Appendix A

Kalman filter

Given the content in Section 3.3, we shall show how to derive the Kalman filtering step by step based on the general expression of a state space model below. The whole process could also be found in Durbin and Koopman, 2012.

$$y_t = Z_t\alpha_t + \epsilon_t \qquad \epsilon_t \sim NID(0, H_t) \tag{A.1}$$

$$\alpha_{t+1} = T_t\alpha_t + R_t\eta_t \qquad \eta_t \sim NID(0, Q_t) \tag{A.2}$$

Before giving the derivation procedure, we post a known conclusion from multivariate analysis:

Lemma A.0.1. Suppose X and Y are jointly normally distributed as following,

$$E[(x \quad y)^T] = (\mu_x \quad \mu_y)^T \qquad Var\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{xy}^T & \Sigma_{yy} \end{pmatrix} \tag{A.3}$$

then the conditional distribution of X given Y is also normal with mean

$$E[x|y] = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(y - \mu_y) \tag{A.4}$$

and variance matrix

$$Var[x|y] = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T \tag{A.5}$$

A.1 Filtering process

It's not hard to show the expectation of v_t given Y_{t-1} is 0, then with Lemma A.0.1 applying on α_t and v_t given Y_{t-1}, we could show

$$a_{t|t} = E(\alpha_t|Y_{t-1}) + Cov(\alpha_t, v_t)Var(v_t)^{-1}v_t$$

where

$$\begin{aligned} Cov(\alpha_t, v_t) &= E(\alpha_t(Z_t\alpha_t + \varepsilon_t - Z_t a_t)'|Y_{t-1}) \\ &= E(\alpha_t(\alpha_t - a_t)'Z_t'|Y_{t-1}) \\ &= P_t Z_t' \\ Var(v_t|Y_{t-1}) &= Var(Z_t\alpha_t + \varepsilon_t - Z_t a_t|Y_{t-1}) \\ &= Z_t P_t Z_t' + H_t \\ &= F_t \end{aligned}$$

thereby,

$$a_{t|t} = a_t + P_t Z_t' F_t^{-1} v_t$$

Similarly, by Lemma A.0.1 we derive another update equation

$$\begin{aligned} P_{t|t} &= Var(\alpha_t|Y_t) \\ &= Var(\alpha_t|Y_{t-1}, v_t) \\ &= Var(\alpha_t|Y_{t-1}) - Cov(\alpha_t, v_t) Var(v_t)^{-1} Cov(\alpha_t, v_t)' \\ &= P_t - P_t Z_t' F_t^{-1} Z_t P_t \end{aligned}$$

Now let's look at how to predict the state at time t+1:

$$\begin{aligned} a_{t+1} &= E(\alpha_{t+1}|Y_t) \\ &= E(T_t\alpha_t + R_t\eta_t|Y_t) \\ &= T_t E(\alpha_t|Y_t) \\ &= T_t a_{t|t} \\ P_{t+1} &= Var(T_t\alpha_t + R_t\eta_t|Y_t) \\ &= T_t Var(\alpha_t|Y_t) T_t' + R_t Q_t R_t' \\ &= T_t P_{t|t} T_t' + R_t Q_t R_t' \end{aligned}$$

With update equations we obtained above and the Kalman gain $K_t = T_t P_t Z_t' F_t^{-1}$, we could have the final version of our prediction equation:

$$\begin{aligned} a_{t+1} &= T_t a_t + K_t v_t \\ P_{t+1} &= T_t P_t (T_t - K_t Z_t)' + R_t Q_t R_t' \end{aligned}$$

Sometimes Z_t, T_t, H_t, R_t and Q_t are time-invariant, then we can show that the variance matrix P_t converges to a constant matrix $\bar{P}$, which is the solution to

$$\bar{P} = T\bar{P}T' - T\bar{P}Z'\bar{F}^{-1}Z\bar{P}T' + RQR' \tag{A.6}$$

where $\bar{F} = Z\bar{P}Z' + H$. Then it is not hard to derive the lemma below

Lemma A.1.1. For the same time series, if two state space models share the same Z_t, T_t, R_t and the same *ratios* of variance matrices, then their decomposition stay the same.

Proof. Take the local level model as an example: suppose we have such a local level model

$$\begin{aligned} y_t &= T_t + \varepsilon_t \qquad & \varepsilon_t &\sim N(0, \sigma_\varepsilon^2) \\ T_{t+1} &= T_t + \eta_t \qquad & \eta_t &\sim N(0, \sigma_\eta^2) \end{aligned} \tag{A.7}$$

Let $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\eta^2}$, corresponding to Equation A.6, we could have

$$\bar{P} = \frac{1 + \sqrt{1+4\lambda}}{2}\sigma_\eta^2 \qquad \bar{F} = \frac{1 + 2\lambda + \sqrt{1+4\lambda}}{2}\lambda_\eta^2 \tag{A.8}$$

At time t, when we update the state by equation $x_{t|t} = x_t + \bar{P}_t Z_t^T \bar{F}_t^{-1} v_t$, we could find no matter what the values of σ_η^2 and σ_ε^2 are, as long as their ratio doesn't change, then the result after updating stay the same. Similar things also happened in the prediction step. Under this lemma, we could reduce our parameters by one if we are looking for the decomposition results. □

A.2 Smoothing process

Appendix B

Other supplement

This is the comparison of the decomposition results from MLEs and random small variances at 1:

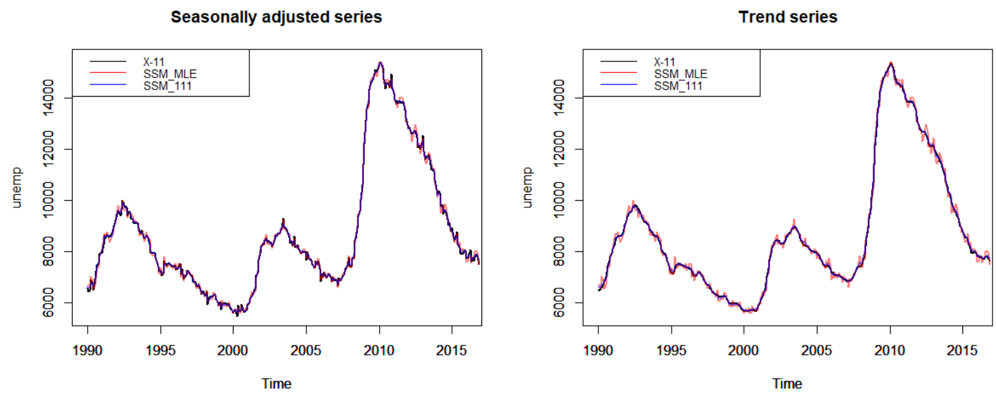

Figure B.1: Comparison of decomposition(Unemployment from 1990 to 2016 in U.S.)

The following is the hypothesis test results between the weakly-informative prior and MLEs:

Friedman rank sum test
Friedman chi-squared = 583.7, df = 1, p-value < 2.2e-16

Wilcoxon rank sum test with continuity correction
W = 570603, p-value = 4.565e-08
alternative hypothesis: true location shift is not equal to 0

The decomposition error comparison among MLEs, the posterior estimators from weakly-informative and empirical priors is:

	MLE	Loss	MAP(hnormal)	MAP(empirical)
Median	775.7	651.7	716.1	690.8
Mean	796.8	662.3	743.8	719.6
sd	213.4	150.9	185.5	171.4

Table B.1: Information of decomposition error(2)

This is the box plot of their decomposition errors:

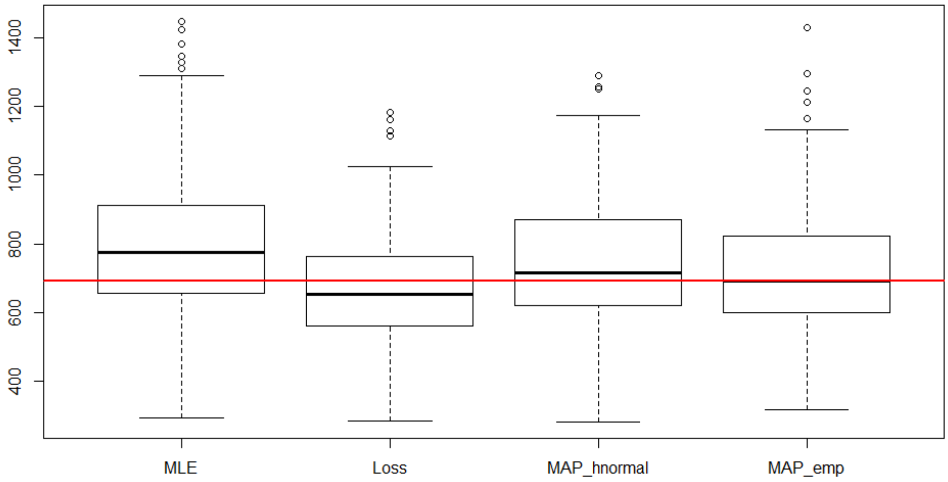

Figure B.2: Boxplot comparison of decomposition errors

Part of testing result w.r.t the posterior estimators from the empirical prior:

Friedman rank sum test
data: MLE,MAP(hnormal),MAP(empirical)
Friedman chi-squared = 186.91, df = 2, p-value < 2.2e-16

Friedman rank sum test
data: MAP(hnormal),MAP(empirical)
Friedman chi-squared = 34.68, df = 1, p-value = 3.886e-09

Friedman rank sum test
data: MLE,MAP(empirical)
Friedman chi-squared = 75, df = 1, p-value < 2.2e-16

Wilcoxon signed rank test with continuity correction
data: MLE,MAP(empirical)
V = 500500, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0

Wilcoxon signed rank test with continuity correction
data: MAP(hnormal),MAP(empirical)
V = 500500, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0

The following figure is the empirical distribution of variances from 8 different groups:

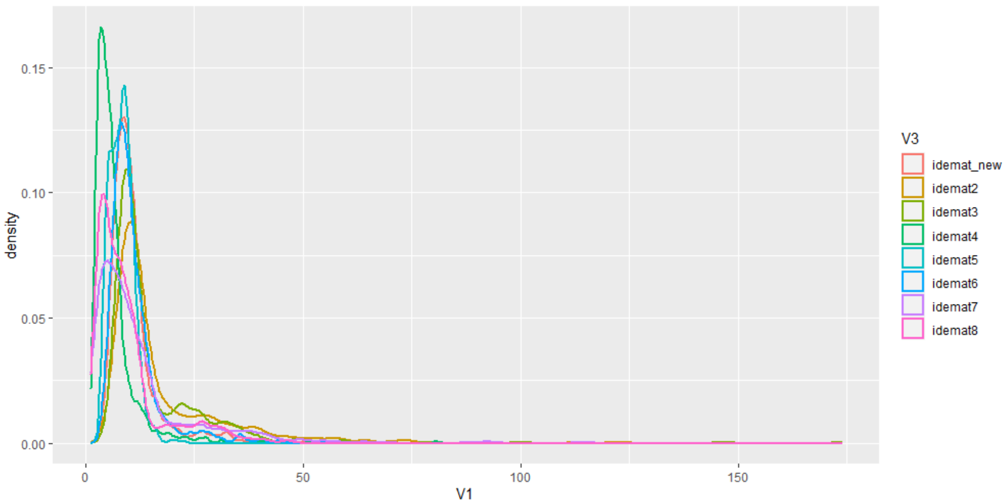

Figure B.3: Empirical distributions of the irregular variance from 8 groups

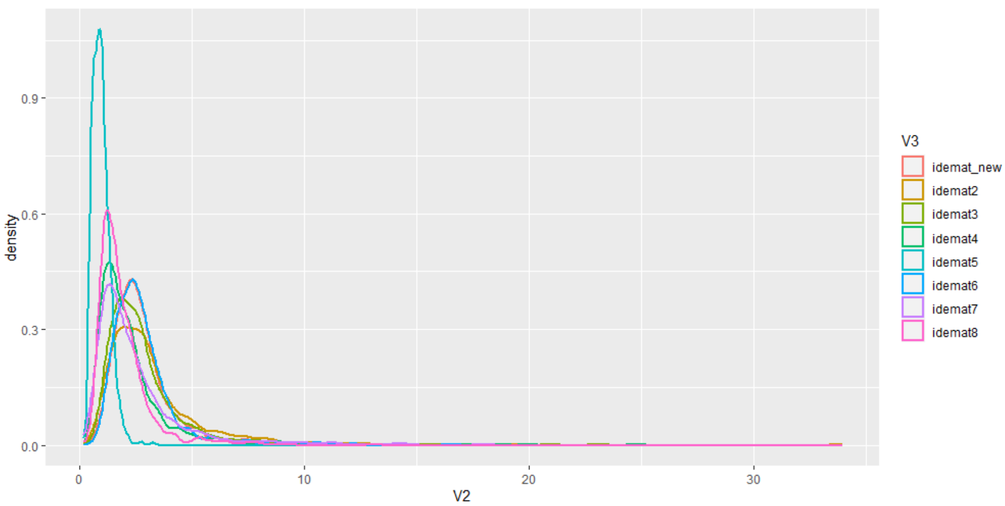

Figure B.4: Empirical distributions of the trend variance from 8 groups

Note: the information of the SSMs used to simulate is contained in Table B.2

Name	Length(yrs)	σ_I^2	σ_T^2	σ_S^2
simlist_new	15	20	10	1
simlist2	15	100	25	1
simlist3	20	100	25	1
simlist4	15	25	100	1
simlist5	15	1	0.25	1
simlist6	15	200	100	10
simlist7	15	$(N(0,10))^2$	$(N(0,10))^2$	1
simlist8	30	$(N(0,10))^2$	$(N(0,10))^2$	1

Table B.2: Information of SSMs used for simulation