

A note on Kalman filtering for the seasonal moving average model

BY ROBERT KOHN AND CRAIG F. ANSLEY

Graduate School of Business, University of Chicago, Chicago, Illinois, U.S.A.

SUMMARY

We show that if we apply the Kalman filter to a pure seasonal moving average model with no missing observations, then, by using a result of Ansley (1979), we can obtain significant computational savings.

Some key words: Kalman filter; Likelihood; Moving average; Seasonal model.

In recent years, the Kalman filter has been used extensively to compute the exact likelihood of autoregressive-moving average (ARMA) and autoregressive integrated moving average (ARIMA) time series models. See, for example, Akaike (1978), Gardner, Harvey & Phillips (1980), Jones (1980) and Harvey & Pierse (1984).

Using the Cholesky factorization, Ansley (1979) gave an alternative computationally efficient algorithm for computing the likelihood of an ARMA model when there are no missing observations. In particular, he showed how to obtain substantial computational savings for the seasonal moving average model by using the Cholesky factorization. By applying his Theorem 4.1, we can obtain similar savings for an algorithm based on the Kalman filter. An extension of the result to a seasonal ARMA model is also given.

We consider the zero mean seasonal moving average model

$$y(t) = \Theta(B^s) \theta(B) e(t), \quad (1)$$

where $e(t)$ is a sequence of $N(0, \sigma^2)$ random variables, s is the seasonal period, and

$$\theta(B) = 1 + \sum_{j=1}^q \theta_j B^j, \quad \Theta(B) = 1 + \sum_{j=1}^Q \Theta_j B^j.$$

For convenience, we define

$$v(B) = 1 + \sum_{j=1}^r v_j B^j = \theta(B) \Theta(B^s),$$

where $r = q + sQ$. Then following Gardner *et al.* (1980) or Jones (1980), we can rewrite (1) in state space form as

$$y(t) = h'x(t), \quad x(t+1) = Fx(t) + ge(t+1), \quad (2)$$

where the state vector $x(t)$ is $(r+1) \times 1$ with

$$x_{j+1}(t) = \sum_{i=j}^r v_i e(t+j-i) \quad (j = 0, \dots, r),$$

where $v_0 = 1$, $h = (1, 0, \dots, 0)'$, $g = (1, v_1, \dots, v_r)'$ and F is a $(r+1) \times (r+1)$ matrix having 1 in the $(i, i+1)$ position ($i = 1, \dots, r$), and zeros elsewhere.

For any $t, s \geq 1$, we define $x(t|s) = E\{x(t)|y(1), \dots, y(s)\}$ and

$$S(t|s) = E\{[x(t) - x(t|s)][x(t) - x(t|s)]'\} / \sigma^2.$$

Then (Gardner *et al.*, 1980; Akaike, 1978), a typical step of the Kalman filter applied to the state space representation (2) can be broken into three parts:

- (a) $x(t|t-1) = Fx(t-1|t-1), S(t|t-1) = FS(t-1|t-1)F' + gg'$;
- (b) $\varepsilon(t) = y(t) - x_1(t|t-1), R(t) = S_{11}(t|t-1)$;
- (c) $x(t|t) = x(t|t-1) + S(t|t-1)h\varepsilon(t)/R(t),$ (3)
 $S(t|t) = S(t|t-1) - S(t|t-1)hh'S(t|t-1)/R(t).$ (4)

Because of the special structure of F , only index shifting is involved in obtaining $Fx(t-1|t-1)$ and $FS(t-1|t-1)F'$ in (a) above. Also gg' can be computed once and for all at the initialization of the Kalman filter. Thus (a) and (b) require a minimal number of arithmetic operations; the real cost lies in (c). Now $S(t|t-1)h$, which appears in both (3) and (4), is the first column of $S(t|t-1)$. By identifying structural zeros in the first column of $S(t|t-1)$, we can substantially reduce the computations required in (c).

Suppose now that we have n consecutive observations from (1). Then

$$\varepsilon(t) = y(t) - E\{y(t) | y(1), \dots, y(t-1)\},$$

with $\varepsilon(1) = y(1)$. It follows that we can write

$$y(t) = \sum_{j=1}^t l(t, j) \varepsilon(j),$$

with $l(t, t) = 1$ for all t .

We now obtain a useful representation of the first row and column of $S(t|t-1)$.

LEMMA 1. *The first row and column of $S(t|t-1)$ is*

$$R(t) [1, l(t+1, t), l(t+2, t), \dots, l(t+r, t)].$$

Proof. From the above $x_1(t) = y(t), x_2(t) = y(t+1) - e(t+1)$, and, for $j = 3, \dots, r+1$,

$$x_j(t) = y(t+j-1) - \left\{ e(t+j-1) + \sum_{i=1}^{j-2} v_j e(t+j-1-i) \right\}.$$

Therefore,

$$x_1(t) - x_1(t|t-1) = \varepsilon(t),$$

$$x_2(t) - x_2(t|t-1) = \varepsilon(t+1) + l(t+1, t) \varepsilon(t) - e(t+1),$$

and, for $j = 3, \dots, r+1$,

$$x_j(t) - x_j(t|t-1) = \varepsilon(t+j-1) + \sum_{i=1}^{j-1} l(t+j-1, t+j-1-i) \varepsilon(t+j-1-i) - \left\{ e(t+j-1) + \sum_{i=1}^{j-2} v_i e(t+j-1-i) \right\}.$$

The required result now follows because the j th element of the first row of $S(t|t-1)$ is

$$E[\varepsilon(t) \{x_j(t) - x_j(t|t-1)\}].$$

Now Ansley (1979) showed that $l(t, j) = 0$ for values of t and j satisfying

$$t - j = hs + i \quad (h = 0, 1, \dots; i = q + 1, \dots, s - 1),$$

$$t = rs + k \quad (r = 0, 1, \dots; k = 1, \dots, s - q).$$

We can make (c) of the Kalman filter step more efficient by using this information.

Also for all t , the first row and column of $S(t|t)$ is zero because $x_1(t) = y(t)$ is known at time t . Further (Gardner *et al.*, 1980), $-2 \log$ likelihood is equal to

$$\sum_{t=1}^n \varepsilon(t)^2/R(t) + \log \left\{ \prod_{t=1}^n R(t) \right\}. \quad (5)$$

Consider now the mixed seasonal ARMA process $\phi(B)y(t) = v(B)e(t)$, where $v(B)$ and $e(t)$ are defined above and $\phi(B) = 1 + \phi_1 B + \dots + \phi_p B^p$. We assume that all roots of $\phi(B) = 0$ lie outside the unit circle.

The above approach cannot be used to compute the likelihood for $y(t)$. However, by following §4 of Ansley (1979) we make the backward differencing transformation

$$z(t) = \begin{cases} \phi(B^{-1})y(t) & (t = 1, \dots, n-p), \\ y(t) & (t = n-p+1, \dots, n). \end{cases}$$

Thus $z(t)$ is a pure moving average process having the structure (1) for $t = 1, \dots, n-p$. Therefore, the pattern of structural zeros obtained above can be exploited for the Kalman filter applied to $z(t)$ ($t = 1, \dots, n-p$). This gives $\varepsilon(t)$ and $R(t)$ for those t . One way of obtaining $\varepsilon(t)$ and $R(t)$ for $t = n-p+1, \dots, n$ is to use Ansley's (1979) Cholesky method to obtain the last p rows of the Cholesky factorization. To do so we need to reconstruct rows $n-p-r-1, \dots, n-p$ of the Cholesky factor L of Ansley (1979). For the t th row ($t = n-p-r-1, \dots, n-p$), the $(t, t-r), \dots, (t, t)$ elements of L are given by $\{l(t, t-r)R(t-r)^{\frac{1}{2}}, \dots, l(t, t)R(t)^{\frac{1}{2}}\}$, and by Lemma 1 these have already been obtained by the Kalman filter recursion. All other elements of the t th row are zero.

It can be shown that the number of arithmetic operations required to compute the likelihood using the Kalman filter is the same as that required by the Cholesky decomposition method of Ansley (1979).

REFERENCES

- AKAIKE, H. (1978). Covariance matrix computation of the state variable of a stationary Gaussian process. *Ann. Inst. Statist. Math.* **30**, 499–504.
- ANSLEY, C. F. (1979). An algorithm for the exact likelihood of a mixed autoregressive moving average process. *Biometrika* **66**, 59–65.
- GARDNER, G., HARVEY, A. C. & PHILLIPS, G. D. A. (1980). An algorithm for exact maximum likelihood estimation of autoregressive moving average models by means of Kalman filtering. *Appl. Statist.* **29**, 311–22.
- HARVEY, A. C. & PIERSE, R. G. (1984). Estimating missing observations in economic time series. *J. Am. Statist. Assoc.* **79**, 125–31.
- JONES, R. H. (1980). Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22**, 389–95.

[Received March 1984. Revised May 1984]