PREDICTION-ERROR DECOMPOSITION OF THE LIKELIHOOD FUNCTION OF AN ARMA MODEL

The likelihood function associated with a dynamic time series model can be expressed in terms of a prediction-error decomposition. The normal probability density function $N(y_1, \ldots, y_T)$ of a sample of T observations can be written as the product of a sequence of conditional density functions. Thus

$$N(y_1, \ldots, y_T) = N(y_1)N(y_2|y_1)\cdots N(y_T|y_1, \ldots, y_{T-1}). \tag{1}$$

Let $\mathcal{I}_t = \{y_1, \ldots, y_t, \mathcal{I}_0\}$ denote all of the information regarding the density function that is available at time t, including previously observed data points and the *a priori* information. Then the decomposition of the density function can be represented by

$$N(y_1, \ldots, y_T; \mathcal{I}_0) = N(y_1; \mathcal{I}_0)\prod_{t=2}^{T} N(y_t|\mathcal{I}_{t-1}). \tag{2}$$

Given that it is a normal density function, the generic factor $N(y_t|\mathcal{I}_{t-1})$ of this decomposition is characterised completely by the conditional mean $E(y_t|\mathcal{I}_{t-1}) = \hat{y}_{t|t-1}$ and the conditional dispersion matrix $D(y_t|\mathcal{I}_{t-1}) = F_t$, which is the dispersion of the prediction error.

The time series model in question can often be represented via a state space model comprising two equations:

$$y_t = H\xi_t + \eta_t, \qquad \textit{Observation Equation} \tag{3}$$

$$\xi_t = \Phi\xi_{t-1} + \nu_t, \qquad \textit{Transition Equation} \tag{4}$$

where y_t is the observation on the system and ξ_t is the state vector. The observation error η_t and the state disturbance ν_t are mutually uncorrelated random vectors of zero mean with dispersion matrices

$$D(\eta_t) = \Omega \qquad \text{and} \qquad D(\nu_t) = \Psi. \tag{5}$$

The Kalman-filter equations determine the state-vector estimates $x_{t|t-1} = E(\xi_t|\mathcal{I}_{t-1})$ and $x_t = E(\xi_t|\mathcal{I}_t)$ and their associated dispersion matrices $P_{t|t-1}$ and P_t. From $x_{t|t-1}$, the prediction $\hat{y}_{t|t-1} = Hx_{t|t-1}$ is formed which has a dispersion matrix F_t. A summary of these equations is as follows:

$$E(\xi_t|\mathcal{I}_{t-1}) = x_{t|t-1} = \Phi x_{t-1}, \qquad \textit{State Prediction} \tag{6}$$

$$D(\xi_t|\mathcal{I}_{t-1}) = P_{t|t-1} = \Phi P_{t-1}\Phi' + \Psi, \qquad \textit{Prediction Dispersion} \tag{7}$$

$$E(y_t|\mathcal{I}_{t-1}) = \hat{y}_{t|t-1} = Hx_{t|t-1}, \qquad \textit{Observation Prediction} \tag{8}$$

$$D(y_t|\mathcal{I}_{t-1}) = F_t = HP_{t|t-1}H' + \Omega, \qquad \textit{Error Dispersion} \tag{9}$$

$$K_t = P_{t|t-1}H'F_t^{-1}, \qquad \textit{Kalman Gain} \tag{10}$$

$$E(\xi_t|\mathcal{I}_t) = x_t = x_{t|t-1} + K_te_t, \qquad \textit{State Estimate} \tag{11}$$

$$D(\xi_t|\mathcal{I}_t) = P_t = (I - K_tH)P_{t|t-1}. \qquad \textit{Estimate Dispersion} \tag{12}$$

The error from predicting y_t on the basis of the information available at time $t-1$ is $e_t = y_t - Hx_{t|t-1}$.

To begin the recursive filtering operation, values are needed for the initial estimate x_0 of the state vector ξ_0 at time $t = 0$ and for its dispersion matrix $D(\xi_0) = P_0$. On the assumption that the process is stationary, these are provided by its unconditional moments. Taking expectations in the transition equation gives $E(\xi_0) = x_0 = 0$. The dispersion of the transition equation is given by $D(\xi_t) = \Phi D(\xi_{t-1})\Phi' + D(\nu_t)$; and the assumption of stationarity implies that $D(\xi_t) = D(\xi_{t-1}) = P_0$. Thus the initial state dispersion matrix is provided by the solution of the equation

$$D(\xi_0) = P_0 = \Phi P_0 \Phi' + \Psi. \tag{13}$$

The log of the likelihood of the sample $y_1, \ldots, y_T$ can be expressed in terms of the prediction errors:

$$\begin{aligned} &L(y_1, \ldots, y_T | \xi_0, P_0, \Phi, \Psi) \\ &\quad = -\frac{1}{2}\log 2\pi - \frac{1}{2}T\log\sigma^2 - \frac{1}{2}\sum_t \log|F_t| - \frac{1}{2}\sum_t e_t' F_t^{-1} e_t. \end{aligned} \tag{14}$$

Now consider an ARMA model in the form of $\alpha(L)y(t) = \mu(L)\varepsilon(t)$. The model can be written in state-space from in a variety of ways. A convenient way is to specify transition equation $\xi_t = \Phi\xi_{t-1} + \eta_t$ as

$$\begin{bmatrix} \xi_0(t) \\ \xi_1(t) \\ \vdots \\ \xi_{r-2}(t) \\ \xi_{r-1}(t) \end{bmatrix} = \begin{bmatrix} -\alpha_1 & 1 & 0 & \ldots & 0 \\ -\alpha_2 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\alpha_{r-1} & 0 & 0 & \ldots & 1 \\ -\alpha_r & 0 & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} \xi_0(t-1) \\ \xi_1(t-1) \\ \vdots \\ \xi_{r-2}(t-1) \\ \xi_{r-1}(t-1) \end{bmatrix} + \begin{bmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_{r-2} \\ \mu_{r-1} \end{bmatrix} \varepsilon(t). \tag{15}$$

In that case, the observation equation is provided by

$$y(t) = [1, 0, \ldots, 0]\xi(t). \tag{16}$$

Here there is $H = [1, 0, \ldots, 0]$ whilst the observation error is $\eta_t = 0$, which implies that $D(\eta_t) = \Omega = 0$.

We shall illustrate this specification with the first-order moving-average MA(1) model model $y(t) = \varepsilon(t) - \theta\varepsilon(t-1)$. For this model, the transition equation is

$$\begin{bmatrix} \xi_0(t) \\ \xi_1(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \xi_0(t-1) \\ \xi_1(t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ -\theta \end{bmatrix} \varepsilon(t) \tag{17}$$

The estimate of the initial state vector is $x_0 = 0$. The initial state dispersion matrix is obtained by solving the equation

$$\begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}_0 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}_0 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + \sigma_\varepsilon^2 \begin{bmatrix} 1 & -\theta \\ -\theta & \theta^2 \end{bmatrix} \tag{18}$$

which gives

$$P_0 = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}_0 = \sigma_\varepsilon^2 \begin{bmatrix} 1+\theta^2 & -\theta \\ -\theta & \theta^2 \end{bmatrix}. \tag{19}$$

Observe that the initial state prediction is $x_{1|0} = 0$, whilst its dispersion is $P_{1|0} = P_0$. The prediction error of the observation is $e_1 = y_1$, and the corresponding prediction-error dispersion is $F_1 = \sigma_\varepsilon^2(1+\theta^2)$. This is the familiar unconditional variance of an MA(1) process.

The Kalman gain at this stage is $K_1 = [1, \theta/(1+\theta)]'$ and so the state estimate and its dispersion are

$$x_1 = \begin{bmatrix} y_1 \\ \dfrac{y_1\theta}{1+\theta^2} \end{bmatrix}, \qquad P_1 = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{\theta^4}{1+\theta^2} \end{bmatrix}. \tag{20}$$