Available online at www.sciencedirect.com
SCIENCE DIRECT®
Statistics & Probability Letters 70 (2004) 119–125
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

The ARMA model in state space form

Piet de Jong[a], Jeremy Penzer[b,*]

[a]*Department of Actuarial Studies, Macquarie University, NSW 2109, Australia*
[b]*Department of Statistics, London School of Economics, Houghton St, WC2A 2AE London, UK*

Received 19 September 2002; received in revised form 15 June 2004

Abstract

This article explores alternative state space representations for ARMA models. We advocate representations that have minimal state order and appealing Kalman filter steady state properties. We derive expressions for smoother output and describe concrete connections to classical infinite sample representations.

Keywords: Filter steady state; Kalman filter smoother; State space model; Time series

1. Introduction

State space forms are used for prediction, smoothing and likelihood evaluation (Schweppe, 1965; Anderson and Moore, 1979; Harvey and Phillips, 1979). There are several practical state space forms for an ARMA process. For invertible models, the corresponding Kalman filter recursions reach or converge to steady state. We advocate the use of a representation with minimal state order and transparent steady state properties. The converged quantities reveal the relationship between Kalman filtering and the classical method of conditioning on presample values. Our approach establishes the nature of smoothing algorithm quantities. The finite sample generalization of methods for dealing with explanatory variables and transfer functions is also discussed.

*Corresponding author. Fax: +44-20-7955-7416.
E-mail address: j.penzer@ise.ac.uk (J. Penzer).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2004.08.006

The results mentioned above can be derived using any state space representation of an ARMA process. However, using the approach advocated in this paper is particularly straightforward, illustrating the usefulness of the correlated form of the state space model and the judicious labelling of state space disturbances.

2. The ARMA model and the state space form

The ARMA(p,q) model for a time series y_t is

$$y_t = \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q},$$

where ε_t is a sequence of uncorrelated random variables with zero mean and variance σ^2. The model is written in the usual lag operator notation as $\phi(B)y_t = \theta(B)\varepsilon_t$. The case where $\phi(B)$ has roots inside or on the unit circle, and hence the model maybe nonstationary (ARIMA), is not excluded.

The time invariant state space form is

$$y_t = Z\alpha_t + G\varepsilon_t, \tag{1}$$

$$\alpha_{t+1} = T\alpha_t + H\varepsilon_t, \quad t = 1, \ldots, n, \tag{2}$$

where $\varepsilon_t \sim (0, \sigma^2 I)$, $\alpha_1 \sim (a_1, \sigma^2 P_1)$ and the ε_t and α_1 are mutually uncorrelated. The matrices Z, T, G and H are nonrandom and typically depend on hyperparameters. For a univariate model with an $m \times 1$ state vector α_t and $s \times 1$ vector of errors ε_t, the matrices Z, T, G and H are $1 \times m$, $m \times m$, $1 \times s$ and $m \times s$, respectively.

Pearlman (1980) puts forward the ARMA state space representation below

$$Z = (1, 0, \ldots, 0), \quad G = 1, \quad T = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ \vdots & 0 & \cdots & 0 & 1 \\ \phi_m & 0 & \cdots & \cdots & 0 \end{pmatrix} \quad \text{and } H = \begin{pmatrix} \theta_1 + \phi_1 \\ \theta_2 + \phi_2 \\ \vdots \\ \vdots \\ \theta_m + \phi_m \end{pmatrix}, \tag{3}$$

where the size of the state vector is $m = \max(p,q)$. For example, the representation for an ARMA(1, 1) is

$$y_t = \alpha_t + \varepsilon_t,$$
$$\alpha_{t+1} = \phi\alpha_t + (\theta + \phi)\varepsilon_t.$$

We refer to (3) as the $\max(p,q)$ representation. Pearlman (1980) gives no references for $\max(p,q)$ form and it is largely ignored in the statistical literature. Two exceptions are Burridge and Wallis (1988) who consider the implications for classical prediction theory and Harvey (1989, p. 103) who discusses the representation for the MA(1) case. Note that the ε_t in (1) and (2) is the same as the ε_t noise in the ARMA model. Also $GH' \neq 0$ implying correlated measurement and state noise.

If $\alpha_{j,t+1}$ is the jth element of α_{t+1} then, for $j = 1, \ldots, m$,

$$\alpha_{j,t+1} = \phi_j y_t + \cdots + \phi_m y_{t-m+j} + \theta_j \varepsilon_t + \cdots + \theta_m \varepsilon_{t-m+j}. \tag{4}$$

In the literature (Brockwell and Davis, 1987; Harvey, 1993; Box et al., 1994; Hamilton, 1994) the max(p, q) representation has been overlooked in favour of forms in which the state vector is of length $m = \max(p, q+1)$. In one version (Harvey, 1993, p. 96), Z and T are as above but $G = 0$ and $H = (1, \theta_1, \ldots, \theta_{m-1})'$. An alternative (Box et al., 1994, p. 163) is

$$Z = (1, 0, \ldots, 0), \quad G = 0, \quad T = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & 1 \\ \phi_m & \phi_{m-1} & \cdots & \cdots & \phi_1 \end{pmatrix} \text{ and } H = \begin{pmatrix} 1 \\ \psi_1 \\ \vdots \\ \vdots \\ \psi_{m-1} \end{pmatrix}, \tag{5}$$

where ψ_j are the leading coefficients in the polynomial expansion of $\theta(B)/\phi(B)$ for $j = 1, 2, \ldots$. The literature does not mention that representation (5) can be adapted to $m = \max(p, q)$ form by leaving Z and T unchanged but taking $G = 1$ and $H = (\psi_1, \ldots, \psi_m)'$. For example, in this new form an ARMA(2, 2) process would be represented as

$$y_t = (1 \quad 0)\, \alpha_t + \varepsilon_t,$$
$$\alpha_{t+1} = \begin{pmatrix} 0 & 1 \\ \phi_2 & \phi_1 \end{pmatrix} \alpha_t + \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \varepsilon_t,$$

where $\psi_1 = \theta_1 + \phi_1$ and $\psi_2 = \theta_2 + \phi_1\theta_1 + \phi_1^2 + \phi_2$. This representation has the useful property that at time $t+1$, the jth element of the state is a predictor of y_{t+j} based on information available at time t,

$$\alpha_{j,t+1} = E(y_{t+j} | y_t, \ldots, y_{t-m+1}, \varepsilon_t, \ldots, \varepsilon_{t-m+1}), \quad j = 1, \ldots, m. \tag{6}$$

3. Steady state properties of Kalman filtering

For $t = 1, \ldots, n$, the Kalman filter transforms the observations y_t to innovations v_t via the equations,

$$\begin{aligned} v_t &= y_t - Za_t, & F_t &= ZP_tZ' + GG', \\ & & K_t &= (TP_tZ' + HG')F_t^{-1}, \\ a_{t+1} &= Ta_t + K_t v_t, & P_{t+1} &= TP_tL_t' + HJ_t', \end{aligned} \tag{7}$$

where $L_t = T - K_tZ$, $J_t = H - K_tG$. In certain circumstances, discussed below, the Kalman filter converges to a steady state, meaning the matrices F_t, K_t and P_t approach constancy.

For the Pearlman max(p, q) representation the steady state is convenient since the state covariance matrix approaches zero. In fact, $P_t \to 0$, $F_t \to 1$ and $K_t \to H$. Thus, once the steady state is attained, $v_t = \varepsilon_t$ in mean square and the Kalman filter (7) reduces to the computation

$$\varepsilon_t = y_t - Z\alpha_t = \frac{\phi(B)}{\theta(B)}\, y_t.$$

In the steady state, the Kalman filter amounts to the computation of the true innovation ε_t and the recursions for F_t, K_t and P_t in (7) can be discarded. Prior to the steady state, the Kalman filter recursions effect an exact initialization corresponding to the finite history. From this perspective, techniques such as "back forecasting" or discarding the stretch of data used for initialization (Box et al., 1994, p. 227) are unwieldy complications which yield only approximate results.

A brief description of the circumstances in which the Kalman filter reaches its steady state is given below. A more detailed explanation is given in the appendix. For an alternative approach based on the Ricatti equation see Burridge and Wallis (1988). Define $Y_{t,-\infty} = [y_t, \ldots, y_1, \ldots]$, the linear span of the entire past of the series, and put $Y_t = [y_t, \ldots, y_1]$. Then, for an invertible model, Eq. (4) indicates $\alpha_t \in Y_{t-1,-\infty}$. For large t, predictions of the state based on Y_{t-1} are approximately equal to those based on $Y_{t-1,-\infty}$. Hence, the components of α_t can be predicted with negligible error, that is,

$$a_t = E(\alpha_t | Y_{t-1}) \approx E(\alpha_t | Y_{t-1,-\infty}) = \alpha_t.$$

Since $P_t = \sigma^{-2}\,\mathrm{var}(\alpha_t - a_t)$, it follows that $P_t \to 0$ in turn implying $F_t \to 1$ and $K_t \to H$ as $t \to \infty$. A similar result holds for the $\max(p, q)$ representation developed in Section 2 from the form used by Box et al. (1994). For the new representation, we see from (6), $\alpha_t \in Y_{t-1,-\infty}$ and hence $P_t \to 0$.

The steady state Kalman filter properties of $\max(p, q+1)$ representations are less easy to work with. For these forms $\alpha_t \in Y_{t,-\infty}$, so the state estimate does not converge to the state. For example, with the representation used by Harvey (1993), $P_t \to HH'$, $F_t \to 1$ and $K_t \to (\phi_1 + \theta_1, \ldots, \phi_{m-1} + \theta_{m-1}, \phi_m)'$. Although all our discussion is stated in terms of scalar y_t, no new issues arise when moving to the vector ARMA model.

4. Steady state properties for smoothing

Under the Pearlman $\max(p, q)$ representation, smoothing quantities also converge to convenient and readily interpretable constructs. The smoothing filter (De Jong, 1988; Kohn and Ansley, 1989), corresponding to the general state space representation (1) and (2), takes the form, for $t = n, \ldots, 1$,

$$u_t = F_t^{-1} v_t - K_t' r_t, \quad M_t = F_t^{-1} + K_t' N_t K_t,$$
$$r_{t-1} = Z' u_t + T' r_t, \quad N_{t-1} = Z' F_t^{-1} Z + L_t' N_t L_t,$$

where $r_n = 0$ and $N_n = 0$. The smoothing errors u_t play a role in smoothing analogous to the innovations in Kalman filtering. The smoothing errors and associated r_t, which have the same dimensions as the state, contain information about departures from the model (De Jong and Penzer, 1998). The smoothing errors have the interpolation characterization $u_t = \{\mathrm{var}(y_t - \tilde{y}_t)\}^{-1}(y_t - \tilde{y}_t)$, where $\tilde{y}_t = E(y_t | y_1, \ldots, y_{t-1}, y_{t+1}, \ldots, y_n)$.

A convenient expression for the smoothing errors is readily derived using the $\max(p,q)$ representation. Assuming the Kalman filter is in steady state at t, the smoothing filter implies

$$r_{t-1} = Z'\varepsilon_t + (T' - Z'H')r_t = \begin{pmatrix} \varepsilon_t - \theta_1 r_{1,t} - \cdots - \theta_m r_{m,t} \\ r_{1,t} \\ \vdots \\ r_{m-1,t} \end{pmatrix}.$$

Repeated substitution, using the fact that $r_{i,t-1} = r_{i-1,t}$ for $i = 2,\ldots,m$, yields $r_{1,t-1} = \{1/\theta(B^{-1})\}\varepsilon_t$ and hence

$$r_{t-1} = \frac{1}{\theta(B^{-1})}\begin{pmatrix} \varepsilon_t \\ \varepsilon_{t+1} \\ \vdots \\ \varepsilon_{t+m-1} \end{pmatrix},$$

where the ε_t are interpreted as zero if $t>n$. Substituting this relation into the expression for the smoothing errors, $u_t = \varepsilon_t - H'r_t$, yields

$$u_t = \varepsilon_t - \phi_1\varepsilon_{t+1} - \cdots - \phi_m\varepsilon_{t+m} - \theta_1 u_{t+1} - \cdots - \theta_m u_{t+m} = \frac{\phi(B^{-1})}{\theta(B^{-1})}\,\varepsilon_t. \tag{8}$$

This expression is exact provided the steady state of the Kalman filter at t, and does not require t to be well below n nor does it require stationarity. Substituting $\varepsilon_t = \{\phi(B)/\theta(B)\}y_t$ yields the familiar expression from the Wiener–Kolmogorov infinite sample theory (Whittle, 1983)

$$u_t = \frac{\phi(B^{-1})\phi(B)}{\theta(B^{-1})\theta(B)}\,\varepsilon_t.$$

Thus, provided the Kalman filter has attained the steady state, the exact finite sample interpolation errors u_t for an ARMA(p,q), as computed by the Kalman filter smoother, are equal to the infinite sample interpolation errors for $t \leqslant n-m$. There is no requirement for stationarity nor smoothing filter convergence.

5. Explanatory variables and transfer function modelling

A regression model with ARMA errors can be written as $y_t = X_t'\beta + w_t$, where X_t is a vector of regressors and w_t is the error process. If w_t is an AR(p) with known parameters, the Cochrane–Orcutt transformation (Johnston, 1984), $y_t^* = \phi(B)y_t$ and $X_t^* = \phi(B)X_t$, is used to reduce the problem of estimating β to ordinary least squares. General ARMA(p,q) errors can be dealt with using an approach based on augmented filtering. The Kalman filter (7) is augmented with the recursions

$$X_t^{*'} = X_t' - ZA_t, \qquad A_{t+1} = TA_t + K_t X_t',$$

where $A_1 = 0$ (Rosenberg, 1973; Wecker and Ansley, 1983; De Jong, 1991). Using the arguments of Section 3, once steady state is attained, augmented filtering performs the transformation

$$X_t^{*'} = \frac{\phi(B)}{\theta(B)} X_t,$$

which is the generalized Cochrane–Orcutt procedure.

The results of Section 3 and augmented Kalman filtering also shed light on transfer function modelling. A transfer function model is $y_t = \psi(B)x_t + \varepsilon_t$, where both y_t and x_t are observed and ε_t is a white noise disturbance. The aim is to estimate the transfer function $\psi(B)$. In general, x_t is assumed to be an ARMA process $\phi(B)x_t = \theta(B)\eta_t$. One proposal (Box et al., 1994, p. 417) is to use prewhitening, that is, apply the linear filter $\phi(B)/\theta(B)$ to both x_t and y_t yielding

$$v_t = \frac{\phi(B)}{\theta(B)} y_t = \psi(B)\eta_t + \frac{\phi(B)}{\theta(B)} \varepsilon_t.$$

Preliminary estimates of the coefficients in $\psi(B)$ are then calculated using the cross correlation function between the v_t and η_t. The transformation $\phi(B)/\theta(B)$ can be applied to y_t using the Kalman filter. Applying the augmented Kalman filter under this setup yields $\eta_t = x_t^* = \{\phi(B)/\theta(B)\}x_t$. Provided convergence, the Kalman filter calculates the prewhitened series v_t and $x_t^* = \eta_t$ exactly. Prior to convergence, the v_t and x_t^* are uncorrelated but heteroskedastic. Hence, for these initial observations the cross product terms of the form $v_t\eta_{t-k}$ in the cross covariance estimate have to be standardized using the filter output.

6. Conclusion

This paper deals with state space representations for ARMA models that have been overlooked in the time series literature. These representations have both computational and conceptual advantages; after the processing of an initial stretch of the data, the Kalman filter collapses to computing the exact state and errors $\varepsilon_t = \{\phi(B)/\theta(B)\}y_t$. Collapsing is analogous to augmented Kalman filtering reducing to ordinary Kalman filtering and emphasizes that the uncollapsed part performs exact initialization. Steady state results for smoothing provide a useful interpretation of smoother output. Transfer function modelling can also be thought of in terms of steady state filtering.

Appendix

The transition equation (2) can be rearranged to give

$$\alpha_{t+1} = (T - HZ)\alpha_t + Hy_t = A\alpha_t + Hy_t = A^k\alpha_{t+1-k} + \sum_{i=0}^{k} A^i Hy_{t-i}, \tag{9}$$

where $A = T - HZ$. It can be shown that for the Pearlman $\max(p, q)$ representation $|I - Az| = \theta(z)$ and so for an invertible model $|I - Az| \neq 0$ for $|z| < 1$. This implies that $|A - \lambda I| \neq 0$ for $|\lambda| > 1$

so $A^k \to 0$ as $k \to \infty$. Thus from (9)

$$\alpha_t = \sum_{i=1}^{\infty} \pi_i y_{t-i}, \tag{10}$$

where $\pi_i = A^{i-1}H$. Recalling that $a_t = E(\alpha_t | Y_{t-1})$ and taking conditional expectations in (10) yields

$$\alpha_t - a_t = \sum_{i=t}^{\infty} \pi_i [y_{t-i} - E(y_{t-i} | Y_{t-1})]$$

which converges to zero as $t \to \infty$. Thus, $a_t \to \alpha_t$ and $P_t \to 0$ as required.

References

Anderson, B.D.O., Moore, J.B., 1979. Optimal Filtering. Prentice-Hall, Englewood Cliffs, NJ.
Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. Time Series Analysis: Forecasting and Control, third ed. Prentice-Hall, Englewood Cliffs, NJ.
Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Models. Springer, New York.
Burridge, P., Wallis, K., 1988. Prediction theory for autoregressive-moving average processes. Econometric Rev. 7 (1), 65–95.
De Jong, P., 1988. The likelihood for a state space model. Biometrika 75 (1), 165–169.
De Jong, P., 1991. The diffuse Kalman filter. Ann. Statist. 19 (2), 1073–1083.
De Jong, P., Penzer, J.R., 1998. Diagnosing shocks in time series. J. Amer. Statist. Assoc. 93 (442), 796–806.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton.
Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
Harvey, A.C., 1993. Time Series Models, second ed. Harvester Wheatsheaf, London.
Harvey, A.C., Phillips, G.D.A., 1979. Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. Biometrika 66 (1), 49–58.
Johnston, J., 1984. Econometric Methods, third ed. McGraw-Hill, Singapore.
Kohn, R., Ansley, C.F., 1989. A fast algorithm for signal extraction, influence and cross-validation in state space models. Biometrika 76 (1), 65–79.
Pearlman, J.G., 1980. An algorithm for the exact likelihood of a high-order autoregressive-moving average process. Biometrika 67 (1), 232–233.
Rosenberg, B., 1973. Random coefficient models: the analysis of a cross-section of time series by stochastically convergent parameter regression. Ann. Econom. Soc. Meas. 2, 399–428.
Schweppe, F., 1965. Evaluation of likelihoods for Gaussian signals. IEEE Trans. Inform. Theory 11, 61–70.
Wecker, W., Ansley, C.F., 1983. The signal extraction approach to nonlinear regression and spline smoothing. J. Amer. Statist. Assoc. 78, 81–89.
Whittle, P., 1983. Prediction and Regulation, second, revised ed. Blackwell, Oxford.