Adjustment of Provisional Mortality Series: The Dynamic Linear Model With Structured Measurement Errors

Author(s): Robert H. Shumway and Myron J. Katzoff
Source: *Journal of the American Statistical Association*, Vol. 86, No. 415 (Sep., 1991), pp. 611-617
Published by: Taylor & Francis, Ltd. on behalf of the American Statistical Association
Stable URL: https://www.jstor.org/stable/2290389
Accessed: 30-07-2022 00:53 UTC

Adjustment of Provisional Mortality Series: The Dynamic Linear Model With Structured Measurement Errors

ROBERT H. SHUMWAY and MYRON J. KATZOFF*

We consider the problem of adjusting provisional time series using a bivariate structural model with correlated measurement errors. Maximum likelihood estimators and a minimum mean squared error adjustment procedure are derived for a provisional and final series containing common trend and seasonal components. The model also includes measurement errors common to both series and errors that are specific to the provisional series. We illustrate the technique by using provisional data to forecast ischemic heart disease mortality.

KEY WORDS: EM algorithm; Forecasting; Mortality surveillance; State-space model; Structural models; Trends.

1. INTRODUCTION

There are many situations that arise in time series analysis where additional outside information may be available for use in forecasting one or more of the given series. These situations occur often in economics [see, for example, Pankratz (1989)], and methods for using such "benchmarks" have been developed by Hillmer and Trabelsi (1987) and Trabelsi and Hillmer (1989). Harvey (1984, 1988) gave examples using structural models in state-space form.

The application considered here involves two systems for collecting monthly cause-specific mortality rates used by the National Center for Health Statistics. One system provides provisional estimates based on a 10% sample of death certificates received during a reference month. A second system provides "final" mortality rates, which are based on a tabulation of all of the death certificates filed in state offices during a given year by cause and month of occurrence. The final rates are not obtained until about 20 months after the provisional rates have already been available.

Provisional death rates may differ from final death rates for practical reasons such as simple sampling variability or reporting errors. Given that the final values lag many months behind the provisional ones, we would like to use the provisional estimates to arrive at the best early estimates for the final rates. The adjustment method, which simply substitutes the provisional rates for the final rates during the missing months, may be inadequate for some or all series. More effective methods may be available for deriving estimators for the final series using the "benchmark" provisional series.

The solution that we will propose here uses provisional series as well as the past of the final series to produce the optimal estimator for the final series over a time span where the final series has not been observed. The optimal estimator for the final series is derived as the conditional expectation given the past of the final series and the whole history of the provisional series. The conditional expectation is shown to be a weighted average of a smooth extension of the final series and the provisional series. We derive the mean squared error of the estimators.

The model that we develop for the provisional and final series is a combination of the structural model used by Kitagawa and Gersch (1984) [see, also, Akaike (1980); Harrison and Stevens (1976); Harvey and Todd (1983)] and a measurement error component model that is new. We assume that both series can be modeled as sums of a common smooth trend function, a seasonal component, and specially structured measurement errors. The error model is unique in that it proposes that the usual irregular error component be composed of (a) an error that the two series have in common and (b) an error that is unique to the provisional series.

The overall model that results is conveniently expressed in tate-psace form. We consider parameter estimation and show that the EM algorithm reduces the maximum likelihood procedure to simple iterative regression computations. We discuss the possibility of using the incomplete-data log-likelihood for model selection. Finally, we apply the procedure to the problem of modeling and adjusting death rates from ischemic heart disease for males in two age groups.

2. STRUCTURAL MODELS WITH CORRELATED MEASUREMENT ERRORS

We begin by examining a set of mortality series for ischemic heart disease (IHD) for males in two age groups (25–34, 65–74) measured over 60 months (1979–1983) for the final series and over 84 months (1979–1985) for the provisional series. An analysis of four other age groups appears in Shumway (1989). Figure 1 shows the observed series, with trend functions estimated by the method of this article shown as the smoother lines.

To develop models for the final and provisional series, we note first that they are, by definition, attempting to measure the same quantity, namely, the final monthly mortality series. Furthermore, the provisional series must by its very nature be a noisier version of the final series. If y_{t1} denotes the final series at time t and y_{t2} denotes the provisional measurement at time t, we argue that the model

$$y_{t2} = y_{t1} + e_{t2} \tag{1}$$

* Robert H. Shumway is Professor, Division of Statistics, University of California, Davis, CA 95616. Myron J. Katzoff is Mathematical Statistician, Office of Research and Methodology, National Center for Health Statistics, Hyattsville, MD 20782. Shumway was supported in part by the Office of Research Methodology, National Center for Health Statistics, Hyattsville, MD 20782. The authors are pleased to acknowledge valuable comments and suggestions from an associate editor and two referees.

Journal of the American Statistical Association
September 1991, Vol. 86, No. 415, Applications and Case Studies

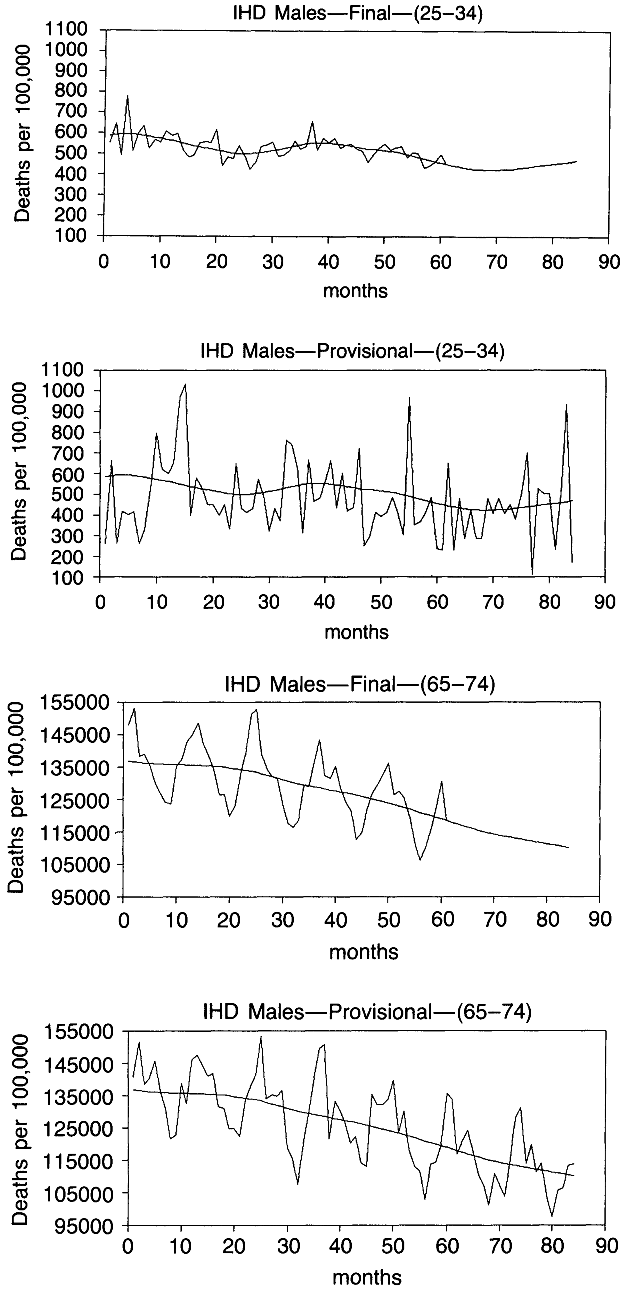

Figure 1. Final and Provisional Monthly Ischemic Heart Disease Rates for Males in Two Age Groups (1979–1984) With Final Series Missing for Last 24 Months. The smoothest lines are the fitted trend terms. Annualized deaths are as defined in the Monthly Vital Statistics Reports, National Center for Health Statistics.

is plausible as a start, where the independent irregular components e_{t2} represent the extra error in the provisional measurements y_{t2}. Such a model has been proposed for economic data revisions by Harvey (1988, sec. 6.4.4). Further justification for the model in (1) is provided by considering the difference series $y_{t2} - y_{t1}$ for the age group 65–74, which should be an independent white noise series if (1) is to hold. The autocorrelation function for the difference series is consistent with the white noise assumption; none of the values exceed two standard errors ($\approx \pm .22$).

The series also tend to confirm visually an impression that the idea of a common nonstationary smooth trend is reasonable. The trend is not purely linear but tends to have some departures characteristic of higher-order variation. There is also some tentative indication that some of the more irregular variation may be regarded as belonging to both of the series. We denote the hypothesized common error appearing on both the provisional and final series as e_{t1}. Now, regarding the time-varying component μ_t as an unobserved component satisfying

$$\nabla^2 \mu_t = w_t, \tag{2}$$

where $\nabla^2 \mu_t = \mu_t - 2\mu_{t-1} + \mu_{t-2}$ is the second difference, has provided reasonable models for trend in general economic series [see Akaike (1980); Gersch and Kitagawa (1983); Kitagawa and Gersch (1984); Wecker and Ansley (1983)]. The assumption that both series can be modeled in terms of a common trend that can be made stationary by differencing leads to a form of the co-integrated models of Engle and Granger (1987). Alternatively, fractionally differenced trend models were used in Carlin and Dempster (1989). The error terms w_t are assumed to be iid with zero mean and variance σ_w^2. The foregoing remarks allow us to formulate the following model.

Model 1: Stochastic Trend With Errors in Measurement. In this model take

$$y_{t1} = \mu_t + e_{t1}$$

and

$$y_{t2} = \mu_t + e_{t1} + e_{t2}. \tag{3}$$

The common errors e_{t1} and the errors specific to the provisional series, e_{t2}, are assumed to be mutually independent as well as being independent of w_t. The parameters in this formulation are the variances σ_1^2 and σ_2^2 for the errors e_{t1} and e_{t2} along with the variance σ_w^2 for the trend μ_t.

To check that the foregoing error structure is reasonable, we note that the difference series $e_{t2} = y_{t2} - y_{t1}$ should not be correlated with y_{t1}. This was checked for both age groups; the absolute values of the cross-correlations at all lags were less than .07.

The series representing the older 65–74 age group in Figure 1 also shows a relatively periodic component corresponding to a seasonal rise and fall in death rates. To suggest a model for this age group, the series was first linearly detrended and then autocorrelation and partial autocorrelation functions (ACF's and PACF's) were computed. The results show a relatively periodic behavior appearing in the ACF's of both the provisional and final series. The lack of repeating isolated peaks in the ACF's at multiples of the seasonal period 12 provides an argument for not using a seasonal difference as, for example, in Kitagawa and Gersch (1984). We prefer to focus on nonzero values in the PACF's of both series that can be eliminated by fitting higher order autoregressive components. The second-order autoregression is the lowest order that can produce a peak in the spectrum at the seasonal frequency matching that observed in the mortality data. The second-order autoregression also fits other age groups (35–44, 45–54, 55–64, and 75–84) rather consistently (see Shumway 1989). Hence it is used to emulate the seasonal in Model 2.

Model 2: Trend Autoregressive-Seasonal Model With Measurement Error. Here, we have

$$y_{t1} = \mu_t + z_t + e_{t1},$$
$$y_{t2} = \mu_t + z_t + e_{t1} + e_{t2}, \tag{4}$$

where μ_t is as before and z_t is a pth-order autoregression defined by

$$z_t = \sum_{k=1}^{p} \varphi_k z_{t-k} + a_t, \tag{5}$$

where σ_a^2 is the variance of the error terms a_t, which are assumed to be mutually independent as well as independent of the other errors in the model. The parameters in this model are those involved in the trend model plus the autoregressive parameters $\varphi_1, \varphi_1, \ldots, \varphi_p$ and the autoregressive error variance σ_a^2.

It is convenient to write both Models 1 and 2 in state–space form as, for example, in Kitagawa and Gersch (1984) or Shumway (1988, p. 175) by noting that the bivariate observation vector $\mathbf{y}_t = (y_{t1}, y_{t2})'$ in the trend autoregressive-seasonal model can be written as

$$\mathbf{y}_t = M_t x_t + \mathbf{v}_t, \tag{6}$$

where $\mathbf{v}_t = (v_{t1}, v_{t2})$ with $v_{t1} = e_{t1}$ and $v_{t2} = e_{t1} + e_{t2}$. The state equation

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t \tag{7}$$

has the $(p + 2) \times 1$ state vector defined as $x_t' = (\mu_t, \mu_{t-1}, z_t, z_{t-1}, \ldots, z_{t-p+1})$. In the observation equation (6) we identify the measurement matrix $M_t' = (\mathbf{m}_t, \mathbf{m}_t)$ with $\mathbf{m}_t' = (1, 0, 1, 0, \ldots, 0)$ and the measurement error covariance matrix

$$R = E(\mathbf{v}_t \mathbf{v}_t') = \begin{pmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 \end{pmatrix}, \tag{8}$$

which illustrates the correlated structure of the measurement errors. The $(p + 2) \times (p + 2)$ transition matrix Φ is of the form

$$\Phi = \begin{pmatrix} \Phi_{11} & \Phi_{12} \\ \Phi_{21} & \Phi_{22} \end{pmatrix}, \tag{9}$$

where

$$\Phi_{11} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix},$$

$$\Phi_{12} = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \end{pmatrix},$$

$$\Phi_{21} = \Phi_{12}',$$

$$\Phi_{22} = \begin{bmatrix} \varphi_1 & \varphi_2 & \ldots & \ldots & \ldots & \varphi_p \\ 1 & 0 & \ldots & \ldots & \ldots & 0 \\ 0 & 1 & 0 & \ldots & \ldots & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{bmatrix}, \tag{10}$$

and $\mathbf{w}_t' = (w_t, 0, a_t, 0, 0, \ldots, 0)$ with covariance matrix Q, where $q_{11} = \sigma_w^2$, $q_{33} = \sigma_a^2$, and the remaining elements of the matrix are 0.

The Kalman smoothed values

$$x_t^T = E(x_t \mid \mathbf{y}_1, \ldots, \mathbf{y}_T) \tag{11}$$

and their covariances are computed by recursions given in a number of references [see Shumway (1988, p. 177)], with the revisions necessary when y_{t1} is not observed due to Shumway and Stoffer (1982) [see also Shumway (1988, p. 182)]. The modification amounts to setting $r_{12} = r_{21} = 0$ in the covariance matrix (8) and the first row of M_t in (6) at 0 when y_{t1} is not observed. Note that parts of the vectors $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_T$ may be missing when the final series is unobserved. The exact definition of the conditioning set Y that we will use is given by Equation (12). The modifications necessary for handling the likelihood function are given in the Appendix.

3. OPTIMAL ADJUSTMENT

The fact that the model can be written in state–space form allows us to obtain fairly easily the minimum mean squared estimator for the final series y_{t1} given the observed data

$$Y = \{y_{t1}; y_{t2}; t = 1, 2, \ldots, T_1; y_{t2}; t = T_1 + 1, \ldots, T\}, \tag{12}$$

which includes the present and past values of the provisional and final series along with the future values of the provisional series. The optimum estimator for the final series y_{t1} will be its expectation conditioned on Y, which we write as the iterated expectation

$$\hat{y}_{t1} = E[y_{t1} \mid Y]$$
$$= E[E(y_{t1} \mid \mathbf{x}_t, Y) \mid Y], \tag{13}$$

$t > T_1$, where the outer expectation is with respect to $\mathbf{x}_t$. Note that the covariance matrix structure (8) gives

$$E(y_{t1} \mid \mathbf{x}_t, Y) = \mathbf{m}_t'\mathbf{x}_t + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}(y_{t2} - \mathbf{m}_t'\mathbf{x}_t). \tag{14}$$

Then, using Equation (13) gives the estimator

$$\hat{y}_{t1} = \omega \mathbf{m}_t' \mathbf{x}_t^T + (1 - \omega) y_{t2}, \tag{15}$$

where

$$\omega = \sigma_2^2 / (\sigma_1^2 + \sigma_2^2) \tag{16}$$

and $(1 - \omega)$ are the weights to be applied to the smoothed extrapolation (11) and the observed provisional value, respectively. A similar formula was given by Harvey (1988) for a different model. His estimators were derived by incorporating the unobserved values into the state vector.

It is interesting that the optimal combination involves the two irregular variances σ_1^2 and σ_2^2 corresponding to the common component e_{t1} and the provisional series component e_{t2}. These two variances determine the relative importance of the smoothed extension of the final series and the provisional values in determining the adjusted final estimators. These weights could be assigned arbitrarily based on prior knowledge of the sampling errors involved, but we prefer to determine them from the prior provisional and final data using maximum likelihood estimation, as detailed in the Appendix.

To derive the variance, note that

$$E(y_{t1}^2 \mid \mathbf{x}_t, Y) = \mathrm{var}(y_{t1} \mid \mathbf{x}_t, Y) + E^2(y_{t1} \mid \mathbf{x}_t, Y)$$
$$= \omega\sigma_1^2 + (\omega\mathbf{m}_t'\mathbf{x}_t + (1 - \omega)y_{t2})^2.$$

Then, using (11) and the definition of P_{tt}^T given in (A.3) in the Appendix leads to

$$E[E(y_{t1}^2 \mid \mathbf{x}_t, Y)] = \hat{y}_{t1}^2 + \omega\sigma_1^2 + \omega^2\mathbf{m}_t'P_{tt}^T\mathbf{m}_t.$$

This leads to expressing the conditional variance as

$$\mathrm{var}(y_{t1} \mid Y) = \omega\sigma_1^2 + \omega^2\mathbf{m}_t'P_{tt}^T\mathbf{m}_t, \qquad (17)$$

which can be used to set probability intervals for the forecast of the final series and to evaluate the forecasting method.

The foregoing procedure is similar in philosophy to that of Trabelsi and Hillmer (1989) but differs significantly with respect to the underlying model and the conditioning method. Trabelsi and Hillmer considered a method for combining pure forecasts of the final series from a univariate autoregressive integrated moving average (ARIMA) model with the future benchmark or provisional values using a joint linear model involving the final values. The forecast errors and the future provisional values are assumed to be uncorrelated in their approach, and one must estimate the covariance matrix of the benchmark errors.

There is still the problem of estimating the parameters in the two models. Details for estimating the parameters by maximum likelihood using the EM algorithm of Dempster, Laird, and Rubin (1977), as adapted to the dynamic linear model by Shumway and Stoffer (1982), are given in the Appendix. Our preference for the EM algorithm in this situation stems from the ease with which it is applied, with successive steps in the iterations [see Eqs. (A.8)–(A.12) in the Appendix] involving only simple variance computations and regression updates. Furthermore, successive steps are guaranteed to increase the likelihood and one avoids the sometimes divergent corrections resulting from other algorithms such as Newton–Raphson. Gersch and Kitagawa (1983) and Kitagawa and Gersch (1984) used a numerical search over a grid spanned by the variance parameters and made Newton–Raphson corrections to the autoregressive coefficients. If there are more than two variance parameters, the simple grid search procedure becomes more complicated.

The log-likelihood function can also be used for model selection. Two models can be compared using the likelihood ratio criterion, which amounts to the difference between the maximized log-likelihoods under two different models; for example, Models 1 and 2 should be compared. Since both models involve long memory trend components (the transition matrix has two unit roots), the usual asymptotic likelihood theory [see Ljung and Caines (1979)] will not necessarily guarantee that the difference between the two maximized log-likelihoods is proportional to a chi-squared random variable with $p + 1$ (the difference between the number of parameters in the full and reduced model) df. This may, however, still be a reasonable approximation, and we shall use it in the next section. Another possibility is to use Akaike's information theory criterion (AIC) (Akaike 1973),

$$\mathrm{AIC} = -2 \text{ log-likelihood} + 2 \text{ (number of parameters)},$$

as in Kitagawa and Gersch (1984). The model to be preferred is the one that minimizes AIC.

4. ADJUSTING MORTALITY SERIES: AN APPLICATION

In this section we consider applying the approach outlined in the previous sections to the problem of modeling and constructing optimal estimators for the final months of the mortality series shown in Figure 1. Note that the final series is given for only the first $T_1 = 60$ months, whereas the provisional series is assumed to be known for an additional 24 months ($T = 84$ months in this example). One would like to estimate values for the final 24 months.

We look first at a preliminary model identification procedure, which selects between Models 1 and 2, specified as "trend" and "trend autoregressive-seasonal" models, respectively. We estimate the parameters of the two models by maximum likelihood and tentatively select one for each data set using the likelihood ratio criterion and AIC. Then, using the estimated parameters, we construct optimal estimators for the final series using Equations (15) and (16). We compare the performance of the models with the performance of two alternate estimators, one derived from the provisional series alone and a second one using Box–Jenkins transfer function modeling of the provisional and the final series.

Table 1 shows the maximum likelihood estimators under Models 1 and 2 for the two age groups. Both data series were scaled by dividing by 1,000. The variance components σ_1^2 and σ_2^2 show that the error specific to the provi-

Table 1. Parameter Estimates for Models 1 and 2 in Two Age Groups

Parameters	Age (25–34) Model 1	Age (25–34) Model 2	Age (65–74) Model 1	Age (65–74) Model 2
σ_1^2	2.13×10^{-5}	1.40×10^{-5}	16.96	.24
σ_2^2	2.43×10^{-3}	4.60×10^{-3}	7.27	6.98
σ_w^2	2.25×10^{-4}	1.70×10^{-4}	24.64	7.66
σ_a^2	—	3.30×10^{-2}		24.66
ϕ_1	—	−.108		1.550
ϕ_2	—	.543		−.825
ω	.931	.949	.772	.763
−2 ln L	−478.41	−481.32	677.15	642.46

sional series is dominant in the 25–34 age group, whereas the common error is more important in the older age group. The weight functions reflect this with very high weights (.931, .949) on the smoothed extension and low weights (.069, .051) on the relatively unreliable provisional series in the 25–34 age group. The situation reverses in the older age group where the smoothed extensions receive somewhat lower values (.772 and .763). The initial mean of the trend component was estimated by maximum likelihood with the initial variance fixed at a reasonable prior value. The initial means of the autoregressive series were set to 0 [$E(z_0) = 0$].

The log-likelihoods indicate that the pure trend model ought to be preferred in the 25–34 age group. This follows either by comparing the log-likelihoods using the likelihood ratio criterion ($-478.41 + 481.32 = 2.91$) with a chi-squared random variable with 3 df or by comparing the AIC values of -472.41 and -469.32 for Models 1 and 2. In the older age group, the situation reverses as we compare the log-likelihood difference 34.69 with a chi-squared random variable with 3 df or the AIC values of 683.15 and 654.46 for Models 1 and 2. Hence the autoregressive seasonal term belongs in the model for the older age group. This is not a surprise since a periodic component is clearly present in the second series of Figure 1.

An interesting phenomenon arises when trend is estimated under Models 1 and 2 in the 65–74 age group. In Figure 2 we show in the top frame the difference between the trend estimators under Models 1 and 2. When the pure trend model is fitted, it adapts to the periodic behavior. Hence the smoothed extrapolation would not have been too far off even if the simpler model had been chosen in this case. The practical impact is that the stochastic trend is somewhat robust to misspecification of the model. The lower frame shows the estimated autoregressive component under Model 2; it is a quite accurate version of the apparent seasonal component as it appears in Figure 1.

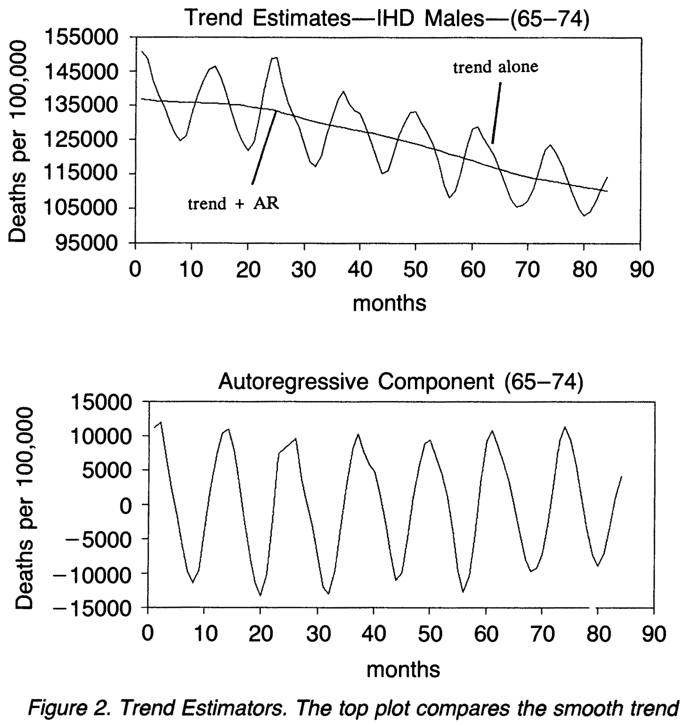

Figure 2. Trend Estimators. The top plot compares the smooth trend estimator when the autoregressive component is in the model with the trend estimator when the model is trend alone. The bottom plot shows the estimated autoregressive component when the time-varying trend is in the model.

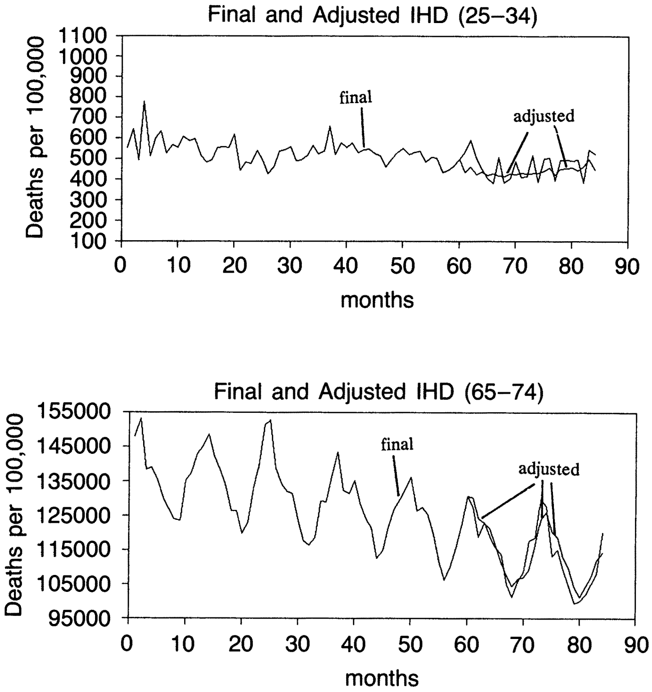

Figure 3. Final and Adjusted Ischemic Heart Disease Compared for Two Age Groups.

Finally, the adjusted final mortality values are shown in Figure 3, where they are compared with the known final values. They are quite accurate, particularly in the 65–74 age group. To compare the performance of these models with several alternative formulations, we also fitted a Box–Jenkins transfer function model [see Box and Jenkins (1976)] to each of four age groups, as suggested in Katzoff (1989). The models resulting in the four age groups 25–34, 45–54, 55–64, and 65–74 were seasonal with period 12 in the provisional and final series with the noise containing autoregressive and autoregressive-seasonal components. Details are given in Shumway (1989). The other alternative considered was simply to forecast the final series with the provisional value. We show the results in terms of absolute percentage errors in Table 2. It is clear that the structural models do best and that the provisional data are the worst estimator with the transfer function model somewhere in between. The structural models are only marginally better

Table 2. Mean and Maximum Absolute Percent Forecast Errors Over 24-Month Forecast Interval

	Provisional		*Structural*		*Box–Jenkins*	
Age	*Mean*	*Maximum*	*Mean*	*Maximum*	*Mean*	*Maximum*
25–34	27.4	71.5	11.4	24.4	15.3	38.5
45–54	6.9	16.1	3.8	9.1	4.2	12.7
55–64	3.9	11.0	2.8	6.9	2.9	7.2
65–74	3.4	11.7	2.9	7.2	3.8	9.1

than the transfer function models, but they seem to be parsimonious and are in terms of the more easily interpreted components model.

It should be noted that Equation (17) provides an estimated standard error for the adjusted final series. This standard error ranges from 55 to 70 deaths for the 25–35-year age group over most of the interpolation interval (61–80) and rises to 90 deaths near the end (81–84). For the older 65–74 age group, the standard error is 3,500 over the time period (61–80) and rises to 4,250 by the end (81–84).

5. DISCUSSION

We have attempted to develop a form of the dynamic linear model that will have broad applicability to the common problem of estimating a time series based on incomplete survey information. Within this framework we have derived an optimal adjustment that improves on pure forecasts or on the practice of replacing final values by their sample-based provisional estimates. The optimal adjustments are conditional expectations of the final values given all observations on the provisional and final series. Along the way we have discussed parameter estimation by maximum likelihood and the model identification problem for the trend and trend-plus-seasonal models with additive structural measurement errors.

The correlated measurement error structure seems to improve on the structural models involving only smooth components and uncorrelated measurement errors. It does this by assuming that a portion of the irregular component is common to both series. It should be noted that this structural components of variance approach may not be as good as one assuming a general error covariance if the model does not hold. There is also the possibility of using a general bivariate state–space model or a bivariate ARIMA model, as in Tiao and Box (1981). This may just add parameters without really improving the forecasts. It should also be noted that the equations for optimal adjustment derived in Section 3 can be easily derived for any model that can be put into state–space form.

APPENDIX: MAXIMUM LIKELIHOOD ESTIMATION

It is convenient to maximize the incomplete data form of the innovations likelihood given, for example, in Shumway (1988, sec. 3.4.3) using the complete-data log-likelihood and the EM algorithm of Dempster et al. (1977). Using the formulation for the state–space version of the additive trend autoregressive-seasonal model in (6)–(10), the complete-data log-likelihood, based on observing e_{t1}, e_{t2}, w_t, and a_t, conditioned on $\mathbf{x}_0$, has the simple form

$$\begin{aligned}\ln L' = &-(T/2)\ln\sigma_1^2 - \frac{1}{2\sigma_1^2}\sum_{t=1}^{T}(y_{t1}-\mathbf{m}_t'\mathbf{x}_t)^2\\ &-(T/2)\ln\sigma_2^2 - \frac{1}{2\sigma_2^2}\sum_{t=1}^{T}(y_{t2}-y_{t1})^2\\ &-(T/2)\ln\sigma_w^2 - \frac{1}{2\sigma_w^2}\sum_{t=1}^{T}(\nabla^2\mu_t)^2\\ &-(T/2)\ln\sigma_a^2 - \frac{1}{2\sigma_a^2}\sum_{t=1}^{T}(z_t-\boldsymbol{\varphi}'\mathbf{Z}_{t-1})^2, \qquad (\text{A.1})\end{aligned}$$

where we define the vectors $\boldsymbol{\varphi}' = (\varphi_1, \varphi_2, \ldots, \varphi_p)$ and $\mathbf{Z}_t' = (z_t, z_{t-1}, \ldots, z_{t-p+1})$ for ease in writing. The terms on the first two lines correspond to the residuals e_{t1} and e_{t2}, whereas the third and fourth lines account for the errors w_t and a_t.

First, we introduce some notation as in Shumway and Stoffer (1982) that is convenient for expressing the estimators. Define

$$\mathbf{x}_t^s = E(\mathbf{x}_t \mid \mathbf{y}_1, \ldots, \mathbf{y}_s) \qquad (\text{A.2})$$

and

$$P_{tu}^s = E[(\mathbf{x}_t - \mathbf{x}_t^s)(\mathbf{x}_u - \mathbf{x}_u^s)' \mid \mathbf{y}_1, \ldots, \mathbf{y}_s]. \qquad (\text{A.3})$$

The matrices

$$A = \sum_{t=1}^{T}(\mathbf{x}_{t-1}^T\mathbf{x}_{t-1}^{T\prime} + P_{t-1,t-1}^T), \qquad (\text{A.4})$$

$$B = \sum_{t=1}^{T}(\mathbf{x}_t^T\mathbf{x}_{t-1}^{T\prime} + P_{t,t-1}^T), \qquad (\text{A.5})$$

and

$$C = \sum_{t=1}^{T}(\mathbf{x}_t^T\mathbf{x}_t^{T\prime} + P_{tt}^T) \qquad (\text{A.6})$$

also play important parts. In addition, define the matrix

$$Q = C - \Phi B' - B\Phi' + \Phi A\Phi'. \qquad (\text{A.7})$$

Now, examining the additional expectation of the log-likelihood (A.1), it follows that the EM algorithm can be updated using the equations

$$\hat\sigma_1^2 = \frac{1}{T}\sum_{t=1}^{T}E[(y_{t1} - \mathbf{m}_t'\mathbf{x}_t)^2 \mid Y], \qquad (\text{A.8})$$

$$\hat\sigma_2^2 = \frac{1}{T}\sum_{t=1}^{T}E[(y_{t2} - y_{t1})^2 \mid Y], \qquad (\text{A.9})$$

$$\hat\sigma_w^2 = \frac{1}{T}\sum_{t=1}^{T}E[(\nabla^2\mu_t)^2 \mid Y], \qquad (\text{A.10})$$

$$\hat{\boldsymbol{\varphi}} = \left(\sum_{t=1}^{T}E(\mathbf{Z}_{t-1}\mathbf{Z}_{t-1}' \mid Y)\right)^{-1}\sum_{t=1}^{T}E(\mathbf{z}_{t-1}z_t \mid Y), \qquad (\text{A.11})$$

and

$$\hat\sigma_a^2 = \frac{1}{T}\sum_{t=1}^{T}E[(z_t - \boldsymbol{\varphi}'\mathbf{Z}_{t-1})^2 \mid Y]. \qquad (\text{A.12})$$

We note first that

$$\begin{aligned}E[(y_{t1} - \mathbf{m}_t'\mathbf{x}_t)^2 \mid Y] &= (y_{t1} - \mathbf{m}_t'\mathbf{x}_t^T)^2 + \mathbf{m}_t'P_{tt}^T\mathbf{m}_t, \quad \text{if } y_{t1} \text{ is observed}\\ &= (\hat y_{t1} - \mathbf{m}_t'\mathbf{x}_t^T)^2 + \omega\sigma_1^2\\ &\quad + (1-\omega)^2\mathbf{m}_t'P_{tt}^T\mathbf{m}_t, \quad \text{if } y_{t1} \text{ is not observed}, \qquad (\text{A.13})\end{aligned}$$

where $\hat y_{t1}$ is given in Equation (15). In addition,

$$\begin{aligned}E[(y_{t2} - y_{t1})^2 \mid Y] &= (y_{t2} - y_{t1})^2, \quad \text{if } y_{t1} \text{ is observed}\\ &= (y_{t2} - \hat y_{t1})^2 + \omega\sigma_1^2\\ &\quad + \omega^2\mathbf{m}_t'P_{tt}^T\mathbf{m}_t, \quad \text{if } y_{t1} \text{ is not observed}. \qquad (\text{A.14})\end{aligned}$$

Then,

$$\hat\sigma_w^2 = \frac{1}{T}q_{11}, \qquad (\text{A.15})$$

where q_{11} is the first element of Q in (A.7). If we denote the lower right $p \times p$ submatrix of A by A_{22} and the $p \times 1$ lower left column of B by $\mathbf{b}_{21}$, then Equation (A.11) becomes

$$\hat{\boldsymbol{\varphi}} = A_{22}^{-1}\mathbf{b}_{21}. \tag{A.16}$$

In addition, Equation (A.12) becomes

$$\hat{\sigma}_a^2 = \frac{1}{T} q_{33}. \tag{A.17}$$

[*Received August 1989. Revised January 1991.*]

REFERENCES

Akaike, H. (1973), "Information Theory and an Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory,* eds. B. N. Petrov and F. Csaki, Budapest: Akademiai Kiado, pp. 267–281.

——— (1980), "Seasonal Adjustment by a Bayesian Modeling," *Journal of Time Series Analysis,* 1, 1–13.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control* (2nd ed.), San Francisco: Holden-Day.

Carlin, J. B., and Dempster, A. P. (1989), "Sensitivity Analysis of Seasonal Adjustments: Empirical Case Studies" (with discussion), *Journal of the American Statistical Association,* 84, 6–32.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977), "Maximum Likelihood From Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society,* Ser. B, 39, 1–38.

Engle, R. F., and Granger, C. W. J. (1987), "Co-Integration and Error Correction: Representation, Estimation and Testing," *Econometrica,* 55, 251–276.

Gersch, W., and Kitagawa, G. (1983), "The Prediction of Time Series With Trends and Seasonalities," *Journal of Business & Economic Statistics,* 1, 253–264.

Harrison, P. G., and Stevens, C. F. (1976), "Bayesian Forecasting," *Journal of the Royal Statistical Society,* Ser. B, 38, 205–247.

Harvey, A. C. (1984), "A Unified View of Statistical Forecasting Procedures," *Journal of Forecasting,* 3, 245–275.

——— (1988), *Forecasting, Structural Time Series Models and the Kalman Filter,* New York: Cambridge University Press.

Harvey, A. C., and Todd, P. H. J. (1983), "Forecasting Economic Series With Structural and Box–Jenkins Models: A Case Study With Comments," *Journal of Business & Economic Statistics,* 1, 299–315.

Hillmer, S. C., and Trabelsi, A. (1987), "Benchmarking of Economic Time Series," *Journal of the American Statistical Association,* 82, 1064–1071.

Katzoff, M. J. (1989), "The Application of Time Series Forecasting Methods to an Estimation Problem Using Provisional Mortality Statistics," *Statistics in Medicine,* 8, 335–341.

Kitagawa, G., and Gersch, W. (1984), "A Smoothness Priors State–Space Modeling of Time Series With Trend and Seasonality," *Journal of the American Statistical Association,* 79, 378–389.

Ljung, L. J., and Caines, P. E. (1979), "Asymptotic Normality of Prediction Error Estimators for Approximate System Models," *Stochastics,* 3, 29–46.

Pankratz, A. (1989), "Time Series Forecasts and Extra-Model Information," *Journal of Forecasting 8,* 75–83.

Shumway, R. H. (1988), *Applied Statistical Time Series Analysis,* Englewood Cliffs, NJ: Prentice-Hall.

——— (1989), "Time Series Forecasting Using Provisional Mortality Statistics," Final Report, Contract 89C072428301D, National Center for Health Statistics.

Shumway, R. H., and Stoffer, D. S. (1982), "An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm," *Journal of Time Series Analysis,* 3, 253–264.

Tiao, G. C., and Box, G. E. P. (1981), "Modeling Multiple Time Series With Applications," *Journal of the American Statistical Association,* 76, 802–816.

Trabelsi, A., and Hillmer, S. C. (1989), "A Benchmarking Approach to Forecasting Combination," *Journal of Business & Economic Statistics,* 7, 353–362.

Wecker, W. E., and Ansley, C. F. (1983), "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing," *Journal of the American Statistical Association,* 78, 71–89.