State Space Models and the Kalman Filter

Eric Zivot

April 9, 2006

1 State Space Models

A state space model for an $N-$dimensional time series $\mathbf{y}_t$ consists of a measurement equation relating the observed data to an $m-$ dimensional state vector $\boldsymbol{\alpha}_t$, and a Markovian transition equation that describes the evolution of the state vector over time. The *measurement equation* has the form

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \mathbf{d}_t + \boldsymbol{\varepsilon}_t, \;\; t = 1, \ldots, T$$

where $\mathbf{Z}_t$ is an $N \times m$ matrix, $\mathbf{d}_t$ is an $N \times 1$ vector and $\boldsymbol{\varepsilon}_t$ is an $N \times 1$ error vector such that

$$\boldsymbol{\varepsilon}_t \sim \text{iid } N(\mathbf{0}, \mathbf{H}_t)$$

The *transition equation* for the state vector $\boldsymbol{\alpha}_t$ is the first order Markov process

$$\boldsymbol{\alpha}_t = \mathbf{T}_t \boldsymbol{\alpha}_{t-1} + \mathbf{c}_t + \mathbf{R}_t \boldsymbol{\eta}_t \;\; t = 1, \ldots, T$$

where $\mathbf{T}_t$ is an $m \times m$ transition matrix, $\mathbf{c}_t$ is an $m \times 1$ vector, $\mathbf{R}_t$ is a $m \times g$ matrix, and $\boldsymbol{\eta}_t$ is a $g \times 1$ error vector satisfying

$$\boldsymbol{\eta}_t \sim \text{iid } N(\mathbf{0}, \mathbf{Q}_t)$$

For most applications, it is assumed that the measurement equation errors ε_t are independent of the transition equation errors

$$E[\boldsymbol{\varepsilon}_t \boldsymbol{\eta}_s'] = \mathbf{0} \text{ for all } s, t = 1, \ldots, T$$

However, this assumption is not required. The state space representation is completed by specifying the behavior of the initial state

$$\begin{aligned} \boldsymbol{\alpha}_0 &\sim N(\mathbf{a}_0, \mathbf{P}_0) \\ E[\boldsymbol{\varepsilon}_t \mathbf{a}_0'] &= \mathbf{0}, \;\; E[\boldsymbol{\eta}_t \mathbf{a}_0'] = \mathbf{0} \text{ for } t = 1, \ldots, T \end{aligned}$$

The matrices $\mathbf{Z}_t, \mathbf{d}_t, \mathbf{H}_t, \mathbf{T}_t, \mathbf{c}_t, \mathbf{R}_t$ and $\mathbf{Q}_t$ are called the *system matrices*, and contain non-random elements. If these matrices do not depend deterministically on

t the state space system is called *time invariant*. Note: If $\mathbf{y}_t$ is covariance stationary, then the state space system will be time invariant.

If the state space model is covariance stationary, then the state vector $\boldsymbol{\alpha}_t$ is covariance stationary. The unconditional mean of $\boldsymbol{\alpha}_t$, $\mathbf{a}_0$, may be determined using

$$E[\boldsymbol{\alpha}_t] = \mathbf{T}E[\boldsymbol{\alpha}_{t-1}] + \mathbf{c} = \mathbf{T}E[\boldsymbol{\alpha}_t] + \mathbf{c}$$

Solving for $E[\boldsymbol{\alpha}_t]$, assuming $\mathbf{T}$ is invertible, gives

$$\mathbf{a}_0 = E[\boldsymbol{\alpha}_t] = (\mathbf{I}_m - \mathbf{T})^{-1}\mathbf{c}$$

Similarly, $\text{var}(\boldsymbol{\alpha}_0)$ may be determined analytically using

$$\begin{aligned} \mathbf{P}_0 &= \text{var}(\boldsymbol{\alpha}_t) = \mathbf{T}\text{var}(\boldsymbol{\alpha}_t)\mathbf{T}' + \mathbf{R}\text{var}(\boldsymbol{\eta}_t)\mathbf{R}' \\ &= \mathbf{T}\mathbf{P}_0\mathbf{T}' + \mathbf{R}\mathbf{Q}\mathbf{R}' \end{aligned}$$

Then, using $\text{vec}(\mathbf{ABC}) = \mathbf{C}' \otimes \mathbf{A}\text{vec}(\mathbf{B})$, where vec is the column stacking operator,

$$\begin{aligned} \text{vec}(\mathbf{P}_0) &= \text{vec}(\mathbf{T}\mathbf{P}_0\mathbf{T}') + \text{vec}(\mathbf{R}\mathbf{Q}\mathbf{R}') \\ &= (\mathbf{T} \otimes \mathbf{T})\text{vec}(\mathbf{P}_0) + \text{vec}(\mathbf{R}\mathbf{Q}\mathbf{R}') \end{aligned}$$

which implies that

$$\text{vec}(\mathbf{P}_0) = (\mathbf{I}_{m^2} - \mathbf{T} \otimes \mathbf{T})^{-1}\text{vec}(\mathbf{R}\mathbf{Q}\mathbf{R}')$$

Here $\text{vec}(\mathbf{P}_0)$ is an $m^2 \times 1$ column vector. It can be easily reshaped to form the $m \times m$ matrix $\mathbf{P}_0$.

Example 1 *AR(2) model*

Consider the AR(2) model

$$\begin{aligned} y_t &= \alpha + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \eta_t \\ & \eta_t \sim \text{iid } N(0, \sigma^2) \end{aligned}$$

One way to represent the AR(2) in state space form is as follows. Define $\boldsymbol{\alpha}_t = (y_t, y_{t-1})'$ so that the transition equation for $\boldsymbol{\alpha}_t$ becomes

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix} + \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \eta_t$$

The transition equation system matrices are

$$\mathbf{T} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix}, \mathbf{R} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{c} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, Q = \sigma^2$$

The measurement equation is

$$y_t = (1,0)\boldsymbol{\alpha}_t$$

which implies that

$$\mathbf{Z}_t = (1,0), d_t = 0, \varepsilon_t = 0, H_t = 0$$

The distribution of the initial state vector is

$$\boldsymbol{\alpha}_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$$

Since $\boldsymbol{\alpha}_t = (y_t, y_{t-1})'$ is stationary

$$\begin{aligned} \mathbf{a}_0 &= E[\boldsymbol{\alpha}_t] = (\mathbf{I}_2 - \mathbf{T})^{-1}\mathbf{c} \\ &= \begin{pmatrix} 1-\phi_1 & -\phi_2 \\ -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \alpha \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} \alpha/(1-\phi_1-\phi_2) \\ \alpha/(1-\phi_1-\phi_2) \end{pmatrix} \end{aligned}$$

The variance of the initial state vector satisfies

$$\text{vec}(\mathbf{P}_0) = (\mathbf{I}_4 - \mathbf{T}\otimes\mathbf{T})^{-1}\text{vec}(\mathbf{RQR}')$$

Here, simple algebra gives

$$\begin{aligned} \mathbf{I}_4 - \mathbf{T}\otimes\mathbf{T} &= \begin{pmatrix} 1-\phi_2^2 & -\phi_1\phi_2 & -\phi_1\phi_2 & -\phi_2^2 \\ -\phi_1 & 1 & -\phi_2 & 0 \\ -\phi_1 & -\phi_2 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix} \\ \text{vec}(\mathbf{RQR}') &= \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix} \end{aligned}$$

and so

$$\text{vec}(\mathbf{P}_0) = \begin{pmatrix} 1-\phi_2^2 & -\phi_1\phi_2 & -\phi_1\phi_2 & -\phi_2^2 \\ -\phi_1 & 1 & -\phi_2 & 0 \\ -\phi_1 & -\phi_2 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Example 2 *AR(2) model again*

Another state space representation of the AR(2) is

$$\begin{aligned} y_t &= \mu + c_t \\ c_t &= \phi_1 c_{t-1} + \phi_2 c_{t-2} + \eta_t \end{aligned}$$

The state vector is $\boldsymbol{\alpha}_t = (c_t, c_{t-1})'$, which is unobservable, and the transition equation is

$$\begin{pmatrix} c_t \\ c_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & \phi_2 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ c_{t-2} \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \eta_t$$

This representation has measurement equation matrices

$$\begin{aligned} \mathbf{Z}_t &= (1, 0), \ \mathbf{d}_t = \mu, \varepsilon_t = 0, H_t = 0 \\ \mu &= \alpha/(1 - \phi_1 - \phi_2) \end{aligned}$$

The initial state vector has mean zero, and the initial covariance matrix is the same as that derived above.

Example 3 *AR(2) model yet again*

Yet another state space representation of the AR(2) model is

$$\begin{aligned} y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_t &= \begin{pmatrix} y_t \\ \phi_2 y_{t-1} \end{pmatrix} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \begin{pmatrix} y_{t-1} \\ \phi_2 y_{t-2} \end{pmatrix} + \begin{pmatrix} \alpha \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \eta_t \end{aligned}$$

Example 4 *MA(1) model*

The MA(1) model

$$y_t = \mu + \eta_t + \theta \eta_{t-1}$$

can be put in state space form in a number of ways. Define $\boldsymbol{\alpha}_t = (y_t - \mu, \theta \eta_t)$ and write

$$\begin{aligned} y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{\alpha}_t + \mu \\ \boldsymbol{\alpha}_t &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \boldsymbol{\alpha}_{t-1} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \eta_t \end{aligned}$$

The first element of $\boldsymbol{\alpha}_t$ is then $\theta \eta_{t-1} + \eta_t$ which is indeed $y_t - \mu$.

Example 5 *ARMA(1,1) model*

The ARMA(1,1) model

$$y_t = \mu + \phi(y_{t-1} - \mu) + \eta_t + \theta \eta_{t-1}$$

can be put in a state space form similar to the state space form for the MA(1). Define $\boldsymbol{\alpha}_t = (y_t - \mu, \theta \eta_t)$ and write

$$\begin{aligned} y_t &= \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{\alpha}_t + \mu \\ \boldsymbol{\alpha}_t &= \begin{pmatrix} \phi & 1 \\ 0 & 0 \end{pmatrix} \boldsymbol{\alpha}_{t-1} + \begin{pmatrix} 1 \\ \theta \end{pmatrix} \eta_t \end{aligned}$$

The first element of $\boldsymbol{\alpha}_t$ is then $\phi(y_{t-1} - \mu) + \theta \eta_{t-1} + \eta_t$ which is indeed $y_t - \mu$.

Example 6 *ARMA(p,q) model*

The general ARMA(p,q) model

$$y_t = \phi y_{t-1} + \cdots \phi_p y_{t-p} + \eta_t + \theta_1 \eta_{t-1} + \cdots + \theta_q \eta_{t-q}$$

may be put in state space form in the following way. Let $m = \max(p, q+1)$ and re-write the ARMA(p,q) model as

$$y_t = \phi y_{t-1} + \cdots \phi_p y_{t-m} + \eta_t + \theta_1 \eta_{t-1} + \cdots + \theta_{m-1} \eta_{t-m+1}$$

where some of the AR or MA coefficients will be zero unless $p = q + 1$. Define

$$\boldsymbol{\alpha}_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} + \cdots + \phi_p y_{t-m+1} + \theta_1 \eta_t + \cdots + \theta_{m-1} \eta_{t-m+2} \\ \vdots \\ \phi_m y_{t-1} + \theta_m \eta_t \end{pmatrix}$$

and set

$$\begin{aligned} y_t &= \begin{pmatrix} 1 & \mathbf{0}'_{m-1} \end{pmatrix} \boldsymbol{\alpha}_t \\ \boldsymbol{\alpha}_t &= \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \phi_{m-1} & 0 & 0 & \cdots & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{pmatrix} \boldsymbol{\alpha}_{t-1} + \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-2} \\ \theta_{m-1} \end{pmatrix} \eta_t \end{aligned}$$

1.1 The Kalman Filter

The *Kalman filter* is a set of recursion equations for determining the optimal estimates of the state vector $\boldsymbol{\alpha}_t$ given information available at time t, I_t. The filter consists of two sets of equations:

1. *Prediction equations*

2. *Updating equations*

To describe the filter, let

$$\begin{aligned} \mathbf{a}_t &= E[\boldsymbol{\alpha}_t | I_t] = \text{ optimal estimator of } \boldsymbol{\alpha}_t \text{ based on } I_t \\ \mathbf{P}_t &= E[(\boldsymbol{\alpha}_t - \mathbf{a}_t)(\boldsymbol{\alpha}_t - \mathbf{a}_t)' | I_t] = \text{ MSE matrix of } \mathbf{a}_t \end{aligned}$$

1.1.1 Prediction Equations

Given $\mathbf{a}_{t-1}$ and $\mathbf{P}_{t-1}$ at time $t-1$, the optimal predictor of $\boldsymbol{\alpha}_t$ and its associated MSE matrix are

$$\begin{aligned}
\mathbf{a}_{t|t-1} &= E[\boldsymbol{\alpha}_t|I_{t-1}] = \mathbf{T}_t\mathbf{a}_{t-1} + \mathbf{c}_t \\
\mathbf{P}_{t|t-1} &= E[(\boldsymbol{\alpha}_t - \mathbf{a}_{t-1})(\boldsymbol{\alpha}_t - \mathbf{a}_{t-1})'|I_{t-1}] \\
&= \mathbf{T}_t\mathbf{P}_{t-1}\mathbf{T}'_{t-1} + \mathbf{R}_t\mathbf{Q}_t\mathbf{R}'_t
\end{aligned}$$

The corresponding optimal predictor of $\mathbf{y}_t$ give information at $t-1$ is

$$\mathbf{y}_{t|t-1} = \mathbf{Z}_t\mathbf{a}_{t|t-1} + \mathbf{d}_t$$

The *prediction error* and its MSE matrix are

$$\begin{aligned}
\mathbf{v}_t &= \mathbf{y}_t - \mathbf{y}_{t|t-1} = \mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1} - \mathbf{d}_t = \mathbf{Z}_t(\boldsymbol{\alpha}_t - \mathbf{a}_{t|t-1}) + \boldsymbol{\varepsilon}_t \\
E[\mathbf{v}_t\mathbf{v}'_t] &= \mathbf{F}_t = \mathbf{Z}_t\mathbf{P}_{t|t-1}\mathbf{Z}'_t + \mathbf{H}_t
\end{aligned}$$

These are the components that are required to form the prediction error decomposition of the log-likelihood function.

1.1.2 Updating Equations

When new observations $\mathbf{y}_t$ become available, the optimal predictor $\mathbf{a}_{t|t-1}$ and its MSE matrix are updated using

$$\begin{aligned}
\mathbf{a}_t &= \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'_t\mathbf{F}_t^{-1}(\mathbf{y}_t - \mathbf{Z}_t\mathbf{a}_{t|t-1} - \mathbf{d}_t) \\
&= \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'_t\mathbf{F}_t^{-1}\mathbf{v}_t \\
\mathbf{P}_t &= \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{Z}_t\mathbf{F}_t^{-1}\mathbf{Z}_t\mathbf{P}_{t|t-1}
\end{aligned}$$

The value $\mathbf{a}_t$ is referred to as the *filtered estimate* of $\boldsymbol{\alpha}_t$ and P_t is the MSE matrix of this estimate. It is the optimal estimate of $\boldsymbol{\alpha}_t$ given information available at time t.

1.2 Prediction Error Decomposition

Let $\boldsymbol{\theta}$ denote the parameters of the state space model. These parameters are embedded in the system matrices. For the state space model with a fixed value of $\boldsymbol{\theta}$, the Kalman Filter produces the prediction errors, $\mathbf{v}_t(\boldsymbol{\theta})$, and the prediction error variances, $\mathbf{F}_t(\boldsymbol{\theta})$, from the prediction equations. The *prediction error decomposition* of the log-likelihood function follows immediately:

$$\ln L(\boldsymbol{\theta}|\mathbf{y}) = -\frac{NT}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\ln|\mathbf{F}_t(\boldsymbol{\theta})| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{v}'_t(\boldsymbol{\theta})\mathbf{F}_t^{-1}(\boldsymbol{\theta})\mathbf{v}_t(\boldsymbol{\theta})$$

1.3 Derivation of the Kalman Filter

The derivation of the Kalman filter equations relies on the following result:
Result 1: Suppose

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} \right)$$

Then, the distribution of $\mathbf{x}$ given $\mathbf{y}$ is observed is normal with

$$\begin{aligned} E[\mathbf{x}|\mathbf{y}] &= \boldsymbol{\mu}_{x|y} = \boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \\ \mathrm{var}(\mathbf{x}|\mathbf{y}) &= \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\boldsymbol{\Sigma}_{yy}^{-1}\boldsymbol{\Sigma}_{yx} \end{aligned}$$

In the linear Gaussian state-space model, the disturbances $\boldsymbol{\varepsilon}_t$, and $\boldsymbol{\eta}_t$ are normally distributed and the initial state vector $\boldsymbol{\alpha}_0$ is also normally distributed. From the transition equation, the state vector at time 1 is

$$\boldsymbol{\alpha}_1 = \mathbf{T}_1\boldsymbol{\alpha}_0 + \mathbf{c}_1 + \mathbf{R}_1\boldsymbol{\eta}_1$$

Since $\boldsymbol{\alpha}_0 \sim N(\mathbf{a}_0, \mathbf{P}_0)$, $\boldsymbol{\eta}_1 \sim N(\mathbf{0}, \mathbf{Q}_1)$ and $\boldsymbol{\alpha}_0$ and $\boldsymbol{\eta}_1$ are independent it follows that

$$\begin{aligned} E[\boldsymbol{\alpha}_1] &= \mathbf{a}_{1|0} = \mathbf{T}_1\mathbf{a}_0 + \mathbf{c}_1 \\ \mathrm{var}(\boldsymbol{\alpha}_1) &= E[(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})'] = \mathbf{P}_{1|0} = \mathbf{T}_1\mathbf{P}_0\mathbf{T}_1 + \mathbf{R}_1\mathbf{Q}_1\mathbf{R}_1' \\ & \boldsymbol{\alpha}_1 \sim N(\mathbf{a}_{1|0}, \mathbf{P}_{1|0}) \end{aligned}$$

Notice that the expression for $\mathbf{a}_{1|0}$ is the prediction equation for $\boldsymbol{\alpha}_1$ at $t = 0$. Next, from the measurement equation

$$\mathbf{y}_1 = \mathbf{Z}_1\boldsymbol{\alpha}_1 + \mathbf{d}_1 + \boldsymbol{\varepsilon}_1$$

Since $\boldsymbol{\alpha}_1 \sim N(\mathbf{a}_{1|0}, \mathbf{P}_{1|0})$, $\boldsymbol{\varepsilon}_1 \sim N(\mathbf{0}, \mathbf{H}_1)$ and $\boldsymbol{\alpha}_1$and $\boldsymbol{\varepsilon}_1$ are independent it follows that $\mathbf{y}_1$ is normally distributed with

$$\begin{aligned} E[\mathbf{y}_1] &= \mathbf{y}_{1|0} = \mathbf{Z}_1\mathbf{a}_{1|0} + \mathbf{d}_1 \\ \mathrm{var}(\mathbf{y}_1) &= E[(\mathbf{y}_1 - \mathbf{y}_{1|0})(\mathbf{y}_1 - \mathbf{y}_{1|0})'] \\ &= E[(\mathbf{Z}_1(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}) + \boldsymbol{\varepsilon}_1)(\mathbf{Z}_1(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}) + \boldsymbol{\varepsilon}_1)'] \\ &= \mathbf{Z}_1\mathbf{P}_{1|0}\mathbf{Z}_1' + \mathbf{H}_1 \end{aligned}$$

Notice that the expression for $\mathbf{y}_{1|0}$ is the prediction equation for $\mathbf{y}_1$ at $t = 0$.

For the updating equations , the goal is to find the distribution of $\boldsymbol{\alpha}_1$ conditional on $\mathbf{y}_1$ being observed. To do this, the joint normal distribution of $(\boldsymbol{\alpha}_1', \mathbf{y}_1')'$ must be determined and then Result 1 can be applied. To determine the joint distribution of $(\boldsymbol{\alpha}_1', \mathbf{y}_1')'$ use

$$\begin{aligned} \boldsymbol{\alpha}_1 &= \mathbf{a}_{1|0} + (\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}) \\ \mathbf{y}_1 &= \mathbf{y}_{1|0} + \mathbf{y}_1 - \mathbf{y}_{1|0} \\ &= \mathbf{Z}_1\mathbf{a}_{1|0} + \mathbf{d}_1 + \mathbf{Z}_1(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}) + \boldsymbol{\varepsilon}_1 \end{aligned}$$

and note that

$$
\begin{aligned}
cov(\boldsymbol{\alpha}_1, \mathbf{y}_1) &= E[(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})(\mathbf{y}_1 - \mathbf{y}_{1|0})'] \\
&= E[(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})\left((\mathbf{Z}_1(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0}) + \boldsymbol{\varepsilon}_1\right)'] \\
&= E[(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})\left((\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})\mathbf{Z}_1' + \boldsymbol{\varepsilon}_1'\right)] \\
&= E[(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})\mathbf{Z}_1'] + E[(\boldsymbol{\alpha}_1 - \mathbf{a}_{1|0})\boldsymbol{\varepsilon}_1'] \\
&= \mathbf{P}_{1|0}\mathbf{Z}_1'
\end{aligned}
$$

Therefore,

$$
\begin{pmatrix} \boldsymbol{\alpha}_1 \\ \mathbf{y}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{a}_{1|0} \\ \mathbf{Z}_1\mathbf{a}_{1|0} + \mathbf{d}_1 \end{pmatrix}, \begin{pmatrix} \mathbf{P}_{1|0} & \mathbf{P}_{1|0}\mathbf{Z}_1' \\ \mathbf{Z}_1\mathbf{P}_{1|0} & \mathbf{Z}_1\mathbf{P}_{1|0}\mathbf{Z}_1' + \mathbf{H}_1 \end{pmatrix}\right)
$$

Now, use Result 1 to determine the mean and variance of the distribution of $\boldsymbol{\alpha}_1$ conditional on $\mathbf{y}_1$ being observed:

$$
\begin{aligned}
& \boldsymbol{\alpha}_1|\mathbf{y}_1 \sim N(\mathbf{a}_1, \mathbf{P}_1) \\
\mathbf{a}_1 &= E[\boldsymbol{\alpha}_1|\mathbf{y}_1] = \mathbf{a}_{1|0} + \mathbf{P}_{1|0}\mathbf{Z}_1'\left(\mathbf{Z}_1\mathbf{P}_{1|0}\mathbf{Z}_1' + \mathbf{H}_1\right)^{-1}(\mathbf{y}_1 - \mathbf{Z}_1\mathbf{a}_{1|0} - \mathbf{d}_1) \\
&= \mathbf{a}_{1|0} + \mathbf{P}_{1|0}\mathbf{Z}_1'\mathbf{F}_1^{-1}\mathbf{v}_1 \\
\mathbf{P}_1 &= \mathrm{var}(\boldsymbol{\alpha}_1|\mathbf{y}_1) = \mathbf{P}_{1|0} - \mathbf{P}_{1|0}\mathbf{Z}_1'\left(\mathbf{Z}_1\mathbf{P}_{1|0}\mathbf{Z}_1' + \mathbf{H}_1\right)^{-1}\mathbf{Z}_1\mathbf{P}_{1|0} \\
&= \mathbf{P}_{1|0} - \mathbf{P}_{1|0}\mathbf{Z}_1'\mathbf{F}_1^{-1}\mathbf{Z}_1\mathbf{P}_{1|0}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{v}_1 &= \mathbf{y}_1 - \mathbf{y}_{1|0} = (\mathbf{y}_1 - \mathbf{Z}_1\mathbf{a}_{1|0} - \mathbf{d}_1) \\
\mathbf{F}_1 &= E[\mathbf{v}_1\mathbf{v}_1'] = \mathbf{Z}_1\mathbf{P}_{1|0}\mathbf{Z}_1' + \mathbf{H}_1
\end{aligned}
$$

Notice that the expressions for $\mathbf{a}_1$ and $\mathbf{P}_1$ are exactly the Kalman filter updating equations for $t = 1$. Repeating the above prediction and updating steps for $t = 2, \ldots, T$ gives the Kalman filter recursion equations.