Spatio-temporal Methods in Environmental Epidemiology

Alexandra M. Schmidt

Session 10 - Dynamic Linear Models

Department of Epidemiology, Biostatistics and Occupational Health
McGill University

alexandra.schmidt@mcgill.ca

Applied Bayesian Statistics School 2025

University of Genova, Department of Architecture and Design

03-06, June 2025

Linear Dynamic Models

The models we have already seen are all particular cases of a more general linear structure. We will see that this structure encompasses not only the models we've already studied, but also a number of other useful models.
The model is again described by two equations:

$$
\begin{aligned}
\text{Observation Eq.: } \mathbf{y}_t &= \mathbf{F}_t'\theta_t + \nu_t \qquad \nu_t \sim N[0, \mathbf{V}_t] \\
\text{System Eq.: } \theta_t &= \mathbf{G}_t\theta_{t-1} + \omega_t \qquad \omega_t \sim N[0, \mathbf{W}_t] \\
\text{Initial Info.: } (\theta_0 \mid D_0) &\sim N[\mathbf{m}_0, \mathbf{C}_0]
\end{aligned} \tag{1}
$$

- $\mathbf{y}_t$ observation vector of size $r \times 1$;
- θ_t parameter vector of size $n \times 1$;
- $\mathbf{F}_t$ known matrix of size $n \times r$;
- $\mathbf{G}_t$ known matrix of size $n \times n$;
- $\mathbf{V}_t$ known covariance matrix of size $r \times r$;
- $\mathbf{W}_t$ known covariance matrix of size $n \times n$;

We can also define the model using the quadruple $\{\mathbf{F}, \mathbf{G}, V, \mathbf{W}\}_t$.

Examples

(i) First-order model:
$F_t = 1$, $G_t = 1$, $\theta_t = \mu_t$

(ii) Dynamic regression model through the origin with n regressors:
$\mathbf{F}'_t = (x_{t,1}, \cdots, x_{n,t})$, $\mathbf{G} = \mathbf{I}_n$, $\theta'_t = (\beta_{t,1}, \cdots, \beta_{t,n})$ where $\mathbf{I}_n$ is the identity matrix of order n.

Inference Cycle

The inference cycle is as follows:

$$
\begin{array}{ccccc}
\theta_{t-1} \mid D_{t-1} & \xrightarrow{\text{EVOLUTION}} & \theta_t \mid D_{t-1} & \xrightarrow{\text{UPDATE}} & \theta_t \mid D_t \\
\text{posterior} & & \text{prior} & & \text{posterior} \\
& & \downarrow & & \\
& & Y_t \mid D_{t-1} & & \\
& & \text{prediction} & &
\end{array}
$$

Inference Cycle

The distributions of interest are obtained through the following theorem:

Theorem

In the univariate DLM, the one-step-ahead forecast and posterior distributions, for all t, are given by:

(a) Posterior at $t-1$:
For some mean $\mathbf{m_{t-1}}$ and variance matrix $\mathbf{C}_{t-1}$,

$$\theta_{t-1} \mid D_{t-1} \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}]$$

(b) Prior at t:

$$\theta_t \mid D_{t-1} \sim N[\mathbf{a}_t, \mathbf{R}_t], \text{ where:}$$

$$\begin{aligned} \mathbf{a}_t &= \mathbf{G}_t \mathbf{m}_{t-1} \\ \mathbf{R}_t &= \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \end{aligned}$$

McGill

Theorem

(continued...)

(c) *One-step-ahead forecast:*

$$\mathbf{Y}_t \mid D_{t-1} \sim N[\mathbf{f}_t, \mathbf{Q}_t], \text{ where:}$$

$$\begin{aligned} \mathbf{f}_t &= \mathbf{F}_t' \mathbf{a}_t \\ \mathbf{Q}_t &= \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t \end{aligned}$$

(d) *Posterior at t:*

$$\theta_t \mid D_t \sim N[\mathbf{m}_t, \mathbf{C}_t], \text{ where:}$$

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t \qquad \mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1} \\ \mathbf{e}_t &= \mathbf{Y}_t - \mathbf{f}_t \\ \mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t \mathbf{A}_t' Q_t \end{aligned}$$

Checking Model Adequacy in the Scalar Y_t Case

The one-step-ahead forecast is given by $f_t = E(Y_t \mid y_{1:t-1})$, and the forecast error is defined as:

$$e_t = Y_t - E(Y_t \mid y_{1:t-1}) = Y_t - f_t.$$

As seen, alternatively we can write:

$$\begin{aligned} e_t &= Y_t - F_t' a_t = F_t' \theta_t + v_t - F_t' a_t \\ &= F_t'(\theta_t - a_t) + v_t \end{aligned}$$

The sequence of forecast errors, $(e_t)_{t \geq 1}$, has interesting properties, the most important of which are summarized in the following proposition.

McGill

Checking Model Adequacy in the Scalar Y_t Case

Proposition: Let $(e_t)_{t\geq 1}$ be the sequence of forecast errors from a DLM. Then the following properties hold:

1. $E(e_t) = 0$.
2. The random vector e_t is uncorrelated with any function of $Y_1, \cdots, Y_{t-1}$.
3. For any $s < t$, e_t and Y_s are uncorrelated.
4. For any $s < t$, e_t and e_s are uncorrelated.
5. e_t is a linear function of $Y_1, \cdots, Y_t$.
6. $(e_t)_{t\geq 1}$ is a Gaussian process.

McGill

Checking Model Adequacy in the Scalar Y_t Case

- When the observations are univariate, the sequence of standardized errors, defined as $\tilde{e}_t = e_t/\sqrt{Q_t}$, follows a standard normal distribution.
- This property can be used to verify the model assumptions: if the model is correct, the sequence $\tilde{e}_1, \cdots, \tilde{e}_t$ should resemble a sample of size t from a standard normal distribution.
- The most useful tools for this verification are: QQ-plots and empirical autocorrelation plots of the standardized error.

Predictive Distributions

Definition

For any current time t, the forecast function $f_t(k)$ is defined for all integers $k \geq 0$ as:

$$f_t(k) = E[\mu_{t+k} \mid D_t] = E[\mathbf{F}'_{t+k}\theta_{t+k} \mid D_t],$$

where

$$\mu_{t+k} = \mathbf{F}'_{t+k}\theta_{t+k}$$

is the mean response function.

McGill

Predictive Distributions

Theorem

(Theorem 7.2) For each time t and $k \geq 1$, the k-step-ahead distributions for θ_{t+k} and Y_{t+k} given D_t are:

(a) *State Distribution:*

$$(\theta_{t+k} \mid D_t) \sim N[\mathbf{a}_t(k), \mathbf{R}_t(k)]$$

(b) *Forecast Distribution:*

$$(Y_{t+k} \mid D_t) \sim N[f_t(k), Q_t(k)]$$

with moments defined recursively by:

$$\begin{aligned}
f_t(k) &= \mathbf{F}_t'\mathbf{a}_t(k) \quad \textit{and} \\
Q_t(k) &= \mathbf{F}_t'\mathbf{R}_t(k)\mathbf{F}_t + V_{t+k}, \quad \textit{where} \\
\mathbf{a}_t(k) &= \mathbf{G}_{t+k}\mathbf{a}_t(k-1), \quad \textit{and} \\
\mathbf{R}_t(k) &= \mathbf{G}_{t+k}\mathbf{R}_t(k-1)\mathbf{G}_{t+k}' + W_{t+k},
\end{aligned}$$

with initial values $\mathbf{a}_t(0) = \mathbf{m}_t$ and $\mathbf{R}_t(0) = \mathbf{C}_t$.

Predictive Distributions

Corolary

In the special case where the evolution matrix $\mathbf{G}_t$ *is constant,* $\mathbf{G}_t = \mathbf{G}$ *for all* t*, then for* $k \geq 0$*,*

$$\mathbf{a}_t(k) = \mathbf{G}^k \mathbf{m}_t \tag{2}$$

such that

$$f_t(k) = \mathbf{F}'_{t+k} \mathbf{G}^k \mathbf{m}_t \tag{3}$$

If, in addition, $\mathbf{F}_t = \mathbf{F}$ *for all* t*, the model is a Time-Invariant DLM (DLMST), and the forecast function takes the form:*

$$f_t(k) = \mathbf{F}' \mathbf{G}^k \mathbf{m}_t.$$

The important point of this result is that forecasts are governed by powers of **G**.

Predictive Distributions

Corolary

If the evolution matrix $\mathbf{G}_t = \mathbf{G}$ *is constant for all* t*, then for* $k, v \geq 0$*,*

$$\mathbf{R}_t(k) = \mathbf{G}^k\mathbf{C}_t + \sum_{i=0}^{k-1} \mathbf{G}^i\mathbf{W}_{t+k-i}(\mathbf{G}^i)'$$

and $\mathbf{C}_t(k+v, k) = \mathbf{G}^v\mathbf{R}_t(k)$*. If, in addition,* $\mathbf{F}_t = \mathbf{F}$ *for all* t*, then:*

$$\mathbf{Q}_t(k) = \mathbf{F}'\mathbf{R}_t(k)\mathbf{F} + V_{t+k}$$

and

$$C[Y_{t+k+v}, Y_{t+k}] = \mathbf{F}'\mathbf{C}_t(k+v, k)\mathbf{F}.$$

Models with Unknown Variances

- So far, it has been assumed in the general case that the observation variance, denoted by V_t, was known at all times
- In practice, this rarely happens, and what needs to be done is to estimate the variance at each time step
- Following the same structure of the method, the variance will be estimated sequentially, and an evolution form will be obtained to relate it over successive time points. From now on, we will work with $V_t = V_{t-1} = V = \phi^{-1}$, for all t, where ϕ is the observational precision.

General Model with Unknown Observational Variance

For all t, the model is defined as:

$$\begin{aligned} Y_t &= \mathbf{F}_t' \theta_t + \nu_t \quad , \quad \nu_t \sim N[0, V] \\ \theta_t &= \mathbf{G}_t \theta_{t-1} + \omega_t \quad , \quad \omega_t \sim N[0, V\mathbf{W}_t^*] \\ (\theta_0 \mid D_0, V) &\sim N[\mathbf{m}_0, V\mathbf{C}_0^*] \\ (V \mid D_0) &\sim IG[n_0/2, d_0/2] \end{aligned} \tag{4}$$

Again, we assume that $V_t = V_{t-1} = V$ for all t.
The usual independence assumptions remain valid, but now conditional on V, or equivalently, $\phi = V^{-1}$. The prior for ϕ has mean $E[\phi \mid D_t] = n_0/d_0 = 1/S_0$, where S_0 is a prior estimate for the observational variance V.

McGill

Inference Procedure

The inferential process once again consists of:

$$\underset{\text{posterior}}{\theta_{t-1}, V \mid D_{t-1}} \overset{\text{EVOLUTION}}{\longrightarrow} \underset{\text{prior}}{\theta_t, V \mid D_{t-1}} \overset{\text{UPDATE}}{\longrightarrow} \underset{\text{posterior}}{\theta_t, V \mid D_t}$$

$$\downarrow$$

$$\underset{\text{prediction}}{Y_t \mid D_{t-1}}$$

McGill

Inference Procedure

Theorem:

With the model specified as in (4), we obtain the following distributions for each time $t \geq 1$:

(a) Conditional on V:

$$
\begin{aligned}
(\theta_{t-1} \mid D_{t-1}, V) &\sim N[\mathbf{m}_{t-1}, V\mathbf{C}^*_{t-1}], \\
(\theta_t \mid D_{t-1}, V) &\sim N[\mathbf{a}_t, V\mathbf{R}^*_t], \\
(Y_t \mid D_{t-1}, V) &\sim N[f_t, VQ^*_t], \\
(\theta_t \mid D_t, V) &\sim N[\mathbf{m}_t, V\mathbf{C}^*_t],
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{a}_t &= \mathbf{G}_t\mathbf{m}_{t-1}, \qquad \mathbf{R}^*_t = \mathbf{G}_t\mathbf{C}^*_{t-1}\mathbf{G}'_t + \mathbf{W}^*_t \\
f_t &= \mathbf{F}'_t\mathbf{a}_t, \qquad Q^*_t = 1 + \mathbf{F}'_t\mathbf{R}^*_t\mathbf{F}_t \\
\mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t, \qquad \mathbf{C}^*_t = \mathbf{R}^*_t - \mathbf{A}_t\mathbf{A}'_t Q^*_t \\
\mathbf{A}_t &= \mathbf{R}^*_t\mathbf{F}_t {Q^*_t}^{-1}, \qquad e_t = Y_t - f_t
\end{aligned}
$$

Inference Procedure

(b) For the precision $\phi = V^{-1}$:

$$\begin{aligned} (\phi \mid D_{t-1}) &\sim G(n_{t-1}/2, d_{t-1}/2), \\ (\phi \mid D_t) &\sim G(n_t/2, d_t/2), \end{aligned}$$

where:
$n_t = n_{t-1} + 1$ and $d_t = d_{t-1} + e_t^2 {Q_t^*}^{-1}$

McGill

Inference Procedure

(c) Marginal (Unconditional on V):

$$
\begin{aligned}
(\theta_{t-1} \mid D_{t-1}) &\sim t_{n_{t-1}}[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}], \\
(\theta_t \mid D_{t-1}) &\sim t_{n_{t-1}}[\mathbf{a}_t, \mathbf{R}_t], \\
(Y_t \mid D_{t-1}) &\sim t_{n_{t-1}}[f_t, Q_t], \\
(\theta_t \mid D_t) &\sim t_{n_t}[\mathbf{m}_t, \mathbf{C}_t],
\end{aligned}
$$

where:

$$
\begin{aligned}
\mathbf{C}_{t-1} &= S_{t-1}\mathbf{C}^*_{t-1}, \quad \mathbf{R}_t = S_{t-1}\mathbf{R}^*_t, \quad Q_t = S_{t-1}Q^*_t \\
\mathbf{C}_t &= S_t\mathbf{C}^*_t, \quad S_{t-1} = d_{t-1}/n_{t-1}, \quad S_t = d_t/n_t
\end{aligned}
$$

We can also obtain the following relationships:

$$
\begin{aligned}
\mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t e_t, \qquad \mathbf{C}_t = \frac{S_t}{S_{t-1}}[\mathbf{R}_t - \mathbf{A}_t\mathbf{A}'_t Q_t] \\
n_t &= n_{t-1} + 1, \qquad d_t = d_{t-1} + S_{t-1} e_t^2 Q_t^{-1} \\
Q_t &= S_{t-1} + \mathbf{F}'_t\mathbf{R}_t\mathbf{F}_t, \qquad \mathbf{A}_t = \mathbf{R}_t\mathbf{F}_t Q_t^{-1}
\end{aligned}
$$

Backward learning: the Kalman smoother

- The Kalman filter only provides posterior distribution of a given state θ_t conditionally on the past observations, $y_1, \ldots, y_t$, $p(\theta_t | D_t, V)$
- We are omitting the dependence on the state variances $\mathbf{W}_1^*, \ldots, \mathbf{W}_n^*$.
- One may want to obtain the posterior distribution of the states given the whole set of observations, $y_1, \ldots, y_n$, i.e. $p(\theta_1, \ldots, \theta_n | V, D_n)$; for instance, to understand the dynamics driving observations as opposed to simply forecasting its future values.

Backward learning: the Kalman smoother

Full joint distribution of states

By the Markov property of Equations (4), this joint posterior can be rewritten as

$$p(\theta_1, \ldots, \theta_n | V, D_n) = p(\theta_n | V, D_n) \prod_{t=1}^{n-1} p(\theta_t | \theta_{t+1}, V, D_t). \qquad (5)$$

First, the forward learning scheme of the previous Section (Kalman filter) is run to obtain $p(\theta_t | V, D_t)$, for $t = 1, \ldots, n$. Then, an analogous backward learning scheme (Kalman smoother) is run to obtain $p(\theta_t | \theta_{t+1}, V, D_t)$, for $t = n - 1, \ldots, 1$.

Backward learning: the Kalman smoother

Full joint distribution of states

More precisely, another (backwards) application of Bayes' theorem leads to

$$\begin{aligned} p(\theta_t|\theta_{t+1}, V, D_t) &\propto f_N(\theta_{t+1}; \mathbf{G_t}\theta_\mathbf{t}, \mathbf{VW_t^*})\mathbf{f_N}(\theta_\mathbf{t}; \mathbf{m_t}, \mathbf{VC_t^*}) \\ &\propto f_N(\theta_{t+1}; \widetilde{\mathbf{m}}_t, V\widetilde{\mathbf{C}}_t^*), \end{aligned} \tag{6}$$

where $\widetilde{\mathbf{m}}_t = \widetilde{\mathbf{C}}_t^*(\mathbf{G}_t'\mathbf{W}_t^{-1}\theta_{t+1} + \mathbf{C_t^{*-1}m_t})$ and $\widetilde{\mathbf{C}}_t^* = (\mathbf{G}_t'\mathbf{W}_t^{*-1}\mathbf{G}_t + \mathbf{C}_t^{*-1})^{-1}$, for $t = n-1, n-2, \ldots, 1$.
From the Kalman filter, $V|D_n \sim IG(n_n/2, n_nS_n/2)$, so it follows that $\theta_n|D_n \sim t_{n_n}(\tilde{m}_n, S_n\widetilde{\mathbf{C}}_n^*)$ and

$$\theta_t|\theta_{t+1}, D_t \sim t_{n_n}(\widetilde{\mathbf{m}}_t, S_n\widetilde{\mathbf{C}}_t^*), \tag{7}$$

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

Marginal distributions of the states

Similarly, it can be shown that the marginal distribution of θ_t is

$$\theta_t | V, D_n \sim t_{n_n}(\overline{\mathbf{m}}_t, S_n \overline{\mathbf{C}}_t^*), \tag{8}$$

where, for $t = n-1, n-2, \ldots, 1$, $\overline{\mathbf{m}}_t = \mathbf{m}_t + \mathbf{C}_t^* \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{*-1} (\overline{\mathbf{m}}_{t+1} - \mathbf{a}_{t+1})$ and $\overline{\mathbf{C}}_t^* = \mathbf{C}_t^* - \mathbf{C}_t^* \mathbf{G}'_{t+1} \mathbf{R}_{t+1}^{*-1} (\mathbf{R}_{t+1}^* - \overline{\mathbf{C}}_{t+1}^*) \mathbf{R}_{t+1}^{*-1} \mathbf{G}_{t+1} \mathbf{C}_t^*$.

Marginal distributions of the states

Full conditional distributions of the states

Let $\theta_{-t} = \{\theta_1, \ldots, \theta_{t-1}, \theta_{t+1}, \ldots, \theta_n\}$ and $t = 2, \ldots, n-1$, it follows that the full conditional distribution of θ_t is

$$
\begin{aligned}
p(\theta_t|\theta_{-t}, V, D_n) &\propto f_N(y_t; \mathbf{F}_t'\theta_t, V)\, f_N(\theta_{t+1}; \mathbf{G_{t+1}}\boldsymbol{\theta_t}, \mathbf{VW^*_{t+1}}) \\
&\times f_N(\theta_t; \mathbf{G}_t\theta_{t-1}, V\mathbf{W}^*_t) = f_N(\theta_t; \mathbf{b}_t, V\mathbf{B}^*_t), \quad (9)
\end{aligned}
$$

where $\mathbf{b}_t = \mathbf{B}^*_t(\mathbf{F}_t y_t + \mathbf{G}'_{t+1}\mathbf{W}^{*-1}_{t+1}\theta_{t+1} + \mathbf{W}^{*-1}_t\mathbf{G}_t\theta_{t-1})$ and $\mathbf{B}^*_t = (\mathbf{F}_t\mathbf{F}'_t + \mathbf{G}'_{t+1}\mathbf{W}^{*-1}_{t+1}\mathbf{G}_{t+1} + \mathbf{W}^{*-1}_t)^{-1}$. The endpoint parameters θ_1 and θ_n also have full conditional distributions $N(\mathbf{b}_1, V\mathbf{B}^*_1)$ and $N(\mathbf{b}_n, V\mathbf{B}^*_n)$, respectively, where $\mathbf{b}_1 = \mathbf{B}^*_1(\mathbf{F}_1 y_1 + \mathbf{G}'_2\mathbf{W}^{*-1}_2\theta_2 + \mathbf{R}^{*-1}\mathbf{a}_1)$, $\mathbf{B}^*_1 = (\mathbf{F}_1\mathbf{F}'_1 + \mathbf{G}'_2\mathbf{W}^{*-1}_2\mathbf{G}_2 + \mathbf{R}^{-1})^{*-1}$, $\mathbf{b}_n = \mathbf{B}^*_n(\mathbf{F}_n y_n + \mathbf{W}^{*-1}_n\mathbf{G}_n\theta_{n-1})$ and $\mathbf{B}^*_n = (\mathbf{F}_n\mathbf{F}'_n + \mathbf{W}^{*-1}_n)^{-1}$. Again,

$$
\theta_t|\theta_{-t}, D_t \sim t_{n_n}(\mathbf{b}_t, S_n\mathbf{B}^*_t) \qquad \text{for all } t. \quad (10)
$$

McGill

Full posterior inference

- When the evolution variances $\mathbf{W}_1^*, \ldots, \mathbf{W}_n^*$ are unknown, closed-form analytical full posterior inference is infeasible and numerical or Monte Carlo approximations are needed
- Numerical integration, in fact, is only realistically feasible for very low dimensional settings
- Markov Chain Monte Carlo methods have become the norm over the last quarter of century for state space modelers
- In particular, the full joint of Equation (5) can be combined with full conditional distributions for V and $\mathbf{W}_1^*, \ldots, \mathbf{W}_n^*$
- This is the well known *forward filtering, backward sampling (FFBS)* algorithm of Carter and Kohn (1994) and Früwirth-Schnatter (1994)
- The FFBS algorithm is commonly used for posterior inference in Gaussian and conditionally Gaussian DLMs
- The main steps needed to fit DLM's using the software `R` are in the package `dlm`, detailed in Petris, Petrone and Campagnoli (2009)

Marginal distribution

Another very important result of the sequential Bayesian depicted above is the derivation of the marginal likelihood of $\mathbf{W}_1^*, \ldots, \mathbf{W}_t^*$ given $y_1, \ldots, y_n$. Without loss of generality, assume that $\mathbf{W}_t = \mathbf{W}$ for all $t = 1, \ldots, n$, where n is the sample size, and that $p(\mathbf{W})$ denotes the prior distribution of $\mathbf{W}$. In this case,

$$p(y_1, \ldots, y_n|\mathbf{W}) = \prod_{t=1}^{n} p(y_t|D_{t-1}, \mathbf{W}), \tag{11}$$

where $y_t|D_{t-1} \sim t_{n_{t-1}}(f_t, S_{t-1}Q_t^*)$. Therefore, the posterior distribution of $\mathbf{W}$, $p(\mathbf{W}|D_n) \propto p(D_n|\mathbf{W})p(\mathbf{W})$, can be combined with $p(V|\mathbf{W}, D_n)$ to produce the joint posterior of $(V, \mathbf{W})$:

$$\begin{aligned} p(V, \mathbf{W}|D_n) &\propto p(D_n|\mathbf{W})p(V|\mathbf{W}, D_n)p(\mathbf{W}) \\ &= \left[\prod_{t=1}^{n} p(y_t; f_t, S_{t-1}Q_t^*, n_{t-1})\right] p(V; n_n/2, n_nS_n/2)p(\mathbf{W}). \end{aligned}$$

McGill

Marginal distribution

- Gaussianity and linearity, leads to a posterior distribution for V and $\mathbf{W}$ by integrating out all state space vectors $\theta_1, \theta_2, \ldots, \theta_n$

- If M independent Monte Carlo draws $\{(V, \mathbf{W})^{(1)}, \ldots, (V, \mathbf{W})^{(M)}\}$ are obtained from $p(V, \mathbf{W}|D_n)$, then M independent Monte Carlo draws from $p(\theta_1, \ldots, \theta_n|D_n)$ are easily obtained by repeating the FFBS of Equation (5) (or Equation (6)) M times

- This leads to M independent Monte Carlo draws from $p(\theta_1, \ldots, \theta_n, V, \mathbf{W}|D_n)$, hence there is no need for iterative Markov chain Monte Carlo (MCMC) schemes.

Some properties of NDLMs

Superposition of models

- More general time series structures can be modeled through a composition of DLM components. This is possible because of a theorem proved in West & Harrison (1997)

- Consider m time series y_{it} generated by NDLMs identified by the quadruples $M_i = \{\mathbf{F}_{it}, \mathbf{G}_{it}, V_{it}, \mathbf{W}_{it}\}$, $i = 1, 2, \ldots, m$

- Under model M_i, the state vector θ_{it} is of dimension k_i, and the observation and evolution error series are respectively ε_{it} and ω_{it}

- The state vectors are distinct and, for all distinct $i \neq j$, the series ε_{it} and ω_{it} are mutually independent of the series ε_{jt} and ω_{jt}

- Then the series $y_t = \sum_{i=1}^{m} y_{it}$ follows a k-dimensional DLM $\{\mathbf{F}_t, \mathbf{G}_t, V_t, \mathbf{W}_t\}$, where $k = k_1 + \ldots + k_m$, the state vector θ_t is given by $\theta_t = (\theta'_{1t}, \ldots, \theta'_{mt})'$, $\mathbf{F}_t = \left(\mathbf{F}'_{1t}, \ldots, \mathbf{F}'_{mt}\right)'$, $\mathbf{G}_t = \text{block diag}(\mathbf{G}_{1t}, \ldots, \mathbf{G}_{mt})$, $\mathbf{W}_t = \text{block diag}(\mathbf{W}_{1t}, \ldots, \mathbf{W}_{mt})$, and $V_t = \sum_{i=1}^{m} V_{it}$.

Some properties of NDLMs

Superposition of models

- **Example:** *(Dynamic regression)* If at each time t a pair of observations (y_t, x_t) is available, and there is a local linear relationship between y_t and x_t, a simple dynamic regression is defined by making $\mathbf{F}_t' = (1\ x_t)$, $\mathbf{G} = \mathbf{I}_2$, the 2-dimensional identity matrix, and $\mathbf{W} = \text{diag}(W_1, W_2)$.

- **Example:** *(Time varying level plus an annual seasonal component)* $\mathbf{F}_t' = (1\ 1\ 0)$, $\mathbf{W} = \text{diag}(W_1, W_2, W_3)$, and

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\pi/12) & \sin(2\pi/12) \\ 0 & -\sin(2\pi/12) & \cos(2\pi/12) \end{pmatrix}.$$

Discount factors

- The sequential updating steps described previously follow normal or Student-t distributions because we assume that $\mathbf{W}_t$ are known for all time t
- In practice, this is rarely true
- One way to avoid estimating the elements of the covariance matrix $\mathbf{W}_t$ is through the use of discounting factors
- Let $\mathbf{P}_t = \mathbf{G}_t\mathbf{C}_{t-1}\mathbf{G}_t' = Var[\mathbf{G}_t\theta_t|D_{t-1}]$, which can be viewed as the prior variance in a NDLM with $\mathbf{W}_t = \mathbf{0}$, that is, with no evolution error term. The usual NDLM leads to $Var[\mathbf{G}_t\theta_t|D_{t-1}] = \mathbf{P}_t + \mathbf{W}_t$
- The idea of discounting factor is introduced by making $\mathbf{W}_t = \mathbf{P}_t(1-\delta)/\delta$, for some $\delta \in (0, 1]$. Then, $\mathbf{R}_t = \mathbf{P}_t/\delta$, resulting in an increase in the prior variance of θ_t, at time $t-1$

Discount factors

- Note that, for given values of δ and $\mathbf{C}_0$ the series $\mathbf{W}_1, \ldots, \mathbf{W}_n$ is identified
- If a NDLM has different components following the superposition theorem, a possible strategy is to specify different values of the discount factor for the different components
- Note that the higher the discount factor, the higher the degree of smoothing. For this reason, the value of δ is typically used in the range $[0.8, 1]$ for polynomial, regression, and seasonal components (West & Harrison, 1997).

Intervention and monitoring

- Because of the sequential learning structure of NDLMs, they naturally accommodate changes in the observed time series. For example, if an observation is missing at time t, then $D_t = D_{t-1}$, and $\theta_t | D_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$, with $\mathbf{m}_t = \mathbf{a}_t$ and $\mathbf{C}_t = \mathbf{R}_t$

- On the other hand, if a change has occurred and it is difficult to attribute it to a particular component, the prior variance matrix $\mathbf{R}_t$ can be changed to reflect the increased uncertainty about all parameters, without a change in the prior mean $\mathbf{a}_t$ that would anticipate the direction of the change

- The adequacy of NDLMs can be monitored online through the comparison between the observed value and the one-step ahead forecast distribution. In particular, Bayes' factors can be used, and are usually based on the predictive densities of the forecast errors, $e_t = y_t - f_t$. See (Chapter 11, West & Harrison, 1997) for more details both on intervention and monitoring in NDLMs.

Dynamic generalized linear models (DGLM)

- The DLM was extended by West, Harrison & Migon (1985) to the case wherein observations belong to the exponential family

- Assume that observations y_t are generated from a dynamic exponential family (EF), defined as

$$p(y_t|\eta_t,\phi) = \exp\{\phi[y_t\eta_t - a(\eta_t)]\}b(y_t,\phi), \tag{12}$$

  where $a(.)$ and $b(.)$ are known functions, η_t is the canonical parameter, and ϕ is a scale parameter, usually time invariant

- We denote the distribution of y_t as $EF(\eta_t,\phi)$

Dynamic generalized linear models (DGLM)

- Let $\mu_t = E(y_t|\eta_t,\phi)$ and $g(.)$ be a known link function assumed at least twice differentiable, which relates the linear predictor with the canonical parameter, that is

$$g(\mu_t) = \mathbf{F}(\psi_1)'\theta_t, \tag{13}$$

  where $\mathbf{F}$, and θ_t are vectors of dimension k, and ψ_1 denotes unknown quantities involved in the definition of $\mathbf{F}(\psi_1)$

- Following the parameterization of the exponential family in equation (12) it follows that the mean and variance of y_t are, respectively, given by $E(y_t|\eta_t,\phi) = \mu_t = \frac{da(\eta_t)}{d\eta_t} = \dot{a}(\eta_t)$ and $V(y_t|\eta_t,\phi) = V_t = \phi^{-1}\ddot{a}(\eta_t)$.

McGill

Dynamic generalized linear models (DGLM)

- Usually, in practice, assuming $g(.)$ to be the natural link function provides good results
- Some examples, where the link function is suggested by the definition of the canonical function, include the log-linear Poisson and logistic-linear Bernoulli models
- The state parameters, θ_t, evolve through time via a Markovian structure, that is

$$\theta_t = \mathbf{G}(\psi_2)\theta_{t-1} + \omega_t, \quad \omega_t \sim N(\mathbf{0}, \mathbf{W}). \tag{14}$$

- The hyperparameter vector ψ_2 represents possible unknown quantities in $\mathbf{G}(\cdot)$, the $k \times k$ evolution matrix

Dynamic generalized linear models (DGLM)

- Lastly, ω_t is the disturbance associated with the system evolution with covariance structure $\mathbf{W}$, commonly a diagonal matrix which can vary over time

- The initial information of the model is denoted by θ_0, and its prior distribution is defined through a k-variate normal distribution, that is, $\theta_0|D_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$, where D_0 denotes the initial information set. Typically, we assume that the components of θ_0 and ω_t are independent for all time periods

Dynamic generalized linear models (DGLM)

Posterior inference

- West *et al.* (1985) specify only the first and second moments of ω_t
- They perform inference taking advantage of conjugate prior and posterior distributions in the exponential family, and use linear Bayes' estimation to obtain estimates of the moments of $\theta_t|D_t$.
- This approach is appealing as no assumption is made about the shape of the distribution of the disturbance component ω_t
- However, we are limited to learn only about the first two moments of the posterior distribution of $\theta_t|D_t$

Dynamic generalized linear models (DGLM)

Posterior inference

- The assumption of normality of ω_t allows us to write down a likelihood function for θ_t and all the other parameters in the model

- Inference procedure can be performed using Markov chain Monte Carlo algorithms. However, care must be taken when proposing an algorithm to obtain samples from the posterior distribution of the state vectors θ_t

- West *et al.* (1985), Kitagawa (1987), and Fahrmeir (1992), among others, proposed different methods to obtain approximations of the posterior distribution of parameters belonging to DGLMs

- Estimation methods based on MCMC can be seen in Shephard and Pitt (1997), Gamerman (1998), Geweke (2001), Migon *et al.* (2013). More recently, Alves *et al.* (2025) propose a procedure for k-parametric DGLM based on information geometry concepts