The Kalman Filter and BVARs

Empirical Analysis II, Econ 311: Topic 4

Prof. Harald Uhlig[1]

[1]University of Chicago
Department of Economics
huhlig@uchicago.edu

Winter 2019

Outline

1. The Kalman Filter
   - Two useful lemmas
   - The state space system
   - The Kalman Smoother
2. Bayesian Vector Autoregressions (BVARs)
   - BVARs per Kalman Filtering
   - BVARs per Normal-Wishart distributions

Outline

1 The Kalman Filter
- Two useful lemmas
- The state space system
- The Kalman Smoother

2 Bayesian Vector Autoregressions (BVARs)
- BVARs per Kalman Filtering
- BVARs per Normal-Wishart distributions

The conditional normal distribution 1

Lemma

Let

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} S_{XX'} & S_{XY'} \\ S_{YX'} & S_{YY'} \end{bmatrix}\right).$$

Then

$$X \mid Y \sim \mathcal{N}(AY, S_{XX'|Y}),$$

where $A = S_{XY'} S_{YY'}^{-1}$*, and*

$$S_{XX'|Y} = S_{XX'} - S_{XY'} S_{YY'}^{-1} S_{YX'} = S_{XX'} - A S_{YY'} A'$$

The conditional normal distribution 2

Lemma

Let $Y \mid H, \xi \sim \mathcal{N}(H\xi, \Sigma)$ *and* $\xi \mid H \sim \mathcal{N}(\hat{\xi}, \Omega)$*. Then*

$$\begin{bmatrix} \xi \\ Y \end{bmatrix} \mid H \sim \mathcal{N}\left(\begin{bmatrix} \hat{\xi} \\ H\hat{\xi} \end{bmatrix}, \begin{bmatrix} S_{\xi\xi'} & S_{\xi Y'} \\ S_{Y\xi'} & S_{YY'} \end{bmatrix} \right).$$

and

$$\begin{aligned} \xi \mid Y, H &\sim \mathcal{N}\left(\hat{\xi} + S_{\xi Y'} S_{YY'}^{-1} (Y - H\hat{\xi}), \Omega - S_{\xi Y'} S_{YY'}^{-1} S_{Y\xi'} \right) \\ &\sim \mathcal{N}\left(\hat{\xi} + G\hat{\epsilon}, \Omega - G S_{YY'} G' \right) \end{aligned}$$

where $S_{\xi\xi'} = \Omega$, $S_{\xi Y'} = \Omega H' = S'_{Y\xi'}$, $S_{YY'} = H\Omega H' + \Sigma$ *and* $G = S_{\xi Y'} S_{YY'}^{-1}$, $\hat{\epsilon} = Y - H\hat{\xi}$.

Kalman Gain Matrix

Outline

1. The Kalman Filter
   - Two useful lemmas
   - The state space system
   - The Kalman Smoother
2. Bayesian Vector Autoregressions (BVARs)
   - BVARs per Kalman Filtering
   - BVARs per Normal-Wishart distributions

The ingredients

- data: $Y_t \in \mathbb{R}^n$, $t = 1, \ldots, T$, observable,
- Unobservable: $\xi_t \in \mathbb{R}^r$.
- Parameters: $H_t, F_t, \Sigma_t, \Phi_t$.
- Parameters may be constant. In some applications: parameters observable or known. In others: to be estimated.
- For now: treat them as known.
- Hamilton, Chapter 13.

The state space system

- observation equation:

$$Y_t = H_t \xi_t + \epsilon_t, \text{ where } \epsilon_t \sim \mathcal{N}(0, \Sigma_t)$$

- state equation:

$$\xi_{t+1} = F_{t+1} \xi_t + \eta_{t+1}, \text{ where } \eta_{t+1} \sim \mathcal{N}(0, \Phi_{t+1})$$

- ϵ_t, η_t: independent.
- The Kalman Filter can be and is used without the normal distribution assumption. In that case, all the formulas amount to linear projections or linear least squares.

Recursive updating

- Begin with a date-$(t-1)$ forecast for ξ_t,

$$\xi_t \sim \mathcal{N}\left(\hat{\xi}_{t|t-1}, \Omega_{t|t-1}\right)$$

- to be found: the Kalman predictor $\hat{\xi}_{t+1|t}$ as well as $\Omega_{t+1|t}$, the Kalman prediction error covariance matrix for ξ_{t+1}, given data up to and including t.
- Three steps:
  1. Forecast Y_t.
  2. Observe Y_t and update inference for ξ_t
  3. Forecast ξ_{t+1} at date t.

Step 1: Forecast Y_t

- Given:

$$\begin{aligned} \xi_t &\sim \mathcal{N}\left(\hat{\xi}_{t|t-1}, \Omega_{t|t-1}\right) \\ Y_t &= H_t \xi_t + \epsilon_t, \text{ where } \epsilon_t \sim \mathcal{N}\left(0, \Sigma_t\right) \end{aligned}$$

- Forecast:

$$Y_t \sim \mathcal{N}(\hat{Y}_t, S_{YY'|t})$$

  where

$$\begin{aligned} \hat{Y}_t &= H_t \hat{\xi}_{t|t-1} \\ S_{YY'|t} &= H_t \Omega_{t|t-1} H_t' + \Sigma_t \end{aligned}$$

Step 2: Update

- One-step ahead forecast error:

$$\hat{\epsilon}_t = Y_t - \hat{Y}_t$$

- Kalman filter equation : $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t}, \Omega_{t|t})$, where

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + G_t \hat{\epsilon}_t$$

- Kalman gain equation:

$$G_t = S_{\xi Y'|t} S_{YY'|t}^{-1}$$

where

$$\begin{aligned} \Omega_{t|t} &= \Omega_{t|t-1} - S_{\xi Y'|t} S_{YY'|t}^{-1} S_{Y\xi'|t} \\ S_{\xi Y'|t} &= \Omega_{t|t-1} H_t' = S_{Y\xi'|t}' \end{aligned}$$

Step 3: Forecast ξ_{t+1}

- $\xi_{t+1} \sim \mathcal{N}(\hat{\xi}_{t+1|t}, \Omega_{t+1|t})$
- Kalman predictor:

$$\hat{\xi}_{t+1|t} = F_{t+1}\hat{\xi}_{t|t}$$

- Kalman prediction error covariance matrix:

$$\Omega_{t+1|t} = F_{t+1}\Omega_{t|t}F'_{t+1} + \Phi_{t+1}$$

Kalman Filter: Summary

The Kalman Filter:

$$Y_t \sim \mathcal{N}\left(H_t \xi_t, \Sigma_t\right), \quad \xi_{t+1} \sim \mathcal{N}\left(F_{t+1} \xi_t, \Phi_{t+1}\right)$$

Given $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t-1}, \Omega_{t|t-1})$,

1. Forecast $Y_t \sim \mathcal{N}(\hat{Y}_t, S_{YY'|t})$, where

$$\hat{Y}_t = H_t \hat{\xi}_{t|t-1}, \; S_{YY'|t} = H_t \Omega_{t|t-1} H_t' + \Sigma_t$$

2. Update $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t}, \Omega_{t|t})$, where

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + G_t \hat{\epsilon}_t, \;\; \hat{\epsilon}_t = Y_t - \hat{Y}_t, \;\; G_t = S_{\xi Y'|t} S_{YY'|t}^{-1}$$

$$\Omega_{t|t} = \Omega_{t|t-1} - S_{\xi Y'|t} S_{YY'|t}^{-1} S_{Y\xi'|t}, \;\; S_{\xi Y'|t} = \Omega_{t|t-1} H_t' = S_{Y\xi'|t}'$$

3. Forecast $\xi_{t+1} \sim \mathcal{N}(\hat{\xi}_{t+1|t}, \Omega_{t+1|t})$, where

$$\hat{\xi}_{t+1|t} = F_{t+1} \hat{\xi}_{t|t}, \;\; \Omega_{t+1|t} = F_{t+1} \Omega_{t|t} F_{t+1}' + \Phi_{t+1}$$

Kalman Filter: Summary

The Kalman Filter:

$$Y_t \sim \mathcal{N}\left(H_t \xi_t, \Sigma_t\right), \quad \xi_{t+1} \sim \mathcal{N}\left(F_{t+1} \xi_t, \Phi_{t+1}\right)$$

Given $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t-1}, \Omega_{t|t-1})$,

1. Forecast $Y_t \sim \mathcal{N}(\hat{Y}_t, S_{YY'|t})$, where

$$\hat{Y}_t = H_t \hat{\xi}_{t|t-1}, \; S_{YY'|t} = H_t \Omega_{t|t-1} H_t' + \Sigma_t$$

2. Update $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t}, \Omega_{t|t})$, where

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + G_t \hat{\epsilon}_t, \quad \hat{\epsilon}_t = Y_t - \hat{Y}_t, \quad G_t = S_{\xi Y'|t} S_{YY'|t}^{-1}$$

$$\Omega_{t|t} = \Omega_{t|t-1} - S_{\xi Y'|t} S_{YY'|t}^{-1} S_{Y\xi'|t}, \quad S_{\xi Y'|t} = \Omega_{t|t-1} H_t' = S_{Y\xi'|t}'$$

3. Forecast $\xi_{t+1} \sim \mathcal{N}(\hat{\xi}_{t+1|t}, \Omega_{t+1|t})$, where

$$\hat{\xi}_{t+1|t} = F_{t+1} \hat{\xi}_{t|t}, \quad \Omega_{t+1|t} = F_{t+1} \Omega_{t|t} F_{t+1}' + \Phi_{t+1}$$

Kalman Filter: Summary

The Kalman Filter:

$$Y_t \sim \mathcal{N}\left(H_t \xi_t, \Sigma_t\right), \quad \xi_{t+1} \sim \mathcal{N}\left(F_{t+1} \xi_t, \Phi_{t+1}\right)$$

Given $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t-1}, \Omega_{t|t-1})$,

1. Forecast $Y_t \sim \mathcal{N}(\hat{Y}_t, S_{YY'|t})$, where

$$\hat{Y}_t = H_t \hat{\xi}_{t|t-1}, \; S_{YY'|t} = H_t \Omega_{t|t-1} H_t' + \Sigma_t$$

2. Update $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t}, \Omega_{t|t})$, where

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + G_t \hat{\epsilon}_t, \quad \hat{\epsilon}_t = Y_t - \hat{Y}_t, \quad G_t = S_{\xi Y'|t} S_{YY'|t}^{-1}$$

$$\Omega_{t|t} = \Omega_{t|t-1} - S_{\xi Y'|t} S_{YY'|t}^{-1} S_{Y\xi'|t}, \quad S_{\xi Y'|t} = \Omega_{t|t-1} H_t' = S_{Y\xi'|t}'$$

3. Forecast $\xi_{t+1} \sim \mathcal{N}(\hat{\xi}_{t+1|t}, \Omega_{t+1|t})$, where

$$\hat{\xi}_{t+1|t} = F_{t+1} \hat{\xi}_{t|t}, \quad \Omega_{t+1|t} = F_{t+1} \Omega_{t|t} F_{t+1}' + \Phi_{t+1}$$

Kalman Filter: Summary

The Kalman Filter:

$$Y_t \sim \mathcal{N}\left(H_t \xi_t, \Sigma_t\right), \quad \xi_{t+1} \sim \mathcal{N}\left(F_{t+1}\xi_t, \Phi_{t+1}\right)$$

Given $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t-1}, \Omega_{t|t-1})$,

1. Forecast $Y_t \sim \mathcal{N}(\hat{Y}_t, S_{YY'|t})$, where

$$\hat{Y}_t = H_t \hat{\xi}_{t|t-1}, \; S_{YY'|t} = H_t \Omega_{t|t-1} H_t' + \Sigma_t$$

2. Update $\xi_t \sim \mathcal{N}(\hat{\xi}_{t|t}, \Omega_{t|t})$, where

$$\hat{\xi}_{t|t} = \hat{\xi}_{t|t-1} + G_t \hat{\epsilon}_t, \quad \hat{\epsilon}_t = Y_t - \hat{Y}_t, \quad G_t = S_{\xi Y'|t} S_{YY'|t}^{-1}$$

$$\Omega_{t|t} = \Omega_{t|t-1} - S_{\xi Y'|t} S_{YY'|t}^{-1} S_{Y\xi'|t}, \quad S_{\xi Y'|t} = \Omega_{t|t-1} H_t' = S_{Y\xi'|t}'$$

3. Forecast $\xi_{t+1} \sim \mathcal{N}(\hat{\xi}_{t+1|t}, \Omega_{t+1|t})$, where

$$\hat{\xi}_{t+1|t} = F_{t+1} \hat{\xi}_{t|t}, \quad \Omega_{t+1|t} = F_{t+1} \Omega_{t|t} F_{t+1}' + \Phi_{t+1}$$

Initializing the Kalman Filter

- What about $\hat{\xi}_{1|0}$, $\Omega_{1|0}$?
- Possibility 1: known starting point. E.g. $\hat{\xi}_{1|0} = 0$, $\Omega_{1|0} = 0_{r \times r}$.
- Possibility 2: nearly flat. $\hat{\xi}_{1|0} = 0$, $\Omega_{1|0} = \omega I_r$ with $\omega \in \mathbb{R}$ very large.
- Possibility 3: stationary distribution. Suppose $F_t \equiv F, \Phi_t \equiv \Phi$. $\hat{\xi}_{1|0} = 0$, $\Omega_{1|0} = \Omega = E[\xi_t \xi_t']$, where we calculate

$$
\begin{aligned}
\xi_{t+1} &= F\xi_t + \eta_{t+1},\ \eta_{t+1} \sim \mathcal{N}(0, \Phi) \\
\Omega &= F\Omega F' + \Phi \\
vec(I_r \Omega I_r - F\Omega F') &= vec(\Phi) \\
(I_{r^2} - F \otimes F)\, vec(\Omega) &= vec(\Phi)
\end{aligned}
$$

so, if all eigenvalues of F are smaller than one in absolute value,

$$vec(\Omega) = (I_{r^2} - F \otimes F)^{-1} vec(\Phi)$$

Initializing the Kalman Filter

- What about $\hat{\xi}_{1|0}$, $\Omega_{1|0}$?
- Possibility 1: known starting point. E.g. $\hat{\xi}_{1|0} = 0$, $\Omega_{1|0} = 0_{r \times r}$.
- Possibility 2: nearly flat. $\hat{\xi}_{1|0} = 0$, $\Omega_{1|0} = \omega I_r$ with $\omega \in \mathbb{R}$ very large.
- Possibility 3: stationary distribution. Suppose $F_t \equiv F, \Phi_t \equiv \Phi$. $\hat{\xi}_{1|0} = 0$, $\Omega_{1|0} = \Omega = E[\xi_t \xi_t']$, where we calculate

$$
\begin{aligned}
\xi_{t+1} &= F\xi_t + \eta_{t+1}, \; \eta_{t+1} \sim \mathcal{N}(0, \Phi) \\
\Omega &= F\Omega F' + \Phi \\
vec(I_r \Omega I_r - F\Omega F') &= vec(\Phi) \\
(I_{r^2} - F \otimes F)\, vec(\Omega) &= vec(\Phi)
\end{aligned}
$$

so, if all eigenvalues of F are smaller than one in absolute value,

$$
vec(\Omega) = (I_{r^2} - F \otimes F)^{-1} vec(\Phi)
$$

Outline

1. The Kalman Filter
   - Two useful lemmas
   - The state space system
   - The Kalman Smoother
2. Bayesian Vector Autoregressions (BVARs)
   - BVARs per Kalman Filtering
   - BVARs per Normal-Wishart distributions

Full sample information

- Note: $\hat{\xi}_{t|t}$ is the "best estimate" of ξ_t, given all information up to and including t ...
- ... but what can we learn about ξ_t from the entire sample?
- Want: $\hat{\xi}_{t|T}$, $\Omega_{t|T}$.
- The Kalman smoother
- "Run the Kalman filter backwards". All sample information is contained in $\hat{\xi}_{t|t-1}$, $\hat{\xi}_{t|t}$, $\Omega_{t|t-1}$, $\Omega_{t|t}$. No need to "consult" the data again.

Full sample information

- Note: $\hat{\xi}_{t|t}$ is the “best estimate” of ξ_t, given all information up to and including t ...
- ... but what can we learn about ξ_t from the entire sample?
- Want: $\hat{\xi}_{t|T}$, $\Omega_{t|T}$.
- The Kalman smoother
- “Run the Kalman filter backwards”. All sample information is contained in $\hat{\xi}_{t|t-1}$, $\hat{\xi}_{t|t}$, $\Omega_{t|t-1}$, $\Omega_{t|t}$. No need to “consult” the data again.

Derivation, part 1

- Date T: $\hat{\xi}_{T|T}$, $\Omega_{T|T}$: done.
- At date t, from forward filtering and $\xi_{t+1} = F_{t+1}\xi_t + \eta_{t+1}$:

$$\left[\begin{array}{c} \xi_t \\ \xi_{t+1} \end{array} \right] \mid t \sim \mathcal{N}\left(\left[\begin{array}{c} \hat{\xi}_{t|t} \\ \hat{\xi}_{t+1|t} \end{array} \right], \left[\begin{array}{cc} \Omega_{t|t} & \Omega_{t|t}F'_{t+1} \\ F_{t+1}\Omega_{t|t} & \Omega_{t+1|t} \end{array} \right] \right).$$

- Suppose, one were to observe ξ_{t+1}. Second Lemma gives

$$\xi_t \mid \xi_{t+1}, t \sim \mathcal{N}\left(\hat{\xi}_{t|t} + J_t \left(\xi_{t+1} - \hat{\xi}_{t+1|t} \right), \Omega_{t|t} - J_t \Omega_{t+1|t} J_t{}' \right)$$

where

$$J_t = \Omega_{t|t} F'_{t+1} \Omega^{-1}_{t+1|t}$$

Derivation, part 2

- Split
$$\xi_{t+1} - \hat{\xi}_{t+1|t} = \hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t} + \nu_{t+1}$$
for some "residual" ν_{t+1}.
- Update $\hat{\xi}_{t|t}$ with $\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}$, not $\xi_{t+1} - \hat{\xi}_{t+1|t}$.
- $\text{Var}(\nu_{t+1}) = \Omega_{t+1|T}$.
- $\text{Var}\left(\xi_{t+1} - \hat{\xi}_{t+1|t}\right) = \Omega_{t+1|t}$.
- Thus,
$$\text{Var}\left(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}\right) = \Omega_{t+1|t} - \Omega_{t+1|T}$$
- Use this to calculate the reduction in variance due to smoothing.

The Kalman Smoother

The Kalman Smoother:

- Date T: $\hat{\xi}_{T|T}$, $\Omega_{T|T}$: done.
- $$\begin{aligned}\hat{\xi}_{t|T} &= \hat{\xi}_{t|t} + J_t\left(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t}\right)\\ \Omega_{t|T} &= \Omega_{t|t} - J_t\left(\Omega_{t+1|t} - \Omega_{t+1|T}\right)J_t'\\ &\quad\text{where}\\ J_t &= \Omega_{t|t}F_{t+1}'\Omega_{t+1|t}^{-1}\end{aligned}$$

Outline

1. The Kalman Filter
   - Two useful lemmas
   - The state space system
   - The Kalman Smoother

2. Bayesian Vector Autoregressions (BVARs)
   - BVARs per Kalman Filtering
   - BVARs per Normal-Wishart distributions

VAR (k)

- VAR(k) in $Y_t \in \mathbb{R}^m$:

$$Y_t = \mu + \sum_{j=1}^{k} B_j Y_{t-j} + A\epsilon_t,\ \epsilon_t \sim \mathcal{N}(0, I), t = 1, \ldots, T$$

- We assume that data is available for t = -k+1,..,0,..,T.
- Let

$$u_t = A\,\epsilon_t,\ \Sigma = AA'$$

- Bayesian Vector Autoregression (BVAR):
  - reduced-form BVAR: from a prior for $\theta = (\mu, (B_j)_{j=1}^k, \Sigma)$ and the data $Y_t, t = -k+1, \ldots, T$, find the posterior.
  - structural BVAR: from a prior for $\theta = (\mu, (B_j)_{j=1}^k, A)$, data $Y_t, t = -k+1, \ldots, T$ and identifying assumptions about A, find the posterior.

VAR(1)

This can be rewritten as a VAR(1): (stacking)

$$X_t = \begin{bmatrix} Y_t \\ Y_{t-1} \\ \vdots \\ Y_{t-k+1} \\ 1 \end{bmatrix} = \mathcal{B} X_{t-1} + \mathcal{A}\epsilon_t,$$

where

$$\mathcal{B} = \begin{bmatrix} B_1 & \dots & B_{k-1} & B_k & \mu \\ I_m & \dots & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots & \vdots \\ 0 & \dots & I_m & 0 & 0 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix} \text{ and } \mathcal{A} = \begin{bmatrix} A \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

Kalman Filter setup, part 1

- Define

$$H_t = \begin{bmatrix} X'_{t-1} & 0 & \dots & 0 \\ 0 & X'_{t-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & X'_{t-1} \end{bmatrix}, \; \xi_t = \begin{bmatrix} (\mathcal{B}_{1\cdot})' \\ (\mathcal{B}_{2\cdot})' \\ \vdots \\ (\mathcal{B}_{m\cdot})' \end{bmatrix}$$

- State space representation:

$$\begin{aligned} Y_t &= H_t \xi_t + u_t, \, u_t \sim \mathcal{N}(0, \Sigma) \\ \xi_{t+1} &= \xi_t \end{aligned}$$

- Easy to generalize: time-varying coefficients, ...

Kalman Filter setup, part 2

- Matrices:

$$F_{t+1} = I_{km^2+m}, \Phi_{t+1} = 0, \Sigma$$

- We treat Σ as given, for now. One possibility: get Σ from the residuals of univariate AR(k) regressions. Better: treat inference about Σ in a second step, condition on Σ for now.
- Information in initial observations $Y_{t-k+1}, \ldots, Y_0$. Condition on them for now.
- Kalman Filter initialization: $\hat{\xi}_{1|0}, \Omega_{1|0}$. Equivalently: a Normal-distribution prior

$$\pi(\xi \mid 0) \sim \mathcal{N}\left(\hat{\xi}_{1|0}, \Omega_{1|0}\right)$$

  for ξ, conditional on Σ and conditional of $Y_{t-k+1}, \ldots, Y_0$.
- Each Kalman-Filter step amounts to updating the posterior,

$$\pi(\xi \mid t-1) \rightarrow \pi(\xi \mid t) \sim \mathcal{N}\left(\hat{\xi}_{t+1|t}, \Omega_{t+1|t}\right)$$

Minnesota prior

- Random walk prior mean, $(\xi_{1|0})_i = 1$, whenever this is the coefficient of a variable on its own first lag, else zero.
- $\Omega_{1|0}$ diagonal. For coefficient in equation i on variable j at lag l: $\left(\Omega_{1|0}\right)_{(i,j,l),(i,j,l)} = S(i,j,l)^2$ where

$$S(i,j,l) = \gamma g(l) f(i,j) \frac{s_i}{s_j},$$

where

$$g(l) = \frac{1.0}{l^\delta}$$

(harmonic decay, $\delta = 1$ or $\delta = 2$ is a common choice) and where

$$f(i,j) = \begin{cases} 1.0 & \text{if } i = j \\ \omega & \text{if } i \neq j \end{cases}$$

for some $\omega \geq 0$.

Outline

1. The Kalman Filter
   - Two useful lemmas
   - The state space system
   - The Kalman Smoother

2. Bayesian Vector Autoregressions (BVARs)
   - BVARs per Kalman Filtering
   - BVARs per Normal-Wishart distributions

A matrix algebra result

Lemma

Consider four matrices

$$
\begin{array}{llll}
A & = \left[A_{ij}\right], & a \times b \, , \\
B & = \left[B_{jk}\right], & b \times c \, , \\
C & = [C_{km}], & c \times d \, , \\
D & = [D_{mn}], & d \times a \, ,
\end{array}
$$

Then, with vec$(\cdot)$ denoting columnwise vectorization,

$$tr(ABCD) = \left(vec(B')\right)' \left(A' \otimes C\right) vec(D)$$

Another useful property: if A is $a \times b$, and B is $b \times a$, then

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

Proof

Proof.

Note $\text{vec}(B')_{(j-1)c+k} = B_{jk}$, $\text{vec}(D)_{(i-1)d+m} = D_{mi}$, $(A' \otimes C)_{(j-1)c+k,(i-1)d+m} = A_{ij}C_{km}$. Thus,

$$\text{tr}(ABCD) = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{d} A_{ij}B_{jk}C_{km}D_{mi}$$

$$= \sum_{i,j,k,m} \text{vec}(B')_{(j-1)c+k}(A' \otimes C)_{(j-1)c+k,(i-1)d+m}\text{vec}(D)_{(i-1)d+m}$$

$$= \left(\text{vec}(B')\right)' \left(A' \otimes C\right) \text{vec}(D)$$

$\square$

The Wishart distribution

- Generalization of Gamma-distribution or χ^2-distribution to multivariate context.
- Let $x_i \sim \mathcal{N}(0, \Omega), i = 1, \ldots, \nu$ i.i.d, $x_i \in \mathbb{R}^m$. Let $X = [x_1, \ldots, x_\nu]$. Then,
$$W = XX' \sim \mathcal{W}_m(\Omega, \nu)$$
- Note: $E[XX'] = \nu\Omega$.
- ν: “degrees of freedom”.
- Density:
$$f(W) = \text{const}(|\,\Omega\,|, \nu, m) \mid W \mid^{(\nu-m-1)/2} \exp\left(-\frac{1}{2}\text{tr}\left(W\Omega^{-1}\right)\right)$$
- Bayesian inference: easier wrt precisions. Replace $S^{-1}/\nu = \Omega, \Sigma^{-1} = W$.

Normal-Wishart distributions

- $\theta = (\mathbf{B}, \Sigma^{-1})$, where $\mathbf{B}$ is $n \times m$ and where Σ is $m \times m$ and positive semidefinite.
- $NW(\bar{\mathbf{B}}, S, \mathbf{N}, \nu)$, parameterized by
  - a "mean coefficient" matrix $\bar{\mathbf{B}}$ of size $n \times m$
  - a positive definite "mean covariance" matrix S of size $m \times m$
  - a positive definite matrix $\mathbf{N}$ of size $n \times n$
  - "degrees of freedom" $\nu \geq 0$.
- The Normal-Wishart distribution specifies,
  - that Σ^{-1} is Wishart, $\Sigma^{-1} \sim \mathcal{W}_m(S^{-1}/\nu, \nu)$. Thus $E[\Sigma^{-1}] = S^{-1}$.
  - Conditionally on Σ, the matrix $\mathbf{B}$ in its columnwise vectorized form, $\text{vec}(\mathbf{B})$ follows a Normal distribution $\mathcal{N}(\text{vec}(\bar{\mathbf{B}}), \Sigma \otimes \mathbf{N}^{-1})$.

Density of a Normal-Wishart

$$
\begin{aligned}
f(\mathbf{B}, \Sigma^{-1}) = \kappa(\mathbf{N}, S, \nu, m) \\
\mid \Sigma^{-1} \mid^{n/2} \exp\left(-\frac{1}{2}(\beta - \bar{\beta})' \left[\Sigma^{-1} \otimes \mathbf{N}\right] (\beta - \bar{\beta})\right) \\
\mid \Sigma^{-1} \mid^{(\nu - m - 1)/2} \exp\left(-\frac{1}{2}\nu \mathrm{tr}(\Sigma^{-1} S)\right)
\end{aligned}
$$

for some integrating constant $\kappa(\mathbf{N}, S, \nu, m)$, where $\beta = \mathrm{vec}(\mathbf{B})$, $\bar{\beta} = \mathrm{vec}(\bar{\mathbf{B}})$

VAR (k)

- VAR(k) in $Y_t \in \mathbb{R}^m$:

$$Y_t = \mu_0 + \mu_1 t + \sum_{j=1}^{k} B_j Y_{t-j} + u_t,\ u_t \sim \mathcal{N}\left(0, \Sigma\right), t = 1, \ldots, T \quad (1)$$

- Stack the system (1) as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{u} \quad (2)$$

where $X_t = [Y_t', Y_{t-1}', \ldots, Y_{t-k+1}', 1, t]'$, $\mathbf{Y} = [Y_1, \ldots, Y_T]'$, $\mathbf{X} = [X_0, \ldots, X_{T-1}]'$, $\mathbf{u} = [u_1, \ldots, u_T]'$ and $\mathbf{B} = [B_1, \ldots, B_k, \mu_0, \mu_1]'$.

Examining the notation

$$Y_t = \mu_0 + \mu_1 t + \sum_{j=1}^{k} B_j Y_{t-j} + u_t,\; u_t \sim \mathcal{N}(0, \Sigma), t = 1, \ldots, T$$

stacked as

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{u}$$

is

$$\underbrace{\begin{bmatrix} Y_1' \\ Y_2' \\ \vdots \\ Y_T' \end{bmatrix}}_{T \times m} = \underbrace{\begin{bmatrix} Y_0' & Y_{-1}' & \ldots & Y_{1-k}' & 1 & 1 \\ Y_1' & Y_0' & \ldots & Y_{2-k}' & 1 & 2 \\ & \vdots & & & & \\ Y_{T-1}' & Y_{T-2}' & \ldots & Y_{T-k}' & 1 & T \end{bmatrix}}_{T \times (mk+2)} \underbrace{\begin{bmatrix} B_1' \\ B_2' \\ \vdots \\ B_k' \\ \mu_0' \\ \mu_1' \end{bmatrix}}_{(mk+2) \times m} + \underbrace{\begin{bmatrix} u_1' \\ u_2' \\ \vdots \\ u_T' \end{bmatrix}}_{T \times m}$$

Summary statistics

- Assume $u_t \sim \mathcal{N}(0, \Sigma)$ iid.
- MLE for $(\mathbf{B}, \Sigma)$:

$$\begin{aligned} \hat{\mathbf{B}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \\ \hat{\Sigma} &= \frac{1}{T}(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) \end{aligned} \tag{3}$$

Constructing the likelihood function

Conditionally on the initial observations,

$$
\begin{aligned}
L &= (2\pi)^{-mT/2} \mid \Sigma \mid^{-T/2} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} u_t' \Sigma^{-1} u_t \right) \\
&= (2\pi)^{-mT/2} \mid \Sigma \mid^{-T/2} \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \operatorname{tr}\left(u_t' \Sigma^{-1} u_t \right) \right) \\
&= (2\pi)^{-mT/2} \mid \Sigma \mid^{-T/2} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\Sigma^{-1} \sum_{t=1}^{T} u_t u_t' \right) \right) \\
&= (2\pi)^{-mT/2} \mid \Sigma \mid^{-T/2} \exp\left(-\frac{1}{2} \operatorname{tr}\left(\Sigma^{-1} \mathbf{u}' \mathbf{u} \right) \right)
\end{aligned}
$$

Examining the pieces, part 1

$$\begin{aligned}\mathbf{u}'\mathbf{u} = (\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}) \\ &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}} - \mathbf{X}(\mathbf{B} - \hat{\mathbf{B}}))'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}} - \mathbf{X}(\mathbf{B} - \hat{\mathbf{B}})) \\ &= (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + (\mathbf{B} - \hat{\mathbf{B}})'\mathbf{X}'\mathbf{X}(\mathbf{B} - \hat{\mathbf{B}}) \\ &= T\hat{\Sigma} + (\mathbf{B} - \hat{\mathbf{B}})'\mathbf{X}'\mathbf{X}(\mathbf{B} - \hat{\mathbf{B}})\end{aligned}$$

since

$$(\mathbf{B} - \hat{\mathbf{B}})'\mathbf{X}'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) = (\mathbf{B} - \hat{\mathbf{B}})\left(\mathbf{X}'\mathbf{Y} - \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) = 0$$

Examining the pieces, part 2

$$
\begin{aligned}
& \operatorname{tr}\left(\Sigma^{-1}\mathbf{u}'\mathbf{u}\right) \\
= \quad & T \operatorname{tr}\left(\Sigma^{-1}\hat{\Sigma}\right) \\
& + \operatorname{tr}\left(\Sigma^{-1}(\mathbf{B}-\hat{\mathbf{B}})'\mathbf{X}'\mathbf{X}(\mathbf{B}-\hat{\mathbf{B}})\right) \\
= \quad & T \operatorname{tr}\left(\Sigma^{-1}\hat{\Sigma}\right) \\
& + \left(\beta-\hat{\beta}\right)'\left[\Sigma^{-1}\otimes\mathbf{X}'\mathbf{X}\right]\left(\beta-\hat{\beta}\right)
\end{aligned}
$$

where $\beta = \text{vec}(\mathbf{B})$, $\hat{\beta} = \text{vec}(\hat{\mathbf{B}})$, with the help of the Lemma above.

The likelihood function

Proposition

Given the data Y_t, $t = -k+1, \ldots, T$, the conditional likelihood function as a function in $\mathbf{B}$ *and* Σ^{-1} *is proportional to a Normal-Wishart density,*

$$
\begin{aligned}
& L(\mathbf{B}, \Sigma^{-1} \mid \mathbf{Y}) \\
= \quad & (2\pi)^{-mT/2} \mid \Sigma \mid^{-T/2} \exp\left(-\frac{1}{2}\left(\beta - \hat{\beta}\right)' \left[\Sigma^{-1} \otimes \mathbf{X}'\mathbf{X}\right]\left(\beta - \hat{\beta}\right)\right) \\
& \qquad \exp\left(-\frac{T}{2} tr\left(\Sigma^{-1}\hat{\Sigma}\right)\right) \\
\propto \quad & NW(\hat{\mathbf{B}}, \mathbf{X}'\mathbf{X}, (T/\nu)\hat{\Sigma}, \nu)
\end{aligned}
$$

where $\nu = T - (k-1)m - 1$ and $\beta = vec(\mathbf{B})$, $\hat{\beta} = vec(\hat{\mathbf{B}})$.

Uhlig, H. , "What Macroeconomists Should Know About Unit Roots," Econometric Theory, vol. 10, no. 3.4, Aug.-Oct. 1994, 645-671.

Summarizing the Likelihood Function

Remark

Given the data Y_t, $t = -k+1, \ldots, T$, the conditional likelihood function as a function in B and Σ^{-1} is proportional to a Normal-Wishart density.

Remark

Conventional t and F statistics and their conventional p-values are meaningful in summarizing the shape of the likelihood function, regardless of whether there are unit roots or not.

Bayesian updating

- Proposition 1 on p. 670 in Uhlig (1994): if the prior for $\theta = (\mathbf{B}, \Sigma^{-1})$ is $NW(\bar{B}_0, N_0, S_0, \nu_0)$, then the posterior is $NW(\bar{B}_T, N_T, S_T, \nu_T)$, where

$$
\begin{aligned}
\nu_T &= \nu_0 + T \\
N_T &= N_0 + \mathbf{X}'\mathbf{X} \\
\bar{B}_T &= N_T^{-1}(N_0 \bar{B}_0 + \mathbf{X}'\mathbf{X}\hat{B}) \\
S_T &= \frac{\nu_0}{\nu_T} S_0 + \frac{T}{\nu_T}\hat{\Sigma} + \frac{1}{\nu_T}(\hat{B} - \bar{B}_0)' N_0 N_T^{-1} \mathbf{X}'\mathbf{X}(\hat{B} - \bar{B}_0)
\end{aligned}
$$

- Recommendation: a "weak" prior. Use $N_0 = 0$, $\nu_0 = 0$, S_0 and $\bar{B}_0$ arbitrary. Then, $\bar{B}_T = \hat{B}$, $S_T = \hat{\Sigma}$, $\nu_T = T$, $N_T = \mathbf{X}'\mathbf{X}$. See also the RATS manual.

In sum:

Remark

If the prior π_0 is given by a Normal-Wishart density, then the posterior π_T is given by a Normal-Wishart density as well

Is this a good procedure?

- Unit roots vs Normal-Wishart prior vs Jeffreys prior. We shall return to this issue later.
- Sims-Zha
- Website of Zha: check the code