matrix	dimensions
F_t	$r \times n$
G_t	$n \times n$
V_t	$r \times r$
W_t	$n \times n$
C_0	$n \times n$

1 Model Terminology

These models are called either state space models (econometrics) or dynamic (linear) models (statistics / Bayesian).

The following defines a *state space model*

$$
\begin{aligned}
y_t &= f(\theta_t | b_t, F_t, \nu_t) \\
\theta_t &= f(\theta_{t-1} | g_t, G_t, \nu_t)
\end{aligned}
$$

If θ_t is continuous then it is a *continuous state space model*, if θ_t is discrete then it is a *discrete state space model*.

If those equations can be written as

$$
\begin{aligned}
y_t &= b_t + F_t \theta_t + \nu_t \\
\theta_t &= g_t + G_t \theta_{t-1} + \omega_t
\end{aligned}
$$

then the model is a *Dynamic Linear Model (DLM)* (linear SSM), otherwise it is a non-linear dynamic model. If ν_t and ω_t are normal distributions, then it is *Guassian* or *Normal Dynamic Linear Model* (GDLM or NDLM).

A dynamic linear model is defined by the following set of equations,

$$
\begin{aligned}
& y_t = b_t + F_t \theta_{t-1} + \nu_t & \nu_t \sim N(0, V_t) \quad (1) \\
& \theta_t = g_t + G_t \theta_{t-1} + \omega_t & \omega_t \sim N(0, W_t) \quad (2) \\
& \theta_0 \sim N(m_0, C_0) & (3)
\end{aligned}
$$

where equation 1 is the observation or measurement equation, equation 2 is the system equation, and equation 3 is the initial information. The number of variables is r and the number of states is n.

1.1 Filtering Equations

See Petris, Petrone, and Campagnoli [2, Chapter 2.7, p. 53] and West and Harrison [4, Chapter 4] for proofs.

vector	dimensions
Y_t	r
θ_t	n
b_t	r
ν_t	r
g_t	n
ω_t	n
m_0	n

Assume the posterior distribution at $t-1$. Let $\theta_{t-1}|y_{1:t-1} \sim N(m_{t-1}, C_{t-1})$. The one step ahead predictive distribution of θ_{t-1} given $y_{1:t-1}$ is $N(a_t, R_t)$,

$$a_t = \mathrm{E}(\theta_t|y_{1:t-1}) = g_t + G_t m_{t-1} \tag{4}$$
$$R_t = \mathrm{Var}(\theta_t|y_{1:t-1}) = G_t C_{t-1} G_t' + W_t \tag{5}$$

The one step ahead predictive distribution of Y_t given $y_{1:t-1}$ is $N(f_t, Q_t)$,

$$f_t = \mathrm{E}(Y_t|y_{1:t-1}) = b_t + F_t a_t \tag{6}$$
$$Q_t = \mathrm{Var}(Y_t|y_{1:t-1}) = F_t R_t F_t' + V_t \tag{7}$$

The filtered distribution of θ_t given $y_{1:t}$ is $N(m_t, C_t)$

$$m_t = \mathrm{E}(\theta_t|y_{1:t}) = a_t + R_t F_t' Q_t^{-1} e_t \tag{8}$$
$$= a_t + K_t e_t \tag{9}$$
$$C_t = \mathrm{Var}(\theta_t|y_{1:t}) = R_t - R_t F_t' Q_t^{-1} F_t R_t \tag{10}$$
$$= R_t - K_t Q_t K_t' \tag{11}$$
$$= (I_n - K_t F_t) R_t (I_n - K_t F_t)' + K_t V_t K_t' \tag{12}$$
$$e_t = y_t - f_t \tag{13}$$
$$K_t = R_t F_t' Q_t^{-1} \tag{14}$$

where e_t is the one step forecast error and K_t is the Kalman gain (adaptative coefficient). Equation 12 is the "Joseph stablized form" [3, p. 3].

A few key identies [4, pp. 106-107] are,

$$K_t = R_t F_t' Q_t^{-1} = C_t F_t' V_t^{-1} \tag{15}$$
$$C_t = R_t^{-1} - K_t Q_t K_t = R_t (I_n - F_t K_t') \tag{16}$$
$$C_t^{-1} = R_t^{-1} + F_t' V_t^{-1} F_t \tag{17}$$
$$Q_t = (I_r - F_t K_t)^{-1} V_t \tag{18}$$
$$F_t K_t = I_r - V_t Q_t^{-1} \tag{19}$$

variable	dim
a_t	n
R_t	n, n
f_t	r
Q_t	r, r
m_t	n
C_t	n, n
e_t	r
K_t	n, r

1.2 Missing Values

If all values in t are missing, replace the filter steps (equations 8 and 10) with,

$$m_t = a_t \tag{20}$$
$$C_t = R_t \tag{21}$$

Suppose some, but not all variables in time t are missing. Let $n_t \in [0, r]$ is the number of non-missing values in each time period, $n_t = \sum_{j=1}^{r}(y_{j,t} \neq \emptyset)$. If some are missing (let $r > r_t > 0$ be observed), let M be a $r_t \times r$ selection matrix and define,

$$y_t^* = M_t y_t \tag{22}$$
$$F_t^* = M_t F_t \tag{23}$$
$$V_t^* = M_t V_t M_t' \tag{24}$$

and replace y_t, F_t and V_t in the filtering equations with those. Note that the smoothing (section 1.4) and backward sampling (section **??**) algorithms do not depend on F, V and y and thus do not need to be altered when there are missing values.

1.3 Likelihood

If there are no missing values, then the log likelihood is,

$$L(y_{1:T}) = -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\left(\log|Q_t| + e_t' Q_t^{-1} e_t\right) \tag{25}$$

If there are missing values,

$$L(y_{1:T}) = -\frac{1}{2}\sum_{t=1}^{T}(n_t > 0)\left(n_t \log(2\pi) + \log|Q_t| + e_t' Q_t^{-1} e_t\right) \tag{26}$$

See Koopman, Shephard, and Doornik [1, Chapter 5, p. 57].

1.4 Smoothing

If $\theta_{t+1}|y_{1:T} \sim N(s_{t+1}, S_{t+1})$, then $\theta_t|y_{1:T} \sim N(s_t, S_t)$ where

$$
\begin{aligned}
s_t &= \mathrm{E}(\theta_t|y_{1:T}) &&= m_t + C_t G'_{t+1} R^{-1}_{t+1}(s_{t+1} - a_{t+1}) && (27)\\
S_t &= \mathrm{Var}(\theta_t|y_{1:T}) &&= C_t - C_t G'_{t+1} R^{-1}_{t+1}(R_{t+1} - S_{t+1}) R^{-1}_{t+1} G_{t+1} C_t && (28)
\end{aligned}
$$

Note that $s_T = m_T$ and $S_T = C_T$. See Petris, Petrone, and Campagnoli [2, Prop 2.4, p. 61] for a proof.

1.5 Backward Sampling

Supposing that $m_{1:T}$, $C_{1:T}$, $a_{1:T}$ and $R_{1:T}$ have been calculated by the filter,[1] To draw $\theta_{1:T}|y_{1:T}$,

1. From the Kalman filter, $\theta_T|y_{1:T} \sim N(m_T, C_T)$
2. For $t = (T-1), \dots, 0$, $\theta_t|y_{1:T} \sim N(h_t, H_t)$ where

$$
\begin{aligned}
h_t &= m_t + C_t G'_{t+1} R^{-1}_{t+1}(\theta_{t+1} - a_{t+1}) && (29)\\
H_t &= C_t - C_t G'_t R^{-1}_{t+1} G_{t+1} C_t && (30)
\end{aligned}
$$

See Petris, Petrone, and Campagnoli [2, Chapter 4.4.1, p. 161] for more details.

2 Sequential Estimation

First, consider the case in which all V_t are diagonal. The vector series $y_1, \dots, y_T$ is treated as a scalar series

$$
y_{1,1}, \dots, y_{1,r}, y_{2,1}, \dots, y_{T,r} \tag{31}
$$

The prior distribution of the state is $\theta_t|y_{1:(t-1)} \sim N(a_t, R_t)$,

$$
\begin{aligned}
a_t &= \mathrm{E}(\theta_t|y_{1:t-1}) &&= g_t + G_t m_{t-1} && (32)\\
R_t &= \mathrm{Var}(\theta_t|y_{1:t-1}) &&= G_t C_{t-1} G'_t + W_t && (33)
\end{aligned}
$$

For each variable, $i = 1, \dots, r$. The prediction equation is $y_{t,i}|y_{1:t-1}, y_{t,j|j<i} \sim N(f_{t,i}, Q_{t,i})$. Note that $f_{t,i}$ and $Q_{t,i}$ are scalars.

$$
\begin{aligned}
f_{t,i} &= \mathrm{E}(Y_{t,i}|y_{1:t-1}, y_{t,j|j<i}) &&= b_{t,i} + F_{t,i} m_{t,i-1} && (34)\\
Q_{t,i} &= \mathrm{Var}(Y_{t,i}|y_{1:t-1}, y_{t,j|j<i}) &&= F_{t,i} C_{t,i-1} F'_{t,i} + v_{t,i} && (35)
\end{aligned}
$$

[1] No additional adjustment for intercepts is required because they are already incorporated in a_t

For the first variable, let $m_{t,0} = a_t$ and $C_{t,0} = R_t$. The filtered distribution of the latent state is $\theta_t | y_{1:t-1}, y_{t,j|j<i} \sim N(m_{t,i}, C_{t,i})$,

$$
\begin{aligned}
m_{t,i} &= \mathrm{E}(\theta_t | y_{1:t-1}, y_{t,j|j\leq i}) && = m_{t,i-1} + C_{t,i-1} F'_{t,i} Q^{-1}_{t,i} e_{t,i} && (36)\\
C_{t,i} &= \mathrm{Var}(\theta_t | y_{1:t-1}, y_{t,j|j\leq i}) && = C_{t,i-1} - C_{t,i-1} F'_{t,i} Q^{-1}_{t,i} F_{t,i} C_{t,i-1} && (37)\\
e_{t,i} & && = y_{t,i} - f_{t,i} && (38)\\
K_{t,i} & && = C_{t,i-i} F'_{t,i} Q^{-1}_{t,i} && (39)
\end{aligned}
$$

Define $m_t = m_{t,r}$ and $C_t = C_{t,r}$. If $y_{t,i}$ is missing, then

$$
m_{t,i} = m_{t,i-1} \quad (40)
$$

$$
C_{t,i} = C_{t,i-1} \quad (41)
$$

The likelihood in the sequential case is

$$
L(y_{1:T}) = -\frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{r} (y_{i,j} \neq \emptyset) \left(\log(2\pi) + \log|Q_{t,i}| + \frac{e^2_{t,i}}{Q_{t,i}} \right) \quad (42)
$$

If V_t is not diagonal, then diagonalize it with the Cholesky decomposition of V_t, such that $V_t = L_t D_t L'_t$ where D_t is diagonal, and L_t is lower triangular.

$$
y^*_t = L^{-1}_t y_t \quad (43)
$$

$$
F^*_t = L^{-1}_t F_t \quad (44)
$$

$$
b^*_t = L^{-1}_t b_t \quad (45)
$$

$$
\epsilon^* = L^{-1}_t \epsilon_t \sim N(0, D_t) \quad (46)
$$

2.1 Univariate Local Level Model

For the univariate local level model, with $r = n = 1$, $F_t = G_t = 1$, and $g_t = b_t = 0$, the calculations can be simplified.

One step ahead predictive distribution of θ_{t-1} given $y_{1:t-1}$ is $N(a_t, R_t)$

$$
\begin{aligned}
a_t &= E(\theta_t | y_{1:t-1}) = m_{t-1}\\
R_t &= Var(\theta_t | y_{1:t-1}) = C_{t-1} + W_t
\end{aligned}
$$

The One step ahead predictive distribution of Y_t given $y_{1:t-1}$ is $N(f_t, Q_t)$

$$
\begin{aligned}
f_t &= E(Y_t | y_{1:t-1}) = a_t = m_{t-1}\\
Q_t &= Var(Y_t | y_{1:t-1}) = R_t + V_t = C_{t-1} + W_t + V_t
\end{aligned}
$$

The posterior distribution of θ_t given $y_{1:t}$ is $N(m_t, C_t)$

$$
\begin{aligned}
f_t &= E(\theta_t | y_{1:t}) = a_t + R_t Q^{-1}_t e_t = m_{t-1} + \frac{C_{t-1} + W_t}{C_{t-1} + W_t + V_t} e_t\\
Q_t &= Var(\theta_t | y_{1:t}) = R_t - \frac{(C_{t-1} + W_t)^2}{C_{t-1} + W_t + V_t}
\end{aligned}
$$

where $e_t = Y_t - f_t$. The Kalman gain is defined as $K_t = R_t / Q_t$.

3 Notes

- Use discounting
  - Use a beta rectangular distribution
  - Beta distribution with priors over α and β

4 Examples

4.1 Nile

The data is included with R as `datasets::Nile`.

1. Standard
2. With missing
3. With intervention

4.2 UK Gas

Example from

This is a model of quarterly consumption of gas in the UK from 1960 to 1986. This data is included with R as `datasets::UKGas`. The model to be estimated is a local linear trend model and a quarterly seasonal factor model.

$$
\begin{aligned}
F &= \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \end{bmatrix} & G &= \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \\
V &= \sigma_y^2 & W &= \operatorname{diag}(0, \sigma_\beta^2, \sigma_s^2, 0, 0)
\end{aligned}
$$

4.3 Industrial Production

A linear trend model of U.S. industrial production of consumption goods, seasonally adjusted. The data is IPCONGD; also available on Quandl.

$$
\begin{aligned}
F &= \begin{bmatrix} 1 & 0 \end{bmatrix} & G &= \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \\
V &= \sigma_y^2 & W &= \operatorname{diag}(\sigma_1^2, \sigma_2^2)
\end{aligned}
$$

4.4 CAPM

The state space are the betas of the stocks,

$$\theta_t = (\beta_{1,t}, \dots, \beta_{m,t})' \tag{47}$$

The system equations are,

$$
\begin{aligned}
F_t &= x_t I \quad G_t = I \\
V_t &= \Sigma_\epsilon \quad W_t = \Sigma_\beta
\end{aligned}
\tag{48}
$$

5 Durbin and Koopmans

$$
\begin{aligned}
y_t &= Z_t \alpha_t + \varepsilon_t & \varepsilon_t &\sim N(0, H_t) \\
\alpha_{t+1} &= T_t \alpha_t + R_t \eta_t & \eta_t &\sim N(0, Q_t) \\
\alpha_1 &\sim N(a_1, P_1)
\end{aligned}
$$

with dimensions r (number of disturbances), p (number of variables) and m (number of states).

Matrices,

matrix	dimensions
Z_t	$p \times m$
T_t	$m \times m$
H_t	$p \times p$
Q_t	$r \times r$
R_t	$m \times r$
P_1	$m \times m$

and vectors,

vector	dimensions
y_t	$p \times 1$
α_t	$m \times 1$
ε_t	$p \times 1$
η_t	$r \times 1$
a_1	$m \times 1$

This is initialized with the filtered states.

References

[1] Siem Jan Koopman, Neil Shephard, and Jurgen A. Doornik. *Statistical Algorithms for Models in State Space Form: SsfPack 3.0*. Timberlake Consultants Press, 2008. ISBN: 9780955707636. URL: http://books.google.com/books?id=hLyfQAAACAAJ.

[2] G. Petris, S. Petrone, and P. Campagnoli. *Dynamic Linear Models with R*. Use R! Springer, 2009. ISBN: 9780387772370. URL: http://books.google.com/books?id=VCt3zVq8TO8C.

[3] Fernando Tusell. "Kalman Filtering in R". In: *Journal of Statistical Software* 39.2 (Mar. 1, 2011), pp. 1–27. ISSN: 1548-7660. URL: http://www.jstatsoft.org/v39/i02.

[4] M. West and J. Harrison. *Bayesian forecasting and dynamic models*. Springer series in statistics. Springer, 1997. ISBN: 9780387947259. URL: http://books.google.com/books?id=jcl8lD75fkYC.