Estimating Missing Observations in Economic Time Series*

A. C. Harvey
London School of Economics, Houghton Street, London, WC2A 2AE, UK

R. G. Pierse
Department of Applied Economics, Cambridge University, Cambridge CB3 9DE, UK

March 1983

Abstract

Two related problems are considered. The first concerns the maximum likelihood estimation of the parameters in an ARIMA model when some of the observations are missing or subject to temporal aggregation. The second concerns the estimation of the missing observations. Both problems can be solved by setting up the model in state space form and applying the Kalman filter.

Key Words: Autoregressive-integrated-moving average processes; Kalman filter; Maximum likelihood estimation; Missing observations; Smoothing; Temporal aggregation.

1 Introduction

It is not unusual to encounter economic time series that are currently published at monthly or quarterly intervals but are only available on an annual basis in earlier periods. For a stock variable, such as the money supply, this means that there are missing observations in the first part of the series, while for a flow, like investment, it means that the early observations are subject to temporal aggregation. Note that a stock is the quantity of something at a particular point in time, while a flow is a quantity that accumulates over a given period of time. The relevance of these concepts is obviously not confined to economics.

This article considers two related problems: the estimation of an autoregressive-integrated-moving average (ARIMA) model based on *all* the available observations and the estimation of the missing values in the first part of the series. The solution of both problems lies in finding a suitable state space representation of the ARIMA model. Maximum likelihood estimation is then possible via the prediction error decomposition, and once this has been done the missing observations can be estimated by smoothing.

Section 2 introduces the state space methodology and shows how it can be applied to series that can be modelled by stationary ARMA processes. The applicataion of this technique to stock variables is already fairly well known (see, e.g. Jones (1980)), but it does not seem to have been used in connection with flows. The extension to the more relevant case of an ARIMA model raises some nontrivial problems and has not been dealt with before, even for a stock variable. Two approaches to estimating ARIMA models with missing observations are described in Section 3. Section 4 deals with the prediction of future observations and describes how the missing observations are estimated by smoothing. The additional complications caused to ARIMA modelling by the logarithmic transformation are considered in Section 5. Some examples of the application of the techniques are given in Section 6, and Section 7 sets out a general solution to the problem considered by Chow and Lin (1971, 1976), namely the estimation of missing observations by regressing on a related series.

*Published in *Journal of the American Statistical Association* (1984), Vol. 79, pp. 125–131. The work was carried out while R. G. Pierse was a Research Assistant on the program in Methodology, Inferrence and Modelling in Econometrics at the London School of Economics. Financial support from the SSRC is gratefully acknowledged. The authors would like to thank an Associate Editor of this journal and a referee for helpful comments on an earlier draft. In addition they have benefited from valuable discussions with P. Pereira and with several of the participants in the Symposium on Time Series Analysis of Irregularly Observed Data held at Texas A&M University in February 1983. Any errors remain the authors' responsibility.

2 Stationary Processes

Since this article is only concerned with univariate time series, attention can be focused on a special case of the state space model. Let $\boldsymbol{\alpha}_t$ be an $m \times 1$ state vector that obeys the transition equation

$$\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \mathbf{R}\epsilon_t, \quad t = 1, \dots, T, \tag{2.1}$$

where $\mathbf{T}$ is a fixed matrix of dimension $m \times m$, $\mathbf{R}$ is an $m \times 1$ vector, and ϵ_t is a sequence of normally distributed independent random variables with mean zero and variance σ^2; that is $\epsilon_t \sim \text{NID}(0, \sigma^2)$. The state vector is related to a single series of observations by the measurement equation

$$y_t = \mathbf{z}_t{}'\boldsymbol{\alpha}_t + \zeta_t, \quad t = 1, \dots, T, \tag{2.1b}$$

where y_t is the observed value, $\mathbf{z}_t$ is a fixed $m \times 1$ vector and ζ_t is a sequence of normally distributed independent random variables with mean zero and variance $\sigma^2 h_t$.

Let $\mathbf{a}_{t-1}$ denote the optimal or minimum mean squared estimator (MMSE) based on all the information available at time $t-1$. Let $\sigma^2\mathbf{P}_{t-1}$ denote the covariance matrix of $\mathbf{a}_{t-1} - \boldsymbol{\alpha}_{t-1}$. Given $\mathbf{a}_{t-1}$ and $\mathbf{P}_{t-1}$, the MMSE of $\boldsymbol{\alpha}_t$, $\mathbf{a}_{t|t-1}$, together with its associated covariance matrix, $\mathbf{P}_{t|t-1}$ is obtained by applying the prediction and updating equations of the Kalman filter (see, e.g., Anderson and Moore (1979), or Harvey (1981b)).

2.1 State space formulation of an ARMA model

A stationary ARMA(p,q) model for a sequence of normally distributed variables $y_1^\dagger, \dots, y_T^\dagger$ can be written as

$$\begin{aligned} y_t^\dagger &= \phi_1 y_{t-1}^\dagger + \dots + \phi_p y_{t-p}^\dagger + \epsilon_t \\ &\quad + \theta_1\epsilon_{t-1} + \dots + \theta_q\epsilon_{t-q}, \quad t = 1, \dots, T, \end{aligned} \tag{2.2}$$

where $\phi, \dots, \phi_p$ are the AR parameters, $\theta_1, \dots, \theta_q$ are the MA parameters and $\epsilon_t \sim \text{NID}(0, \sigma^2)$. The model can be put in state space form that obeys a transition equation of the form (2.1). The transition matrix $\mathbf{T}$, has ϕ_i, $i = 1, \dots, p$ as the ith element in its first column, unity as element $(j, j+1)$, $j = 1, \dots, m-1$ and all other elements zero. The $m \times 1$ vector $\mathbf{R}$ is defined as $\mathbf{R} = (1, \theta_1, \dots, \theta_{m-1})'$, where $\theta_{q+1}, \dots, \theta_{m-1}$ are zero if $m > q+1$. This particular matrix and vector will be denoted $\boldsymbol{\Phi}$ and $\boldsymbol{\theta}$ in future sections. Given these definitions the first element in $\boldsymbol{\alpha}_t$ is identically equal to $y_t^\dagger$. Thus in the measurement equation, (2.1b), $\mathbf{z}_t = (1, 0, \dots, 0)'$ for $t = 1, \dots, T$, and if $y_t^\dagger$ is observed without error, $y_t = y_t^\dagger$ and $h_t = 0$.

Because the model is stationary, the initial conditions for the Kalman filter are given by $\mathbf{a}_{1|0} = \mathbf{0}$ and $\mathbf{P}_{1|0} = \sigma^{-2}\,\text{E}(\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t{}')$. Given that an observation is available in every time period, the Kalman filter produces a set of T prediction errors or innovations, $\nu_t = y_t - \mathbf{z}_t{}'\mathbf{a}_{t|t-1}$ for $t = 1, \dots, T$. These can be used to construct the likelihood function by the prediction error decomposition; that is,

$$\log L(y_1, \dots, y_T; \boldsymbol{\Phi}, \boldsymbol{\theta}, \sigma^2) = -\frac{T}{2}\log 2\pi - \frac{T}{2}\log\sigma^2 - \frac{1}{2}\sum_{t=1}^{T}\log f_t - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\nu_t^2/f_t. \tag{2.3}$$

The quantities $f_1, \dots, f_T$ each of which is proportional to the variance of the corresponding innovation, are also produced by the Kalman filter. The parameter σ^2 does not appear in the Kalman filter, and it can be concentrated out of (2.3).

Careful programming of the Kalman filter recursions leads to a very efficient algorithm for evaluating the exact likelihood function of an ARMA model; compare the evidence presented in Gardner et al. (1980). Various modifications of the Kalman filter can also be used for this purpose (see, e.g. Pearlman (1980)).

2.2 Missing observations on a stock variable

A missing observation on a stock variable can be handled very easily simply by bypassing the corresponding updating equation. Skipping the missing observations in this way makes no difference to the validity of the prediction error decomposition provided that when an observation is missing the corresponding $\log f_t$ term is omitted from the likelihood. Thus the likelihood function is of the form given in (2.3) but with the summations covering only those values of t for which the variable is actually observed. The T appearing in the first two terms is replaced by the number of observations.

2.3 Temporal aggregation of a flow variable

Let n denote the maximum number of time periods over which a flow variable is aggregated and let y^*_{t-1} be the $(n-1)\times 1$ vector $(y^\dagger_{t-1},\ldots,y^\dagger_{t-n+1})'$. Define an $(m+n-1)\times 1$ augmented state vector, $\boldsymbol{\alpha}_t = (\overline{\boldsymbol{\alpha}}'_t \;\; y^*_{t-1})'$, where $\overline{\boldsymbol{\alpha}}_t$ is the state vector appropriate to the ARMA model for $y^\dagger_t$. The transition equation for the augmented state vector is

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \boldsymbol{\Phi} & \mathbf{0} & \\ 1 & \mathbf{0}' & \mathbf{0}' \\ \mathbf{0} & \mathbf{I}_{n-2} & \mathbf{0} \end{bmatrix} \boldsymbol{\alpha}_{t-1} + \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix} \epsilon_t, \quad t = 1,\ldots,T \tag{2.4}$$

(When $q = 0$, a more economical state space representation is obtained by redefining m as $\max(p,n)$ and letting the state vector be $\boldsymbol{\alpha}_t = (y^\dagger_t,\ldots,y^\dagger_{t-m+1})'$. The matrix $\mathbf{T}$ in the transition equation is then the transpose of $\boldsymbol{\Phi}$.) If at time t, the aggregate y_t of the previous $n(t)$ terms in the series is observed, the measurement equation is

$$y_t = (1, \mathbf{0}_{m-1}{}', \mathbf{i}_{n(t)-1}{}', \mathbf{0}_{n-n(t)}{}')\boldsymbol{\alpha}_t = \sum_{j=0}^{n(t)-1} y^\dagger_{t-j}, \quad 1 \le t \le T, \tag{2.5}$$

where $\mathbf{i}$ is an $(n(t)-1)\times 1$ vector of ones. In periods when there is no observation the updating equation can be skipped in the same way as for a stock variable. Note that the definition of the vector $\mathbf{z}_t$ changes as the basis upon which the variable is aggregated changes.

The starting values for the augmented state space model are $\mathbf{a}_{1|0} = 0$ and $\mathbf{P}_{1|0} = \sigma^{-2}\,\mathrm{E}(\boldsymbol{\alpha}_t\boldsymbol{\alpha}_t{}')$. However, if a run of disaggregated observations is available at the end of the series, the calculations can be simplified by working backwards. This is quite legitimate since if $y^\dagger_t$, $t = 1,\ldots,T$ is generated by an ARMA $(p,\, q)$ process, the observations taken from $t = T$ to $t = 1$ can be regarded as being generated by exactly the same process; see (Box and Jenkins, 1976, pp. 197–198). The calculations begin with the state model appropriate to $y^\dagger_t$ and only when the agregate observations start to arrive is a transfer to the augmented model made. The advantage of this approach is that the initial $m \times m$ matrix $\widehat{\mathbf{P}}_{1|0} = \sigma^{-2}\,\mathrm{E}(\widehat{\boldsymbol{\alpha}}_t\widehat{\boldsymbol{\alpha}}_t{}')$ can be evaluated using standard algorithms. When the run of disaggregate observations comes to an end, the MMSE of the augmented state vector and the associated $\mathbf{P}_t$ matrix can be formed immediately since all the observations in the vector corresponding to y^*_{t-1} are known. Note that when the observations are processed in reverse, the aggregate observations must be regarded as arising at the beginning of the period of aggregation rather than at the end.

3 Nonstationary Process

In general, economic time series are nonstationary and the usual approach is to fit an ARIMA model. Thus if Δ denotes the first difference operator and Δ_s denotes the seasonal difference operator (for a season of s periods), the series $\Delta^d\Delta^D_s y^\dagger_t = w^\dagger_t$ is modelled as a stationary (seasonal) ARMA process.

There are basically two ways of constructing the likelihood function for an ARIMA model with missing observations. The first approach formulates the state space model in such a way that the observations are in levels, while the second has the observations in differences. The choice between them depends on the pattern of missing observations. If they are missing at regular intervals, the algorithm based on differenced observations may be preferrable. The levels formulation is, however, more flexible. In addition it forms the basis for the smoothing algorithm.

3.1 Levels formulation

Let L be the lag operator and let $-\delta_j$ be the coefficient of L^j in the expansion of $\Delta^d\Delta^D_s = (1-L)^d(1-L^s)^D$. Let $\overline{\boldsymbol{\alpha}}_t$ be the state vector in the state space model for the stationary ARMA$(p,\, q)$ process, $w^\dagger_t$, and define the augmented state vector, $\boldsymbol{\alpha}_t = (\overline{\boldsymbol{\alpha}}_t{}' \;\; y^*_{t-1}{}')'$, where $y^*_{t-1} = (y^\dagger_{t-1},\ldots,y^\dagger_{t-d-sD})'$. The transition equation for the augmented state vector is

$$\boldsymbol{\alpha}_t = \begin{bmatrix} \overline{\boldsymbol{\alpha}}_t \\ y^*_{t-1} \end{bmatrix} = \begin{bmatrix} & \boldsymbol{\Phi} & \mathbf{0}' & \\ 1 & 0\cdots 0 & \delta_1\cdots\delta_k & \\ & \mathbf{0}' & \mathbf{I}_{k-1} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \overline{\boldsymbol{\alpha}}_{t-1} \\ y^*_{t-2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{0} \end{bmatrix} \epsilon_t, \quad t = 1,\ldots,T, \tag{3.1}$$

where $k = d + sD$.

If $y_t^\dagger$ is observed for all $t = 1, \dots, T$, the measurement equation is

$$y_t = (1, \mathbf{0}_{m-1}{}', \delta_1 \cdots \delta_k)\boldsymbol{\alpha}_t, \quad t = 1, \dots, T, \tag{3.2}$$

and the Kalman filter can be initialised at $t = k$ with $\mathbf{a}_{k+1|k} = (\mathbf{0}', y_{k+1}^*{}')'$ and

$$\mathbf{P}_{k+1|k} = \begin{bmatrix} \overline{\mathbf{P}}_{1|0} & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.3}$$

where $\overline{\mathbf{P}}_{1|0} = \sigma^{-2}\,\mathrm{E}(\overline{\boldsymbol{\alpha}}_t\overline{\boldsymbol{\alpha}}_t{}')$. It is not difficult to show that the likelihood function constructed from the Kalman filter is identical to the likelihood function that would result from applying the Kalman filter to the differenced observations $\Delta^d\Delta_s^D y_t$; compare a similar argument in Harvey (1981a).

For stock variables, missing observations can be handled in the same way as described in Section 2. For a flow variable, the measurement equation for an aggregate observation of the form

$$y_t = \sum_{j=0}^{n(t)-1} y_{t-j}^\dagger$$

is

$$y_t = (1, \mathbf{0}_{m-1}{}', \delta_1 + 1, \dots, \delta_{n(t)-1} + 1, \delta_{n(t)}, \dots, \delta_k)'\boldsymbol{\alpha}_t. \tag{3.4}$$

This assumes $n - 1 \leq d + sD$; if this is not the case, y_{t-1}^* must be redefined as $(y_{t-1}^\dagger, \dots, y_{t-n+1}^\dagger)'$.

The only problem with the levels formulation concerns starting values since in most applications the complete set of values $y_1^\dagger, \dots, y_k^\dagger$ will not be available. However, if at least k consecutive observations are available at the end of the series the problem can be solved by reversing the order of the observations.

3.2 Difference formulation

Suppose that observations on a stock variable are available evey n time period. An immediate difficulty arises with an ARIMA model because it may not be possible to construct the differenced observations $\Delta^d\Delta_s^D y_t^\dagger$, from such a sequence. Thus, for example, first differences cannot be formed if the variable in question is only obseerved every other time period. The solution to the problem is to construct a differenced series that can be observed. If $s \geq n$ and s/n is an integer, the observable differenced series is

$$y_t = \Delta_n^d\Delta_s^D y_t^\dagger, \quad t = nd + SD, n(d+1) + sD, \dots, \tag{3.5}$$

where

$$\Delta_n^d = (1 - L^n)^d = (1 - L)^d(1 + L + \cdots + L^{n-1})^d. \tag{3.6}$$

Expression (3.5) becomes

$$y_t = (1 + L + \cdots + L^{n-1})^d w_t^\dagger, \tag{3.7}$$

where $w_t^\dagger$ is the underlying ARMA process, $\Delta^d\Delta_s^D y_t^\dagger$. Thus although $y_t^\dagger$ is a stock, the observable differenced series, y_t, is a flow when considered in terms of $w_t^\dagger$. The techniques described for flow variables in Section 2 can be applied directly, although if $d > 1$ the weights for different lags of $w_t^\dagger$ are not the same, and the measurement equation must be amended accordingly. Similar methods can be applied when the original variable is itself a flow.

4 Predicting Future Observations and Estimating Missing Observations

Once the parameters of the ARIMA model have been estimated, optimal predictions of future observations, together with their conditional mean squared errors (MSE's), can be made by repeated application of the Kalman filter prediction equations. Similarly, MMSE's of the missing observations can be computed by smoothing. The levels form of the model will normally be appropriate for both purposes.

There are a number of smoothing algorithms available. The best known is probably the fixed interval algorithm; see (Anderson and Moore, 1979, pp. 187–190) or (Harvey, 1981b, Ch. 4). It is based on a set of backward recursions starting at time T, but it has two drawbacks in the present context. The first is that it requires the storage of a large number of covariance matrices, the $\mathbf{P}_t$ and $\mathbf{P}_{t|t-1}$'s computed from the initial pass through the Kalman filter. The second is that the inverse of $\mathbf{P}_{t|t-1}$ is needed and if some of the elements in $\boldsymbol{\alpha}_t$ are known at time $t-1$, this matrix will be singular. Neither of these problems is insurmountable, but for the kind of situations with which we are concerned, a fixed-point smoothing algorithm is more attractive.

Fixed-point smoothing can be applied by proceeding with the Kalman filter and augmenting the state vector each time a missing observation is encountered. Once all the observations have been processed, the components added to the state vector will contain the MMSE's of the missing observations. The corresponding MSE's can be obtained directly from the associated augmented covariance matrix.

The recursions for the augmented parts of the state vector can, in fact, be separated from the Kalman filter recursions for the original state vector. This is extremely useful since it means that a new series of recursions can simply be started off with each missing observation, leaving the basic Kalman filter undisturbed. The form of these recursions is as follows. Suppose that the underlying varaible $y_t^\dagger$, is not observed at time $t = \tau$. The state vector is augmented by $y_\tau^\dagger$, which is, it should be noted, a linear combination of the elements of $\boldsymbol{\alpha}_\tau$; that is,

$$y_\tau^\dagger = \mathbf{z}'\boldsymbol{\alpha}_\tau, \tag{4.1}$$

where $\mathbf{z}$ is constant throughout the series. (Note that for a flow variable $\mathbf{z}_t$ will not be the same as $\mathbf{z}$ when there is temporal aggregation. However, for a stock variable, $\mathbf{z}_t = \mathbf{z}$ whenever there is an observation.) Modifying the formulas in (Anderson and Moore, 1979, 172–173) to take account of the fact that only a linear combination of $\boldsymbol{\alpha}_\tau$ is to be estimated leads to the smoothing recursions

$$\overline{y}_{\tau|t} = \overline{y}_{\tau|t-1} + \mathbf{p}_{\tau|t-1}{}'\mathbf{z}_t f_t^{-1}\nu_t, \quad t = \tau, \ldots, T \tag{4.2a}$$

and

$$\mathbf{p}_{\tau|t} = \mathbf{T}(\mathbf{I} - \mathbf{g}_t\mathbf{z}_t{}')\mathbf{p}_{\tau|t-1}, \quad t = \tau, \ldots, T, \tag{4.2b}$$

where $\mathbf{g}_t = \mathbf{P}_{t|t-1}\mathbf{z}_t f_t^{-1}$. The initial values are $\overline{y}_{\tau|\tau-1} = \mathbf{z}'\mathbf{a}_{\tau|\tau-1}$ and $\mathbf{p}_{\tau|\tau-1} = \mathbf{P}_{\tau|\tau-1}\mathbf{z}$. The quantities f_t, ν_t, and $\mathbf{g}_t$ are all produced by the Kalman filter for the original state space model, the vector $\mathbf{z}_t$ being used to define the measurement equation. If there is no observation at time t, then (4.2a) and (4.2b) collapse to $\overline{y}_{\tau|t} = \overline{y}_{\tau|t-1}$ and $\mathbf{p}_{\tau|t} = \mathbf{T}\mathbf{p}_{\tau|t-1}$, respectively. The MSE of $\overline{y}_{\tau|T}$ is given by $\sigma^2 f_{\tau|T}$, where $f_{\tau|T}$ is obtained from the recursion

$$f_{\tau|t} = f_{\tau|t-1} - \mathbf{p}_{\tau|t-1}{}'\mathbf{z}_t f_t^{-1}\mathbf{z}_t{}'\mathbf{p}_{\tau|t-1}, \quad t = \tau, \ldots, T \tag{4.3}$$

with $f_{\tau|\tau-1} = z'\mathbf{P}_{\tau|\tau-1}\mathbf{z}$.

When set up in this way, the fixed-point smoothing algorithm is extremely efficient. The storage requirements are negligible and in a typical application, the time taken to run the augmented Kalman filter is usually less than twice the time taken for a normal run. This is trivial compared with the time taken to compute the ML estimators of the unknown parameters.

5 Logarithmic Transformations

It is very common to take logarithms of a variable before fitting an ARIMA model. This creates no difficulties whatsoever for a stock variable. For a flow variable, however, an immediate problem arises because it is the sum of the original variables that is observed and the logarithm of a sum is not equal to the sum of the logarithms. Assuming that the logarithms of the aggregated variables are normally distributed is then inconsistent with the assumption that the corresponding disaggregated variables are normal. Notwithstanding this point, one way to proceed is to assume that the logarithms of all variables actually observed are normally distributed. The logarithm of the observed aggregate at time t is

$$y_t = \log \sum_{j=0}^{n(t)-1} \exp(y_{t-j}^\dagger), \tag{5.1}$$

where $y_t^\dagger$ is the underlying ARIMA process. Adopting the notation of (3.4), but assuming that $n = k$ for simplicity, the measurement equation can be written as

$$y_t = \log[\exp\{(1, \mathbf{0}_{m-1}{}', \delta_1, \ldots, \delta_n)\boldsymbol{\alpha}_t\} + \sum_{j=m+1}^{m+n} \exp(\alpha_{jt})], \tag{5.2}$$

where α_{jt} is the jth element in $\boldsymbol{\alpha}_t$. this equation is obviously nonlinear but by using the extended Kalman filter, as in (Anderson and Moore, 1979, pp. 193–195), an approximation to the likelihood function can be computed by the prediction error decomposition.

6 Example

The airline passenger data given in (Box and Jenkins, 1976, p. 531) consists of 144 monthly observations on the number of passengers carried by airlines over the period 1949 to 1960. It is highly seasonal and (Box and Jenkins, 1976, Ch. 9) fitted the following model to the logarithms of the observations:

$$\Delta\Delta_{12}y_t = (1+\theta_1 L)(1+\theta_{12}L^{12})\epsilon_t. \tag{6.1}$$

The above model was estimated using four variations of the data set: (i) all 144 observations; (ii) the observations from January to November deleted for the last six years; (iii) the logarithms of the observations of each of the last six years aggregated and assigned to December; (iv) the raw observations for each of the last six years aggregated and assigned to December. The second data set represents an example of missing observations, with the variable treated as though it were a stock observed annually, rather than monthly, for part of the sample period. The third and fourth data sets are examples of temporal aggregation. Data set (iii) would be relevant if the observations used in the ARIMA model were original observations rather than logarithms. Note that placing the missing or temporally aggregated observations at the end of the series, whereas in practice they might come at the beginning, is unimportant. As already noted, the order of the observations can always be reversed without affecting the underlying ARIMA model.

Table 1: Maximum Lkelihood Estimates of Parameters for Airline Passenger Model, (6.1)

	Parameters[a]	
Data set	θ_1	θ_{12}
(i) Full Data	−.402	−.557
	(.090)	(.073)
(ii) Missing Observations	−.457	−.758
	(.121)	(.236)
(iii) Temporal Aggregation (logs)	−.475	−.741
	(.114)	(.223)
(iv) Temporal Aggregation (raw data)	−.477	−.738
	(.114)	(.221)

[a] Figures in parentheses are asymptotic standard errors.

The computer program we wrote is a fairly general one that can handle both missing observations and temporal aggregation. The pattern of missing observations need not be regular. The only proviso is that there should be a run of $d + sD$ observed values at either the beginning or the end of the series. In writing the program, considerable care was taken to devise a computationally efficient routine for evaluating the likelihood function. Particular attention was paid to the evaluation of $\overline{\mathbf{P}}_{1|0}$, the matrix used to initialise the Kalman filter, and the algorithm adopted is described in some detail in our original research report (Harvey and Pierse (1982)). Maximisation of the likelihood function was carried out by one of the Gill-Murray-Pitfield numerical optimisation routines in the UK NAG library, E04JBF. This is a Quasi-Newton algorithm that uses numerical derivatives and allows simple bounds to be placed on the parameters. By choosing a suitable parameterisation, we were able to devise a very effective procedure in which we were able to constrain the roots of the MA polynomial to lie outside or on the unit circle; again see (Harvey and Pierse, 1982, Appendix B). The reason for allowing strictly noninvertible MA processes is set out in Harvey (1981b).

The results of exact ML estimation are shown in Table 1. The estimates obtained with data sets (ii), (iii) and (iv) are quite close to the estimates obtained with the full set of observations. The higher asymptotic standard errors are a reflection of the smaller number of observations.

Table 2: Estimates of Logarithms of Missing Observations and
Associated Root Mean Squared Errors for 1957

	Month											
Data set	Jan.	Feb.	March	April	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
(ii) Missing Observations	5.733	5.738	5.893	5.850	5.843	5.951	6.051	6.055	5.938	5.8123	5.680	—
	(.045)	(.049)	(.052)	(.054)	(.055)	(.055)	(.055)	(.054)	(.052)	(.049)	(.045)	
(iii) Temporal Aggregation (logs)	5.770	5.778	5.937	5.896	5.890	5.997	6.094	6.093	5.971	5.839	5.700	5.818
	(.041)	(.040)	(.039)	(.038)	(.037)	(.037)	(.037)	(.037)	(.038)	(.039)	(.040)	(.041)
(iv) Temporal Aggregation (raw data)	5.772	5.779	5.939	5.848	5.893	6.001	6.098	6.099	5.976	5.844	5.704	5.823
	(.041)	(.041)	(.039)	(.038)	(.037)	(.036)	(.036)	(.036)	(.037)	(.039)	(.041)	(.041)
Actual Values	5.753	5.707	5.875	5.852	5.872	6.045	6.146	6.146	6.001	5.849	5.720	5.817

Table 2 shows the estimates of the missing observations for 1957 computed by the smoothing algorithm described in Section 4. The root mean squared errors associated with each estimate are

$$\mathrm{RMSE}(\overline{y}_{t|T}) = \widetilde{\sigma}\sqrt{f_{t|T}},$$

where $f_{t|T}$ is given by (4.3) and $\widetilde{\sigma}$ is the square root of the ML estimator of σ^2. For data sets (ii) and (iii) the estimates are all remarkably close to the actual values. Similar results were obtained for the other years in which there were missing observations. The theoretical justification for the estimates obtained when there is temporal aggregation but the model is in logarithms—case (iv)— is somewhat weaker because of the approximation involved in the use of the extended Kalman filter. However, the results presented in Table 2 lend some support to the use of this device in smoothing, as well as estimation. The figures presented for data set (iv) are virtually indistinguishable from those derived for data set (iii).

The results in Table 2 refer to estimates of the logarithms of the missing observations. If x_t denotes the original observation, a direct estimate of a missing x_t is given by $\overline{x}_{t|T} = \exp(\overline{y}_{t|T})$. However, the relationship between the normal and lognormal distributions suggests the modified estimator

$$\overline{x}_{t|T}{}^{*} = \exp\left\{\overline{y}_{t|T} + \frac{1}{2}\,\mathrm{MSE}(\overline{y}_{t|T})\right\}. \tag{6.2}$$

The estimator is unbiased in the sense that the expectation of $x_t - \overline{x}_{t|T}{}^{*}$ is zero when the parameters of the underlying ARIMA model are known. For the airline passenger data, the use of the modification in (6.2) made very little difference. In the case of (ii), for example, the direct and modified estimates for May 1957 were 344.8 and 345.4 respectively. The true value is 355. For the same data point, the 95% prediction interval was 309.5 to 384.1.

7 Regression

Chow and Lin (1971, 1976) approach the problem of estimating missing observations by assuming that $y_t^{\dagger}$ is related to a set of k nonstochastic variables that are observed in all time periods. They assume a linear relationship of the form

$$y_t^{\dagger} = \mathbf{x}_t{}'\boldsymbol{\beta} + u_t, \quad t = 1, \ldots, T, \tag{7.1}$$

where $\mathbf{x}_t$ is the $k \times 1$ vector of related variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters, and u_t is a stationary stochastic disturbance term.

Given the covariance matrix of the disturbances, finding estimates of the missing observations is basically an exercise in best linear unbiased estimation and prediction. However, it does require the construction and inversion of the covariance matrix associated with the variables (aggregates in the flow case) actually observed. Furthermore the covariances between the missing values and the observed values must also be found. In solving the problem in this way Chow and Lin concentrate on situations where the observations are missing at regular intervals and the disturbances are either serially uncorrelated or generated by an AR(1) process.

The Kalman filter can be applied directly to (7.1) by using the techniques described in Section 2, with $y_t^{\dagger}$ replaced by $y_t^{\dagger} - \mathbf{x}_t{}'\boldsymbol{\beta}$. The likelihood function must then be maximised nonlinearly with respect to $\boldsymbol{\beta}$ as well as the ARMA parameters and any estimates of MSE's obtained in a subsequent smoothing operation will only be conditional on $\boldsymbol{\beta}$. This constitutes a possible disadvantage comapred to the Chow-Lin approach although it does avoid the necessity for repeated inversions of relatively large covariance matrices. However, it is possible to conserve the advantages of the Chow-Lin approach while applying the Kalman filter by redefining the state space model to include $\boldsymbol{\beta}$ in the state vector (cf. Harvey and Phillips (1979)). This formulation yields the BLUE of $\boldsymbol{\beta}$ for given values of the ARMA parameters and enables $\boldsymbol{\beta}$ to be concentrated out of the likelihood function. (If consistent estimators of the ARMA parameters can be obtained, this estimator of $\boldsymbol{\beta}$ will be asymptotically efficient under suitable regularity conditions.) Note that it will normally be necessary to process the observations backwards to obtain starting values.

The above techniques can also be used if the regression model has been framed in first differences as suggested by Denton (1971) and Fernandez (1981). In this case the model is

$$\Delta y_t^{\dagger} = (\Delta \mathbf{x}_t)'\boldsymbol{\beta} + u_t, \quad t = 2, \ldots, T. \tag{7.2}$$

The Kalman filter can handle models in which u_t is an ARMA process rather than the serially uncorrelated process assumed in the references just cited. In fact even when u_t is serially uncorrelated, the Kalman filter may still be advantageous since the inversion of the $(T-1) \times (T-1)$ matrix in expressions (3) and (4) of the paper by Fernandez is avoided.

Finally, it is worth noting that one suggestion made by Chow and Lin is to use a time trend and seasonal dummies as the explanatory variables in (7.1). Computing estimates of the missing values from such a model can also be carried within the ARIMA framework described in Sections 3 and 4. If the disturbance in (7.1) is a serially uncorrelated process, ϵ_t, the model is a special case of (6.1) in which $\theta_1 = \theta_{12} = -1$. Although this model is strictly noninvertible, starting off the Kalman filter in the manner suggested in Section 3 ensures that estimates of missing values and predictions of future observations are exactly the same as if the model had been estimated within the regression framework of Chow and Lin (cf. Harvey (1981a)).

8 Conclusion

The results reported in Section 6 show that maximum likelihood estimation of ARIMA models can be carried out efficiently when there are missing or temporally aggregated observations. Furthermore, mimimum mean squared estimates of the missing observations together with their conditional root mean squared errors, can be computed at vey little extra cost. Additional complications arise with temporal aggregation when the ARIMA model is based on logarithms, but an approximate solution can be obtained by the extended Kalman filter. This solution is not altogether satisfactory from the theoretical point of view, although it does seem to give quite reasonable results with the airline data.

Although the approach developed here can handle most configurations of missing values, it does need an unbroken run of observations at the beginning or end of the series. One way of relaxing this requirement is by modifying an algorithm given in Rosenberg (1973). An indication of how this may be done can be found in Harvey and McKenzie (1983).

References

Anderson, B. D. A. and J. B. Moore (1979). *Optimal Filtering*. Englewood Cliffs, N. J.: Prentice Hall.

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

Chow, G. C. and A. Lin (1971). Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *Review of Economics and Statistics 53*, 372–375.

Chow, G. C. and A. Lin (1976). Best linear unbiased estimation of missing observations in an economic time series. *Journal of the American Statistical Association 71*, 719–721.

Denton, F. T. (1971). Adjustment of monthly or quarterly series to annual, totals; an approach based on quadratic minimization. *Journal of the American Statistical Association 66*, 99–102.

Fernandez, R. (1981). A methodological note on the estimation of time series. *Review of Economics and Statistics 63*, 471–475.

Gardner, G., A. C. Harvey, and G. D. A. Phillips (1980). An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of Kalman filtering. *Applied Statistics 29*, 311–322.

Harvey, A. C. (1981a). Finite sample prediction and overdifferencing. *Journal of Time Series Analysis 2*, 221–232.

Harvey, A. C. (1981b). *Time Series Models.* New York: John Wiley.

Harvey, A. C. and C. R. McKenzie (1983). Missing observations in dynamic econometric models. In E. Parzen (Ed.), *Proceedings of Symposium on Time Series Analysis of Irregularly Observed Data, Texas A&M University, February 1983.* New York: Springer-Verlag. (forthcoming).

Harvey, A. C. and G. D. A. Phillips (1979). Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika 66*, 49–58.

Harvey, A. C. and R. G. Pierse (1982). *Estimating missing observations in economic time series.* London: LSE. Econometrics Programme Discussion Paper A33.

Jones, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics 22*, 389–395.

Pearlman, J. G. (1980). An algorithm for the exact likelihood of a high-order autoregressive-moving average process. *Biometrika 67*, 232–233.

Rosenberg, B. (1973). Random coefficient models: the analysis of a cross-section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement 2*, 399–428.