Kalman filter and state-space representations

Ellen L. Hamaker
Department of Psychology
University of Amsterdam.

The Fortran programm MKFM6 (which can be freely downloaded at http://users.fmg.uva.nl/cdolan/) is an implementation of Harvey's Kalman filter (1989). The Kalman filter is an algorithm that can be used to estimate the parameters of a time series model. To use the filter, one has to (re)formulate the time series model to fit the state-space representation. In this paper the Kalman filter algorithm is briefly described. Then, various multivariate stationary time series models are presented. Finally, these models are represented in state-space format to suit the Kalman filter method.

The Kalman filter algorithm Both the Kalman filter and the Kalman smoother can be used to estimate the parameters of a time series model. The Kalman filter method results in approximations of the maximum likelihood estimates, while the smoother results in exact maximum likelihood estimates. Both filter and smoother are a form of prediction-error decomposition, where the goal is to minimize the error between prediction and observation. In the filter, each observation $\mathbf{y}_t$ is predicted based on *all prior observations* $\mathbf{Y}_{t-1}$, that is, $\mathbf{y}_1$ to $\mathbf{y}_{t-1}$. The difference between the prediction and the actual value is called the error. In the Kalman smoother, the prediction of $\mathbf{y}_t$ is not only based on all previous observations $\mathbf{Y}_{t-1}$, but also on all later observations $\mathbf{Y}_{t+1}$, that is $\mathbf{y}_{t+1}$ to $\mathbf{y}_T$. Hence, the filter can be referred to as an online estimation procedure, because it can be used when one wants to estimate the parameters online when new observations are coming in only after they have been estimated. In contrast, the Kalman smoother can be thought of as an offline procedure, since it can only be used after the total series have been observed.

The use of the Kalman filter and smoother require the time series models to be formulated in state-space representation. We use the term *state-space representation* instead of *state-space model*, because all the time series models discussed here can be represented in state-space format, just as they can all be represented in raw data likelihood format or Toeplitz format. The state-space representation consists of two equations: the measurement equation and the transition equation.

The *measurement equation* (also known as the observations equation) is:

$$\mathbf{y}_t = \mathbf{S}\mathbf{a}_t + \mathbf{d} + \mathbf{e}_t \tag{1}$$

where $\mathbf{y}_t$ is an (NY×1) vector with observations at occasion t, $\mathbf{S}$ is an (NY×NE) matrix with factor loadings, $\mathbf{a}_t$ is a (NE×1) state-vector, and $\mathbf{d}$ is an (NY×1) vector with constants. The (NY×1) vector $\mathbf{e}_t$ contains the measurement errors, which is multivariate normally distributed with mean zero and covariance matrix $\mathbf{R}$.

The *transition equation* (also known as the state equation) is:

$$\mathbf{a}_t = \mathbf{H}\mathbf{a}_{t-1} + \mathbf{c} + \mathbf{G}\mathbf{z}_t \tag{2}$$

where $\mathbf{H}$ is a (NE×NE) matrix with structural coefficients, $\mathbf{c}$ is a (NE×1) vector with constants, and $\mathbf{G}$ is a (NE×NE) weight matrix. The (NE×1) vector with residuals $\mathbf{z}_t$ has mean zero and covariance matrix $\mathbf{Q}$.

When the Kalman filter is used to estimate the parameters of a time series model, the goal is to minimize the difference between the observation $\mathbf{y}_t$ and the prediction based on the previous observations $\mathbf{y}_{t|t} = \mathrm{E}[\mathbf{y}_t|\mathbf{Y}_{t-1}]$, where $\mathbf{Y}_{t-1} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1}\}$. The recursive nature of the Kalman filter is explained below.

To use the Kalman filter, one has to guess the value of the initial state $\mathbf{a}_0$ to start up the process. Often, the initial state is fixed at the unconditional expectation of $\mathbf{a}$, thus, $\mathbf{a}_{0|0} = \mathbf{0}$. This fixed initial state $\mathbf{a}_{0|0}$ may differ from the real initial state $\mathbf{a}_0$, the latter which can not be observed. In addition to fixing $\mathbf{a}_{0|0}$, one also has to fix the covariance matrix associated with the difference between the real state $\mathbf{a}_0$, and one's estimate of it $\mathbf{a}_{0|0}$. This covariance matrix

$$\mathbf{P}_{0|0} = \mathrm{E}\big[(\mathbf{a}_0 - \mathbf{a}_{0|0})(\mathbf{a}_0 - \mathbf{a}_{0|0})'\big] \tag{3}$$

has to be defined to start up the recursive algorithm. Most typically the diagonal of this matrix $\mathbf{P}_{0|0}$ are fixed at large values, while the off-diagonal elements are fixed at zero.

From the initial (estimated) state $\mathbf{a}_{0|0}$, a prediction is made of the state at the first occasion conditional on the previous occasion through use of Equation 2, that is,

$$\mathbf{a}_{1|0} = \mathbf{H}\mathbf{a}_{0|0} + \mathbf{c}\ . \tag{4}$$

This $\mathbf{a}_{1|0}$ is said to be the optimal estimate of the true state $\mathbf{a}_1$ based on information up to $t = 0$. The covariance matrix of the estimation error associated with this estimate is

$$\begin{aligned}
\mathbf{P}_{1|0} &= \mathrm{E}\big[(\mathbf{a}_1 - \mathbf{a}_{1|0})(\mathbf{a}_1 - \mathbf{a}_{1|0})'\big] \\
&= \mathrm{E}\Big[\Big(\{\mathbf{H}\mathbf{a}_0 + \mathbf{c} + \mathbf{G}\mathbf{z}_1\} - \{\mathbf{H}\mathbf{a}_{0|0} + \mathbf{c}\}\Big)\Big(\{\mathbf{H}\mathbf{a}_0 + \mathbf{c} + \mathbf{G}\mathbf{z}_1\} - \{\mathbf{H}\mathbf{a}_{0|0} + \mathbf{c}\}\Big)'\Big] \\
&= \mathrm{E}\Big[\Big(\mathbf{H}\mathbf{a}_0 - \mathbf{H}\mathbf{a}_{0|0} + \mathbf{G}\mathbf{z}_1\Big)\Big(\mathbf{H}\mathbf{a}_0 - \mathbf{H}\mathbf{a}_{0|0} + \mathbf{G}\mathbf{z}_1\Big)'\Big] \\
&= \mathrm{E}\Big[\Big(\mathbf{H}\mathbf{a}_0 - \mathbf{H}\mathbf{a}_{0|0}\Big)\Big(\mathbf{H}\mathbf{a}_0 - \mathbf{H}\mathbf{a}_{0|0}\Big)'\Big] + \mathrm{E}\big[(\mathbf{G}\mathbf{z}_1)(\mathbf{G}\mathbf{z}_1)'\big] \\
&= \mathbf{H}\mathrm{E}\big[(\mathbf{a}_0 - \mathbf{a}_{0|0})(\mathbf{a}_0 - \mathbf{a}_{0|0})'\big]\mathbf{H}' + \mathbf{G}\mathrm{E}\big[\mathbf{z}_1\mathbf{z}_1'\big]\mathbf{G}' \\
&= \mathbf{H}\mathbf{P}_{0|0}\mathbf{H}' + \mathbf{G}\mathbf{Q}\mathbf{G}'\ .
\end{aligned} \tag{5}$$

Equations 4 and 5 are known as the *prediction equations.*

Now that we have a prediction of the state at occasion 1 based on occasion 0, we can predict our observation at occasion 1 based on occasion 0. That is,

$$\mathbf{y}_{1|0} = \mathbf{S}\mathbf{a}_{1|0} + \mathbf{d}\ . \tag{6}$$

Then, as the observation at occasion 1 is coming in, we can compare our prediction of **y** with the observed **y**, that is,

$$\begin{aligned}\mathbf{e}_{1|0} &= \mathbf{y}_1 - \mathbf{y}_{1|0} \\ &= \big(\mathbf{S}\mathbf{a}_1 + \mathbf{d} + \mathbf{e}_1\big) - \big(\mathbf{S}\mathbf{a}_{1|0} + \mathbf{d}\big) \\ &= \mathbf{S}(\mathbf{a}_1 - \mathbf{a}_{1|0}) + \mathbf{e}_1\ ,\end{aligned} \tag{7}$$

which is called the *one-step-ahead prediction error.* At every occasion t there is a prediction error, and the goal is to chose the parameters so that these errors are minimized. The covariance matrix of the prediction error $\mathbf{e}_{1|0}$ (which we need in the next step) is given by

$$\begin{aligned}\mathbf{F}_1 &= \mathrm{E}\Big[\Big(\mathbf{S}\{\mathbf{a}_1 - \mathbf{a}_{1|0}\} + \mathbf{e}_1\Big)\Big(\mathbf{S}\{\mathbf{a}_1 - \mathbf{a}_{1|0}\} + \mathbf{e}_1\Big)'\Big] \\ &= \mathbf{S}\mathrm{E}\big[(\mathbf{a}_1 - \mathbf{a}_{1|0})(\mathbf{a}_1 - \mathbf{a}_{1|0})'\big]\mathbf{S}' + \mathrm{E}[\mathbf{e}_t\mathbf{e}_t'] \\ &= \mathbf{S}\mathbf{P}_{1|0}\mathbf{S}' + \mathbf{R}\ .\end{aligned} \tag{8}$$

The next step consists of updating the prediction for $\mathbf{a}_1$ in the light of the observation $\mathbf{y}_1$. Remember that the actual state $\mathbf{a}_1$ can not be observed. We had estimated this state based on the information prior to occasion 1, which resulted in $\mathbf{a}_{1|0}$. With the information available at occasion 1 we can update this estimate to results in $\mathbf{a}_{1|1}$. To this end we make use of the following (see Durbin & Koopman, 2001, p. 37; this is based on assuming a multinormal distribution):

$$\begin{aligned}\mathrm{E}[a|b,c] &= \mathrm{E}[a|b] + \Sigma_{ac}\Sigma_{cc}^{-1}c \\ \mathrm{Var}[a|b,c] &= \mathrm{Var}[a|b] - \Sigma_{ac}\Sigma_{cc}^{-1}\Sigma_{ac}'\ .\end{aligned}$$

Therefore we can write

$$\begin{aligned}\mathbf{a}_{1|1} &= \mathrm{E}\big[\mathbf{a}_1|\mathbf{Y}_1\big] = \mathrm{E}\big[\mathbf{a}_1|\mathbf{Y}_0, \mathbf{e}_{1|0}\big] \\ &= \mathrm{E}\big[\mathbf{a}_1|\mathbf{Y}_0\big] + \mathrm{Cov}[\mathbf{a}_1, \mathbf{e}_{1|0}]\mathrm{Var}[\mathbf{e}_{1|0}]^{-1}\mathbf{e}_{1|0} \\ &= \mathbf{a}_{1|0} + \mathbf{L}_1\mathbf{F}_1^{-1}\mathbf{e}_{1|0}\ ,\end{aligned} \tag{9}$$

where $\mathbf{a}_{1|0}$, $\mathbf{F}_1$, and $\mathbf{e}_{1|0}$ where already presented above. By use of Equation 7 the covariance between the true state $\mathbf{a}_1$ and the prediction error $\mathbf{e}_{1|0}$ can be written as,

$$\begin{aligned}\mathbf{L}_1 &= \mathrm{Cov}\big[\mathbf{a}_1, \mathbf{e}_{1|0}\big] \\ &= \mathrm{E}\Big[\Big\{\mathbf{a}_1 - \mathrm{e}[\mathbf{a}_1]\Big\}\Big\{\big(\mathbf{S}\mathbf{a}_1 - \mathbf{S}\mathbf{a}_{1|0} + \mathbf{e}_1\big) - \mathrm{E}[\mathbf{e}_{1|0}]\Big\}'\Big] \\ &= \mathrm{E}\Big[\big\{\mathbf{a}_1\big\}\big\{\mathbf{S}\mathbf{a}_1 - \mathbf{S}\mathbf{a}_{1|0} + \mathbf{e}_1\big\}'\Big] \\ &= \mathrm{E}\Big[\mathbf{a}_1\mathbf{a}_1'\mathbf{S}' - \mathbf{a}_1\mathbf{a}_{1|0}'\mathbf{S}' + \mathbf{a}_1\mathbf{e}_1'\big\}\Big] \\ &= \mathrm{E}\big[\mathbf{a}_1\mathbf{a}_1'\big]\mathbf{S}' - \mathbf{0} + \mathbf{0} \\ &= \mathbf{P}_{1|0}\mathbf{S}'\ .\end{aligned} \tag{10}$$

The covariance matrix of the updated state $\mathbf{a}_{1|1}$ is

$$\begin{aligned}\mathbf{P}_{1|1} &= \mathrm{Var}[\mathbf{a}_1|\mathbf{Y}_1] = \mathrm{Var}[\mathbf{a}_1|\mathbf{Y}_0, \mathbf{e}_{1|0}] \\ &= \mathrm{Var}[\mathbf{a}_1|\mathbf{Y}_0] - \mathrm{Cov}[\mathbf{a}_1, \mathbf{e}_{1|0}]\Big\{\mathrm{Var}[\mathbf{e}_{1|0}]\Big\}^{-1}\mathrm{Cov}[\mathbf{a}_1, \mathbf{e}_{1|0}]' \\ &= \mathbf{P}_{1|0} - \mathbf{L}_1\mathbf{F}_1^{-1}\mathbf{L}_1' \\ &= \mathbf{P}_{1|0} - \mathbf{P}_{1|0}\mathbf{S}'\mathbf{F}_1^{-1}\mathbf{S}\mathbf{P}_{1|0}' \ .\end{aligned} \tag{11}$$

Equations 9 and 11 (using Equations 8, 10, and 7) are known as the *update equations*.

Next, one can make a prediction of $\mathbf{a}_2$ based on $\mathbf{a}_{1|1}$, that is, $\mathbf{a}_{2|1}$. This $\mathbf{a}_{2|1}$ is then used to predict the observations at occasion 2, that is, $\mathbf{y}_{2|1}$. Using the covariance matrix $\mathbf{P}_{1|1}$ we can also obtain the covariance matrix of the estimation error of $\mathbf{a}_{2|1}$, that is $\mathbf{P}_{2|1}$. As the observation $\mathbf{y}_2$ becomes available, we can determine the one-step-ahead prediction error $\mathbf{e}_{2|1}$. Also, we can obtain the covariance matrix of this error by using the covariance matrix $\mathbf{P}_{2|1}$. Then, we can update our prediction of $\mathbf{a}_2$ based on $\mathbf{y}_2$, resulting in $\mathbf{a}_{2|2}$. In addition, we can update our covariance matrix of the estimation error, which gives $\mathbf{P}_{2|2}$. And so we can predict $\mathbf{a}_3$ based on $\mathbf{a}_{2|2}$ resulting in $\mathbf{a}_{3|2}$, and so on.

In short, the Kalman filter consists of the prediction equations:

$$\mathbf{a}_{t|t-1} = \mathbf{H}\mathbf{a}_{t-1|t-1} + \mathbf{c} \tag{12}$$

$$\mathbf{P}_{t|t-1} = \mathbf{H}\mathbf{P}_{t-1|t-1}\mathbf{H}' + \mathbf{G}\mathbf{Q}\mathbf{G}' \ . \tag{13}$$

Based on the prediction of $\mathbf{a}_t$, the observation $\mathbf{y}_t$ is predicted,

$$\mathbf{y}_{t|t-1} = \mathbf{S}\mathbf{a}_{t|t-1} + \mathbf{d} \ . \tag{14}$$

Then, as the observation $\mathbf{y}_t$ becomes available, the one-step-ahead prediction error is determined, along with it's covariance matrix, that is,

$$\mathbf{e}_{t|t-1} = \mathbf{y}_t - \mathbf{y}_{t|t-1} \tag{15}$$

$$\mathbf{F}_t = \mathbf{S}\mathbf{P}_{t|t-1}\mathbf{S}' + \mathbf{R} \ . \tag{16}$$

Finally, the estimate of the state $\mathbf{a}_t$ is updated, as well as the covariance matrix of the estimation error of $\mathbf{a}_t$, that is,

$$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{S}'\mathbf{F}_t^{-1}\mathbf{e}_{t|t-1} \tag{17}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{S}'\mathbf{F}_t^{-1}\mathbf{S}\mathbf{P}_{t|t-1}' \ . \tag{18}$$

The one-step-ahead prediction errors $\mathbf{e}_{t|t-1}$ for $t = 1, 2, \ldots, T$ are independently and identically distributed. This implies that the joint log likelihood function of $\{\mathbf{e}_{t|t-1}\}_{t=1}^{T}$ can be written as the sum of the log likelihoods at each occasion, that is,

$$\log L = -\frac{NY * T}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{F}_t| - \frac{1}{2}\sum_{t=1}^{T}\mathbf{e}_{t|t-1}'\mathbf{F}_t^{-1}\mathbf{e}_{t|t-1} \tag{19}$$

This whole procedure is also represented in Figure 1. Optimizing Equation 19 with respect to the parameters in $\mathbf{S}$, $\mathbf{d}$, $\mathbf{R}$, $\mathbf{H}$, $\mathbf{c}$, $\mathbf{G}$, and $\mathbf{Q}$, results in the maximum likelihood estimates of these estimates.

Multivariate stationary models In this section several multivariate stationary models are presented: the vector autoregressive moving average (VARMA) model, and the dynamic factor model (DFM), which includes the the multiple indicator (MI) VARMA model as a specific case.

VARMA models The VARMA (NY, p, q) model is a straightforward extension of the univariate ARMA (p, q) model. Suppose one has observed NY separate time series within one system, e.g., a subject or a dyad of two individuals, and all series are of length T. Because these series represent different aspects of the same system, they might have (lagged) cross-covariances that differ significantly from zero. The VARMA (NY, p, q) model can be written as

$$\begin{aligned}\mathbf{y}_t &= \boldsymbol{\Phi}_1\mathbf{y}_{t-1} + \cdots + \boldsymbol{\Phi}_p\mathbf{y}_{t-p} + \mathbf{u}_t - \boldsymbol{\Theta}_1\mathbf{u}_{t-1} - \cdots - \boldsymbol{\Theta}_q\mathbf{u}_{t-q} \\ &= \sum_{j=1}^{p}\boldsymbol{\Phi}_j\mathbf{y}_{t-j} + \mathbf{u}_t - \sum_{j=1}^{q}\boldsymbol{\Theta}_j\mathbf{u}_{t-j} \;, \end{aligned} \tag{20}$$

where $\mathbf{u}_t$ is a vector of length NY with random innovations, i.e., $\mathbf{u}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_u)$. Although not strictly necessary, $\boldsymbol{\Sigma}_u$ is often constrained to be diagonal, that is, the innovations $\{u_{i,t}\}$ are assumed to be independently distributed, where $\{u_{i,t}\}$ are the innovations of variable i. Note that there are no contemporary cross-relations between the observations, that is, the vector $\mathbf{y}_t$ is not regressed upon itself. For the model in Equation 20 to be identified, either the vector $\mathbf{y}_t$ is not regressed upon itself, or the innovations are not correlated, i.e. $\boldsymbol{\Sigma}_u$ is diagonal. It can be shown that these two restrictions are actually equivalent. However, often both restrictions are imposed.

The (NY$\times$NY) matrices $\boldsymbol{\Phi}$ contain the autoregressive parameters on the main diagonal (i.e., the parameters that are used to regress the series upon itself at an earlier occasion), while the off-diagonal elements represent the cross-relations between the series (i.e., $\phi_{ij,k}$ is used to regress a series i at occasion t on another series j at $t-k$). These matrices are not necessarily symmetric, as series i can be regressed upon series j at previous occasions, whereas series j is not regressed on series i. The matrices $\boldsymbol{\Theta}$ relate the observations at a certain occasion to innovations at prior occasions and these matrices are not necessarily symmetric.

As with the univariate ARMA, the VARMA model can also be formulated using the backward shift operator, i.e.,

$$\begin{aligned}\mathbf{y}_t - \boldsymbol{\Phi}_1\mathbf{y}_{t-1} - \cdots - \boldsymbol{\Phi}_p\mathbf{y}_{t-p} &= \mathbf{u}_t - \boldsymbol{\Theta}_1\mathbf{u}_{t-1} - \cdots - \boldsymbol{\Theta}_q\mathbf{u}_{t-q} \\ \boldsymbol{\Phi}(B)\mathbf{y}_t &= \boldsymbol{\Theta}(B)\mathbf{u}_t \;, \end{aligned} \tag{21}$$

where $\boldsymbol{\Phi}(B) = \sum_{j=0}^{p}\boldsymbol{\Phi}_j B^j$ is the AR operator which is a set of NY polynomials in the back shift operator, and $\boldsymbol{\Theta}(B) = \sum_{j=0}^{q}\boldsymbol{\Theta}_j B^j$ is the MA operator which is a set of NY polynomials in the back shift operator. We can multiply both sides by the inverse of the AR operator, which results in,

$$\mathbf{y}_t = \boldsymbol{\Phi}(B)^{-1}\boldsymbol{\Theta}(B)\mathbf{u}_t \;, \tag{22}$$

where $\mathbf{\Phi}(B)^{-1}$ is a set of polynomials of infinite order in the lag operator so that Equation 22 is a pure VMA of infinite order. Put differently, a mixed VARMA model of finite order can always be rewritten as a pure VMA model of infinite order.

The model in Equation 20 is often simplified to a vector autoregression (VAR) model, in which case all $\mathbf{\Theta}$ matrices are zero matrices (e.g., Hamilton, 1994). An even more restricted version of the general VARMA model is the autoregression model. This model is relatively often applied in psychology and it is suited if one variable is assumed to be dependent on itself and other variables (simultaneously or at previous occasions), whereas the predictors are assumed to be white noise sequences. Hence, all $\mathbf{\Theta}$ matrices are zero, and in the $\mathbf{\Phi}$ matrices contain only a single row with nonzero elements, whereas all other rows are filled with zeros, so that only one variable is regressed upon itself, and the predictors at previous occasions, whereas the predictors are modeled as white noise sequences.

Another possibility is that the dependent variable is regressed upon itself at previous occasions, and on simultaneous observations of the predictors. Then, the (NY×1) vector $\mathbf{y}_t$ in Equation 20 has to be premultiplied with a matrix with regression coefficients on one line and further ones on the diagonal. SPSS includes this latter autoregression model. It is not possible to choose the lag of the relationship between the predictors and the dependent variable. Hence, this relationship is always simultaneous, even though one can argue that it is impossible to assume a “causal” relationship within the same measurement occasion.

Dynamic factor model The DFM is characterized by three options to model sequential relationships: (a) through lagged factor loadings; (b) through the covariance structure of the latent series; and (c) through the covariance structure of the measurement errors. These three options can be used simultaneously.

First, to model sequential relations in the DFM, one can use lagged factor loadings between the NE latent series $\{\mathbf{w}_t\}$ and the NY observed series $\{\mathbf{y}_t\}$, that is (Molenaar, 1985):

$$\begin{aligned}\mathbf{y}_t &= \mathbf{\Lambda}_0\mathbf{w}_t + \mathbf{\Lambda}_1\mathbf{w}_{t-1} + \cdots + \mathbf{\Lambda}_r\mathbf{w}_{t-r} + \mathbf{v}_t \\ &= \sum_{h=0}^{r} \mathbf{\Lambda}_h\mathbf{w}_{t-h} + \mathbf{v}_t \ , \end{aligned} \tag{23}$$

where the (NY×1) vector $\mathbf{y}_t$ contains the observations at occasion t that are regressed upon $\mathbf{w}_t$ to $\mathbf{w}_{t-r}$, through the factor loadings in the (NY×NE) matrices $\mathbf{\Lambda}_h$. The vector $\mathbf{w}_t$ is of dimensions (NE×1). The (NY×1) vector $\mathbf{v}_t$ contains the measurement errors at occasion t, i.e., the part of $\mathbf{y}_t$ that could not be predicted by the latent variables $\mathbf{w}_t$ to $\mathbf{w}_{t-r}$. Equation 23 can also be expressed by use of the backward shift operator, resulting in

$$\mathbf{y}_t = \mathbf{\Lambda}(B)\mathbf{w}_t + \mathbf{v}_t \ , \tag{24}$$

where $\mathbf{\Lambda}(B)$ represents the r-th order polynomial in the backward shift operator, i.e., $\mathbf{\Lambda}(B) = \sum_{h=0}^{r} \mathbf{\Lambda}_h B^h$.

A second possibility to model sequential relationships in the DFM is through the covariance structure of the latent series. The DFM does not necessarily include a specific model for the latent series themselves, but the covariance structure of these latent series is simply defined as

$$\mathrm{E}[\mathbf{w}_t\mathbf{w}'_{t-\tau}] = \mathbf{\Omega}_\tau \ . \tag{25}$$

This implies that the latent series are covariance stationary, i.e., the auto- and cross-covariances are dependent on lag size τ only. One possible model for the latent series giving rise to such a stationary covariance structure is the VARMA model similar to Equation 22, that is,

$$\begin{aligned}\mathbf{w}_t &= \mathbf{\Phi}_1\mathbf{w}_{t-1} + \cdots + \mathbf{\Phi}_p\mathbf{w}_{t-p} + \mathbf{u}_t - \mathbf{\Theta}_1\mathbf{u}_{t-1} - \cdots - \mathbf{\Theta}_{t-q}\mathbf{u}_t \\ &= \mathbf{\Phi}(B)^{-1}\mathbf{\Theta}(B)\mathbf{u}_t \ ,\end{aligned} \tag{26}$$

where $\{\mathbf{u}_t\}$ are NE uncorrelated white noise sequences, with mean zero and a diagonal covariance matrix $\mathbf{M}$. The $\mathbf{\Phi}$ matrices and the $\mathbf{\Theta}$ matrices are (NE×NE) matrices.

When $r = 0$ (i.e., if there are only lag 0 factor loadings), the DFM reduces to what could be termed a multiple indicator VARMA model. Moreover, when $r = 0$ and no lagged relationships are specified at the latent level (i.e.,if $\mathbf{\Omega}_\tau$ is a null matrix for all $\tau \neq 0$), Equations 23 and 26 reduce to Cattell's P-technique.

A third possibility to model sequential relationships in the DFM is by the covariance structure of the measurement errors $\mathbf{v}_t$. As with the latent series, no specific model is defined for the lagged relationships of the measurement errors, but the covariance structure of the measurement errors is defined as

$$\mathrm{E}[\mathbf{v}_t\mathbf{v}'_{t-\tau}] = \mathbf{\Upsilon}_\tau \ , \tag{27}$$

which implies that the series are covariance stationary, i.e., the auto- and cross-covariances are dependent on lag size τ only. The covariance matrices $\mathbf{\Upsilon}_\tau$ are defined to be diagonal, implying that the measurement errors of variable i and j (where i and $j = 1, \ldots, \mathrm{NY}$, and $i \neq j$) are independent of each other, but the measurement errors may contain autocorrelation. Such an autocorrelation indicates that the measurement error (that is, the part of a variable that is not explained by the common factor) is not entirely random.

One way of modeling the sequential dependency of the measurement errors, while ensuring covariance stationarity, is by defining a VAR for the errors $\mathbf{v}_t$, that is:

$$\begin{aligned}\mathbf{v}_t &= \mathbf{\Xi}_1\mathbf{v}_{t-1} + \cdots + \mathbf{\Xi}_s\mathbf{v}_{t-s} + \mathbf{b}_t \\ &= \sum_{j=1}^{s} \mathbf{\Xi}_g\mathbf{v}_{t-g} + \mathbf{b}_t\end{aligned} \tag{28}$$

where the (NY×NY) matrices $\mathbf{\Xi}$ are diagonal matrices, and $\mathbf{b}_t$ is the vector of length M containing the unpredictable part of the measurement errors. The (lag zero) covariation matrix of $\mathbf{b}_t$ is the diagonal matrix $\mathbf{\Sigma}_b$.

State-space representations of multivariate stationary time series models Here, the models presented in the previous section are represented in state-space format. To ease presentation, we assume all variables are centered (hence, $\mathbf{d} = \mathbf{0}$ and $\mathbf{c} = \mathbf{0}$).

VARMA model: state-space representation An NY-variate mixed VARMA model with arbitrary p and q, measurement equation can be written as

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_t \\ \dots \\ \mathbf{y}_{t-p+1} \\ \mathbf{u}_t \\ \dots \\ \mathbf{u}_{t-q+1}\ , \end{bmatrix}$$

where the (NY×NY) matrix $\mathbf{I}$ is an identity matrix whereas the other $p+q-1$ blocks in $\mathbf{J}$ are (NY×NY) null-matrices. There are no measurement errors in this model, hence the vector $\mathbf{e}_t$ is a null-vector and the model matrix $\mathbf{R}$ is a zero-matrix. The state-vector $\mathbf{a}_t$ has dimensions $(\{p+q\} * \mathrm{NY}) \times 1$. The transition equation is

$$\begin{bmatrix} \mathbf{y}_t \\ \mathbf{y}_{t-1} \\ \dots \\ \mathbf{y}_{t-p+2} \\ \mathbf{y}_{t-p+1} \\ \mathbf{u}_t \\ \mathbf{u}_{t-1} \\ \mathbf{u}_{t-2} \\ \dots \\ \mathbf{u}_{t-q+1} \end{bmatrix} = \begin{bmatrix} \mathbf{\Phi}_1 & \mathbf{\Phi}_2 & \dots & \mathbf{\Phi}_{p-1} & \mathbf{\Phi}_p & \mathbf{\Theta}_1 & \mathbf{\Theta}_2 & \dots & \mathbf{\Theta}_{q-1} & \mathbf{\Theta}_q \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \dots & & & & & & & & & \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} & \mathbf{0} \\ \dots & & & & & & & & & \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{y}_{t-1} \\ \mathbf{y}_{t-2} \\ \dots \\ \mathbf{y}_{t-p+1} \\ \mathbf{y}_{t-p} \\ \mathbf{u}_{t-1} \\ \mathbf{u}_{t-2} \\ \mathbf{2}_{t-3} \\ \dots \\ \mathbf{u}_{t-q} \end{bmatrix}$$

$$+ \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \\ \dots & & & & & & & & \\ \mathbf{0} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \\ \dots & & & & & & & & \\ \mathbf{0} & \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{u}_t \\ \mathbf{0} \\ \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ \dots \\ \mathbf{0} \end{bmatrix},$$

where the matrix $\mathbf{H}$ with structural coefficients contains a row existing of the (NY×NY) matrices with the VARMA matrices, as defined in Equation 20, and a sub-diagonal of blocks formed by $\mathbf{I}$, which is an (NY×NY) identity matrix. The $\mathbf{0}$'s represent (NY×NY) null-matrices. Hence, in this transition equation, the (NY×1) vector $\mathbf{y}_t$ is computed based on the prior $\mathbf{y}$'s and the innovations $\mathbf{u}$'s and the innovations of this occasion $\mathbf{u}_t$. Simultaneously, the (NY×1) vectors $\mathbf{y}_{t-p}$ and $\mathbf{u}_{t-q}$ are removed from the state-vector, in order to create room for the new vectors $\mathbf{y}_t$ and $\mathbf{u}_t$. The latter comes into the state vector due to the second identity matrix in the first block-column of the weight matrix $\mathbf{G}$ (i.e., the (p+1)th block position). All the other vectors are moved one block-position (that is NY positions) down in the state-vector. Because the innovations $\mathbf{u}_t$ are not correlated (neither over time, nor across series), the covariance matrix $\mathbf{Q}$ contains an (NY×NY) diagonal covariance matrix of $\mathbf{u}$ in the upper left.

DFM without lagged factor loadings: state-space representation As stated above, an DFM without factor loadings beyond lag zero can be called an MI VARMA model. The measurement equation of an MI VARMA $(\mathrm{NY}, \mathrm{NE}, p, q)$ can be written as:

$$\begin{bmatrix}\mathbf{y}_t\end{bmatrix} = \begin{bmatrix}\mathbf{\Lambda} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{w}_t \\ \dots \\ \mathbf{w}_{t-p+1} \\ \mathbf{u}_t \\ \dots \\ \mathbf{u}_{t-q+1}\end{bmatrix} + \begin{bmatrix}\mathbf{v}_t\end{bmatrix} ,$$

where the (NY×NE) matrix $\mathbf{\Lambda}$ with factor loadings is as defined in Equation 23, with $r = 0$. The other $p + q - 1$ blocks with dimensions (NY×NE) of $\mathbf{S}$ are null-matrices. In an MI VARMA there are measurement errors, hence the matrix $\mathbf{e}_t$ is not a null-vector. Since these measurement errors are uncorrelated, the covariance matrix $\mathbf{R}$ is an (NY×NY) diagonal matrix.

The transition equation of the state-vector $\mathbf{a}_t$ is similar to that of an VARMA, except that the observed $\mathbf{y}_t$ is replaced by the latent $\mathbf{w}_t$, and hence the blocks have dimensions (NE×NE).

DFM with latent white noise: state-space representation Another version of the DFM includes lagged factor loadings, while the latent series are white noise. The measurement equation of a DFM $(\mathrm{NY}, \mathrm{NE}, r)$ is:

$$\begin{bmatrix}\mathbf{y}_t\end{bmatrix} = \begin{bmatrix}\mathbf{\Lambda}_0 & \mathbf{\Lambda}_1 & \dots & \mathbf{\Lambda}_r\end{bmatrix} \begin{bmatrix}\mathbf{u}_t \\ \mathbf{u}_{t-1} \\ \dots \\ \mathbf{u}_{t-r}\end{bmatrix} + \begin{bmatrix}\mathbf{v}_t\end{bmatrix}$$

where the (NY×NE) matrix $\mathbf{\Lambda}_h$ with factor loadings at lag h (where $h = 0, \dots, r$), is as defined in Equation 23. The DFM includes measurement errors, hence the vector $\mathbf{e}_t$ is not a null-vector. The covariance matrix $\mathbf{R}$ is a diagonal matrix, implying there is no dependency between the measurement errors of different observed variables.

If one models a DFM with latent white noise sequences, the transition equation of the state-vector $\mathbf{a}_t$ is:

$$\begin{bmatrix}\mathbf{u}_t \\ \mathbf{u}_{t-1} \\ \dots \\ \mathbf{u}_{t-r+1} \\ \mathbf{u}_{t-r}\end{bmatrix} = \begin{bmatrix}\mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \dots & & & & \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{u}_{t-1} \\ \mathbf{u}_{t-2} \\ \dots \\ \mathbf{u}_{t-r} \\ \mathbf{u}_{t-r-1}\end{bmatrix} + \begin{bmatrix}\mathbf{I} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \dots & & & & \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{u}_t \\ \mathbf{0} \\ \dots \\ \mathbf{0} \\ \mathbf{0}\end{bmatrix}$$

where the lower block-subdiagonal of the matrix $\mathbf{H}$ consists of (NE×NE) identity matrices $\mathbf{I}$, which shifts the block-elements in the state vector one block-position down. This results in the deletion of the most ancient vector $\mathbf{u}_{t-r-1}$ from the state vector $\mathbf{a}_t$. The weight matrix $\mathbf{G}$ contains an (NY×NY) identity matrix in the upper left corner. The covariance matrix $\mathbf{Q}$ contains the (NE×NE) diagonal covariance matrix of $\mathbf{u}$ in the upper left corner, while the rest of the matrix is filled with zeros.

DFM with lagged factor loadings and latent VARMA model: state space representation If one wishes to fit a DFM with a latent VARMA model, the state equation is as for an MI VARMA, while the measurement equation becomes:

$$\begin{bmatrix}\mathbf{y}_t\end{bmatrix} = \begin{bmatrix}\mathbf{\Lambda_0} & \dots & \mathbf{\Lambda}_r & \mathbf{0} & \dots & \mathbf{0}\end{bmatrix} \begin{bmatrix}\mathbf{w}_t \\ \dots \\ \mathbf{w}_{t-p+1} \\ \mathbf{u}_t \\ \dots \\ \mathbf{u}_{t-q+1}\end{bmatrix} + \begin{bmatrix}\mathbf{v}_t\end{bmatrix}$$

DFM with autocorrelated errors: state space representation If one wants to estimate first order autocorrelated errors, this can be done by including the errors in the state-vector and regressing them on their previous value. The measurement equation then becomes

$$\mathbf{y}_t = \begin{bmatrix}\mathbf{S}^* & \mathbf{I}\end{bmatrix} \begin{bmatrix}\mathbf{a}_t^* \\ \mathbf{v}_t\end{bmatrix}$$

where $\mathbf{S}^*$ and $\mathbf{a}^*$ are the same as in the equivalent model without autocorrelated errors. However, since the measurement errors $\mathbf{v}_t$ are included in the state-vector $\mathbf{a}_t$, it is no longer a separate vector in the measurement equation. The matrix $\mathbf{I}$ is a (NY×NY) identity matrix.

The state equation then becomes

$$\begin{bmatrix}\mathbf{a}_{t+1}^* \\ \mathbf{v}_{t+1}\end{bmatrix} = \begin{bmatrix}\mathbf{H}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{\Xi}\end{bmatrix} \begin{bmatrix}\mathbf{a}_t^* \\ \mathbf{v}_t\end{bmatrix} + \begin{bmatrix}\mathbf{G}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{I}\end{bmatrix} \begin{bmatrix}\mathbf{z}_t^* \\ \mathbf{b}_t\end{bmatrix}$$

where $\mathbf{I}$ is an (NY×NY) identity matrix. Again, $\mathbf{a}_t^*$, $\mathbf{H}^*$, $\mathbf{G}^*$ and $\mathbf{z}_t^*$ are as in the equivalent model without autocorrelated measurement errors. The lower right (NY×NY) block of the matrix $\mathbf{H}$ is formed by the first order diagonal AR matrix $\mathbf{\Xi}$, as in Equation 28. Thus, the measurement errors $\mathbf{v}_t$ are regressed upon the measurement errors at the previous occasion. The unique part of the measurement errors $\mathbf{b}_t$ are estimated as the extra M elements in the residual vector $\mathbf{z}_t$. Hence the lower right (NY×NY) block of the covariance matrix $\mathbf{Q}$ of $\mathbf{z}_t$ is formed by $\mathbf{\Sigma}_b$, as defined above.

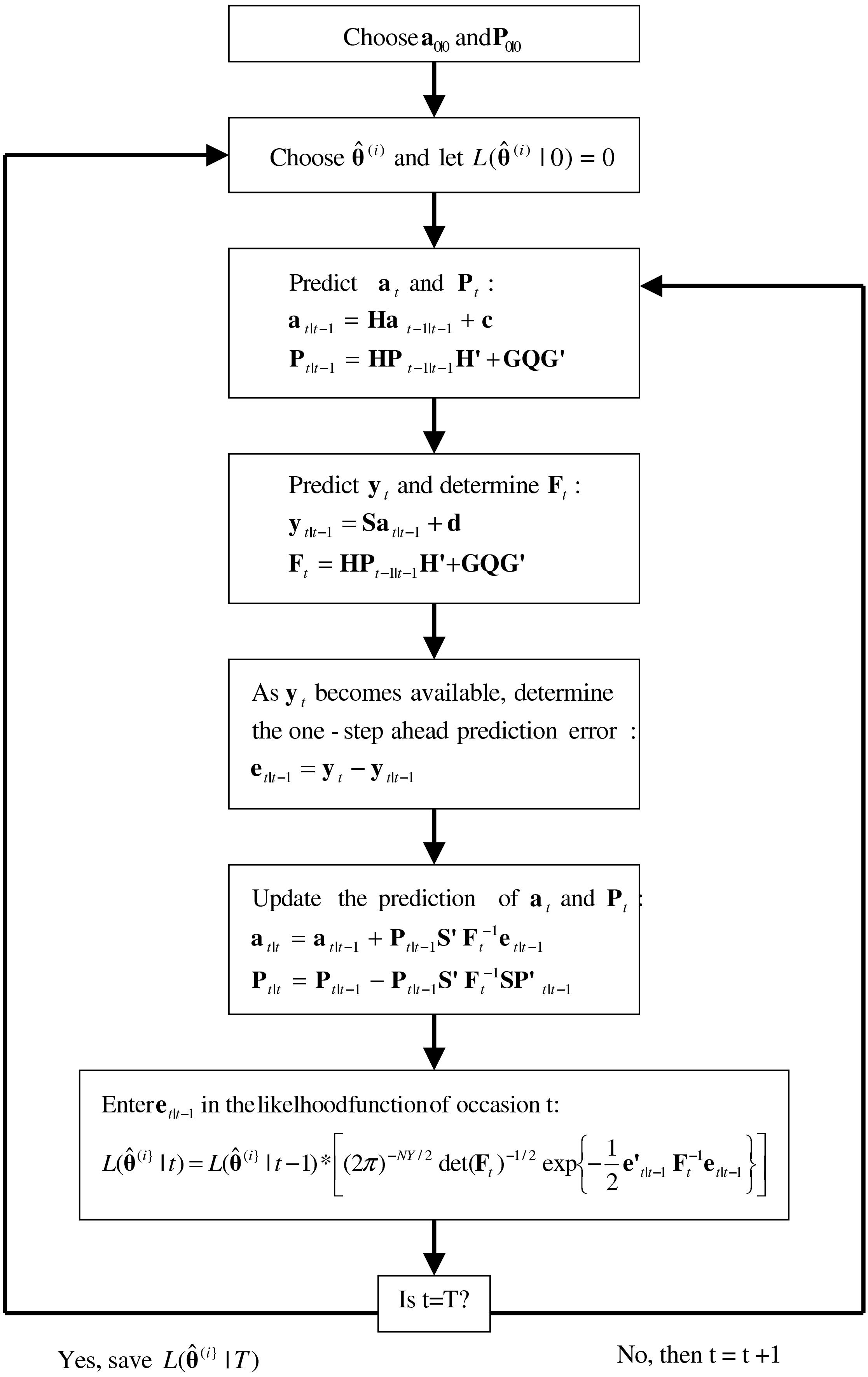
References

Durbin, J., & Koopman, S. J. (2001). *Time series analysis by state space methods*. New York: Oxford University Press.

Hamilton, J. D. (1994). *Time series analysis*. Princeton, New Jersey: Princeton University Press.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge: University Press.

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, *50*, 181–202.


Choose $\mathbf{a}_{0|0}$ and $\mathbf{P}_{0|0}$
Choose $\hat{\boldsymbol{\theta}}^{(i)}$ and let $L(\hat{\boldsymbol{\theta}}^{(i)} \mid 0) = 0$
Predict $\mathbf{a}_t$ and $\mathbf{P}_t$:
$\mathbf{a}_{t|t-1} = \mathbf{Ha}_{t-1|t-1} + \mathbf{c}$
$\mathbf{P}_{t|t-1} = \mathbf{HP}_{t-1|t-1}\mathbf{H}' + \mathbf{GQG}'$
Predict $\mathbf{y}_t$ and determine $\mathbf{F}_t$:
$\mathbf{y}_{t|t-1} = \mathbf{Sa}_{t|t-1} + \mathbf{d}$
$\mathbf{F}_t = \mathbf{HP}_{t-1|t-1}\mathbf{H}' + \mathbf{GQG}'$
As $\mathbf{y}_t$ becomes available, determine the one - step ahead prediction error :
$\mathbf{e}_{t|t-1} = \mathbf{y}_t - \mathbf{y}_{t|t-1}$
Update the prediction of $\mathbf{a}_t$ and $\mathbf{P}_t$:
$\mathbf{a}_{t|t} = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{S}'\mathbf{F}_t^{-1}\mathbf{e}_{t|t-1}$
$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{S}'\mathbf{F}_t^{-1}\mathbf{SP}'_{t|t-1}$
Enter $\mathbf{e}_{t|t-1}$ in the likelhood function of occasion t:
$L(\hat{\boldsymbol{\theta}}^{(i)} \mid t) = L(\hat{\boldsymbol{\theta}}^{(i)} \mid t-1) * \left[(2\pi)^{-NY/2} \det(\mathbf{F}_t)^{-1/2} \exp\left\{ -\frac{1}{2} \mathbf{e}'_{t|t-1} \mathbf{F}_t^{-1} \mathbf{e}_{t|t-1} \right\} \right]$
Is t=T?
Yes, save $L(\hat{\boldsymbol{\theta}}^{(i)} \mid T)$
No, then t = t +1