BUREAU OF THE CENSUS

STATISTICAL RESEARCH DIVISION REPORT SERIES

SRD Research Report Number: CENSUS/SRD/RR-89/01

MODELING TIME SERIES SUBJECT TO SAMPLING ERROR

by

William Bell
Statistical Research Div.
Bureau of the Census
Washington, D.C. 20233

Steven Hillmer
University of Kansas
301 Summerfield Hall
Lawrence, KS 66045

Recommended by: Nash J. Monsour

Report completed: January 3, 1989

Report issued: January 3, 1989

1. Introduction

Papers by Scott and Smith (1974), and Scott, Smith, and Jones (1977) suggested the use of signal extraction results from time series analysis to improve estimates in periodic surveys. Given models for the true unobserved time series (population quantities) and the sampling errors, these results produce estimates of the population quantities that have minimum mean squared error among estimates that are linear functions of the observed time series of survey estimates. To apply these results in practice one must model the time series structure of both the population quantities and the sampling errors. This presents certain difficulties, which have impeded the adoption of signal extraction techniques by government agencies doing periodic surveys. Research efforts have expanded in recent years to attempt to address some of the problems involved. See, e.g., Hausman and Watson (1985), Miazaki (1985), Rao, Srinath, and Quenneville (1986), Bell and Hillmer (1987), Tam (1987), and Binder and Dick (1986).

This paper expands on our previous work (Bell and Hillmer 1987) to address some of the modeling and computational issues involved in applying signal extraction techniques to periodic surveys. Section 2 reviews some theoretical results obtained in Bell and Hillmer (1987). Section 3 briefly discusses modeling of the signal (population quantities) and noise (sampling error) components. Section 4 discusses the use of the Kalman filter and smoother for doing the computations needed for model estimation and signal extraction. Finally, in section 5 we apply these ideas and results to two Census Bureau time series.

2. Theoretical Results

We let Y_t denote the time series of the usual survey estimates, S_t denote the population quantities being estimated, and N_t denote the sampling error in Y_t as an estimate of S_t. The basic decomposition is

$$Y_t = S_t + N_t \ . \tag{2.1}$$

This differs from the notation originally used by Scott and Smith (1974), Scott, Smith, and Jones (1977), and in Bell and Hillmer (1987), where θ_t and e_t are used instead of S_t and N_t. Our choice of notation here is made to conform to the time series signal extraction literature, where S_t denotes the signal and N_t the noise. We will also want to use the multiplicative decomposition

$$Y_t = S_t \cdot U_t = S_t(1 + \tilde{U}_t) \tag{2.2}$$

where $\tilde{U}_t = N_t/S_t$ and $U_t = 1 + \tilde{U}_t$. Taking logs transforms (2.2) into an additive decomposition for $\ln(Y_t)$:

$$\ln(Y_t) = \ln(S_t) + \ln(U_t) \qquad (2.3).$$

Working with $\ln(Y_t)$ is often useful, as will be discussed in section 3.

The theoretical results that follow are obtained under fairly general conditions in Bell and Hillmer (1987), where the assumptions are stated explicitly and proofs are given. Here we briefly state the results and review their implications.

Result 2.1: If Y_t is design unbiased for all t, then S_t and N_t are uncorrelated time series.

Result 2.2: If Y_t is design unbiased for all t, then $\ln(S_t)$ and $\ln(U_t)$ are approximately uncorrelated time series.

These results are useful since time series signal extraction results (see section 4) typically assume S_t and N_t are uncorrelated time series.

Design consistency of signal extraction estimates is established under the superpopulation framework of Fuller and Isaki (1981). Let Y_t^ℓ (from the ℓ^{th} sample at time t) be a sequence of estimators of the characteristic S_t^ℓ of the ℓ^{th} population at time t, where the populations and samples for $\ell = 1,2,\ldots$ are nested. (See their paper for details.) Let $\hat{S}_t^\ell$ be the signal extraction estimate of S_t^ℓ using the time series Y_t^ℓ -- see section 4 for formulas. We have the following results:

Result 2.3: If $Y_t^\ell \to S_t^\ell$ in mean square as $\ell \to \infty$, then $\hat{S}_t^\ell \to S_t^\ell$ in mean square as $\ell \to \infty$.

Result 2.4: If $Y_t^\ell \to S_t^\ell$ in probability as $\ell \to \infty$ and there exist random variables ζ_t with finite variance such that $|N_t^\ell| \leq \zeta_t$ (almost surely) uniformly in ℓ, then $\hat{S}_t^\ell \to S_t^\ell$ in probability as $\ell \to \infty$.

If we take logarithms, we similarly define $\ln(\hat{S}_t^\ell)$ as a signal extraction estimate of $\ln(S_t^\ell)$ using the time series $\ln(Y_t^\ell)$. We have, Result 2.5: If $Y_t^\ell \to S_t^\ell$ in mean square as $\ell \to \infty$, then $\ln(Y_t^\ell) \to \ln(S_t^\ell)$ and $\ln(\hat{S}_t^\ell) \to \ln(S_t^\ell)$ in mean square as $\ell \to \infty$.

As before, a convergence in probability result can be obtained by imposing a boundedness condition on the $\ln(U_t^\ell)$. We can use $\exp[\ln(\hat{S}_t^\ell)]$ as an estimate of $S_t^\ell = \exp[\ln(S_t^\ell)]$ and have the following corollary to Result 2.5.

Corollary: If $Y_t^\ell \to S_t^\ell$ in mean square as $\ell \to \infty$, then $\exp[\ln(\hat{S}_t^\ell)] \to S_t^\ell$ in probability as $\ell \to \infty$.

What these consistency results show is that if the errors in the original estimates Y_t of S_t are small then the errors $\hat{S}_t - S_t$ will be small as well. This is because when there is little error in the original estimates, Y_t, the time series approach will not change them much. Binder and Dick (1986) have noted this phenomenon, and also pointed out that in this case it does not matter what time series model is used. Thus, the consistency results apply to the use of models with estimated parameters. While it is reassuring to know that the time series estimates behave sensibly in the situation of small error in the original estimates, the gains from the time series approach will come in the opposite case -- when $\mathrm{Var}(N_t)$ is large.

3. Component Modeling

Fundamental to our development of models for both the S_t and N_t components will be the use of ARIMA (autoregressive-integrated-moving average) models. Some familiarity with ARIMA models (as in Box and Jenkins 1970) is assumed here. Since such models have often been used successfully for the analysis of observed series Y_t from periodic surveys with little or no sampling error, it seems likely that they should also prove useful for modeling S_t. Using ARIMA models with the sampling error component, N_t, may be thought of more as a useful approximation, though, in some cases, aspects of the survey design or the form of the estimates Y_t may suggest some use of ARIMA structure for N_t. Use of ARIMA models facilitates model estimation and signal extraction, since techniques for these are well-developed for ARIMA models. In what follows we assume the Y_t are design unbiased estimates of S_t, so that (Results 2.1 and 2.2) S_t and N_t are time series uncorrelated with each other.

3.1 Modeling the Signal Component (Population Quantities), S_t

Since we shall assume the model for N_t is estimated using survey microdata and knowledge of the survey design, and not using the time series Y_t, we can use Y_t in developing and estimating the model for S_t. One approach to modeling S_t is to model Y_t directly and deduce the model for S_t from those for Y_t and N_t. This would use the fact that since S_t and N_t are time series uncorrelated with each other

$$\text{Cov}(Y_t, Y_{t+k}) = \text{Cov}(S_t, S_{t+k}) + \text{Cov}(N_t, N_{t+k}) \ . \qquad (3.1)$$

Three potential problems can arise with this approach. One is that the covariance function for S_t resulting from (3.1) may not be positive semi-definite. Another is that, especially with seasonal data, even relatively simple models for S_t and N_t can lead through (3.1) to a relatively complicated model for Y_t of the sort that would not be developed when modeling Y_t directly. Thus, taking a simple model for Y_t, estimating it, and solving for the model for S_t, could miss important features of S_t and yield a bad approximation to the covariance structure of S_t. The third problem occurs if N_t exhibits known nonstationarities such as a variance changing over time. It is difficult to account for such nonstationarities in N_t when modeling Y_t directly, which can lead to an inferior model for S_t as well.

(3.1) might be more useful in model identification. One could use it to obtain autocovariance estimates for S_t given those for N_t and sample autocovariances for Y_t (the same could be done for differenced S_t using differenced Y_t), and the resulting autocorrelation estimates could be used in model identification for S_t. Partial autocorrelations could also be computed. The problem of $Cov(S_t,S_{t+k})$ not being positive semi-definite can arise here as well, though it may not be important in identification.

Experience with modeling time series Y_t suggests that dealing with nonstationarity in S_t will be very important. Nonlinear transformations, differencing, and use of regression mean functions can be quite useful methods for dealing with the usual types of nonstationarity in S_t.

The logarithm is a common transformation used in time series analysis. It is particularly convenient here because of the decomposition (2.3). Fortunately, it often makes sense to work with $\ln(U_t)$ as well as $\ln(S_t)$, as we shall discuss in section 3.2. Other transformations than the logarithm could be used, though they will not generally lead to an additive decomposition for transformed Y_t in terms of transformed S_t and sampling error that can easily be used to produce an estimate of S_t. It seems likely that a choice of either taking logarithms or not transforming will be sufficient to deal with many cases.

Most observed time series appear to require differencing to produce a stationary series. We shall assume here that the model for N_t does not involve differencing, so that Y_t and S_t will then require the same differencing operator. Common choices of differencing operator are 1-B, $1-B^{12}$, and $(1-B)(1-B^{12})$, the latter two for monthly seasonal data. (B is the backshift operator, $BY_t = Y_{t-1}$). If N_t does not require differencing then $\mathrm{Corr}(N_t, N_{t+k})$ should die out with increasing k while $\mathrm{Corr}(S_t, S_{t+k})$ and $\mathrm{Corr}(Y_t, Y_{t+k})$ do not. Thus, sample correlations of Y_t could be used to identify the differencing operator for Y_t and S_t, or we could construct estimates of autocorrelations for S_t as suggested earlier and use these to identify differencing for S_t.

If Y_t is design unbiased, then $E(Y_t) = E(S_t)$ and $E(N_t) = 0$. It is often useful to allow $E(Y_t) = E(S_t) = \mu_t$, say, to vary over time. This is conveniently done with a parametric form for μ_t, such as the linear regression function $\mu_t = \beta_1 X_{it} + \ldots + \beta_m X_{mt}$. Examples of useful regression variables are trading-day and holiday variables for modeling calendar variation (Bell and Hillmer 1983), seasonal indicator variables for a stable seasonal pattern, and intervention and outlier variables to model unusual behavior of the series due to known or unknown causes (see Hillmer, Bell, and Tiao 1983). Other types of regression variables may be suggested in particular applications. We recommend <u>against</u> the use of polynomial functions of time for the X_{it} since differencing seems generally more appropriate, and use of polynomial regression on time can have bad consequences in this case. (See Nelson and Kang 1984 and the references given there.)

The models we shall use for S_t can be written in the form

$$S_t = \sum_{i=1}^{m} \beta_i X_{it} + \frac{\theta_S(B)}{\phi_S(B)\delta(B)} b_t \tag{3.2}$$

where $\delta(B)$, $\phi_S(B)$, $\theta_S(B)$ are the (possibly multiplicative) differencing, autoregressive (AR), and moving average (MA) operators, and b_t is a white noise series (iid $N(0, \sigma_b^2)$). This can be thought of as convenient notation for

$$\delta(B)[S_t - \sum_{i=1}^{m} \beta_i X_{it}] = \frac{\theta_S(B)}{\phi_S(B)} b_t \tag{3.3}$$

showing that the X_{it} must be differenced in the same way as S_t. We can also substitute $\ln(S_t)$ for S_t in (3.2) and (3.3). Several approaches can be used to specify the AR and MA operators in (3.2). We could use estimated autocorrelations and partial autocorrelations for S_t, and follow the scheme in Box and Jenkins (1970). We could just pick simple AR and MA operators, and (1) estimate and diagnostic check the resulting model, modifying the model if it seems inadequate, or (2) try several different models and use a model selection criterion, such as AIC (Akaike 1973) to choose among them. An approach we have found useful is to first model Y_t directly, and then use the resulting form of this model for S_t, but reestimating the parameters taking the sampling error component into account. If there is not an excessive amount of sampling error present, the model for Y_t should at least provide a useful starting point in modeling S_t. Details of the modeling are best illustrated in the examples of section 5.

3.2 Modeling the Sampling Errors, N_t

The first step in modeling N_t is to estimate the sampling error covariances over time, $Cov(N_t, N_{t+k})$. In principle, this is the same problem as estimation of sampling variances (k=0), which is routinely done for periodic surveys and for which many methods are available (Wolter 1985). In practice, there may be difficulties in linking survey microdata over time to do this. We shall not address these problems here, but will assume estimates $c_N(t,t+k)$ and $r_N(t,t+k)$ of $Cov(N_t,N_{t+k})$ and $Corr(N_t,N_{t+k})$ are available.

We could attempt to use the $c_N(t,t+k)$ directly in signal extraction, but this runs into two problems. The first is that $c_N(t,t+k)$ may not be positive semi-definite. The second is that the estimates $c_N(t,t+k)$ and $r_N(t,t+k)$ are likely to be highly variable if $Var(N_t)$ is large. This is the situation where signal extraction can make substantial difference in the estimates of S_t (see section 2). If we can assume stationarity of N_t, so that its covariances and correlations depend only on the lag so that $Cov(N_t,N_{t+k}) = \gamma_N(k)$ for all t, then we can average the estimates of covariances or correlations over time to improve them. For example, with estimates $r_N(t,t+k)$ for $t = 1,\ldots,T-k$ $(k > 0)$ of $Corr(N_t,N_{t+k}) = \rho_N(k)$, we could use

$$\hat{\rho}_N(k) = \frac{1}{T-k} \sum_{t=1}^{T-k} r_N(t,t+k) \quad . \tag{3.4}$$

We can do this averaging over segments of the time series if N_t is not stationary over the full length of the series, as discussed later.

In order to parsimoniously represent the correlation structure of N_t, we shall make use of ARMA models. We could pick the parameters of the ARMA model so the correlations from the model reasonably approximate the $\hat{\rho}_N(k)$ (possibly to minimize some measure of the discrepancy), or we could solve for the model parameters to exactly produce some of the $\hat{\rho}_N(k)$. We should also use any relevant knowledge we have regarding the survey design or estimators Y_t in developing the ARMA model for N_t. One important fact is that $\rho_N(k) = 0$ at any lags k where the samples do not overlap and are drawn independently. Miazaki (1985) used such knowledge in modeling sampling errors in the National Crime Survey, and Hausman and Watson (1985) did so with the Current Population Survey (CPS). However, one must be careful in doing this. Train, Cahoon, and Makens (1978) estimated sampling error correlations in the CPS and obtained apparently nonzero correlations at lags with no sample overlap. These seemed due to the fact that when a housing unit leaves CPS it is generally replaced with one from the same neighborhood. D. G. Steel and R. G. De Mel of the Australian Bureau of Statistics, in unpublished work, developed sampling error models that attempt to account for this phenomenon in the Australian Labor Force Survey.

We will often need to allow N_t to have a variance that changes over time. In some cases it may be appropriate to assume that the <u>relative variance</u>, $R_t = \text{Var}(N_t)/S_t^2$ (where Var is taken with respect to the sampling distribution), is stable over time. Consider the decomposition (2.3). A simple Taylor expansion argument suggests that

$$\mathrm{Var}[\ln(U_t)] \approx \mathrm{Var}(\tilde{U}_t) = E(R_t)$$

if $\tilde{U}_t$ is not too large. If $E(R_t)$ can be assumed constant over time, and if it makes sense to model $\ln(S_t)$, then we take $\ln(Y_t)$ and use (2.3) with models for $\ln(S_t)$ and $\ln(U_t)$. In modeling $\ln(U_t)$ we use the fact that correlations are approximately unchanged by transformations such as the logarithm. (This again follows by approximating the transformation with a first order Taylor series.)

If neither the original nor relative variance of N_t is constant over time, or if the same transformation is not appropriate for both S_t and N_t, then we need to let $\mathrm{Var}(N_t)$ change over time. This could occur, for example, if the population variance (over the individual units) was constant over time but the sample size changed. We shall then use a model for N_t of the form

$$N_t = h(t)\,\dot{N}_t \qquad \phi_N(B)\dot{N}_t = \theta_N(B)c_t \tag{3.5}$$

where c_t is white noise (iid $N(0,\sigma_c^2)$), $\dot{N}_t$ has constant variance over time, and $h(t) \geq 0$ is such that $\mathrm{Var}(N_t) = h(t)^2\,\mathrm{Var}(\dot{N}_t)$. If $\mathrm{Var}(N_t)$ does not change over time we use $h(t) \equiv 1$.

Another problem that can occur is that of modeling the effect of sample redesigns. For example, suppose we are modeling data at times $t=1,\ldots,T$, and at $t = T_1+1$ the sample is independently redrawn, with the first sample in effect from $t = 1,\ldots,T_1$ and the second from $t = T_1+1,\ldots,T$. We handle this by generalizing (3.5) to

$$N_t = h_1(t)N_{1t} + h_2(t)N_{2t} \qquad \phi_N(B)N_{it} = \theta_N(B)c_{it} \qquad i = 1,2$$

$$h_1(t) = \begin{cases} h(t) & t \le T_1 \\ 0 & t > T_1 \end{cases} \qquad h_2(t) = \begin{cases} 0 & t \le T_1 \\ h(t) & t > T_1 \end{cases}, \qquad (3.6)$$

$$c_{it} \text{ iid } N(0, \sigma_c^2) \quad i = 1,2$$

Here N_{1t} is assumed independent of N_{2j} for all t,j. Thus, N_{1t} is the sampling error from $t = 1,\ldots,T_1$ and N_{2t} is the sampling error from $t = T_1+1,\ldots,T$. We have assumed the same ARMA model for N_{1t} and N_{2t} though this could be generalized. We can also obviously generalize (3.6) to handle more than two segments of the series (more than one redrawing of the sample.)

We do not mean to imply that the techniques described here can perfectly represent the covariance structure of N_t. We are merely suggesting these as useful tools for first approximations. Much fundamental work still remains in developing time series models for sampling errors.

4. Computations for Model Estimation and Signal Extraction

Three general approaches to deriving time series results and doing computations might be called the classical approach, the matrix approach, and the Kalman filter approach. Each has its relative advantages and disadvantages. We shall use the Kalman filter approach here (see Anderson and Moore 1979 for a general discussion) because it is particularly convenient for handling component models with such features as changing variances over time. The classical approach, which works directly with linear filters and difference equation forms of models, is not well suited to estimation of component models, nor to signal extraction with changing variances. The matrix approach can be readily used for component model estimation and signal extraction, as discussed in Bell and Hillmer (1988), though dealing with variances changing over time presents some problems not considered there.

Before developing the Kalman filter approach we give classical and matrix signal extraction results for the case where N_t is stationary. This shows what is being calculated by the Kalman smoother recursions for this case. The classical results are most easily expressed for the case where the entire doubly infinite realization $\{Y_t: t=0,\pm1,\pm2,\ldots\}$ is available. Then the signal extraction estimate $\hat{S}_t$, and autocovariance generating function (ACGF) of the error $S_t - \hat{S}_t$, are given by

$$\hat{S}_t = \mu_t + \frac{\gamma_u(B)}{\gamma_w(B)}(Y_t - \mu_t)$$

$$\gamma_{S-\hat{S}}(B) = \gamma_N(B) - \gamma_N(B)^2\delta(B)\delta(F)/\gamma_w(B) \tag{4.1}$$

In (4.1) $\mu_t = \sum_i^m \beta_i X_{it}$ is the mean function, $\gamma_u(B) = \sum_{-\infty}^{\infty} \gamma_u(k)B^k$ is the ACGF of $u_t = \delta(B)(S_t - \mu_t)$, $\gamma_w(B)$ is the ACGF of $w_t = \delta(B)(Y_t - \mu_t)$, $\gamma_N(B)$ is the ACGF of N_t, $\delta(B)$ is the differencing operator needed for S_t and Y_t, and $F = B^{-1}$. We can alternatively express $\hat{S}_t$ as

$$\hat{S}_t = Y_t - \hat{N}_t, \qquad \hat{N}_t = \frac{\gamma_N(B)\delta(B)\delta(F)}{\gamma_w(B)}(Y_t - \mu_t) \tag{4.2}$$

$\hat{S}_t$ has long been known to be the "optimal" (minimum mean squared error (MMSE) linear) estimator when S_t and N_t are both stationary, i.e. $\delta(B) = 1$ (see, e.g. Whittle 1963). Bell (1984) notes it is also optimal when S_t requires differencing under certain assumptions about starting values for S_t and Y_t, the series that need to be differenced. The results can be modified to deal with semi-infinite data (Y_t for t = T, T-1, T-2,...) or finite data. The classical results were used by Scott and Smith (1974), and Scott, Smith, and Jones (1977).

Now assume Y_t is available for $t = 1,\ldots,T$ and let $\underset{\sim}{Y} = (Y_1,\ldots,Y_T)'$. Similarly define $\underset{\sim}{S}$, $\underset{\sim}{N}$, and $\underset{\sim}{\mu}$, let $\underset{\sim}{w} = (w_{d+1},\ldots,w_T)'$ be the differences of the Y_t data with means removed, (where d is the degree of the differencing operator $\delta(B)$), and let $\Sigma_w = \mathrm{Var}(\underset{\sim}{w})$ and $\Sigma_N = \mathrm{Var}(\underset{\sim}{N})$. Define the $(n-d) \times n$ matrix Δ by

$$\Delta = \begin{bmatrix} -\delta_d & \cdots & -\delta_1 & 1 & & & \\ & \ddots & & \ddots & \ddots & & \\ & & -\delta_d & \cdots & -\delta_1 & & 1 \end{bmatrix}$$

so $\underset{\sim}{w} = \Delta(\underset{\sim}{Y} - \underset{\sim}{\mu})$. Then (Bell and Hillmer 1988)

$$\hat{\underset{\sim}{S}} = \underset{\sim}{Y} - \hat{\underset{\sim}{N}} \qquad \hat{\underset{\sim}{N}} = \Sigma_N \Delta' \Sigma_w^{-1} \Delta(\underset{\sim}{Y} - \underset{\sim}{\mu}) = \Sigma_N \Delta' \Sigma_w^{-1} \underset{\sim}{w}$$

$$\mathrm{Var}(\underset{\sim}{S} - \hat{\underset{\sim}{S}}) = \Sigma_N - \Sigma_N \Delta' \Sigma_w^{-1} \Delta \Sigma_N \, . \tag{4.3}$$

Notice the analogy with (4.1) and (4.2). However, this $\hat{\underset{\sim}{S}}$ is obtained using the "transformation approach" of Ansley and Kohn (1985); it is optimal among linear estimators of $\underset{\sim}{S}$ such that the error $\underset{\sim}{S} - \hat{\underset{\sim}{S}}$ does not depend on starting values for the S_t series (Kohn and Ansley 1987). It is globally optimal under the same assumptions about starting values referred to above (Bell and Hillmer 1988). The results in (4.3) were given without a sound justification in R. G. Jones (1980). If no differencing is required $\Delta = I$ and (4.3) reduces to the usual linear projection results.

4.1 State Space Form and the Kalman Filter

An ARIMA model can be put in state space form (see, e.g. Ansley and Kohn 1985), which we illustrate as follows. For the moment let Z_t be a zero mean time series following the ARIMA model $\delta(B)\phi(B)Z_t = \theta(B)a_t$. First let $\tilde{\phi}(B) = \delta(B)\phi(B) = 1-\tilde{\phi}_1 B - \ldots - \tilde{\phi}_r B^r$ and $\theta(B) = 1-\theta_1 B \ \ldots - \theta_q B^q$. Define the f × 1 "state vector" $\underset{\sim}{X}(t) = (X_1(t),\ldots,X_f(t))'$, where $f = \max(r,q+1)$, by $X_1(t) = Z_t$ and

$$X_i(t) = \sum_{j=i}^{r} \tilde{\phi}_j Z_{t-1+i-j} - \sum_{j=i-1}^{q} \theta_j a_{t-1+i-j} \quad i=2,\ldots,f .$$

($\sum_{j=i}^{k}$ is 0 if $i > k$.) Let the f × 1 vector $\underset{\sim}{H} = (1,0,\ldots,0)'$, the f × 1 vector $\underset{\sim}{G} = (1,-\theta_1,\ldots, -\theta_{f-1})'$, and the f × f matrix

$$F = \begin{bmatrix} \tilde{\phi}_1 & 1 & & \\ \vdots & & \ddots & \\ \vdots & & & 1 \\ \tilde{\phi}_f & \cdots & \cdots & 0 \end{bmatrix}$$

Then we can write the state space form of the model for Z_t as

$$\underset{\sim}{X}(t+1) = \underset{\sim}{F}\,\underset{\sim}{X}(t) + \underset{\sim}{G}\,a_{t+1}$$

$$Z_{t+1} = \underset{\sim}{H}'\,\underset{\sim}{X}(t+1) \,. \qquad (4.4)$$

Returning now to our component set up with $Y_t = S_t + N_t = S_t + h(t)\dot{N}_t$, we now let $Z_t = Y_t - \mu_t$, $\tilde{S}_t = S_t - \mu_t$, and define state space representations analogous to (4.4) for the models for $\tilde{S}_t$ and $\dot{N}_t$:

$$\underset{\sim}{X}_S(t+1) = F_S\,\underset{\sim}{X}_S(t) + \underset{\sim}{G}_S\,b_{t+1} \qquad \underset{\sim}{X}_N(t+1) = F_N\,\underset{\sim}{X}_N(t) + \underset{\sim}{G}_N\,c_{t+1}$$

$$\tilde{S}_{t+1} = \underset{\sim}{H}_S'\,\underset{\sim}{X}_S(t+1) = X_{S1}(t+1) \qquad \dot{N}_{t+1} = H_N'\,\underset{\sim}{X}_N(t+1) = X_{N1}(t+1)$$

where $\underset{\sim}{X}_S(t)$ has dimension fs = max(rs,qs+1) and $\underset{\sim}{X}_N(t)$ has dimension fn = max(pn,qn+1). Now let $\underset{\sim}{X}(t) = (X_S(t)', X_N(t)')'$ be $f \times 1$ with f = fs+fn, let

$$F = \begin{bmatrix} F_S & 0 \\ 0 & F_N \end{bmatrix} \qquad G = \begin{bmatrix} \underset{\sim}{G}_S & \underset{\sim}{0} \\ \underset{\sim}{0} & \underset{\sim}{G}_N \end{bmatrix}$$

and let $\underset{\sim}{H}(t) = [1,0,\ldots,0,h(t),0,\ldots,0]'$ where h(t) is the fs+1 element of $\underset{\sim}{H}(t)$. Then we have the state space form of the model for $Z_t = Y_t - \mu_t$:

$$\underset{\sim}{X}(t+1) = F\,\underset{\sim}{X}(t) + \underset{\sim}{G}\begin{bmatrix} b_{t+1} \\ c_{t+1} \end{bmatrix}$$

$$Z_{t+1} = \underset{\sim}{H}(t+1)\,\underset{\sim}{X}(t+1) \tag{4.5}$$

The Kalman filter is a recursive scheme for producing "optimal" estimates of the state vectors $\underset{\sim}{X}(t)$ and $\underset{\sim}{X}(t+1)$ using data $Z_1,\ldots,Z_t$, along with covariance matrices of the errors in these estimates. Let $\hat{\underset{\sim}{X}}(t|t)$ and $\hat{\underset{\sim}{X}}(t+1|t)$ be the estimates, and let $P(t|t) = \mathrm{Var}(\underset{\sim}{X}(t) - \hat{X}(t|t))$ and $P(t+1|t) = \mathrm{Var}(X(t+1) - \hat{X}(t+1|t))$. The Kalman filter recursive equations (Anderson and Moore 1979, p. 40) can be written for our problem as

$$\hat{\underset{\sim}{X}}(t+1|t) = F\,\hat{\underset{\sim}{X}}(t|t)$$

$$P(t+1|t) = F\,P(t|t)F' + \begin{bmatrix} \sigma_b^2\,\underset{\sim}{G}_S\,\underset{\sim}{G}_S' & 0 \\ 0 & \sigma_c^2\,\underset{\sim}{G}_N\,\underset{\sim}{G}_N' \end{bmatrix}$$

$$\epsilon_{t+1} = Z_{t+1} - \underset{\sim}{H}(t+1)'\,\hat{\underset{\sim}{X}}(t+1|t)$$

$$v_{t+1} = \mathrm{Var}(\epsilon_{t+1}) = \underset{\sim}{H}(t+1)'\,P(t+1|t)\,\underset{\sim}{H}(t+1) \tag{4.6}$$

$$\hat{\underset{\sim}{X}}(t+1|t+1) = \hat{\underset{\sim}{X}}(t+1|t) + P(t+1|t)\,\underset{\sim}{H}(t+1)\,\epsilon_{t+1}/v_{t+1}$$

$$P(t+1|t+1) = P(t+1|t) - P(t+1|t)\,\underset{\sim}{H}(t+1)\,\underset{\sim}{H}(t+1)'\,P(t+1|t)/v_{t+1}$$

Another way to look at the Kalman filter is that it linearly transforms the sequence Z_t into the uncorrelated sequence of innovations ϵ_t. If some data points Z_t are missing, the Kalman filter can handle this as discussed in R. H. Jones (1980).

4.2 Initializing the Kalman Filter

To start the recursive equations (4.6) we need $\hat{\underset{\sim}{X}}(t|t)$ and $P(t|t)$ for some t (the initialization problem). For stationary models the initialization is typically at $t = 0$ with the unconditional mean of $\underset{\sim}{X}(t)$ ($\underset{\sim}{0}$ when using the mean corrected series Z_t) and unconditional variance. Computation of the latter is relatively straightforward, and is discussed by R. H. Jones (1980), though for a different choice of state vector than we use here. The problem is much more difficult in the nonstationary ($\delta(B)\neq 1$) case.

For the nonstationary case Ansley and Kohn (1986) define a "modified Kalman filter", involving some auxiliary recursions to (4.6). This is initialized at time zero and produces "transformation approach" estimates of $\underset{\sim}{X}(t)$ using $Z_1,\ldots,Z_t$ for $t = d, d+1, \ldots, T$ (assuming none of $Z_1, \ldots, Z_d$ are missing -- if they are, the situation is more complex). These transformation approach estimates are analogous to those for the signal extraction problem mentioned earlier. Bell and Hillmer (1989) show that an alternative approach that produces the same results is to initialize at $t = d$ with the transformation approach estimate of $\underset{\sim}{X}(d)$ using $Z_1,\ldots,Z_d$ (let this be $\hat{\underset{\sim}{X}}(d|d)$), and with $P(d|d) = \text{Var}(\underset{\sim}{X}(d) - \hat{\underset{\sim}{X}}(d|d))$. This initialization is actually closely related to the initialization of the modified Kalman filter at t = 0, but this approach avoids the need to do recursions for $t=1,\ldots,d$ and avoids the need to use the modified Kalman filter.

We do not have space to derive the initialization at $t=d$ here, but refer the reader to Bell and Hillmer (1989) for this. We shall just state the results for our particular problem that incorporate some simplifications and one generalization (for $\mathrm{Var}(N_t)$ changing over time) relative to the results given in Bell and Hillmer (1989). First, define the quantities for $i=1,\ldots,d$

$$A_{it} = \begin{cases} 1 & i=t \quad t=1,\ldots,d \\ 0 & i\neq t \\ \delta_1 A_{i,t-1} + \cdots + \delta_d A_{i,t-d} & t > d \\ \delta_1 A_{i,t+1} + \cdots + \delta_d A_{i,t+d} & t \leq 0 \end{cases}$$

and let $\underset{\sim}{A}_t = (A_{1t},\ldots,A_{dt})'$. Assume for now $rs > 0$ and $pn > 0$. Define the matrices

$$\underset{rs \times d}{A_{rs}} = \begin{bmatrix} \underset{\sim}{A}'_{1-ps} \\ \vdots \\ \underset{\sim}{A}'_0 \\ I_d \end{bmatrix}, \qquad D = \mathrm{diag}[h(1),\ldots,h(d)]$$

$$\underset{fs \times rs}{\Phi_S} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ \tilde{\phi}_{S,rs} & \tilde{\phi}_{S,rs-1} & \cdots & \tilde{\phi}_{S,2} & 0 \\ & \tilde{\phi}_{S,rs} & & \vdots & \vdots \\ & & \ddots & \tilde{\phi}_{S,rs} & 0 \\ & & 0_{(fs-rs) \times rs} & & \end{bmatrix},$$

$$\underset{fn \times pn}{\Phi_N} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ \phi_{N,pn} & \phi_{N,pn-1} & \cdots & \phi_{N,2} & 0 \\ & \phi_{N,pn} & \ddots & \vdots & \vdots \\ & & & \phi_{N,pn} & 0 \\ & & 0_{(fn-pn) \times pn} & & \end{bmatrix},$$

$$\underset{fs \times qs}{\theta_S} = \begin{bmatrix} 0 & \cdots & 0 \\ -\theta_{S,qs} & \cdots & -\theta_{S,1} \\ & \ddots & \vdots \\ & & -\theta_{S,qs} \\ & 0_{(fs-qs-1) \times qs} & \end{bmatrix}, \quad \underset{fn \times qn}{\theta_N} = \begin{bmatrix} 0 & \cdots & 0 \\ -\theta_{N.qn} & \cdots & -\theta_{N,1} \\ & \ddots & \vdots \\ & & -\theta_{N,qn} \\ & 0_{(fn-qn-1) \times qn} & \end{bmatrix}$$

$$\underset{rs \times ps}{C_{rs}} = (\pm) \begin{bmatrix} \xi_0 & \xi_1 & \cdots & \xi_{ps-1} \\ 0 & \xi_0 & \cdots & \xi_{ps-2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & \xi_0 \\ & 0_{d \times ps} & & \end{bmatrix}$$

where $\xi(B) = \xi_0 + \xi_1 B + \xi_2 B^2 + \ldots = \delta(B)^{-1}$, and (+) is chosen for C_{rs} if 1-B occurs an even number of times in $\delta(B)$ and (-) is chosen otherwise. (We assume here all the zeros of $\delta(B)$ are on the unit circle, as is the case with differencing operators.) If rs = 0 or pn = 0 we redefine them to be max(rs,1) or max(pn,1), respectively, in the above. For i<j define the notation $\tilde{Z}_j^i = (Z_i, \ldots, Z_j)'$, and similarly with any other of the time series involved. Then the initialization is given by

$$\hat{\underset{\sim}{X}}(d|d) = \begin{bmatrix} \Phi_S \; A_{rs} \; \underset{\sim d}{Z}^{1\prime} \\ \underset{\sim}{0}_{fn \times 1} \end{bmatrix} \tag{4.7}$$

$$P(d|d) = P_1 + P_2 + P_3 + P_3' \tag{4.8}$$

where

$$P_1 = \begin{bmatrix} \Phi_S & 0 \\ 0 & \Phi_N \end{bmatrix} \begin{bmatrix} P_1(1,1) & P_1(2,1)' \\ P_1(2,1) & P_1(2,2) \end{bmatrix} \begin{bmatrix} \Phi_S' & 0 \\ 0 & \Phi_N' \end{bmatrix}$$

$$P_1(1,1) = C_{rs} \mathrm{Var}(\underset{\sim d}{u}^{d+1-ps}) C_{rs}' + A_{rs} D \; \mathrm{Var}(\dot{\underset{\sim d}{N}}^{1}) \; D \; A_{rs}'$$

$$P_1(2,1) = - \left[0_{pn \times (d-pn)} \quad I_{pn}\right] \mathrm{Var}(\dot{\underset{\sim d}{N}}^{1}) \; D \; A_{rs}'$$

$$P_1(2,2) = \mathrm{Var}(\dot{\underset{\sim d}{N}}^{d+1-pn})$$

$$P_2 = \begin{bmatrix} \sigma_b^2 \; \theta_S \theta_S' & 0 \\ 0 & \sigma_c^2 \; \theta_N \theta_N' \end{bmatrix}$$

$$P_3 = \begin{bmatrix} \Phi_S & 0 \\ 0 & \Phi_N \end{bmatrix} \begin{bmatrix} P_3(1,1) & P_3(1,2) \\ P_3(2,1) & P_3(2,2) \end{bmatrix} \begin{bmatrix} \theta_S' & 0 \\ 0 & \theta_N' \end{bmatrix}$$

$$P_3(1,1) = C_{rs} \mathrm{Cov}(\underset{\sim d}{u}^{d+1-ps}, \; \underset{\sim d}{b}^{d+1-qs})$$

$$P_3(1,2) = - A_{rs} D \; \mathrm{Cov}(\dot{\underset{\sim d}{N}}^{1}, \; \underset{\sim d}{c}^{d+1-qn}) \qquad P_3(2,1) = 0$$

$$P_3(2,2) = \mathrm{Cov}(\dot{\underset{\sim d}{N}}^{d+1-pn}, \; \underset{\sim d}{c}^{d+1-qn})$$

The autocovariances of u_t and $\dot{N}_t$ needed for (4.8) can be computed using a method of McLeod (1975,1977) since u_t and $\dot{N}_t$ follow the ARMA models $\phi_S(B)u_t = \theta_S(B)b_t$ and $\phi_N(B)\dot{N}_t = \theta_N(B)c_t$. Also, $\mathrm{Cov}(\underset{\sim}{u}_d^{d+1-ps}, \underset{\sim}{b}_d^{d+1-qs})$ is a $ps \times qs$ matrix with (t,j)th element

$$\mathrm{Cov}(u_t,b_j) = \begin{cases} 0 & \text{if } j > t \\ \sigma_b^2\, \psi_{j-t}^S & \text{if } j \le t \end{cases}$$

where $\psi^S(B) = \psi_0^S + \psi_1^S B + \psi_2^S B^2 + \ldots = \theta_S(B)/\phi_S(B)$. We similarly obtain the $\mathrm{Cov}(\dot{N}_t, c_j)$ needed.

4.3 Model Estimation

Model estimation proceeds by maximum likelihood assuming normality. Ignoring the regression terms for the moment and working with $Z_t = Y_t - \mu_t$, the Kalman filter can be used to evaluate the likelihood as suggested by R. H. Jones (1980). This is done using the innovations ϵ_t, which are independent with mean zero and variance v_t. With initialization at $t=d$, the likelihood is the joint density of $\epsilon_{d+1},\ldots,\epsilon_T$ so that, apart from constants, the log-likelihood is

$$\ell = -\frac{1}{2}\sum_{t=d+1}^{T} \ln(v_t) - \frac{1}{2}\sum_{t=d+1}^{T} \epsilon_t^2/v_t \qquad (4.9)$$

This can be numerically maximized over the parameters of the model for $\tilde{S}_t$, keeping those of the model for N_t fixed, since the latter are estimated elsewhere.

We can handle the regression terms as follows. Let X be the $T \times m$ matrix of regression variables, and $\underset{\sim}{\beta}$ the $m \times 1$ vector of regression parameters. For given values for the ARMA parameters, $\underset{\sim}{\beta}$ can be estimated by generalized least squares (GLS). Following R. H. Jones (1985) this can be achieved by running the Kalman filter recursions with $\underset{\sim}{Y} = (Y_1,\ldots,Y_T)'$ as the data vector to get the standardized innovations $\epsilon_t/(v_t)^{1/2}$ for $t = d+1,\ldots,T$. Call the $(T-d) \times 1$ vector of these $\underset{\sim}{\tilde{Y}}$. Similarly do this with each column of the matrix X and collect the resulting "filtered" columns in the $(T-d) \times m$ matrix $\tilde{X}$. Then GLS results in

$$\hat{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\underset{\sim}{\tilde{Y}} \qquad \mathrm{Var}(\underset{\sim}{\hat{\beta}}) = (\tilde{X}'\tilde{X})^{-1} . \tag{4.10}$$

To estimate both $\underset{\sim}{\beta}$ and the ARMA parameters, the likelihood may be efficiently maximized jointly by iterating between the GLS regression (4.10) for fixed values of the ARMA parameters, and maximization of (4.9) for given $\underset{\sim}{\beta}$ using $Z_t = Y_t - \sum_{1}^{m} \beta_i X_{it}$ as the data. This iterative GLS scheme is investigated in Otto, Bell, and Burman (1987) for regression models with ARIMA errors.

4.4 Signal Extraction Using the Fixed Point Smoother

Anderson and Moore (1979) discuss three "smoothers" (fixed point, fixed lag, and fixed interval) that can be used in conjunction with the Kalman filter to produce estimates of the state vectors $\underset{\sim}{X}(t)$ using all the available data. These estimates are denoted $\hat{X}(t|T)$. We wish to estimate S_t using $Y_1,\ldots,Y_T$, which we do by estimating $\tilde{S}_t$ using $Z_1,\ldots,Z_T$ and then adding back μ_t to the estimate of $\tilde{S}_t$. For simplicity in the notation here let us assume means are zero so we can work directly with S_t. Then S_t is simply the first element of $X(t)$, i.e. $S_t = [1\ 0\ \ldots\ 0]X(t)$.

Since we are only interested in one element of the state vector, the fixed point smoother seems best suited to our problem. The fixed point smoother results of Anderson and Moore (1979, pp. 172-173) can be simplified by multiplying through by the vector [1 0 ... 0] as appropriate to produce estimates and variances for S_t alone. The resulting recursive scheme goes as follows.

We first run the Kalman filter to produce for $t = d, d+1,\ldots,T$:

(1) $\hat{S}_{t|t}$ = first element of $\hat{X}(t|t)$

(2) $\pi_{t|t}$ = (1,1) element of $P(t|t)$

(3) v_t and ϵ_t as defined in (4.6) (not needed for t=d)

(4) $\alpha_t = P(t|t-1)\ H(t)/\sqrt{v_t}$ (not needed for t=d or t=T)

(5) $p_{t|t}$ = first column of $F\ P(t|t)$ (not needed for t=T).

Then to apply the fixed point smoother at each of $t=d,\ldots,n-1$:

(0) Start with $\hat{S}_{t|t}$, $\pi_{t|t}$, and $p^a_{t+1|t} = p_{t|t}$

Compute recursively for k = t+1,...,T as follows:

(1) Compute $\lambda_k = H(k)' \underset{\sim}{p}^a_{k|k-1} / \sqrt{v_k}$

(2) Compute $\hat{S}_{t|k} = \hat{S}_{t|k-1} + \lambda_k \epsilon_k / \sqrt{v_k}$

$$\pi_{t|k} = \pi_{t|k-1} - \lambda_k^2$$

$$\underset{\sim}{p}^a_{k+1|k} = F\,[\underset{\sim}{p}^a_{k|k-1} - \lambda_k \underset{\sim}{a}_k] \quad \text{(not needed for k=T).}$$

5. Example: Retail Trade Survey -- Sales of Eating and Drinking Places

As an illustrative example we analyze time series of sales (in millions of dollars) of Eating Places and of Drinking Places which are estimated in the monthly Retail Trade Survey. The Retail Trade Survey has a panel of large businesses that are selected into the sample with certainty and report sales every month, and 3 rotating panels of smaller businesses that are selected into the sample by stratified simple random sampling. Each rotating panel reports current month and previous month sales at intervals of three months. Horvitz-Thompson (HT) estimates of current and previous months sales are constructed; the resulting time series shall be denoted Y'_t and Y''_{t-1}. From these, composite estimators are constructed as described in Wolter (1979). The final composite estimates will make up our time series Y_t. (While it might be interesting to instead analyze Y'_t and Y''_{t-1} directly, these estimates are not saved for a long enough period of time for seasonal time series modeling.) Sampling variances are estimated by the random group method (Wolter 1985) using 16 random groups. Further information on the survey is given in Isaki, et. al. (1976), Wolter, et. al. (1976), and Wolter (1979).

There are several complicating factors in the survey. The sample is redesigned and independently redrawn every five years, with new samples having been introduced in January of 1972, 1977, 1982, and 1987. This could be handled as discussed in section 3.2, but this will not be done here as our software does not presently allow this. (We shall use data from January, 1977 through December, 1986, so there is one redrawing of the sample exactly in the middle of our series.) When a new sample is introduced there is a three month transition period where the composite estimates are not used, which we shall also ignore. Finally, the monthly estimates are benchmarked to annual totals estimated from an annual survey and from the economic census taken every five years. To avoid this complication we use data that are not benchmarked. The reader should be aware, however, that for this reason the data used here do not agree with published estimates.

Development of Sampling Error Models

Our first step will be to develop a model for the correlation structure of the sampling errors. Let us write $Y_t' = S_t + N_t'$ for the current month (t) HT estimate, and $Y_{t-1}'' = S_{t-1} + N_{t-1}''$ for the previous month (t-1) HT estimate. We shall use the same models for N_t' and N_{t-1}''. Estimates of $Corr(N_t', N_{t-1}'')$ are extremely high -- typically .98 or higher. While this is partly artificial (due to businesses reporting the same figure for current and previous month sales, and possibly to the way missing values are imputed), in the absence of other information it is difficult to distinguish characeteristics of N_t' from those of N_{t-1}''.

Since the three rotating panels in the survey are drawn (approximately) independently (Wolter 1979), correlations for N_t' and N_{t-1}'' will be nonzero only for lags that are multiples of 3. Estimates of such lag correlations can be averaged over time (assuming correlation stationarity) and used to produce estimates of model parameters. While estimates of lag correlations are not regularly produced for the Retail Trade Survey, this was done as part of a special study using data from January, 1973 through March, 1975, albeit at a time when the survey had four rotating panels. Lacking more recent data, we "averaged" the correlations at lags 4, 8, 12, 16, 20, and 24 for N_t' and N_{t-1}'' (this was done after applying Fisher's transformation .5 ln((1+r)/(1-r)) and then transforming the result back). The results are shown in Table 1. They show fairly strong positive correlation in the sampling errors, and evidence of seasonality in the correlations at lag 12. A possible model given such data is

$$(1-\phi^m B^m)(1-\phi_{12}B^{12})N_t' = \nu_{1t} \qquad (5.1)$$

where m = 4 for the 4-panel survey, with the same model assumed for N_{t-1}'' with $\nu_{2,t-1}$ replacing ν_{1t}.

A particularly convenient property of (5.1) is that if the sampling error in each panel would follow (5.1) with m = 1 if it were observed every month, then for any number m (<12) of independent panels reporting successively, N_t' follows (5.1). This allows us to use the 4-panel survey results in Table 1 to estimate ϕ^4 and ϕ_{12}, and (assuming $\phi > 0$) convert these to estimates of ϕ^3 and ϕ_{12}, the parameters of the model for the current 3-panel survey. This was done by finding ϕ^4 and ϕ_{12} to minimize the sum of squared deviations of the correlations from (5.1) with those of Table 1. (Lags 20 and 24 were dropped, and lag 16 given a weight of .5, due to the smaller number of correlation estimates that were averaged together at these higher lags.) The resulting estimates then produced $\hat{\phi}^3 = .685$, $\hat{\phi}_{12} = .723$ for Eating Places, and $\hat{\phi}^3 = .664$, $\hat{\phi}_{12} = .714$ for

Drinking Places. The resulting correlations for m = 4 from (5.1) are shown in Table 1, and may be compared to the averaged correlations.

If we make the further assumption that $\mathrm{Corr}(N'_t, N''_{t-1-k}) = \rho\, \mathrm{Corr}(N'_t, N'_{t-k})$ for all k, which seems reasonable given the high correlation for k = 0 (ρ), then (5.1) leads to the following bivariate model for (N'_t, N''_{t-1}):

$$(1-\phi^3 B^3)(1-\phi_{12}B^{12}) \begin{bmatrix} N'_t \\ N''_{t-1} \end{bmatrix} = \begin{bmatrix} \nu_{1t} \\ \nu_{2,t-1} \end{bmatrix} \qquad \mathrm{Var} \begin{bmatrix} \nu_{1t} \\ \nu_{2,t-1} \end{bmatrix} = \sigma_\nu^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \qquad (5.2)$$

with $\rho = \mathrm{Corr}(\nu_{1t}, \nu_{2,t-1}) = \mathrm{Corr}(N'_t, N''_{t-1})$. Estimates of $\mathrm{Corr}(N'_t, N''_{t-1})$ are regularly produced and were available for 1/82 - 12/86. Averaging these (with Fisher's transformation) produced $\hat{\rho}$ = .985 for Eating Places and $\hat{\rho}$ = .986 for Drinking Places.

We can now use (5.2) to derive a model for the sampling error of the linear form of the composite estimator, which is given by

$$Y'''_t = (1-\beta)Y'_t + \beta(Y'''_{t-1} + Y'_t - Y''_{t-1}) \qquad \text{(preliminary estimator)}$$

$$Y_{t-1} = (1-\alpha)Y''_{t-1} + \alpha Y'''_{t-1} \qquad \text{(final estimator)} \qquad (5.3)$$

(See Wolter 1979.) In the retail trade survey, vales of α = .8, β = .75 are used. It is easily seen that (5.3) also holds for the sampling errors, i.e. with Y relaced by N. We can use the resulting relations to derive the following equation for N_t in terms of N'_t and N''_{t-1}:

$$(1-.75B)N_t = .2\, N''_t - .75\, N''_{t-1} + .8\, N'_t \qquad (5.4)$$

Using (5.2) and (5.4) we then get

$$(1-.75B)(1-\phi^3 B^3)(1-\phi_{12}B^{12})\, N_t = .2\, \nu_{2t} - .75\, \nu_{2,t-1} + .8\, \nu_{1t} \qquad (5.5)$$

The right hand side is a first order moving average process whose parameters could be determined given estimates of σ_ν^2 and ρ. Thus, (5.5) would yield an ARMA model for N_t.

Rather than pursue this further, we shall instead make the rather strong assumption that a model of the same form holds for $\ln(U_t)$ in $\ln(Y_t) = \ln(S_t) + \ln(U_t)$, thus

$$(1-.75B)(1-\phi^3B^3)(1-\phi_{12}B^{12}) \ln(U_t) = (1-\eta B)\ c_t \qquad (5.6)$$

We do this because estimates of sampling variance for these series are highly dependent on the level of the series; estimates of relative variance are much more stable over time. We also assume we can use estimates of relative variance and of ρ in determining η and σ_c^2. Estimates Y_t', Y_{t-1}'' and $\hat{Var}(N_t')$, $\hat{Var}(N_{t-1}'')$ were available for 1/82 - 12/86. The resulting relative variance estimates were used in the spirit of maximum likelihood estimation for the lognormal distribution -- taking the average of the logs of the relative variance estimates, adding one half of the sample variance of the logged estimates to this, and exponentiating the results. This was done separately for Rel $Var(Y_t')$ and Rel $Var(Y_{t-1}'')$, and these two results were then averaged, producing a common relative variance estimate that is constant over time. The results are shown in Table 2 under Horvitz-Thompson. Using these and the $\hat{\rho}$'s given earlier, one can solve for η and σ_c^2 for the right side of (5.6). The resulting sampling error models are

$$(1-.75B)(1-.685B^3)(1-.723B^{12}) \ln(U_t) = (1+.130B)c_t \qquad (5.7a)$$

(Eating Places) $\qquad \hat{\sigma}_c^2 = 1.948 \times 10^{-5}$

$$(1-.75B)(1-.664B^3)(1-.714B^{12}) \ln(U_t) = (1+.134B)c_t \qquad (5.7b)$$

(Drinking Places) $\qquad \hat{\sigma}_c^2 = 9.301 \times 10^{-5}$

One can use the method of McLeod (1975,1977) to solve for $Var(\ln(U_t))$ in these models, which is an estimate of the relative variance of the final composite estimator. The results are shown in Table 2. The corresponding coefficients of variation, .025 for Eating Places and .052 for Drinking Places, are quite close to published estimates that are obtained more directly.

Time Series Modeling and Signal Extraction

Figures 1a,b show plots of the time series of final composite estimates Y_t for Eating Places and for Drinking Places, respectively. To develop models for S_t we shall begin by modeling the Y_t series directly. Both series show trends and strong seasonality, with the magnitude of the seasonal fluctuations larger the higher the level of the series. This suggests taking logarithms and the need for differencing; both are typical for economic time series. Examination of sample autocorrelations for $\ln(Y_t)$ and its differences suggested the difference operator $(1-B)(1-B^{12})$ for both series. Retail trade series are known to contain trading-day variation. This can be modeled by including seven regression variables in the model: X_{1t} = number of Mondays in month t, ... , X_{7t} = number of Sundays in month t. Following Bell and Hillmer (1983), a more convenient reparameterization is obtained by using instead the variables T_{1t} = $X_{1t} - X_{7t}$ (number of Mondays - number of Sundays), ... , $T_{6t} = X_{6t} - X_{7t}$ (number of Saturdays - number of Sundays), $T_{7t} = \sum_1^7 X_{it}$ (length of month t). To identify the ARMA structures, the ACFs and PACFs of the residuals from regressions of $(1-B)(1-B^{12})$ $\ln(Y_t)$ on $(1-B)(1-B^{12})$ T_{it} $i = 1,\ldots,7$ were examined. This suggested an ARIMA $(0,1,2)(0,1,1)_{12}$ model for Eating Places, and an ARIMA $(0,1,3)(0,1,1)_{12}$ model for Drinking Places. The resulting estimated models were

$$(1-B)(1-B^{12})\ [\ln(Y_t) - \sum_i \beta_i T_{it}] = (1-.22B-.28B^2)(1-.77B^{12})\ a_t$$

(Eating Places) $\hat{\sigma}_a^2 = .000261$ (5.8a)

$$(1\text{-}B)(1\text{-}B^{12})\ [\ln(Y_t) - \sum_i \beta_i T_{it}] = (1\text{-}.23B\text{-}.17B^2\text{+}.004B^3)(1\text{-}.59B^{12})\ a_t$$

(Drinking Places) $\hat{\sigma}_a^2 = .000577$ (5.8b)

For brevity, we omit the estimates of the trading-day parameters. While neither the lag 2 nor lag 3 moving average parameters in (5.8b) is statistically significant, we shall retain them since we shall only use (5.8a,b) as starting points for modeling $\ln(S_t)$ for both series.

Taking models of the form of (5.8a,b) for $\ln(S_t)$ with models (5.7a,b) for $\ln(U_t)$, the parameters of the models for $\ln(S_t)$ were reestimated. For both series the seasonal moving average parameters were reestimated to be 1, implying deterministic seasonality that can be modeled by cancelling a $(1\text{-}B^{12})$ from both sides of the model and instead including a regression function of the form $\sum_{1}^{11} \gamma_i M_{it}$, where M_{1t} is 1 in January, -1 in December, and 0 otherwise, ... , M_{11t} is 1 in November, -1 in December, and 0 otherwise. Estimation of the resulting models produced the following:

$$(1\text{-}B)\ [\ln(S_t) - \sum_i \hat{\beta}_i T_{it} - \sum_i \hat{\gamma}_i M_{it}] = .00769 + (1\text{-}.26B\text{-}.28B^2)\ b_t$$

(Eating Places) $\hat{\sigma}_b^2 = .000160$ (5.9a)

$$(1\text{-}B)\ [\ln(S_t) - \sum_i \hat{\beta}_i T_{it} - \sum_i \hat{\gamma}_i M_{it}] = .00330 + (1\text{-}.18B\text{-}.36B^3)\ b_t$$

(Drinking Places) $\hat{\sigma}_b^2 = .000261$ (5.9b)

The lag 2 moving average parameter for drinking places was insignificant and so was dropped. We again omit the estimates of the trading-day parameters (which did not change much from the estimation of (5.8)) and also of the seasonal parameters. The trend constants were included when the seasonal difference was dropped. Examination of standardized residuals produced by the Kalman filter, and of their autocorrelations, suggested no major inadequacies with the models for either series.

The estimated models, (5.7a,b) with (5.9a,b), were used to produce signal extraction estimates of $\ln(S_t)$, which were then exponentiated to produce estimates of S_t. The results are shown in Figures 2a,b for the series with the estimated seasonal and trading-day effects removed. Notice that signal extraction makes only slight differences in the estimates for Eating Places, which contained relatively little sampling error, but it makes a considerable difference in the estimates for Drinking Places, which contained much more sampling error. Signal extraction variances for $\ln(S_t)$ were also produced; these are relative variances for the estimates of S_t. These results are summarized in Table 2. They show signal extraction produces about an 8% improvement in CV over the final composite estimates for Eating Places, and nearly a 20% improvement in CV for Drinking Places. These results are somewhat optimistic, since they assume the true component models are those that were estimated.

Conclusions

The results for Eating and Drinking Places depend on several assumptions and approximations, and should be regarded as preliminary. More confidence could be placed in results that used more recent data for estimates of sampling error correlations. The examples are intended primarily to illustrate the application of time series models that allow for sampling error, and the potential for signal extraction to improve survey estimates. An interesting result in these examples is that when sampling error is allowed for, the seasonality in both series appears to be deteministic (a fixed pattern over time), rather than the stochastically varying seasonality implicit with models (5.8a,b). If similar results could be found for other retail sales series or other economic time series, this could have important implications for seasonal adjustment.

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Likelihood Principle," 2nd International Symposium on Information Theory, eds. B.N. Petrov and F. Czaki. Budapest: Akademia Kiado, 267-287.

Anderson, B.D.O. and Moore, J. B. (1979), Optimal Filtering, Englewood Cliffs: Prentice-Hall.

Ansley, C.F. and Kohn, R. (1985), "Estimation, Filtering, and Smoothing in State-Space Models with Incompletely Specified Initial Conditions," Annals of Statistics, 13, 1286-1316.

Bell, W. R. (1984), "Signal Extraction for Nonstationary Time Series," Annals of Statistics, 12, 646-664.

Bell, W. R. and Hillmer, S. C. (1983), "Modeling Time Series with Calendar Variation," Journal of the American Statistical Association, 78, 526-534.

_______ (1987), "Time Series Methods for Survey Estimation," SRD Research Report No. 87/20, Bureau of the Census.

_______ (1988), "A Matrix Approach to Likelihood Evaluation and Signal Extraction for ARIMA Component Time Series Models," SRD Research Report No. 88/22, Bureau of the Census.

_______ (1989), "Initializing the Kalman Filter in the Non-stationary Case: With Application to Signal Extraction," SRD Research Report No. 89/ , Bureau of the Census.

Binder, D. A. and Dick, J. P. (1986), "Modelling and Estimation for Repeated Surveys," Statistics Canada Technical Report, Social Survey Methods Division.

Box, G.E.P. and Jenkins, G. M. (1976), Time Series Analysis: Forecasting and Control, San Francisco: Holden Day.

Fuller, W. A. and Isaki, C. T. (1981), "Survey Design Under Superpopulation Models," in Current Topics in Survey Sampling, ed. D. Krewski, R. Platek, and J.N.K. Rao, New York: Academic Press, 199-226.

Hausman, J. A. and Watson, M. W. (1985), "Errors in Variables and Seasonal Adjustment Procedures," Journal of the American Statistical Association, 80, 531-540.

Hillmer, S. C., Bell, W. R., and Tiao, G. C. (1983), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series," in Applied Time Series Analysis of Economic Data, ed. Arnold Zellner, U. S. Department of Commerce, Bureau of the Census.

Isaki, C. T., Wolter, K.M., Sturdevant, T. R., Monsour, N. J., and Trager, M. L. (1976), "Sample Redesign of the Census Bureau's Monthly Business Surveys," Proceedings of the American Statistical Association, Business and Economic Statistics Section, 90-98.

Jones, R. G. (1980), "Best Linear Unbiased Estimators for Repeated Surveys," Journal of the Royal Statistical Society, Series B, 42, 221-226.

Jones, R. H. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations," Technometrics, 22, 389-395.

_______ (1985), "Time Series Analysis With Unequally Spaced Data," Handbook of Statistics, Vol. 5, ed. E.J. Hannan, P.R. Krishnaiah, and M.M. Rao, Elsevier Science, 157-177.

Kohn, R. and Ansley, C. F. (1986), "Estimation, Prediction, and Interpolation for ARIMA Models With Missing Data," Journal of the American Statistical Association, 81, 751-761.

_______ (1987), "Signal Extraction for Finite Nonstationary Time Series," Biometrika, 74, 411-421.

McLeod, I. (1975), "Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series," Applied Statistics, 24, 255-256.

_______ (1977), "Correction to Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Timer Series," Applied Statistics, 26, 194.

Miazaki, E. S. (1985), "Estimation for Time Series Subject to the Error of Rotation Sampling," unpublished Ph.D. thesis, Department of Statistics, Iowa State University.

Nelson, C. R. and Kang, H. (1984), "Pitfalls in the Use of Time as an Explanatory Variable in Regression," Journal of Business and Economic Statistics, 2, 73-82.

Otto, M.C., Bell, W.R., and Burman, J.P. (1987), "An Iterative GLS Approach to Maximum Likelihood Estimation of Regression Models with ARIMA Errors," SRD Research Report No. 87/34, Bureau of the Census.

Rao, J.N.K., Srinath, K. P., and Quenneville, B. (1986), "Optimal Estimation of Level and Change Using Current Preliminary Data," paper presented at the International Symposium on Panel Surveys, Washington, D.C., November, 1986.

Scott, A. J. and Smith, T.M.F. (1974), "Analysis of Repeated Surveys Using Time Series Methods," Journal of the American Statistical Association, 69, 674-678.

Scott, A. J., Smith, T.M.F., and Jones, R. G. (1977), "The Application of Time Series Methods to the Analysis of Repeated Surveys," International Statistical Review, 45, 13-28.

Train, G., Cahoon, L., and Makens, P. (1978), "The Current Population Survey Variances, Inter-Relationships, and Design Effects," American Statistical Association, Proceedings of the Survey Research Methods Section, 443-448.

Whittle, P. (1963), Prediction and Regulation by Linear Least-Square Methods, Princeton: Van Nostrand.

Wolter, K. M. (1979), "Composite Estimation in Finite Populations," Journal of the American Statistical Association, 74, 604-613.

______ (1985), Introduction to Variance Estimation, New York: Springer-Verlag.

Wolter, K. M., Isaki, C. T., Sturdevant, T. R., Monsour, N. J., and Mayes, F. M. (1976), "Sample Selection and Estimation Aspects of the Census Bureau's Monthly Business Surveys," Proceedings of the American Statistical Association, Business and Economic Statistics Section, 99-109.

Table 1

Sampling Error Correlations for Horvitz-Thompson Estimates

	Lag					
	4	8	12	16	20	24
Eating Places						
Averaged[1]	.72	.71	.79	.63	.65	.77
From (5.1)[2]	.75	.69	.81	.60	.53	.61
Drinking Places						
Averaged[1]	.70	.67	.78	.60	.60	.61
From (5.1)[2]	.72	.66	.80	.56	.50	.59
Number of Correlations Averaged	23	19	15	11	7	3
Weights Used in Determining $\hat{\phi}$'s	1	1	1	.5	0	0

[1]Raw estimates of Corr(N'_t, N'_j) and Corr(N''_{t-1}, N''_{j-1}) were available for all pairs of months from January, 1973 through March, 1975. Averages of the correlations for the lags shown were taken after applying Fisher's transformation, and the results then transformed back.

[2]Correlations are shown from model (5.1) for m=4 with parameters $\hat{\phi}^4$ = .604, $\hat{\phi}_{12}$ = .723 (Eating Places) and $\hat{\phi}^4$ = .580, $\hat{\phi}_{12}$ = .714 (Drinking Places). These parameter values were determined to minimize the weighted sum of squared deviations of the correlations from model (5.1) and the averaged correlations using the weights shown. Lags 20 and 24 were not used because of the small number of correlation estimates available at these lags.

Table 2

Relative Variances (Rel Var) and Coefficients of Variation (CV)[1]
for Retail Sales Estimates

	Horvitz-Thompson		Final Composite[2]		Signal Extraction[3]	
	Rel Var	CV	Rel Var	CV	Rel Var	CV
Eating Places	.00180	.042	.000638	.025	.000508	.023
Drinking Places	.00776	.088	.00267	.052	.00178	.042

[1] CV = $(\text{Rel Var})^{.5}$.

[2] The values for the final composite estimator are obtained using models (5.7a,b).

[3] The values for signal extraction actually vary over time, being highest at the end of the series and lowest in the middle. We show the midpoint of the range of values. CV's range from .022 - .023 for Eating Places (Rel Var from .000483 - .000532) and from .041 - .0435 for Drinking Places (Rel Var from .00167 - .00189).

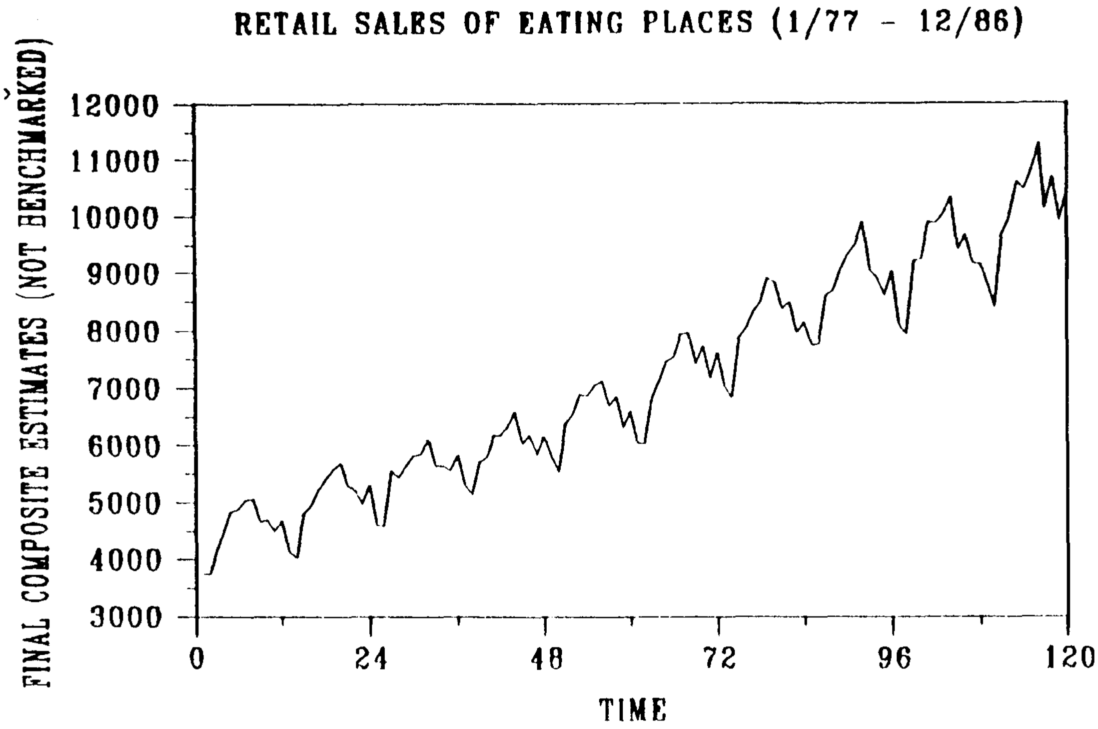

Figure 1a.

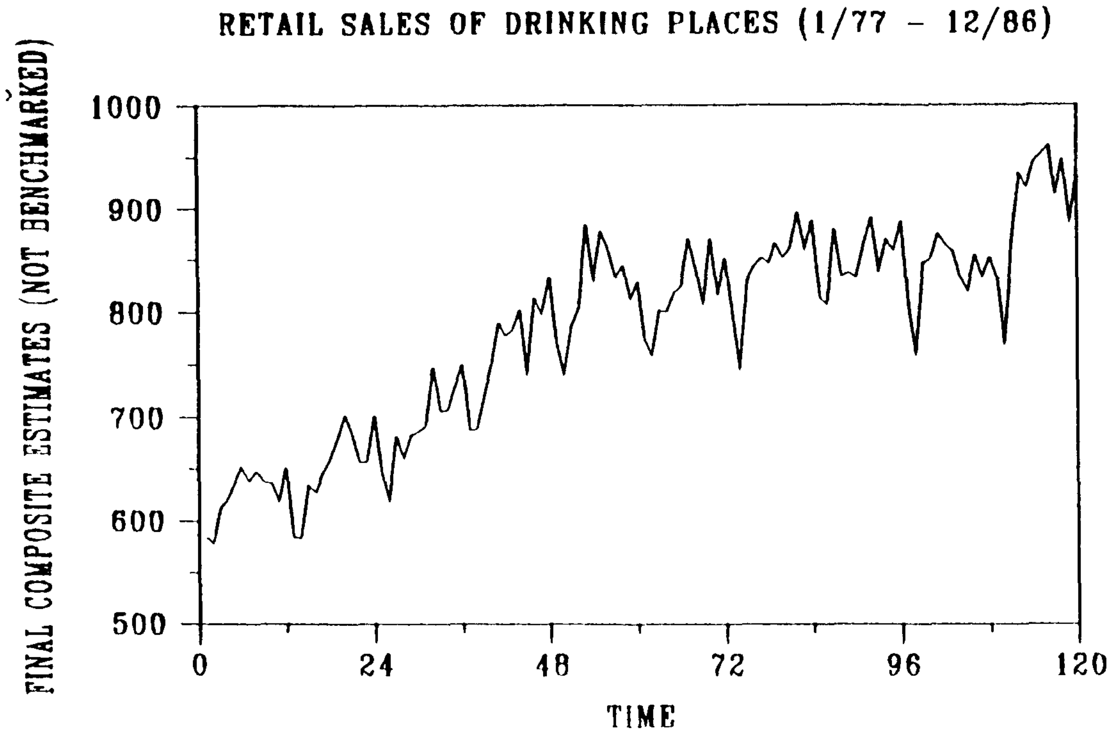

Figure 1b.

EATING PLACES -- FINAL COMPOSITE AND SIGNAL EXTRACTION ESTIMATES

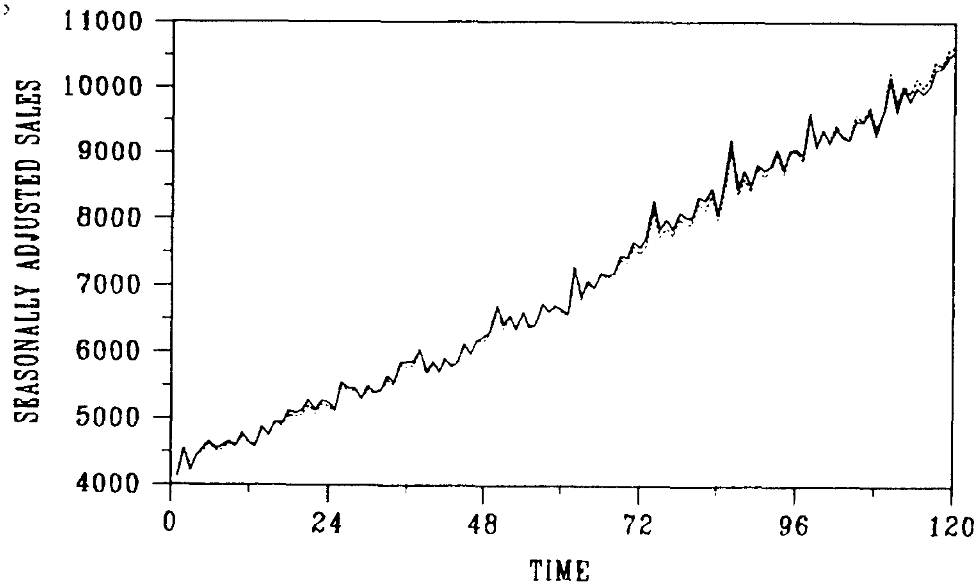

Figure 2a.

DRINKING PLACES -- FINAL COMPOSITE AND SIGNAL EXTRACTION ESTIMATES

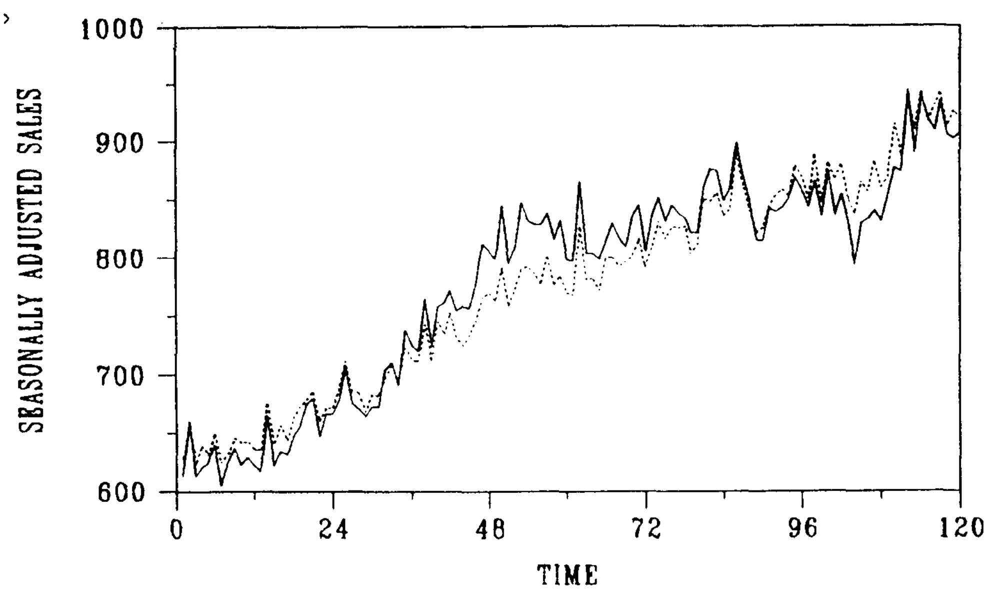

Figure 2b.