Dynamic model: Normal dynamic linear model

Hedibert F. Lopes
Professor of Statistics and Econometrics

Insper Institute of Education and Research

https://www.hedibert.org

March 2026

Outline

NDLMs

Large class of models with time-varying parameters.

Dynamic linear models are defined by a pair of equations, the *observation equation* and the *evolution/system equation*:

$$
\begin{aligned}
y_t &= F_t'\beta_t + \epsilon_t, \qquad \epsilon_t \sim N(0, V) \\
\beta_t &= G_t\beta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W)
\end{aligned}
$$

- y_t: sequence of observations;
- F_t: vector of explanatory variables;
- β_t: d-dimensional state vector;
- G_t: $d \times d$ evolution matrix;
- $\beta_1 \sim N(a, R)$.

Linear growth model

The linear growth model is slightly more elaborate by incorporation of an extra time-varying parameter β_2 representing the growth of the level of the series:

$$
\begin{aligned}
y_t &= \beta_{1,t} + \epsilon_t \quad \epsilon_t \sim N(0, V) \\
\beta_{1,t} &= \beta_{1,t-1} + \beta_{2,t} + \omega_{1,t} \\
\beta_{2,t} &= \beta_{2,t-1} + \omega_{2,t}
\end{aligned}
$$

where $\omega_t = (\omega_{1,t}, \omega_{2,t})' \sim N(0, W)$ and

$$
\begin{aligned}
F_t &= (1, 0)' \\
G_t &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}
\end{aligned}
$$

Prior, updated and smoothed distributions

Prior distributions

$$p(\beta_t|y^{t-k}) \qquad k > 0$$

Updated/online distributions

$$p(\beta_t|y^t)$$

Smoothed distributions

$$p(\beta_t|y^{t+k}) \qquad k > 0$$

Forward learning

Let $y^t = \{y_1, \ldots, y_t\}$.

Posterior at time $t-1$:

$$\beta_{t-1}|y^{t-1} \sim N(m_{t-1}, C_{t-1})$$

Prior at time t:

$$\beta_t|y^{t-1} \sim N(a_t, R_t)$$

with $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t' + W$.

predictive at time t:

$$y_t|y^{t-1} \sim N(f_t, Q_t)$$

with $f_t = F_t' a_t$ and $Q_t = F_t' R_t F_t + V$.

Posterior at time t

$$p(\beta_t|y^t) = p(\beta_t|y_t, y^{t-1}) \propto p(y_t|\beta_t)\, p(\beta_t|y^{t-1})$$

The resulting posterior distribution is

$$\beta_t|y^t \sim N(m_t, C_t)$$

with

$$\begin{aligned} m_t &= a_t + A_t e_t \\ C_t &= R_t - A_t A_t' Q_t \\ A_t &= R_t F_t / Q_t \\ e_t &= y_t - f_t \end{aligned}$$

By induction, these distributions are valid for all times.

Backward smoothing

In dynamic models, the smoothed distribution $\pi(\beta|y^n)$ is more commonly used:

$$\begin{aligned}\pi(\beta|y^n) &= p(\beta_n|y^n)\prod_{t=1}^{n-1} p(\beta_t|\beta_{t+1},\ldots,\beta_n,y^n) \\ &= p(\beta_n|y^n)\prod_{t=1}^{n-1} p(\beta_t|\beta_{t+1},y^t)\end{aligned}$$

Integrating with respect to $(\beta_1,\ldots,\beta_{t-1})$:

$$\begin{aligned}\pi(\beta_t,\ldots,\beta_n|y^n) &= p(\beta_n|y^n)\prod_{k=t}^{n-1} p(\beta_k|\beta_{k+1},y^t) \\ \pi(\beta_t,\beta_{t+1}|y^n) &= p(\beta_{t+1}|y^n)p(\beta_t|\beta_{t+1},y^t)\end{aligned}$$

for $t=1,\ldots,n-1$.

Smoothing: $p(\beta_t|y^n)$

It can be shown that

$$\beta_t|V,W,y^n \sim N(m_t^n, C_t^n)$$

where

$$\begin{aligned} m_t^n &= m_t + C_t G'_{t+1} R_{t+1}^{-1}(m_{t+1}^n - a_{t+1}) \\ C_t^n &= C_t - C_t G'_{t+1} R_{t+1}^{-1}(R_{t+1} - C_{t+1}^n) R_{t+1}^{-1} G_{t+1} C_t \end{aligned}$$

Smoothing: $p(\beta|y^n)$

It can be shown that

$$(\beta_t|\beta_{t+1}, V, W, y^n)$$

is normally distributed with mean

$$(G_t' W^{-1} G_t + C_t^{-1})^{-1}(G_t' W^{-1}\beta_{t+1} + C_t^{-1} m_t)$$

and variance $(G_t' W^{-1} G_t + C_t^{-1})^{-1}$.

Forward filtering, backward sampling (FFBS)

Sampling from $\pi(\beta|y^n)$ can be performed by

- Sampling β_n from $N(m_n, C_n)$ and then
- Sampling β_t from $(\beta_t|\beta_{t+1}, V, W, y^t)$, for $t = n-1, \ldots, 1$.

The above scheme is known as the forward filtering, backward sampling (FFBS) algorithm (Carter and Kohn, 1994 and Frühwirth-Schnatter, 1994).

Individual sampling from $\pi(\beta_t|\beta_{-t}, y^n)$

Let $\beta_{-t} = (\beta_1, \ldots, \beta_{t-1}, \beta_{t+1}, \ldots, \beta_n)$.

For $t = 2, \ldots, n-1$

$$\begin{aligned}
\pi(\beta_t|\beta_{-t}, y^n) &\propto p(y_t|\beta_t)\, p(\beta_{t+1}|\beta_t)\, p(\beta_t|\beta_{t-1}) \\
&\propto f_N(y_t; F_t'\beta_t, V) f_N(\beta_{t+1}; G_{t+1}\beta_t, W) \\
&\times f_N(\beta_t; G_t\beta_{t-1}, W) \\
&= f_N(\beta_t; b_t, B_t)
\end{aligned}$$

where

$$\begin{aligned}
b_t &= B_t(\sigma^{-2}F_t y_t + G_{t+1}' W^{-1}\beta_{t+1} + W^{-1}G_t\beta_{t-1}) \\
B_t &= (\sigma^{-2}F_t F_t' + G_{t+1}' W^{-1} G_{t+1} + W^{-1})^{-1}
\end{aligned}$$

for $t = 2, \ldots, n-1$.

For $t = 1$ and $t = n$,

$$\pi(\beta_1|\beta_{-1}, y^n) = f_N(\beta_1; b_1, B_1)$$

and

$$\pi(\beta_n|\beta_{-n}, y^n) = f_N(\beta_t; b_n, B_n)$$

where

$$\begin{aligned}
b_1 &= B_1(\sigma_1^{-2}F_1y_1 + G_2'W^{-1}\beta_2 + R^{-1}a) \\
B_1 &= (\sigma_1^{-2}F_1F_1' + G_2'W^{-1}G_2 + R^{-1})^{-1} \\
b_n &= B_n(\sigma_n^{-2}F_ny_n + W^{-1}G_n\beta_{n-1}) \\
B_n &= (\sigma_n^{-2}F_nF_n' + W^{-1})^{-1}
\end{aligned}$$

Sampling from $\pi(V, W|y^n, \beta)$

Assume that

$$\begin{array}{rcl} \phi = V^{-1} & \sim & Gamma(n_\sigma/2, n_\sigma S_\sigma/2) \\ \Phi = W^{-1} & \sim & Wishart(n_W/2, n_W S_W/2) \end{array}$$

Full conditionals

$$\begin{array}{rcl} \pi(\phi|\beta, \Phi) & \propto & \prod_{t=1}^{n} f_N(y_t; F_t'\beta_t, \phi^{-1}) \; f_G(\phi; n_\sigma/2, n_\sigma S_\sigma/2) \\ & \propto & f_G(\phi; n_\sigma^*/2, n_\sigma^* S_\sigma^*/2) \\ \pi(\Phi|\beta, \phi) & \propto & \prod_{t=2}^{n} f_N(\beta_t; G_t\beta_{t-1}, \Phi^{-1}) \; f_W\left(\Phi; \frac{n_W}{2}, \frac{n_W S_W}{2}\right) \\ & \propto & f_W(\Phi; n_W^*/2, n_W^* S_W^*/2) \end{array}$$

where $n_\sigma^* = n_\sigma + n$, $n_W^* = n_W + n - 1$,

$$\begin{array}{rcl} n_\sigma^* S_\sigma^* & = & n_\sigma S_\sigma + \sigma(y_t - F_t'\beta_t)^2 \\ n_W^* S_W^* & = & n_W S_W + \Sigma_{t=2}^n (\beta_t - G_t\beta_{t-1})(\beta_t - G_t\beta_{t-1})' \end{array}$$

MCMC scheme to sample from $p(\beta, V, W|y^n)$

- Sample V^{-1} from its full conditional

$$f_G(\phi; n_\sigma^*/2, n_\sigma^* S_\sigma^*/2)$$

- Sample W^{-1} from its full conditional

$$f_W(\Phi; n_W^*/2, n_W^* S_W^*/2)$$

- Sample β from its full conditional

$$\pi(\beta|y^n, V, W)$$

by the FFBS algorithm.

Likelihood for (V, W)

It is easy to see that

$$p(y^n|V, W) = \prod_{t=1}^{n} f_N(y_t|f_t, Q_t)$$

which is the integrated likelihood of (V, W).

Jointly sampling (β, V, W)

(β, V, W) can be sampled jointly by

- Sampling (V, W) from its marginal posterior

  $$\pi(V, W|y^n) \propto l(V, W|y^n)\pi(V, W)$$

  by a rejection or Metropolis-Hastings step;
- Sampling β from its full conditional

  $$\pi(\beta|y^n, V, W)$$

  by the FFBS algorithm.

Jointly sampling (β, V, W) avoids MCMC convergence problems associated with the posterior correlation between model parameters (Gamerman and Moreira, 2002).

Example: Comparing sampling schemes[1]

First order DLM with $V = 1$

$$
\begin{aligned}
y_t &= \beta_t + \epsilon_t\,, \qquad \epsilon_t \sim N(0,1)\\
\beta_t &= \beta_{t-1} + \omega_t\,, \quad \omega_t \sim N(0,W),
\end{aligned}
$$

with $(n, W) \in \{(100, .01), (100, .5), (1000, .01), (1000, .5)\}$.

400 runs: 100 replications per combination.

Priors: $\beta_1 \sim N(0, 10)$ and V and W have inverse Gammas with means set at true values and coefficients of variation set at 10.

Posterior inference: based on 20,000 MCMC draws.

[1]Gamerman, Reis and Salazar (2006) Comparison of sampling schemes for dynamic linear models. *International Statistical Review*, 74, 203-214.

Effective sample size

For a given θ, let $t^{(n)} = t(\theta^{(n)})$, $\gamma_k = Cov_\pi(t^{(n)}, t^{(n+k)})$, the variance of $t^{(n)}$ as $\sigma^2 = \gamma_0$, the autocorrelation of lag k as $\rho_k = \gamma_k/\sigma^2$ and $\tau_n^2/n = Var_\pi(\bar{t}_n)$. It can be shown that, as $n \to \infty$,

$$\tau_n^2 = \sigma^2 \left(1 + 2\sum_{k=1}^{n-1} \frac{n-k}{n}\rho_k\right) \to \sigma^2 \underbrace{\left(1 + 2\sum_{k=1}^{\infty} \rho_k\right)}_{\text{inefficiency factor}}.$$

The *inefficiency factor* measures how far $t^{(n)}$s are from being a random sample and how much $Var_\pi(\bar{t}_n)$ increases because of that.

The *effective sample size* is defined as

$$n_{\text{eff}} = \frac{n}{1 + 2\sum_{k=1}^{\infty} \rho_k}$$

or the size of a random sample with the same variance.

Schemes

Scheme I: Sampling $\beta_1, \ldots, \beta_n, V$ and W individually.
Scheme II: Sampling β, V and W from their conditionals.
Scheme III: Jointly sampling (β, V, W).

Scheme	n=100	n=1000
II	1.7	1.9
III	1.9	7.2

Computing times relative to scheme I. For instance, when $n = 100$ it takes almost 2 times as much to run scheme III.

		Scheme		
W	n	I	II	III
0.01	1000	242	8938	2983
0.01	100	3283	13685	12263
0.50	1000	409	3043	963
0.50	100	1694	3404	923

Sample averages (based on the 100 replications) of effective sample size n_{eff} based on V (see the explanation over the next few pages).

References

Carlin, Polson and Stoffer (1992) A Monte Carlo approach to nonnormal and nonlinear state space modeling. *Journal of the American Statistical Association*, 87, 493-500.

Carter and Kohn (1994) On Gibbs Sampling for State Space Models. *Biometrika*, 81(3), 541-553.

Frühwirth-Schnatter (1994) Data augmentation and dynamic linear models,, *Journal of time series analysis*, 15(2), 183-202.

West and Harrison (1997) *Bayesian Forecasting and Dynamic Models (2nd edition)*. New York: Springer-Verlag.

Migon, Gamerman, Lopes and Ferreira (2005) Dynamic models, In *Handbook of Statistics, Volume 25: Bayesian Thinking, Modeling and Computation* (Eds. D. Dey and C. R. Rao), Amsterdam: Elsevier, 553-588.

Gamerman and Lopes (2006) *MCMC: Stochastic Simulation for Bayesian Inference*. Baton Rouge: Chapman & Hall/CRC.

Petris, Petrone and Campagnoli (2009) *Dynamic Linear Models with R*. New York: Springer.

Schmidt and Lopes (2019) Dynamic models. In Gelfand, Fuentes, Hoeting and Smith (Eds.), *Handbook of Environmental and Ecological Statistics*, pages 57-80. Chapman & Hall.

Prado, Ferreira and West (2021) *Time Series: Modeling, Computation and Inference (2nd edition)*. Baton Rouge: Chapman & Hall/CRC.