THE ESTIMATION OF DYNAMIC MODELS WITH MISSING OBSERVATIONS

A. C. Harvey*

Pedro Luiz Valls Pereira**

ABSTRACT

An ARMA model can be put in state space form and its exact likelihood function calculated by the Kalman filter. The same technique can be extended to handle missing observations, including cases where the data are initially available at an annual level and subsequently become available on a quarterly, or monthly, basis. The Kalman filter enables the likelihood function to be computed for both stock and flow data. Once a suitable model has been fitted, the missing observations may be estimated by "smoothing".

The paper first sets out the Kalman filter approach to missing observations for an ARMA time series model and discusses the implementation of an efficient algorithm. The results are then extended to cover static regression models with ARMA disturbances and dynamic models. A series of Monte Carlo experiments comparing the efficiency of different estimation procedures are reported.

* (London School of Economics)

** (I.M.P.A.)

RESUMO

Um modelo ARMA pode ser escrito em espaço de estado e sua função de verossimilhança exata pode ser calculada pelo filtro de Kalman. Esta técnica pode ser estendida para modelos com falta de observação, incluindo casos onde os dados estão disponíveis em diferentes períodos de agregação, por exemplo, anuais-trimestrais.

O filtro de Kalman é usado para calcular a função de verossimilhança quando os dados são do tipo estoque ou fluxo. As observações faltando podem ser estimadas por suavizamento após o ajuste de um modelo adequado.

Este artigo apresenta o filtro de Kalman para modelo ARMA com falta de observação e discute a implementação de algoritmos eficientes. Os resultados são estendidos para modelos de regressões estatísticas com erros ARMA e para modelos dinâmicos. Um experimento de Monte Carlo é usado para comparar a eficiência dos diversos métodos de estimação.

1. MISSING OBSERVATIONS

Missing observations can occur in time series in a number of different ways. One important case which arises frequently in economics is when the observations are aggregated over time. Thus the basic model may be in terms of observations made at quarterly intervals, but only annual observations are available. Later on observations may become available on a quarterly basis, and the problem is then to combine the annual and quarterly data. The methods proposed here are likely to be most useful for handling mixed observations, since if data are only available at the aggregated level it becomes difficult to identify an appropriate disaggregated model.

Although the discussion is framed entirely in terms of a situation where observations are missing at regular intervals, the proposed techniques also apply to irregular missing observations.

The discussion is restricted to models for single series of observations, i.e. univariate ARMA models, regression models with ARMA disturbances, and dynamic regression models. A method of computing the exact likelihood function for an ARMA model with missing observations is first set out. The algorithm is based on the Kalman filter and it may be regarded as an extension of the one proposed in Gardner, Harvey and Phillips (1980); c.f. Jones (1979). The same technique is then applied to regression models. It would also be possible to extend the theory to (non-simultaneous) systems of equations, but this is not done explicitly.

Previous work in the area of missing observations in econometrics has been carried out by Sargan and Drettakis (1974), Chow and Lim (1971, 1976) and Gilbert (1977). The Sargan-Drettakis paper is primarily concerned with stock variables, although it can

be extended to cover flows including mixed annual and quarterly observations. However, the approach becomes rather unwieldy. Gilbert's paper is primarily concerned with models with independent disturbances. His methods for handling dynamic models are different to ours, and his proposals for handling AR disturbances do not extend to more general ARMA processes in a convenient way. The papers by Chow and Lim are concerned with a slightly different problem, namely estimating the values of missing observations using a related series. The methods in this paper have some bearing on the problem of estimating missing observations in that 'smoothing' can be applied to the ARMA model appropriate for the disaggregated series. Thus missing observations are estimated without using another series.

2. THE KALMAN FILTER

A 'state space' model for a single series of observations takes the form

$$y_t = z_t' \alpha_t + \xi_t \quad (2.1a)$$

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t \quad t=1, \dots, T \quad (2.1b)$$

where α_t is an $s \times 1$ vector of state variables, z_t is an $s \times 1$ vector of known values, T_t and R_t are known matrices of order $s \times s$ and $m \times n$ respectively, ξ_t is a disturbance term with mean zero and variance $\sigma^2 h_t$ and η_t is a vector of disturbances with mean zero and covariance matrix $\sigma^2 Q_t$ which is uncorrelated with ξ_t in all time periods.

Given an initial state vector, a_0 , with covariance matrix $\sigma^2 P_0$, where P_0 is known, the state vector may be updated with the arrival of each new observation. This is effected by means of the Kalman filter, which consists of two sets of equations. These are the predictions equations,

$$a_{t/t-1} = T_t a_{t-1} \quad (2.2a)$$

$$P_{t/t-1} = T_t P_{t-1} T_t' + R_t Q_t R_t' \quad t=1, \dots, T \quad (2.2b)$$

and the updating equations

$$a_t = a_{t/t-1} + P_{t/t-1} z_t' (y_t - z_t a_{t/t-1}') / f_t \quad (2.3a)$$

$$P_t = P_{t/t-1} - P_{t/t-1} z_t' z_t P_{t/t-1} / f_t \quad (2.3b)$$

$$f_t = z_t' P_{t/t-1} z_t + h_t \quad t=1, \dots, T \quad (2.3c)$$

see, for example, Jazwinski (1970), Duncan and Horn (1972) or Harvey and Phillips (1979).

The set of prediction errors,

$$v_t = y_t - z_t' a_{t/t-1} \quad t=1, \dots, T \quad (2.4)$$

play an important role in constructing likelihood functions via the 'prediction error decomposition'.

3. STATE SPACE FORMULATION OF ARMA MODELS WITH MISSING OBSERVATIONS

An ARMA (p,q) model,

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t + \dots + \theta_q \varepsilon_{t-q} \quad t=1, \dots, T \quad (3.1)$$

may be put in state space form by letting $s = \max(p, q+1)$, and defining an $s \times 1$ vector α_t which obeys a transition equation of the form, (2.1b). The matrices T_t and R_t are time invariant, the former depending on $\phi_1, \dots, \phi_p$, the latter in $\theta_1, \dots, \theta_q$; see, for example, Gardner, Harvey and Phillips (1980). The first element in α_t is y_t . Hence the associated measurement equation is of the form (2.1a) with $z_t = (1 \ 0_{t-1}')$ and $h_t = 0$.

The exact likelihood function of a set of T observations, $y_1, \dots, y_T$ may be expressed as a prediction error decomposition, i.e.

$$\begin{aligned} \log L(y) = & -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma^2 \\ & - \frac{1}{2} \sigma^{-2} \sum_{t=1}^T \log f_t \\ & - \frac{1}{2} \sigma^{-2} \sum_{t=1}^T v_t^2 / f_t \end{aligned} \quad (3.2)$$

By setting $a_0 = 0$ and obtaining P_0 from the solution to the equations

$$P_0 = T P_0 T' + R Q R' \quad (3.3)$$

the v_t 's and f_t 's may be computed from the Kalman filter. An algorithm is given in Gardner, Harvey and Phillips (1980).

Missing Observations

In order to simplify matters it will be assumed that observations are only available every m time periods and that $T/m = T^*$ is an integer. If y is a stock variable, $y_m, y_{2m}, \dots, y_T$ are observed. While the transition equation is defined for all $t = 1, \dots, T$, the measurement equation is only defined for $t = m, 2m, \dots, T$. For other values of t , the Kalman filter updating equations are by-passed. The likelihood function then emerges in terms of T^* m -step ahead prediction errors.

For a flow variable, the observations at $t = m, 2m, \dots, T$ are defined by

$$y_t^* = \sum_{j=0}^{m-1} y_{t-j} \quad t=m, 2m, \dots, T \quad (3.4)$$

The state vector must be augmented by a component, $\bullet_t^+ = (y_{t-1}, \dots, y_{t-m+1})'$ which

tion at $t = m, 2m, \dots, T$, is then of the form (2.1a) for $t = m, 2m, \dots, T$, with $h_t = 0$, $z_t = (1 \ 0'_{t-1} \ 1'_{m-1})'$ and $\alpha_t = (\alpha_t^*, \alpha_t^+)'$. The vector α_t^* is defined in the same way as the state vector in the standard state space representation of the ARMA model. The transaction equation with, say, $m = 4$, is

$$\begin{bmatrix} \alpha_t^* \\ \alpha_t^+ \end{bmatrix} = \left[\begin{array}{c|ccc} T^* & 0 & & \\ \hline 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array} \right] \begin{bmatrix} \alpha_{t-1}^* \\ \alpha_{t-1}^+ \end{bmatrix} + \begin{bmatrix} R^* \\ 0 \\ 0 \\ 0 \end{bmatrix} \varepsilon_t \quad t=1, \dots, T \quad (3.5)$$

where T^* , R^* , and ε_t are defined as for an ARMA model with no missing observations.

The definitions of the transition and measurement equations change if the observations are available at unequal intervals but the overall principle remains the same.

When observations are missing at regular intervals the Kalman filter will converge to a steady-state as t increases if the model is stationary and invertible. By monitoring the progress of f_t it is possible to speed up the calculations by fixing P_t and $P_{t/t-1}$ once f_t appears to be constant. A similar device, leading to 'quick recursions', was adopted in Gardner, Harvey and Phillips (1980).

Example

The AR(1) model,

$$y_t = \phi y_{t-1} + \varepsilon_t \quad t=1, \dots, T \quad (3.6)$$

can be put in state space by letting $T = \phi$, $R = 1$, $y_t = \varepsilon_t$ and $z_t = 1$. If y is a stock variable observed every other period, i.e. $m = 2$, then

$$\begin{aligned} \log L(y_2, y_4, \dots, y_T) &= -\frac{T^*}{2} \log 2\pi - \frac{T^*}{2} \log \sigma^2 \\ &\quad - \frac{1}{2} \sum_{\tau=1}^{T^*} \log f_{2\tau} - \frac{1}{2} \sigma^{-2} \sum_{\tau=1}^{T^*} y_{2\tau}^2 / f_{2\tau} \end{aligned}$$

$$\begin{aligned}
&= -\frac{T^*}{2} \log 2\pi - \frac{T^*}{2} \log \sigma^2 - \frac{(T^*-1)}{2} \log (1 + \phi) \\
&\quad - \frac{1}{2} \sigma^{-2} \sum_{\tau=2}^{T^*} (y_{2\tau} - \phi^2 y_{2\tau-2})^2 / (1 + \phi^2) \\
&\quad + \frac{1}{2} \log (1 - \phi^2) - \frac{1}{2} \sigma^{-2} y_2 (1 - \phi^2)^{\frac{1}{2}}
\end{aligned}$$

In this case a steady state is reached after the first observation since $f_{2\tau} = 1 + \phi^2$, $\tau = 2, \dots, T^*$. The prediction errors can therefore be derived without using the Kalman filter.

For flow data, $\alpha_t = (y_t \ y_{t-1})$ and so

$$\alpha_t = \begin{bmatrix} \phi & \alpha_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_t \end{bmatrix} \quad (3.7)$$

The measurement equation is of the form (3.5a) with $z_t^* = (1 \ 1)'$. The starting values for the Kalman filter are

$$a_{1/0} = 0$$

and

$$\begin{aligned}
P_{1/0} &: E(\alpha_1 \ \alpha_1') = E \left[\begin{pmatrix} y_1 \\ y_0 \end{pmatrix} (y_1 \ y_0) \right] \\
&= \frac{1}{1-\phi^2} \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}
\end{aligned} \quad (3.8)$$

If the first observation is $y_2^* = y_2 + y_1$ the first prediction error is

$$v_2 = y_2 + y_1$$

with

$$f_2 = z_2' P_{2/1} z_2 = (1 \ 1) P_{1/0} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{2}{1-\phi}$$

The Kalman filter tends towards a steady state as observations are included. As $\tau \rightarrow \infty$, $\sigma^2 \lim f_{m\tau}$ tends to the variance of the

one-step ahead prediction error for the aggregated series. This value can be obtained by solving an equation given in Amemiya and Wu (1972).

4. STATIC REGRESSION MODELS

Consider a static regression model of the form

$$y_t = x_t' \beta + v_t \quad t=1, \dots, T \quad (4.1)$$

where x_t is a $k \times 1$ vector of observations on explanatory variables, β is a $k \times 1$ vector of parameters and v_t is an ARMA (p,q) disturbance term.

If observations are missing, the likelihood function can be constructed exactly as in the previous section by replacing y_t by $y_t - x_t' \beta$. Thus $\log L$ is computed conditional on β as well as on the ARMA parameters, ϕ and θ . Note that we must have observations in the x 's corresponding to all the observations on y .

5. DYNAMIC REGRESSION MODELS

The treatment of missing observations in the dynamic model

$$y_t = \phi y_{t-1} + \beta x_t + u_t \quad t=1, \dots, T \quad (5.1)$$

is more complicated than in the static case, and a number of different approaches are possible.

The approach favoured here is to cast the model directly into state space form by bringing the exogenous variable, x_t , into the transition equation. Thus if $m = 2$ and the disturbance term is white noise, i.e. $u_t = \epsilon_t$ the transition equation is:

$$\alpha_t = \begin{bmatrix} \phi & 0 \\ 1 & 0 \end{bmatrix} \alpha_{t-1} + \begin{bmatrix} \beta x_t \\ 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_t \quad (5.2)$$

with $\alpha_t = (y_t \ y_{t-1})'$. The measurement is of the form (2.1a) with $h_t = 0$ and $z_t = (1 \ 1)'$ for flow data.

If no disaggregated observations are available the Kalman filter recursions can be started at $t = m$, with each element in a_m set equal to y_m^*/m for flow data. If these values are regarded as fixed, then $p_m = 0$. When some disaggregated observations are available, as in the case of mixed annual and quarterly data, various modifications can be made to this procedure. These are described in a later section.

Conditional on ϕ and β the Kalman filter yields the likelihood and this may then be maximized by numerical optimization. The transition equation, (5.2), can be extended to models where μ_t follows an ARMA process, since the model can always be converted to ARMAX form with restrictions on the parameters. Further lagged values of y_t can also be introduced in a fairly obvious fashion.

6. EFFICIENCY OF ESTIMATORS FOR MIXED ANNUAL/QUARTERLY OBSERVATIONS

Suppose that the last $T/2$ observations are available on a disaggregated basis - say quarterly - while over the period $t = 1, \dots, T/2$ there are only annual observations. The methods proposed above can handle this situation quite easily with $m = 4$ up to $t = T/2$ and $m = 1$ thereafter.

When observations are available at two levels of aggregation one possibility is to discard those at the higher level of aggregation. Thus estimation is carried out with the quarterly data.

The mixture of annual and quarterly observations can be handled quite easily by the Kalman filter. It is assumed that a sui-

table model has been found for the quarterly observations and so the part of the series containing annual observations only has missing observations in terms of a quarterly model. A state space model is therefore set up in the way described in the sections above, and once the quarterly observations become available the measurement equation is simply redefined.

All the Monte Carlo results reported in this section are based on flow data with $T = 80$ and $m = 4$. The switch from annual to quarterly data occurs at $T = 41$.

ARMA Models

One hundred replications from an ARMA (1,1) model

$$y_t = \phi y_{t-1} + \varepsilon_t + \theta \varepsilon_{t-1} \quad \varepsilon_t \sim \text{NID}(0, \sigma^2) \quad (6.1)$$

were simulated with $\phi = 0.5$ and $\theta = 0.1$. The parameters were estimated using (a) only the disaggregated observations, and (b) using all the observations. In both cases full ML was carried out using the Kalman filter to evaluate the likelihood conditional on ϕ and θ . The maximum of the likelihood function was then found using the algorithm of Gill, Murray and Pitfield.

The gain in accuracy of estimation can be judged by the ratio of the RMSE of the disaggregated estimator (quarterly observations only) to the aggregated/disaggregated estimator. For ϕ this ratio was 1.179 while for θ it was 1.098.

Regression Parameters

Estimating a sample mean can be regarded as a special case of (4.1) in which y_t is regressed on a constant term. For stock data, including the annual observations effectively increases the sample size by 25%. However, for flow data aggregation does not lead to any loss of information, and including the annual observations effectively doubles the sample size. More formally, if

$$Y_t = \mu + \varepsilon_t \quad t=1, \dots, T \quad (6.2)$$

where $\varepsilon_t \sim \text{NID}(0, \sigma^2)$, the gain in efficiency from including the annual observations is the flow case in

$$\frac{\text{Var}(\mu^*)}{\text{Var}(\bar{\mu})} = 2$$

Now consider the regression model

$$Y_t = \delta + \beta x_t + \varepsilon_t \quad t = 1, \dots, T \quad (6.3)$$

Gilbert (1977) has shown that estimation with mixed annual and quarterly observations can be carried out efficiently by replacing the missing quarterly observations by their annual averages and applying OLS. Thus

$$\bar{x}_t = x_m^*/m \quad t=1, \dots, m \quad (6.4)$$

and so on for $t=m+1, \dots, T/2$.

If the observations are omitted the loss in efficiency can be quite large, particularly if x is subject to a trend. Consider the case of flow data with $x_t = t$. If terms of $O(T^{-3})$ are omitted from the expressions for the variances of the estimators, β^* and $\bar{\beta}$, it can be shown that $\text{Eff} = 8$. It is therefore vital to use the annual observations.

Dynamic Models

The arguments above suggest that it will be important to include annual observations in the estimation of a dynamic model, (5.1), particularly if x_t is subject to a trend. In order to investigate the gains in efficiency resulting from using both annual and quarterly observations a series of Monte Carlo experiments were carried out. The explanatory variable was generated in three different ways:

- (i) $x_t \sim \text{NID}(0, 4)$
- (ii) $x_t = 0.5 x_{t-1} + n_t$ with $n_t \sim \text{NID}(0, 4)$
- (iii) $x_t = \exp(-0.04t) + \varepsilon_t$ with $\varepsilon_t \sim \text{NID}(0, 0.5)$

The ratio of the RMSE of the disaggregated estimator to the aggregated/disaggregated estimator is shown in Table 1 for each model. Although only 50 replications were used the results nevertheless give a clear idea of the gains which are likely to arise. As expected the gain is greatest for Model (iii).

TABLE 1
RATIO OF RMSE'S OF ESTIMATORS FOR DYNAMIC MODEL

PARAMETER	MODEL		
	(i)	(ii)	(iii)
ϕ	1.09	1.29	1.45
β	1.28	1.30	2.29

If y_t shows a strong seasonal pattern or a strong upward or downward trend, it may be advisable to modify the procedure whereby each of the initial values in a_m is set equal to y_m^*/m . One possibility is to proceed as follows. If y_t is assumed to be fixed and $\varepsilon_2 = \varepsilon_3 = \dots = \varepsilon_m = 0$, the observations are generated from

$$y_t = \phi y_{t-1} + \beta x_t \quad t = 2, \dots, m \quad (6.5)$$

Since

$$\sum_{t=1}^m y_t = y_m^* \quad (6.6)$$

(6.5) and (6.6) give m equations which may be solved for the m unknowns, $y_1, \dots, y_m$. These values are then used to construct the starting vector a_m with $P_m = 0$ as before.

The above device was employed for a model in which x_t took the values $\lambda, 0.5\lambda, 1.5\lambda$, and $.75\lambda$, in the first, second, third and fourth quarters respectively for all years. Table 2 shows the ratio of the RMSE's of each of the annual/quarterly estimators to the quarterly estimator for different values of λ .

Table 2
RATIO OF RMSE'S OF ESTIMATORS WITH DIFFERENT STARTING
VALUES FOR SEASONAL MODEL

	λ		
	1.0	5.0	10.0
Original ϕ	1.39	1.35	1.29
	β 1.08	1.15	1.33
Modified ϕ	1.41	1.36	1.34
	β 1.12	1.17	1.33

Although only 25 replications were used the results nevertheless give an idea of the gain. As expected the gain for the parameter of the lagged endogenous variable decreases with the increase on the seasonal pattern and the parameter of the exogenous variable presents the opposite pattern.

REFERENCES

- AMEMIYA, T. and WU, R.Y. The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 67:628-32, 1972.
- CHOW, G.C. and LIN, A. Best linear unbiased interpolating distribution and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53:372-375, 1971.
- CHOW, G.C. and LIN, A. Best linear unbiased estimation of missing observation in an economic time series. *Journal of the American Statistical Association*, 71:719-721, 1976.
- DUNCAN, A.B. and HORN, S.D. Linear dynamic regression estimation from the viewpoint of regression analysis. *Journal of the American Statistical Association*, 61:815-821, 1972.
- GARDNER, G., HARVEY, A.C. and PHILLIPS, G.D.A. An algorithm for exact maximum likelihood estimation of ARMA models by means of Kalman filtering. *Applied Statistics* (forthcoming), 1980.
- GILBERT, C.L. Regression using mixed annual and quarterly data. *Journal of Econometrics*, 5:221-240, 1977.
- HARVEY, A.C. and PHILLIPS, G.D.A. Maximum likelihood estimation of regression models with autoregressive - moving average disturbances. *Biometrika*, 66:49-58, 1979.
- JONES, R.H. Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, (forthcoming), 1980.
- SARGAN, J.D. and DRETTAKIS, E.G. Missing data in an autoregressive model. *International Economic Review*, 15:39-58, 1974.