Practicum: Dynamic Linear Models

Carlo Zaccardi and Alexandra M. Schmidt

ABS25 Applied Bayesian Statistics School

03-06, June 2025

General DLM

- **Dynamic Linear Models (DLM)** are a class of models characterized by their ability to model time-varying relationships.
- DLMs consist of two main components: the **observation equation** and the **state equation**.

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + \boldsymbol{z}_t' \gamma + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{1}$$

$$\boldsymbol{\beta}_t = \boldsymbol{F}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\beta}_t, \quad \boldsymbol{\beta}_t \sim N(0, \boldsymbol{W}) \tag{2}$$

where:

- y_t is the observed time series at time $t = 1, \ldots, T$.
- $\boldsymbol{x}_t$ is a p-dimensional vector of covariates at time t (including the intercept).
- $\boldsymbol{F}_t$ is a $p \times p$ evolution matrix.
- $\boldsymbol{\beta}_t$ is the state vector at time t, initialized with $\boldsymbol{\beta}_1 \sim N(0, \boldsymbol{D})$.
- $\boldsymbol{z}_t$ is a q-dimensional vector of covariates whose effect is constant over time.
- γ is a vector of coefficients for $\boldsymbol{z}_t$.
- ϵ_t is the observation error.
- $\boldsymbol{W}$ is the covariance matrix of the state evolution error.

Bayesian Inference

- To perform Bayesian inference in DLMs, the **Forward Filtering and Backward Smoothing (FFBS)** algorithm is often used [Carter and Kohn, 1994, Frühwirth-Schnatter, 1994].
- However, the FFBS algorithm can be computationally expensive for large datasets.
- Chan and Jeliazkov [2009] proposed an efficient version of the FFBS that uses **block-banded and sparse matrix algorithms**.
- Their techniques are scalable in the dimension of the model and do not involve the Kalman filtering and smoothing recursions.

A Simple Derivation

Rewrite the observation equation (1) in seemingly unrelated regression (SUR) form:

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Z}\gamma + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim N(0, \sigma_\epsilon^2 \boldsymbol{I}) \tag{3}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_T)'$,

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1' & 0 & \cdots & 0 \\ 0 & \boldsymbol{x}_2' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{x}_T' \end{bmatrix}, \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{pmatrix}, \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{z}_1' \\ \boldsymbol{z}_2' \\ \vdots \\ \boldsymbol{z}_T' \end{bmatrix}$$

A Simple Derivation

The directed conditional structure in (2) implies a joint density for $\boldsymbol{\beta}$ which can be obtained by defining

$$\boldsymbol{H} = \begin{pmatrix} I_p & & & & \\ -F_2 & I_p & & & \\ & -F_3 & I_p & & \\ & & \ddots & \ddots & \\ & & & -F_T & I_p \end{pmatrix} \quad \text{and} \quad \boldsymbol{S} = \begin{pmatrix} D & & & & \\ & W & & & \\ & & W & & \\ & & & \ddots & \\ & & & & W \end{pmatrix}$$

so that (2) can be written as $\boldsymbol{H}\boldsymbol{\beta} = \boldsymbol{\eta}$, where $\boldsymbol{\eta} \sim N(0, \boldsymbol{S})$.

Hence, the prior distribution is $\boldsymbol{\beta} \sim N(0, \boldsymbol{K}^{-1})$ where $\boldsymbol{K} = \boldsymbol{H}'\boldsymbol{S}^{-1}\boldsymbol{H}$ is is block-banded and contains only a small number of non-zero elements.

Some Remarks

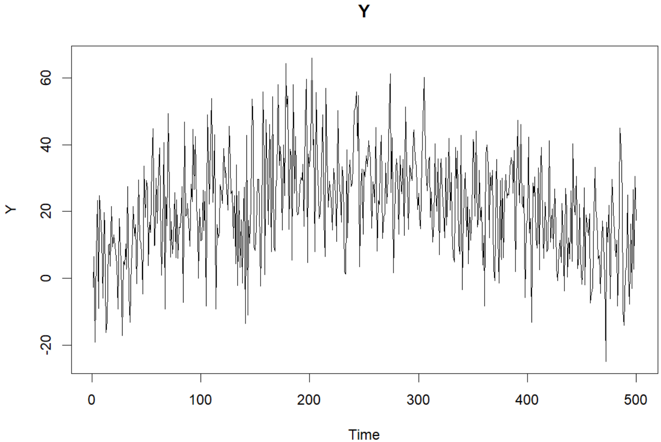
- Inference proceeds following the standard Gibbs sampling algorithm for linear models with normal errors.
- The posterior distribution of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}|\boldsymbol{y}, \gamma, \boldsymbol{W}, \sigma_\epsilon^2 \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{P}^{-1})$$

  where $\boldsymbol{\mu}_\beta = \boldsymbol{P}^{-1}(\sigma_\epsilon^{-2}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{Z}\gamma))$, and $\boldsymbol{P} = \boldsymbol{K} + \sigma_\epsilon^{-2}\boldsymbol{X}'\boldsymbol{X}$.
- $\boldsymbol{P}$ is block-banded and symmetric, thus storage costs are low and the computational benefits of working with banded or sparse matrix algorithms can be exploited.
- Chan and Jeliazkov [2009] discuss an efficient state smoothing and simulation algorithm, which is based on the Cholesky factorization of $\boldsymbol{P}$.

Example: Simulated Data

- We simulate a DLM with $T = 500$ observations, $p = 3$ states, and $q = 1$ covariate with fixed effect.
- The covariates $\boldsymbol{x}_t$ and z_t are generated from normal distributions independently.



Fitting DLMs in R

There are several packages in R for Bayesian inference in (generalized) DLM framework, including:

- `dlm`
- `bsts`
- `bssm`

We will use the first two packages, and also my own implementation of the Chan and Jeliazkov (2009) algorithm in R/C++.

Question: Pollution Data

- You are going to fit a DLM to the daily average PM2.5 concentrations in Milan from 2018 to 2022 (see file `pollution_milan.txt`).
- Pollutant data are collected from the Copernicus Atmosphere Monitoring Service's (CAMS) Atmosphere Data Store, whereas climate data are obtained from the ERA5 reanalyses dataset available at the Copernicus Climate Change Service's (C3S) Climate Data Store.
- Research question: ***Is there any time-varying effect of temperature and relative humidity on PM2.5?***

Question: Pollution Data

- Plot the PM2.5 time series and compare it to the predictor's time series. Is there any correlation?
- Consider whether a log-linear model is appropriate.
- Fit a DLM with the structure discussed in the previous slides. Use the `bsts` package and Chan and Jeliazkov's algorithm.
- Plot the time-varying coefficients with their 95% CI and comment the results. Compare the DLM to a simple linear model using WAIC, and summarize your findings in a concise paragraph.

References I

Chris K Carter and Robert Kohn. On gibbs sampling for state space models. *Biometrika*, 81(3): 541–553, 1994.

Joshua CC Chan and Ivan Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2): 101–120, 2009.

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