might be argued that the robustness of the census procedure would be enhanced if the use of the alternative 9- or 23-term trend filters were taken into account. While this is undoubtedly true to a certain extent, one ought to remember that these alternatives were designed mainly for somewhat different evolutionary trend patterns. Thus, one would expect that the models corresponding to these alternative filters would be similar to (3.2) but with different values for the parameters, so that the use of these alternative filters on the telephone data would lead to a similar pattern of the residual autocorrelations as that shown in Figure F.

APPENDIX

For the general additive model (3.1), we here derive the conditional expectations $E(p_t|\mathbf{y})$ and $E(s_t|\mathbf{y})$, as well as the variance and autocorrelations of the residuals $\hat{e}_t = y_t - E(p_t + s_t|\mathbf{y})$.

To facilitate presentation, we first consider the situation

$$y_t = z_t + e_t \ , \quad t = 1, \ldots, T \ , \qquad \text{(A.1)}$$

where

$$\phi_r(B)(z_t - \mu) = \psi_q(B)b_t \ ,$$

e_t are i.i.d. $N(0, \sigma_e^2)$, b_t are i.i.d. $N(0, \sigma_b^2)$ and independent of e_t, $\psi_q(B)$ is a polynomial in B of degree q having its zeroes lying outside the unit circle, and $\phi_r(B)$ is a polynomial in B of degree r having its zeroes lying on or outside the unit circle. Now, z_t can be written as a function of r values in the past (see, e.g., [1, p. 115]).

$$z_t = c_t + \sum_{j=m}^{t} b_j \pi_{t-j} \ , \quad m \leq -\max(r, q) + 1 \qquad \text{(A.2)}$$

where (i) $c_t = \sum_{l=1}^{r} A_l \alpha_l^{t-m} + \mu$, α_l^{-1} are the zeroes of $\phi_r(B)$ assumed distinct, and A_l depend on the starting values of the series at time m; (ii) π_j are obtained by equating coefficients of B^j from the relation

$$\phi_r(B)(1 + \pi_1 B + \pi_2 B^2 + \ldots) = \psi_q(B) \ . \qquad \text{(A.2a)}$$

Thus the vector $\mathbf{z}$, where $\mathbf{z} = (z_1, \ldots, z_T)'$, is distributed as normal with mean vector $\mathbf{n} = (c_1, \ldots, c_T)'$ and covariance matrix $\sigma_b^2 \mathbf{\Sigma}_z$ whose elements can be readily obtained from (A.2).

A.1 The Conditional Distribution of z Given y

Given the observation vector $\mathbf{y} = (y_1, \ldots, y_T)'$, $\mathbf{z}$ is distributed as normal with

$$E(\mathbf{z}|\mathbf{y}) = (\mathbf{I} + v\mathbf{\Sigma}_z^{-1})^{-1}(\mathbf{y} + v\mathbf{\Sigma}_z^{-1}\mathbf{n})$$

and $\qquad \text{(A.3)}$

$$\text{cov}\,(\mathbf{z}|\mathbf{y}) = \sigma_e^2(\mathbf{I} + v\mathbf{\Sigma}_z^{-1})^{-1}$$

where $v = \sigma_e^2 \sigma_b^{-2}$ and $\mathbf{I}$ is the identity matrix. Note that in $E(\mathbf{z}|\mathbf{y})$, $(\mathbf{I} + v\mathbf{\Sigma}_z^{-1})^{-1}(\mathbf{h} + v\mathbf{\Sigma}_z^{-1}\mathbf{h}) = \mathbf{h}$, where $\mathbf{h}$ is a vector of ones, so that for each z_t, $E(z_t|\mathbf{y})$ is a weighted average of $\mathbf{y}$ and $\mathbf{n}$ with weights summing to one.

In practice, it is often appropriate to suppose that the series began at some remote past point of time. That is, m is a large negative integer. We now distinguish between two situations.

(I) *All Zeroes of $\phi_r(B)$ Lie Outside the Unit Circle.* In this case, $\mathbf{z}$ can be thought of as a vector of observations from a stationary Gaussian process. That is, for large $-m$, $\mathbf{n}$ approaches $\mathbf{h}\mu$, and $\mathbf{\Sigma}_z$ tends to its stationary value $\dot{\mathbf{\Sigma}}_z$ the (i, j)th element of which is $\sigma_{ij} = \sum_{l=0}^{\infty} \pi_l \pi_{l-(i-j)}$ where $\pi_0 = 1$ and $\pi_l = \pi_{-l}$. Thus, the expressions in (A.3) become

$$E(\mathbf{z}|\mathbf{y}) = (\mathbf{I} + v\dot{\mathbf{\Sigma}}_z^{-1})^{-1}(\mathbf{y} + v\dot{\mathbf{\Sigma}}_z^{-1}\mathbf{h}\mu)$$

and $\qquad \text{(A.4)}$

$$\text{cov}\,(\mathbf{z}|\mathbf{y}) = (\mathbf{I} + v\dot{\mathbf{\Sigma}}_z^{-1})^{-1}\sigma_e^2 \ .$$

(II) *Zeroes of $\phi_r(B)$ Lie on or Outside the Unit Circle.* To see the implications here, consider first the simplest case $\phi_r(B) = \phi_{r^*}(B)(1 - B)$ where $r^* = r - 1$ and the zeroes of $\phi_{r^*}(B)$ lie outside the unit circle. Then c_t in (A.2) is of the form

$$c_t = A_1 + \sum_{l=2}^{r} A_l \alpha_l^{t-m} \ . \qquad \text{(A.5)}$$

Consider now the transformation

$$\begin{bmatrix} z_1 \\ \mathbf{w} \end{bmatrix} = J\mathbf{z} \ , \quad J = \begin{bmatrix} 1 & & & \\ -1 & 1 & & \\ & \ddots & \ddots & \\ & & -1 & 1 \end{bmatrix} ,$$

i.e., $\mathbf{w} = (w_2, \ldots, w_T)'$ where $w_t = z_t - z_{t-1}$ follow the model

$$\phi_{r^*}(B)w_t = \psi_q(B)b_t \ . \qquad \text{(A.6)}$$

Making use of the partitioned inverse of a matrix, we obtain

$$\boldsymbol{\Sigma}_z^{-1} = \mathbf{J}'\mathbf{Q}\mathbf{J} \ , \quad \mathbf{Q} = \begin{bmatrix} \sigma_{z_1}^{-2} - \sigma_{z_1}^{-4}\boldsymbol{\sigma}_{z_1w}'\mathbf{G}\boldsymbol{\sigma}_{z_1w} & -\sigma_{z_1}^{-2}\boldsymbol{\sigma}_{z_1w}'\mathbf{G} \\ -\sigma_{z_1}^{-2}\mathbf{G}\boldsymbol{\sigma}_{z_1w} & G \end{bmatrix}$$

where $\sigma_{z_1}^{-2} = \sigma_b^2/\mathrm{Var}\,(z_1)$, $\boldsymbol{\sigma}_{z_1w}' = \sigma_b^{-2}\,\mathrm{cov}\,(z_1\mathbf{w}')$,

$$G = [\boldsymbol{\Sigma}_w - \sigma_{z_1}^{-2}\boldsymbol{\sigma}_{z_1w}\boldsymbol{\sigma}_{z_1w}']^{-1}$$

and $\boldsymbol{\Sigma}_w$ is the covariance matrix of w. Letting

$$\phi_{r^*}(B)(1 + \bar{\pi}_1 B + \bar{\pi}_2 B^2 + \ldots) = \psi_q(B) \ ,$$

we have from (A.2a) that $\pi_j - \pi_{j-1} = \bar{\pi}_j$, $j \geq 1$. It readily follows from (A.2) that

$$\mathrm{Var}\,(z_1) = \sigma_b^2 \sum_{j=0}^{-m} (\sum_{l=0}^{j} \bar{\pi}_l)^2 \ , \quad \bar{\pi}_0 = 1$$

and

$$\mathrm{cov}\,(z_1 w_t) = \sigma_b^2 \sum_{j=t-1}^{t-1-m} \bar{\pi}_j \pi_{j-(t-1)} \ .$$

Note that since the zeroes of $\phi_{r^*}(B)$ lie outside the unit circle, $\sum_{l=0}^{\infty} |\bar{\pi}_j| < \infty$. Thus, for large $-m$, $\sigma_{z_1}^{-2}$ and $\sigma_{z_1}^{-2}\,\mathrm{cov}\,(z_1 w_t)$ are of order m^{-1}; also, $\mathbf{G} = \dot{\boldsymbol{\Sigma}}_w^{-1} + \mathbf{R}$, where $\mathbf{R}$ is a matrix whose elements are of order m^{-1} and $\dot{\boldsymbol{\Sigma}}_w^{-1}$ is the stationary value of $\boldsymbol{\Sigma}_w$, the (ij)th element of which is $\sum_{l=0}^{\infty} \bar{\pi}_l \bar{\pi}_{l-(i-j)}$ with $\bar{\pi}_{-l} = \bar{\pi}_l$. Hence, to order O(1),

$$\boldsymbol{\Sigma}_z^{-1} = \dot{\boldsymbol{\Sigma}}_z^{-1} = J' \begin{bmatrix} 0 & 0 \\ 0 & \dot{\boldsymbol{\Sigma}}_w^{-1} \end{bmatrix} \mathbf{J} \qquad \text{(A.7)}$$

To this order of approximation, then, $\dot{\boldsymbol{\Sigma}}_z^{-1}$ is finite, nonnegative definite and its rank is $T - 1$. It follows from (A.5) and (A.7) that, for large $-m$, the expressions in (A.3) becomes

$$E(\mathbf{z}\,|\,\mathbf{y}) = (I + v\dot{\boldsymbol{\Sigma}}_z^{-1})^{-1}y \ ,$$

and

$$\mathrm{cov}\,(\mathbf{z}\,|\,\mathbf{y}) = \sigma_e^2(I + v\dot{\boldsymbol{\Sigma}}_z^{-1})^{-1} \ . \qquad \text{(A.8)}$$

The preceding result can be readily extended to the situation in which $\phi_r(B)$ in (A.1) takes the form

$$\phi_r(B) = \phi_{r^*}(B)(1 - B)^{d_1}(1 - B^c)^{d_2} \qquad \text{(A.9)}$$

where d_1, d_2, c are positive integers and $r^* + d_1 + cd_2 = r$. In this case, for large $-m$, the expressions in (A.8) still hold except that the rank of $\dot{\boldsymbol{\Sigma}}_z^{-1}$ will be reduced to $T - (d_1 + cd_2)$.

A.2 The Asymptotic Form of $\dot{\boldsymbol{\Sigma}}_z^{-1}$

When the number of observations T is large, the asymptotic elements of $\dot{\boldsymbol{\Sigma}}_z^{-1}$ can be obtained from the generating function

$$\frac{\phi_r(B)\phi_r(F)}{\psi_q(B)\psi_q(F)} = X_0 + \sum_{k=1}^{\infty} X_k(B^k + F^k) \qquad \text{(A.10)}$$

where $F = B^{-1}$. Specifically, let σ^{ij} be the (i, j)th element of $\dot{\boldsymbol{\Sigma}}_z^{-1}$. Then, for i not close to 1 or T, $\lim_{T\to\infty} \sigma^{ij} = X_{|i-j|}$. This result was obtained by Wise [18] for when the model of z_t in (A.1) is stationary and invertible. It also holds when $\phi_r(B)$ takes the form (A.9). This can be seen by considering the simplest nonstationary case $\phi_r(B) = \phi_{r^*}(B)(1 - B)$. Since w_t follows the stationary model in (A.6), the generating function of the asymptotic elements of $\dot{\boldsymbol{\Sigma}}_w^{-1}$ can be written

$$\frac{\phi_{r^*}(B)\phi_{r^*}(F)}{\psi_q(B)\psi_q(F)} = \tilde{x}_0 + \sum_{k=1}^{\infty} \tilde{x}_k(B^k + F^k) \ , \quad \text{say} \ . \qquad \text{(A.11)}$$

From (A.7), the elements σ^{ij} of $\dot{\boldsymbol{\Sigma}}_z^{-1}$ and the elements $\tilde{\sigma}^{ij}$ of $\dot{\boldsymbol{\Sigma}}_w^{-1}$ are related by

$$\sigma^{ij} = \tilde{\sigma}^{ij} - \tilde{\sigma}^{i+1,j} - \tilde{\sigma}^{i,j+1} + \tilde{\sigma}^{i+1,j+1} \ .$$

Whence, from (A.11), for i not close to 1 or T

$$\lim_{T\to\infty} \sigma^{ij} = 2\tilde{x}_{|j-i|} - \tilde{x}_{|j-i-1|} - \tilde{x}_{|j-i+1|} \ .$$

Thus, the generating function of σ^{ij} is

$$2(\tilde{x}_0 - \tilde{x}_1) + \sum_{k=1}^{\infty} (2\tilde{x}_k - \tilde{x}_{k-1} - \tilde{x}_{k+1})(B^k + F^k)$$
$$= \frac{(1 - B)\phi_{r^*}(B)\phi_{r^*}(F)(1 - F)}{\psi_q(B)\psi_q(F)} = \frac{\phi_r(B)\phi_r(F)}{\psi_q(B)\psi_q(F)}$$

which gives (A.10). In a similar way, one can show that the generating function holds for any $\phi_r(B)$ of the form (A.9).

A.3 The Asymptotic Form of $E(z_t|y)$ and the Covariance Generating Function of $\hat{e}_t = y_t - E(z_t|y)$

From (A.1), the overall model of y_t can be written

$$\phi_r(B)(y_t - \mu) = \theta_u(B)c_t \qquad \text{(A.12)}$$

where the c_t are i.i.d. $N(0, \sigma_c^2)$ and $\theta_u(B)$ is a polynomial in B of degree u, $u \leq \max\,(p, q)$, having its zeroes lying outside the unit circle. The quantities $\theta_u(B)$, σ_c^2, $\psi_q(B)$, $\phi_r(B)$, σ_b^2 and σ_e^2 are related by the covariance generating function of $\phi_r(B)(y_t - \mu)$, namely,

$$\psi_q(B)\psi_q(F)\sigma_b^2 + \phi_r(B)\phi_r(F)\sigma_e^2 = \theta_u(B)\theta_u(F)\sigma_c^2 \ . \qquad \text{(A.13)}$$

Making use of (A.10), it follows from (A.4) and (A.8) that, for large T the asymptotic form of the conditional expectation of z_t given y is

$$E(z_t\,|\,\mathrm{y}) = [\theta_u(B)\theta_u(F)\sigma_c^2]^{-1}[\sigma_b^2\psi_q(B)\psi_q(F)y_t + \sigma_e^2\phi_r(B)\phi_r(F)\mu] \ . \qquad \text{(A.14)}$$

In particular, if $\phi_r(B)$ contains the factor $(1 - B)$, then $\phi_r(B)\mu = 0$ and

$$E(z_t\,|\,\mathrm{y}) = [\theta_u(B)\theta_u(F)\sigma_c^2]^{-1}\sigma_b^2\psi_q(B)\psi_q(F)y_t \ , \qquad \text{(A.15)}$$

which is a symmetric moving average of y_t. Further, by setting $B = 1$ in (A.13) we see that $[\psi_q(1)]^2\sigma_b^2 = [\theta_u(1)]^2\sigma_c^2$ so that the weights of (A.15) sum to one.

For the autocorrelations of $\hat{e}_t = y_t - E(z_t|y)$ it is readily verified from (A.13) and (A.14) that the autocovariance generating function of $\hat{e}_t$ is

$$\mathrm{C.G.F.}(\hat{e}_t) = \sigma_e^2 \frac{\sigma_e^2\phi_r(B)\phi_r(F)}{\sigma_c^2\theta_u(B)\theta_u(F)} \qquad \text{(A.16)}$$

from which the variance and autocorrelations can be readily calculated.

If z_t is stationary, expressions (A.14) and (A.16) are given in [17, p. 58] for $\mu = 0$. We have shown that these expressions are equally applicable when z_t is nonstationary with $\phi_r(B)$ assuming the form (A.9).

A.4 The Conditional Expectations $E(p_t|y)$ and $E(s_t|y)$

In (3.1), let $z_t = p_t + s_t$ and let $\phi^*(B)$ be the factor common to $\phi_{r_1}(B)$ and $\phi_{r_2}(B)$ and write

$$\phi_{r_1}(B) = \phi_{r^*_1}(B)\phi^*(B) \ , \quad \phi_{r_2}(B) = \phi_{r^*_2}(B)\phi^*(B) \qquad \text{(A.17)}$$

where

$$r_1^* \leq r_1 \quad \text{and} \quad r_2^* \leq r_2 \ .$$

It follows that the model for z_t in (A.1) is related to the models for p_t and s_t by

$$\phi_r(B) = \phi_{r^*_1}(B)\phi_{r^*_2}(B)\phi^*(B) \qquad \text{(A.18)}$$

and

$$\sigma_b^2\psi_q(B)\psi_q(F) = \sigma_{b_2}^2\phi_{r^*_1}(B)\phi_{r^*_1}(F)\psi_{q_2}(B)\psi_{q_2}(F) + \sigma_{b_1}^2\phi_{r^*_2}(B)\phi_{r^*_2}(F)\psi_{q_1}(B)\psi_{q_1}(F) \ ,$$

where $q \leq \max\,(r_1^* + q_2, r_2^* + q_1)$.

We now obtain the conditional expectation $E(s_t|\mathrm{y})$, by first finding the expectation $E(s_t|\mathbf{z})$. From (3.1) and adopting an approach analogous to the derivation of (A.14), the asymptotic form is

$$E(s_t|z) = [\psi_q(B)\psi_q(F)\sigma_b{}^2]^{-1}\sigma_{b_2}{}^2\psi_{q_2}(B)\psi_{q_2}(F)\phi_{r^*_1}(B)\phi_{r^*_1}(F)(z_t - \mu) \tag{A.19}$$

Since p_t, s_t and e_t are assumed independent, we see that

$$E(s_t|\mathbf{y}) = E[E(s_t|\mathbf{z})|\mathbf{y}] \ .$$

On substituting $E(z_t|\mathbf{y})$ of (A.14) for z_t on the right side of (A.19), we obtain

$$E(s_t|\mathbf{y}) = \frac{\sigma_{b_2}{}^2\psi_{q_2}(B)\psi_{q_2}(F)\phi_{r^*_1}(B)\phi_{r^*_1}(F)}{\sigma_c{}^2\theta_u(B)\theta_u(F)}(y_t - \mu) \ , \tag{A.20}$$

where $\sigma_c{}^2\theta_u(B)\theta_u(F)$ can be readily obtained from (A.13) and (A.18).

Finally, the conditional expectation $E(p_t|\mathbf{y})$ is simply the difference,

$$E(p_t|\mathbf{y}) = E(z_t|\mathbf{y}) - E(s_t|\mathbf{y}) \ . \tag{A.21}$$

Note that when $\phi_{r^*_1}(B)$ contains the factor $(1 - B)$, the constant μ in (A.14) and (A.20) vanishes. In this case, (i) $E(s_t|\mathbf{y})$ is a symmetric moving average with weights summing to zero, (ii) $E(z_t|\mathbf{y})$ and therefore $E(p_t|\mathbf{y})$ are symmetric moving averages with weights summing to one. This situation occurs for the model (3.2) because $\phi_{r_1}(B)$ and $\phi_{r_2}(B)$ have a common factor $(1 - B)$ so that $\phi_{r^*_1}(B) = (1 - B)$, and it corresponds precisely to the nature of the moving average operators for $\hat{s}$ and $\hat{p}$ in the census procedure.

[Received January 1974. Revised September 1975.]

REFERENCES

[1] Box, G.E.P. and Jenkins, G.M., *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day, 1970.

[2] Brown, R.G., *Smoothing, Forecasting, and Prediction of Discrete Time Series*, Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1963.

[3] Burman, J.P., "Moving Seasonal Adjustment of Economic Time Series," *Journal of the Royal Statistical Society, Ser. A*, 128, No. 4 (1965), 534–58.

[4] Cleveland, W.P., "Analysis and Forecasting of Seasonal Time Series," Unpublished Ph.D. thesis, Statistics Department, University of Wisconsin, August 1972.

[5] Durbin, J., "Trend Elimination by Moving-Average and Variate Difference Filters," *Bulletin of the International Statistical Institute*, 39, No. 2 (1962), 131–41.

[6] ———, "Trend Elimination for the Purpose of Estimating Seasonal and Periodic Components of Time Series," in M. Rosenblatt, ed., *Time Series Analysis*, New York: John Wiley & Sons, Inc., 1963, Ch. 1.

[7] Hannan, E.J., "The Estimation of Seasonal Variation in Economic Time Series," *Journal of the American Statistical Association*, 58, (March 1963), 31–44.

[8] ———, *Multiple Time Series*, New York: John Wiley & Sons, Inc., 1970.

[9] Kendall, M.G. and Stuart, A., *The Advanced Theory of Statistics, Vol. 3.*, London: Charles W. Griffin & Co., Ltd., 1966.

[10] Leong, Y.S., "Use of an Iterated Moving Average in Measuring Seasonal Variations," *Journal of the American Statistical Association*, 57, (March 1962), 149–71.

[11] Nerlove, M., "Spectral Analysis of Seasonal Adjustment Procedures," *Econometrica*, 32, No. 3 (1964), 241–86.

[12] Shiskin, J. and Eisenpress, H., "Seasonal Adjustments by Electronic Computer Methods," *Journal of the American Statistical Association*, 52 (December 1957), 415–49.

[13] ———, Young, A.H. and Musgrave, J.C., "The X-11 Variant of Census Method II Seasonal Adjustment Program," Technical Paper No. 15, Bureau of the Census, U.S. Dept. of Commerce, 1967.

[14] Thompson, H.E. and Tiao, G.C., "Analysis of Telephone Data: A Case Study of Forecasting Seasonal Time Series," *The Bell Journal of Economics and Management Science*, No. 2 (1971), 515–41.

[15] Wallis, K.F., "Seasonal Adjustment and Relations Between Variables," *Journal of the American Statistical Association*, 69 (March 1974), 18–31.

[16] Whittaker, Sir Edmond and Robinson, G., *The Calculus of Observations*, 4th ed., London: Blackie and Son, Ltd., 1944.

[17] Whittle, P., *Prediction and Regulation by Linear Least Squares Methods*, Princeton, N.J.: D. Van Nostrand, Inc., 1963.

[18] Wise, J., "The Autocorrelation Function and Spectral Density Function," *Biometrika*, 42, Nos. 1, 2 (1955), 151–9.