Bayesian Dynamic Linear Model - Shared Variance Case

Daniel Zhou, Sudipto Banerjee

2023-01-23

Contents

Chapter 1

Preface

This is a first tutorial for Bayesian Linear Regression assembled in book form.

1.1 Acknowledgements

Chapter 2

Basics of Bayesian linear regression

2.1 Bayes' theorem

Theorem 2.1 (Bayes' theorem)**.** *For events A, B and $P(B) \neq 0$, we have*

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

We denote U as unknown parameters and K as known parameters. We call $P(U)$ prior and $P(K|U)$ likelihood. The Bayes' theorem gives us the posterior distribution of unknown parameters given the known parameters

$$P(U \mid K) \propto P(U) \cdot P(K \mid U)$$

Let $K = \left\{y_{n\times 1}, X_{n\times p}\right\}$ and assume $y \sim N\left(X\beta, \sigma^2 V\right)$, where V is known and $U = \left\{\beta, \sigma^2\right\}$ is unknown. The likelihood is given by

$$P(K \mid U) = N\left(y \mid X\beta, \sigma^2 V\right) \tag{2.1}$$

2.2 Normal Inverse-Gamma (NIG) prior

2.2.1 Joint distribution of NIG prior

Definition 2.1 (Normal Inverse-Gamma Distribution)**.** Suppose

$$x \mid \sigma^2, \mu, M \sim \mathrm{N}(\mu, \sigma^2 M)$$
$$\sigma^2 \mid \alpha, \beta \sim \mathrm{IG}(\alpha, \beta)$$

Then (x, σ^2) has a Normal-Inverse-Gamma distribution, denoted as $(x, \sigma^2) \sim \mathrm{NIG}(\mu, M, \alpha, \beta)$.

We use a Normal Inverse-Gamma prior for (β, σ^2)

$$\begin{aligned}
P(\beta, \sigma^2) &= NIG\left(\beta, \sigma^2 \mid m_0, M_0, a_0, b_0\right) && (2.2)\\
&= IG\left(\sigma^2 \mid a_0, b_0\right) \cdot N\left(\beta \mid m_0, \sigma^2 M_0\right) && (2.3)\\
&= \frac{b_0^{a_0}}{\Gamma\left(a_0\right)}\left(\frac{1}{\sigma^2}\right)^{a_0+1} e^{-\frac{b_0}{\sigma^2}} \frac{1}{(2\pi)^{\frac{p}{2}}\left|\sigma^2 M_0\right|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2} Q(\beta, m_0, M_0)} && (2.4)\\
&= \frac{b_0^{a_0}}{\Gamma\left(a_0\right)}\left(\frac{1}{\sigma^2}\right)^{a_0+1} e^{-\frac{b_0}{\sigma^2}} \frac{1}{(2\pi\sigma^2)^{\frac{p}{2}}\left|M_0\right|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2} Q(\beta, m_0, M_0)} && (2.5)\\
& && (2.6)
\end{aligned}$$

where $Q(x, m, M) = (x - m)^\top M^{-1}(x - m)$

Note: the Inverse-Gamma (IG) distribution has a relationship with Gamma distribution. $X \sim Gamma(\alpha, \beta)$, the density function of X is $f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$. Let $Y = \frac{1}{X} \sim IG(\alpha, \beta)$, the density function of Y is $f(y) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}$.

2.2.2 Marginal distribution of NIG prior

As for marginal priors, we can can get it by integration

$$P(\sigma^2) = \int NIG\left(\beta, \sigma^2 \mid m_0, M_0, a_0, b_0\right)\, d\beta = IG\left(\sigma^2 \mid a_0, b_0\right)$$
$$P(\beta) = \int NIG\left(\beta, \sigma^2 \mid m_0, M_0, a_0, b_0\right)\, d\sigma^2 = t_{2a_0}(m_0, \frac{b_0}{a_0} M_0)$$

Click to show or hide details

$$
\begin{aligned}
P\left(\sigma^2\right) &= \int NIG\left(\beta, \sigma^2 \mid m_0, M_0, a_0, b_0\right)\ d\beta &(2.7)\\
&= IG\left(\sigma^2 \mid a_0, b_0\right) \int N\left(\beta \mid m_0, \sigma^2 M_0\right)\ d\beta &(2.8)\\
&= IG\left(\sigma^2 \mid a_0, b_0\right) &(2.9)
\end{aligned}
$$

$$
\begin{aligned}
P(\beta) &= \int NIG\left(\beta, \sigma^2 \mid m_0, M_0, a_0, b_0\right) \ d\sigma^2 && (2.10)\\
&= \int \frac{b_0^{a_0}}{\Gamma\left(a_0\right)} \left(\frac{1}{\sigma^2}\right)^{a_0+1} e^{-\frac{b_0}{\sigma^2}} \frac{1}{\left(2\pi\sigma^2\right)^{\frac{p}{2}} \left|M_0\right|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2} Q(\beta, m_0, M_0)} \ d\sigma^2 && (2.11)\\
&= \frac{b_0^{a_0}}{\Gamma\left(a_0\right) (2\pi)^{\frac{p}{2}} \left|M_0\right|^{\frac{1}{2}}} \int \left(\frac{1}{\sigma^2}\right)^{a_0+\frac{p}{2}+1} e^{-\frac{1}{\sigma^2}\left(b_0+\frac{1}{2}Q(\beta, m_0, M_0)\right)} \ d\sigma^2 && (2.12)\\
&\quad (\text{let } u = \frac{1}{\sigma^2}, \left|d\sigma^2\right| = \frac{1}{u^2} du) && (2.13)\\
&= \frac{b_0^{a_0}}{\Gamma\left(a_0\right) (2\pi)^{\frac{p}{2}} \left|M_0\right|^{\frac{1}{2}}} \int u^{a_0+\frac{p}{2}+1} e^{-\left(b_0+\frac{1}{2}Q(\beta, m_0, M_0)\right)u} \frac{1}{u^2} \ du && (2.14)\\
&= \frac{b_0^{a_0}}{\Gamma\left(a_0\right) (2\pi)^{\frac{p}{2}} \left|M_0\right|^{\frac{1}{2}}} \int u^{a_0+\frac{p}{2}-1} e^{-\left(b_0+\frac{1}{2}Q(\beta, m_0, M_0)\right)u} \ du && (2.15)\\
&\quad (\text{by Gamma integral function: } \int x^{\alpha-1} exp^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^{\alpha}}) && (2.16)\\
&= \frac{b_0^{a_0}}{\Gamma\left(a_0\right) (2\pi)^{\frac{p}{2}} \left|M_0\right|^{\frac{1}{2}}} \frac{\Gamma\left(a_0+\frac{p}{2}\right)}{\left[b_0+\frac{1}{2}Q(\beta, m_0, M_0)\right]^{\left(a_0+\frac{p}{2}\right)}} && (2.17)\\
&= \frac{b_0^{a_0}\Gamma\left(a_0+\frac{p}{2}\right)}{\Gamma\left(a_0\right) (2\pi)^{\frac{p}{2}} \left|M_0\right|^{\frac{1}{2}}} \left[b_0\left(1+\frac{1}{2b_0}Q(\beta, m_0, M_0)\right)\right]^{-\left(a_0+\frac{p}{2}\right)} && (2.18)\\
&= \frac{b_0^{a_0}\Gamma\left(a_0+\frac{p}{2}\right) b_0^{-\left(a_0+\frac{p}{2}\right)}}{\Gamma\left(a_0\right) (2\pi)^{\frac{p}{2}} \left|M_0\right|^{\frac{1}{2}}} \left[1+\frac{1}{2b_0}\left(\beta-m_0\right) \top M_0^{-1} \left(\beta-m_0\right)\right]^{-\left(a_0+\frac{p}{2}\right)} && (2.19)\\
&= \frac{\Gamma\left(a_0+\frac{p}{2}\right)}{(2\pi)^{\frac{p}{2}} b_0^{\frac{p}{2}} \Gamma\left(a_0\right) |M|^{\frac{1}{2}}} \left[1+\frac{1}{2b_0}\left(\beta-m_0\right) \top M_0^{-1} \left(\beta-m_0\right)\right]^{-\left(a_0+\frac{p}{2}\right)} && (2.20)\\
&= \frac{\Gamma\left(a_0+\frac{p}{2}\right)}{(2\pi)^{\frac{p}{2}} \left(a_0 \cdot \frac{b_0}{a_0}\right)^{\frac{p}{2}} \Gamma\left(a_0\right) |M|^{\frac{1}{2}}} \left[1+\frac{1}{2a_0 \cdot \frac{b_0}{a_0}}\left(\beta-m_0\right) \top M_0^{-1} \left(\beta-m_0\right)\right]^{-\left(a_0+\frac{p}{2}\right)} && (2.21)\\
&= \frac{\Gamma\left(a_0+\frac{p}{2}\right)}{\left(2a_0\pi\right)^{\frac{p}{2}} \Gamma\left(a_0\right) \left|\frac{b_0}{a_0}M\right|^{\frac{1}{2}}} \left[1+\frac{1}{2a_0}\left(\beta-m_0\right)^{\top} \left(\frac{b_0}{a_0}M_0\right)^{-1} \left(\beta-m_0\right)\right]^{-\left(a_0+\frac{p}{2}\right)} && (2.22)\\
&= t_{2a_0}(m_0, \frac{b_0}{a_0}M_0) && (2.23)
\end{aligned}
$$

Note: the density of multivariate t-distribution is given by

$$t_v(\mu, \Sigma) = \frac{\Gamma\left(\frac{v+p}{2}\right)}{(v\pi)^{\frac{p}{2}}\Gamma\left(\frac{v}{2}\right)|\Sigma|^{\frac{1}{2}}}\left[1 + \frac{1}{v}(x-\mu)^\top\Sigma^{-1}(x-\mu)\right]^{-\frac{v+p}{2}}$$

2.3 Posterior distribution

The posterior distribution of (β, σ^2) is given by

$$P\left(\beta, \sigma^2 \mid y\right) = NIG\left(\beta, \sigma^2 \mid M_1 m_1, M_1, a_1, b_1\right) \tag{2.24}$$

where

$$\begin{aligned}
M_1 &= (M_0^{-1} + X^\top V^{-1}X)^{-1} \ ; && (2.25)\\
m_1 &= M_0^{-1}m_0 + X^\top V^{-1}y \ ; && (2.26)\\
a_1 &= a_0 + \frac{p}{2} \ ; && (2.27)\\
b_1 &= b_0 + \frac{c^*}{2} = b_0 + \frac{1}{2}\left(m_0^\top M_0^{-1}m_0 + y^\top V^{-1}y - m_1^\top M_1 m_1\right) \ . && (2.28)
\end{aligned}$$

Click to show or hide details

$$\begin{aligned}
P\left(\beta, \sigma^2 \mid y\right) &\propto NIG\left(\beta, \sigma^2 \mid m_0, M_0, a_0, b_0\right) \cdot N\left(y \mid X\beta, \sigma^2 V\right)\\
&\propto IG\left(\sigma^2 \mid a_0, b_0\right) \cdot N\left(\beta \mid m_0, \sigma^2 M_0\right) \cdot N\left(y \mid X\beta, \sigma^2 V\right)\\
&\propto \frac{b_0^{a_0}}{\Gamma\left(a_0\right)}\left(\frac{1}{\sigma^2}\right)^{a_0+1} e^{-\frac{b_0}{\sigma^2}} \frac{1}{(2\pi\sigma^2)^{\frac{p}{2}}\left|M_0\right|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}Q(\beta, m_0, M_0)} \frac{1}{(2\pi\sigma^2)^{\frac{p}{2}}\left|V\right|^{\frac{1}{2}}} e^{-\frac{1}{2\sigma^2}Q(y, X\beta, V)}\\
&\propto \left(\frac{1}{\sigma^2}\right)^{a_0+p+1} e^{-\frac{b_0}{\sigma^2}} e^{-\frac{1}{2\sigma^2}(Q(\beta, m_0, M_0) + Q(y, X\beta, V))} && (2.29)
\end{aligned}$$

where

$$\begin{aligned}
Q\left(\beta, m_0, M_0\right) + Q\left(y, X\beta, V\right) &= (\beta - m_0)^\top M_0^{-1}(\beta - m_0) + (y - X\beta)^\top V^{-1}(y - X\beta)\\
&= \beta^\top M_0^{-1}\beta - 2\beta^\top M_0^{-1}m_0 + m_0^\top M_0^{-1}m_0\\
&\quad + \beta^\top X^\top V^{-1}X\beta - 2\beta^\top X^\top V^{-1}y + y^\top V^{-1}y\\
&= \beta^\top\left(M_0^{-1} + X^\top V^{-1}X\right)\beta - 2\beta^\top\left(M_0^{-1}m_0 + X^\top V^{-1}y\right)\\
&\quad + m_0^\top M_0^{-1}m_0 + y^\top V^{-1}y\\
&= \beta^\top M_1^{-1}\beta - 2\beta^\top m_1 + c\\
&= (\beta - M_1 m_1)^\top M_1^{-1}(\beta - M_1 m_1) - m_1^\top M_1 m_1 + c\\
&= (\beta - M_1 m_1)^\top M_1^{-1}(\beta - M_1 m_1) + c^* && (2.30)
\end{aligned}$$

where M_1 is a symmetric positive definite matrix, m_1 is a vector, and c & c^* are scalars given by

$$\begin{aligned} M_1 &= (M_0^{-1} + X^\top V^{-1} X)^{-1} \ ; && (2.31)\\ m_1 &= M_0^{-1} m_0 + X^\top V^{-1} y \ ; && (2.32)\\ c &= m_0^\top M_0^{-1} m_0 + y^\top V^{-1} y \ ; && (2.33)\\ c^* &= c - m^\top M m = m_0^\top M_0^{-1} m_0 + y^\top V^{-1} y - m_1^\top M_1 m_1 \ . && (2.34) \end{aligned}$$

Note: M_1, m_1 and c do not depend upon β.

Then, we have

$$\begin{aligned} P\left(\beta, \sigma^2 \mid y\right) &\propto \left(\frac{1}{\sigma^2}\right)^{a_0+p+1} e^{-\frac{b_0}{\sigma^2}} e^{-\frac{1}{2\sigma^2}((\beta - M_1 m_1)^\top M_1^{-1}(\beta - M_1 m_1) + c^*)} && (2.35)\\ &\propto \left(\frac{1}{\sigma^2}\right)^{a_0+p+1} e^{-\frac{b_0 + \frac{c^*}{2}}{\sigma^2}} e^{-\frac{1}{2\sigma^2}(\beta - M_1 m_1)^\top M_1^{-1}(\beta - M_1 m_1)} && (2.36)\\ &\propto \left(\frac{1}{\sigma^2}\right)^{a_0+\frac{p}{2}+1} e^{-\frac{b_0 + \frac{c^*}{2}}{\sigma^2}} (\frac{1}{\sigma^2})^{\frac{p}{2}} e^{-\frac{1}{2\sigma^2}(\beta - M_1 m_1)^\top M_1^{-1}(\beta - M_1 m_1)} && (2.37)\\ &= IG\left(\sigma^2 \mid a_0 + \frac{p}{2}, b_0 + \frac{c^*}{2}\right) \cdot N\left(\beta \mid M_1 m_1, \sigma^2 M_1\right) && (2.38)\\ &= IG\left(\sigma^2 \mid a_1, b_1\right) \cdot N\left(\beta \mid M_1 m_1, \sigma^2 M_1\right) && (2.39)\\ &= NIG\left(\beta, \sigma^2 \mid M_1 m_1, M_1, a_1, b_1\right) && (2.40) \end{aligned}$$

where

$$\begin{aligned} M_1 &= (M_0^{-1} + X^\top V^{-1} X)^{-1} \ ; && (2.41)\\ m_1 &= M_0^{-1} m_0 + X^\top V^{-1} y \ ; && (2.42)\\ a_1 &= a_0 + \frac{p}{2} \ ; && (2.43)\\ b_1 &= b_0 + \frac{c^*}{2} = b_0 + \frac{1}{2}\left(m_0^\top M_0^{-1} m_0 + y^\top V^{-1} y - m_1^\top M_1 m_1\right) \ . && (2.44) \end{aligned}$$

From derivation in marginal priors, the marginal posterior distributions can be easily get by updating corresponding parameters

$$P\left(\sigma^2 \mid y\right) = IG\left(\sigma^2 \mid a_1, b_1\right)$$
$$P(\beta \mid y) = t_{2a_1}(M_1 m_1, \frac{b_1}{a_1} M_1)$$

2.4 Bayesian prediction

Assume $V = I_n$. Let $\tilde{y}$ denote an $\tilde{n} \times 1$ vector of outcomes. $\tilde{X}$ is corresponding predictors. We seek to predict $\tilde{y}$ based upon y

$$P(\tilde{y} \mid y) = t_{2a_1}(\tilde{X} M_1 m_1, \frac{b_1}{a_1}(I_{\tilde{n}} + \tilde{X} M_1 \tilde{X}^\top)) \tag{2.45}$$

Click to show or hide details

$$\begin{aligned}
P\left(\beta, \sigma^2, \tilde{y} \mid y\right) &= P\left(\beta, \sigma^2 \mid y\right) \cdot P\left(\tilde{y} \mid \beta, \sigma^2, y\right) && (2.46)\\
&\propto P\left(\beta, \sigma^2\right) \cdot P\left(y \mid \beta, \sigma^2\right) \cdot P\left(\tilde{y} \mid \beta, \sigma^2, y\right) && (2.47)\\
&= NIG\left(\beta, \sigma^2 \mid m_0, M_0, a_0, b_0\right) \cdot N\left(y \mid X\beta, \sigma^2 I_n\right) \cdot N\left(\tilde{y} \mid \tilde{X}\beta, \sigma^2 I_{\tilde{n}}\right) && (2.48)\\
&= NIG\left(\beta, \sigma^2 \mid M_1 m_1, M_1, a_1, b_1\right) \cdot N\left(\tilde{y} \mid \tilde{X}\beta, \sigma^2 I_{\tilde{n}}\right) && (2.49)\\
&= IG(\sigma^2 \mid a_1, b_1) \cdot N\left(\beta \mid M_1 m_1, \sigma^2 M_1\right) \cdot N\left(\tilde{y} \mid \tilde{X}\beta, \sigma^2 I_{\tilde{n}}\right) && (2.50)
\end{aligned}$$

Then we can calculate posterior predictive density $P(\tilde{y} \mid y)$ from $P\left(\beta, \sigma^2, \tilde{y} \mid y\right)$

$$\begin{aligned}
P(\tilde{y} \mid y) &= \iint P\left(\beta, \sigma^2, \tilde{y} \mid y\right)\, d\beta\, d\sigma^2 && (2.51)\\
&= \iint IG(\sigma^2 \mid a_1, b_1) \cdot N\left(\beta \mid M_1 m_1, \sigma^2 M_1\right) \cdot N\left(\tilde{y} \mid \tilde{X}\beta, \sigma^2 I_{\tilde{n}}\right)\, d\beta\, d\sigma^2 && (2.52)\\
&= \int IG(\sigma^2 \mid a_1, b_1) \int N\left(\beta \mid M_1 m_1, \sigma^2 M_1\right) \cdot N\left(\tilde{y} \mid \tilde{X}\beta, \sigma^2 I_{\tilde{n}}\right)\, d\beta\, d\sigma^2 && (2.53)\\
& && (2.54)
\end{aligned}$$

As for $\int N\left(\beta \mid M_1 m_1, \sigma^2 M_1\right) \cdot N\left(\tilde{y} \mid \tilde{X}\beta, \sigma^2 I_{\tilde{n}}\right)\, d\beta$, we provide an easy way to derive it avoiding any integration at all. Note that we can write the above model as

$$\begin{aligned}
\tilde{y} &= \tilde{X}\beta + \tilde{\epsilon}, \text{ where } \tilde{\epsilon} \sim N(0, \sigma^2 I_{\tilde{n}}) && (2.55)\\
\beta &= M_1 m_1 + \epsilon_{\beta|y}, \text{ where } \epsilon_{\beta|y} \sim N(0, \sigma^2 M_1) && (2.56)
\end{aligned}$$

where $\tilde{\epsilon}$ and $\epsilon_{\beta|y}$ are independent of each other. It then follows that

$$\tilde{y} = \tilde{X}M_1m_1 + \tilde{X}\epsilon_{\beta|y} + \tilde{\epsilon} \sim N(\tilde{X}M_1m_1, \sigma^2(I_{\tilde{n}} + \tilde{X}M_1\tilde{X}^\top)) \tag{2.57}$$

As a result

$$P(\tilde{y} \mid y) = \int IG(\sigma^2 \mid a_1, b_1) \cdot N(\tilde{X}M_1m_1, \sigma^2(I_{\tilde{n}} + \tilde{X}M_1\tilde{X}^\top)) \; d\sigma^2 \tag{2.58}$$

$$= t_{2a_1}(\tilde{X}M_1m_1, \frac{b_1}{a_1}(I_{\tilde{n}} + \tilde{X}M_1\tilde{X}^\top)) \tag{2.59}$$

2.5 Sampling process

We can get the joint posterior density $P\left(\beta, \sigma^2, \tilde{y} \mid y\right)$ by sampling process

1) Draw $\hat{\sigma}^2_{(i)}$ from $IG\left(a_1, b_1\right)$
2) Draw $\hat{\beta}_{(i)}$ from $N\left(M_1m_1, \hat{\sigma}^2_{(i)}M_1\right)$
3) Draw $\tilde{y}_{(i)}$ from $N\left(\tilde{X}\hat{\beta}_{(i)}, \hat{\sigma}^2_{(i)}I_{\tilde{n}}\right)$

Chapter 3

The Divide & Conquer Algorithm

The divide and conquer algorithm is a strategy of solving a large problem by breaking it down into two or more smaller and more manageable sub-problems of the same or of a similar genre. The solutions to the sub-problems are then combined to get the desired output: the solution to the original problem.

In this section, we apply the divide and conquer algorithm to the Bayesian linear regression framework.

3.1 The Problem

Let $y = (y_1, y_2, ..., y_n)^T$ be an $n \times 1$ random vector of outcomes and X be a fixed $n \times p$ matrix of predictors with full column rank. Consider the following Bayesian (hierarchical) linear regression model,

$$IG(\sigma^2 \mid a_0, b_0) \times N(\beta \mid m_0, \sigma^2 M_0) \times N(y \mid X\beta, \sigma^2 I_n).$$

The posterior density is $p(\beta, \sigma^2 \mid y) = IG(\sigma^2 \mid a_0^*, b_0^*) \times N(\beta \mid Mm, \sigma^2 M)$, and to carry out Bayesian inference, we first sample $\sigma^2 \sim IG(a^*, b^*)$ and then, for each sampled σ^2, we sample $\beta \sim N(Mm, \sigma^2 M)$.

Assume that n is so large that we are unable to store or load y or X into our CPU to carry out computations for them. We decide to divide our data set into K mutually exclusive and exhaustive subsets, each comprising a manageable number of points. Note that p is small, so computations involving $p \times p$ matrices are fine. Let y_k denote the $m_k \times 1$ sub-vector of y, and X_k be the $m_k \times p$ sub-matrix of X in subset k, where each m_k has been chosen by us so that $m_k > p$, and is small enough such that we can fit the above model on $\{y_k, X_k\}$. This section will clearly explain how we can still compute a^*, b^*, M and m without ever having to store or compute with y or X, but with quantities computed using only the subsets $\{y_k, X_k\}$ for $k = 1, 2, ..., K$.

3.2 The Solution

Using the multivariate completing the square method, we know that the explicit expressions for a^*, b^*, m and M are given by:

- $a_0^* = a_0 + \frac{n}{2}$
- $b_0^* = b_0 + \frac{c^*}{2}$
- $c = m_0^T M_0^{-1} m_0 + y^T y - m^T M m$
- $m = (X^T y + M_0^{-1} m_0)$
- $M = (X^T X + M_0^{-1})^{-1}$

This implies that the explicit expressions for the a_0^*, b_0^*, m and M for the $i_t h$ subset are given by:

- $a_{0i}^* = a_0 + \frac{m_i}{2}$
- $b_{0i}^* = b_0 + \frac{c^*}{2}$
- $c_i^* = m_0^T M_0^{-1} m_0 + y_i^T y_i - m_i^T M_i m_i$
- $m_i = (X_i^T y_i + M_0^{-1} m_0)$
- $M_i = (X_i^T X_i + M_0^{-1})^{-1}$

We can express the posteriors of the entire data set as a function of the posteriors of the subsets with some math, resulting in:

- $a_0^* = \sum_{i=1}^k a_{0i}^* + (k-1) a_0$
- $b_0^* = b_0 + m_0^T M_0^{-1} m_0 + \sum_{i=1}^k y_i^T y_i - m^T M m$
- $m = \sum_{i=1}^k m_i - (k-1) M_0^{-1} m_0$
- $M = \sum_{i=1}^k (X_i^T X_i + M_0^{-1})^{-1} = (\sum_{i=1}^k M_i^{-1} - (k-1) M_0^{-1})^{-1}$

Chapter 4

Dynamic Linear Model - Shared Variance

This chapter discusses the Dynamic Linear Model with a scale factor for the variance shared across time and its derivations at each step. The approach taken in this chapter is borrowed from West and Harrison (1997), with some details derived from Petris et al (2009). For full generality and to maintain a multivariate normal system in both the data and parameter matrices, we assume all $Y_t \in \mathbb{R}^n$, $\beta_t \in \mathbb{R}^p$, and $t \in \{1, \ldots, T\}$ for some integer T.

4.1 Background

The model we are concerned with studying is a class of time-varying models called the Dynamic Linear Model. The setup for the equation follows:

$$\begin{aligned}
Y_t|\beta_t, \sigma^2 &\sim N(F_t^T \beta_t, \sigma^2 V_t) \\
\beta_t|\beta_{t-1}, \sigma^2 &\sim N(G_t \beta_{t-1}, \sigma^2 W_t) \\
\sigma^{-2} &\sim \Gamma(a_{t-1}, b_{t-1}) \\
\beta_{t-1}|\sigma^2 &\sim N(m_{t-1}, \sigma^2 C_{t-1})
\end{aligned}$$

Alternatively, using Normal-Inverse Gamma notation, where, if $\sigma^{-2} \sim \Gamma(a_{t-1}, b_{t-1})$, $\sigma^2 \sim IG(a_{t-1}, b_{t-1})$, where IG denotes an inverse Gamma distribution, we may write the above set of equations as the following:

$$\begin{aligned} Y_t, \sigma^2 | \beta_t &\sim NIG(F_t^T \beta_t, V_t, a_{t-1}, b_{t-1}) \\ \beta_t, \sigma^2 | \beta_{t-1} &\sim NIG(G_t \beta_{t-1}, W_t, a_{t-1}, b_{t-1}) \\ \beta_{t-1}, \sigma^2 &\sim NIG(m_{t-1}, C_{t-1}, a_{t-1}, b_{t-1}) \end{aligned}$$

The task is to acquire estimates for $\beta_{0,\ldots,T}$ and σ^2. This task may be divided into the forward filter and backwards sampling steps (collectively referred to as the Forward Filter-Backwards Sampling (FFBS) algorithm): The forward filter to acquire sequential estimates, and the backwards sampling step to retroactively "smooth" our initial estimates. We are given a set of observations $Y_{t,j}$, and known parameters F_t, G_t, V_t, W_t, and n_{t-1}, although Frankenburg and Banerjee also apply FFBS to cases where F_t and G_t are not pre-specified.

Note that while the setup is borrowed from Petris, the notation is retained from West and Harrison to be fully consistent across documentation where the variance is dampened.

4.1.1 Derivation

We proceed for some arbitrary t:

$$\begin{aligned} \beta_t &= G_t \beta_{t-1} + \omega_t, \omega_t \sim N(0, \sigma^2 W_t) \\ \beta_t | \sigma^2 &\sim N(G_t m_{t-1}, \sigma^2 (G_t C_{t-1} G_t^T + W_t)) \end{aligned}$$

Now, let $m_t^* = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t^T + W_t$. We then have:

$$\begin{aligned} Y_t &= F_t^T \beta_t + \nu_t, \nu_t \sim N(0, \sigma^2 V_t) \\ Y_t | \sigma^2 &\sim N(F_t^T m_t^*, \sigma^2 (F_t^T R_t F_t + V_t)) \end{aligned}$$

Since $\sigma^2 \sim IG(a_{t-1}, b_{t-1})$, we marginalize it out of $Y_t | \sigma^2$ to get

$$Y_t \sim T_{2a_{t-1}}(F_t^T m_t^*, \frac{b_{t-1}}{a_{t-1}}(F_t^T R_t F_t + V_t))$$

We now have the apparatus needed to compute the sequential posterior $\beta_t | Y_t$ and $\sigma^2 | Y_t$:

4.1.1.1 Deriving $\beta_t|Y_t$

$$\begin{aligned}
p(\beta_t|Y_t,\sigma^2) &\propto p(\beta_t,Y_t|\sigma^2)\\
&\propto p(Y_t|\beta_t,\sigma^2)p(\beta_t|\sigma^2)\\
&\propto \sigma^{-n}\exp(-\frac{1}{2\sigma^2}(y_t-F_t^T\beta_t)^TV_t^{-1}(y_t-F_t^T\beta_t))\sigma^{-p}\exp(-\frac{1}{2\sigma^2}(\beta_t-m_t^*)^TR_t^{-1}(\beta_t-m_t^*))\\
&\propto \sigma^{-(n+p)}\exp(-\frac{1}{2\sigma^2}[(y_t-F_t^T\beta_t)^TV_t^{-1}(y_t-F_t^T\beta_t)+(\beta_t-m_t^*)^TR_t^{-1}(\beta_t-m_t^*)])
\end{aligned}$$

Note next that

$$\begin{bmatrix}Y_t\\ \beta_t\end{bmatrix}|\sigma^2 \quad\sim\quad N\left(\begin{bmatrix}F_t^Tm_t^*\\ m_t^*\end{bmatrix},\sigma^2\begin{bmatrix}F_t^TR_tF_t+V_t & F_t^TR_t\\ R_tF_t & R_t\end{bmatrix}\right)$$

with the cross-terms $\mathrm{Cov}(Y_t,\beta_t) = \mathrm{Cov}(F_t^T\beta_t+\nu_t,\beta_t) = F_t^T\mathrm{Cov}(\beta_t,\beta_t) = F_t^TR_t$.

Since, for the following block-normal system

$$\begin{bmatrix}x_1\\ x_2\end{bmatrix} \quad\sim\quad N\left(\begin{bmatrix}\mu_1\\ \mu_2\end{bmatrix},\begin{bmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{21} & \Sigma_{22}\end{bmatrix}\right)$$

we have

$$x_2|x_1 \quad\sim\quad N(\mu_2+\Sigma_{21}\Sigma_{11}^{-1}(x_1-\mu_1),\Sigma_{22}-\Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$$

(The derivation of the density of $x_2|x_1$ can be found in the Appendix.)

We arrive at,

$$\begin{aligned}
\beta_t|\sigma^2,Y_t &\sim N(m_t^*+R_tF_t(F_t^TR_tF_t+V_t)^{-1}(Y_t-F_t^Tm_t^*),R_t-R_tF_t(F_t^TR_tF_t+V_t)^{-1}F_t^TR_t)\\
&\sim N(m_t^*+R_tF_tQ_t^{-1}(Y_t-F_t^Tm_t^*),R_t-R_tF_tQ_t^{-1}F_t^TR_t)
\end{aligned}$$

where $Q_t = F_t^TR_tF_t+V_t$.

(Note that Petris's expression for the variance suffers from a typo; to see this, simply take their $\widetilde{C}_t^T$.)

4.1.1.2 Deriving $\sigma^2|Y_t$

We next deduce the density of $\sigma^2|Y_t$. Note before we begin that since $Y_t \sim T_{2a_{t-1}}(F_t^T m_t^*, Q_t) = \int NIG_{Y_t}(F_t^T m_t^*, Q_t, a_{t-1}, b_{t-1}) d\sigma^2$, we can write $Y_t|\sigma^2 \sim N(F_t m_t^*, \sigma^2 Q_t)$. Hence:

$$\begin{aligned}
p(\sigma^2|Y_t) &\propto p(Y_t|\sigma^2)p(\sigma^2) \\
&\propto \sigma^{-n}\exp(-\frac{1}{2\sigma^2}(y_t - F_t^T m_t^*)^T Q_t^{-1}(y_t - F_t^T m_t^*))\sigma^{-2(a_{t-1}+1)}\exp(-b_{t-1}\sigma^{-2}) \\
&\propto \sigma^{-2(a_{t-1}+\frac{n}{2}+1)}\exp(-\sigma^{-2}[\frac{1}{2}(y_t - F_t^T m_t^*)^T Q_t^{-1}(y_t - F_t^T m_t^*) + b_{t-1}])
\end{aligned}$$

We conclude that $\sigma^{-2}|Y_t \sim \Gamma(a_t, b_t)$, where $a_t = a_{t-1} + \frac{n}{2}$ and $b_t = b_{t-1} + \frac{1}{2}(y_t - F_t^T m_t^*)^T Q_t^{-1}(y_t - F_t^T m_t^*)$.

This gives us the set of updating equations according to Petris Proposition 4.1.

4.1.1.3 Final Commentary on the Derivation

Note that we have derived the forward filtering step for the set of equations for time t given the parameters for the distributions at time $t-1$. Hence the equation's setup is Markovian, i.e. the state of this set of equations only depends on that of the preceding time point. Nevertheless, applications where the forward filter's equations propagate from an initial time point $t = 0$ are written so that the dependence of the parameters' values β_t and σ^2 on the data up to time $t-1$ or time t are made explicit. Specifically, letting $D_t = \{Y_\tau\}_{\tau=1,\ldots,t}$, we may write the set of equations in our setup as:

$$\begin{aligned}
Y_t, \sigma^2|\beta_t, D_{t-1} &\sim NIG(F_t^T\beta_t, V_t, a_{t-1}, b_{t-1}) \\
\beta_t, \sigma^2|\beta_{t-1}, D_{t-1} &\sim NIG(G_t\beta_{t-1}, W_t, a_{t-1}, b_{t-1}) \\
\beta_{t-1}, \sigma^2|D_{t-1} &\sim NIG(m_{t-1}, C_{t-1}, a_{t-1}, b_{t-1})
\end{aligned}$$

and the sequential posteriors we have derived, $\beta_t|Y_t$ and $\sigma^2|Y_t$, as $\beta_t|D_t$ and $\sigma^2|D_t$ respectively.

4.2 Appendix

4.2.1 Deriving $x_2|x_1$ when $(x_1 x_2)^T$ is a block-normal multivariate random variable.

Recall our block normal system:

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right)$$

Assuming that Σ_{11} is invertible (though unless x_1 contains degenerate terms, we have nothing to worry about), we then have

$$\begin{aligned} p(x_2|x_1) &= \frac{p(x_1, x_2)}{p(x_1)} \\ &\propto \exp\left(-\frac{1}{2} \left[\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}^T \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} \begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix} - (x_1 - \mu_1)^T \Sigma_{11}^{-1} (x_1 - \mu_1) \right] \right) \end{aligned}$$

Now, one of the expressions we may use to invert the block covariance matrix is:

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_{11}^{-1} + \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & -\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \\ -(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1} & (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} \end{bmatrix}$$

Hence,

$$\begin{aligned} p(x_2|x_1) &\propto \exp\left(-\frac{1}{2} \left[(x_1 - \mu_1)^T \Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1}\Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1) - 2(x_1 - \mu_1)\Sigma_{11}^{-1}\Sigma_{12}(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1} \right.\right. \\ &\propto \exp\left(-\frac{1}{2} \left[((x_2 - \mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1))^T (\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})^{-1} ((x_2 - \mu_2) - \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1)) \right] \right) \end{aligned}$$

i.e. $x_2|x_1 \sim N(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_1 - \mu_1), \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})$. $\square$