State-Space Models and Kalman Filter

Frank Schorfheide

University of Pennsylvania

Gerzensee Ph.D. Course on Bayesian Macroeconometrics

May 28, 2019

A Basic State-Space Model

- Consider

$$\begin{aligned} y_t &= \Psi s_t + u_t \quad \text{measurement} \\ s_t &= \Phi s_{t-1} + \epsilon_t \quad \text{state transition} \end{aligned}$$

  where $\epsilon_t \sim iidN(0, \Sigma)$ and $u_t \sim iidN(0, H)$.

- y_t's are observed.
- s_t's are unobserved.
- Model generates joint density for the observations and latent states:

$$\begin{aligned} p(Y_{1:T}, S_{1:T}|\theta) &= \prod_{t=1}^{T} p(y_t, s_t | Y_{1:t-1}, S_{1:t-1}, \theta) \\ &= \prod_{t=1}^{T} p(y_t|s_t, \theta) p(s_t|s_{t-1}, \theta). \end{aligned}$$

- Problem: likelihood inference requires the marginal $p(Y_{1:T}|\theta)$.

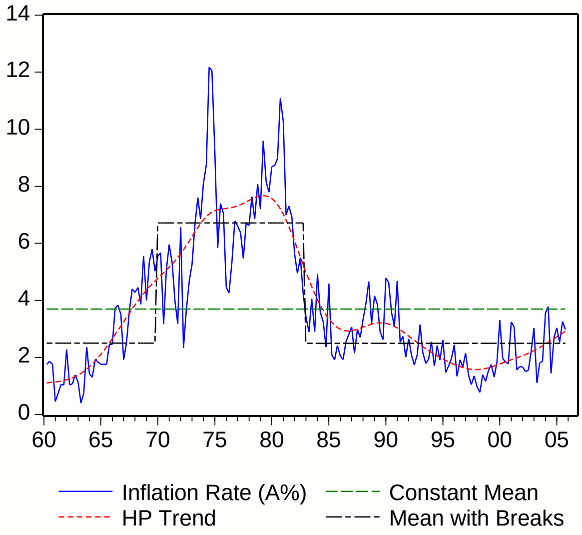
Application 1: Time-varying Coefficients

- We previously considered constant-coefficient autoregressive models.
- For instance, a simple model for inflation could be

$$\pi_t = \pi^* + \tilde{\pi}_t, \quad \tilde{\pi}_t = \rho \tilde{\pi}_{t-1} + \sigma_\epsilon \epsilon_t$$

- where
  - π^* is time-invariant steady state or target inflation;
  - $\tilde{\pi}_t$ captures fluctuations around the target

Application 1: Time-varying Coefficients



Application 1: Time-varying Coefficients

- Consider the following alternative model of inflation:

$$\pi_t = \pi_t^* + \tilde{\pi}_t$$

where π_t^* is a time-varying inflation target:

$$\tilde{\pi}_t = \rho \tilde{\pi}_{t-1} + \sigma_\epsilon \epsilon_t, \quad \pi_t^* = \pi_{t-1}^* + \sigma_\eta \eta_t.$$

- This looks like a state-space model:

$$\begin{aligned} y_t &= \begin{bmatrix} 1 & 1 \end{bmatrix} s_t \\ s_t &= \begin{bmatrix} \pi_t^* \\ \tilde{\pi}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \rho \end{bmatrix} s_{t-1} + \begin{bmatrix} \sigma_\eta & 0 \\ 0 & \sigma_\epsilon \end{bmatrix} \begin{bmatrix} \eta_t \\ \epsilon_t \end{bmatrix}. \end{aligned}$$

Application 2: Mixed-Frequency VAR

In macroeconomic applications, vector autoregressions (VARs) are typically estimated either exclusively based on

- quarterly observations
  $\Longrightarrow$ large set of macroeconomic series is available
- monthly information
  $\Longrightarrow$ VAR is able to track the economy more closely in real time

Application 2: Mixed-Frequency VAR

- State-Transition Equation
  - Economy evolves at monthly frequency according to the following VAR(p) dynamics:

$$x_t = \Phi_1 x_{t-1} + \ldots + \Phi_p x_{t-p} + \Phi_c + u_t, \quad u_t \sim iidN\big(0, \Sigma\big) \tag{1}$$

  - Partition: $x_t' = [x_{m,t}', \; x_{q,t}']$

- Measurement Equation
  - Actual observations are denoted by y_t and subscript indicates the observation frequency

$$\begin{aligned} y_{m,t} &= x_{m,t} \\ y_{q,t} &= \frac{1}{3}(x_{q,t} + x_{q,t-1} + x_{q,t-2}) \quad \text{if observed in period } t \end{aligned} \tag{2}$$

Application 3: Factor Models

- Extract "true" GDP growth from income and expenditure-side GDP measures.
- Measurement equation:

$$\begin{bmatrix} GDP_{Et} \\ GDP_{It} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} GDP_t + \begin{bmatrix} \epsilon_{Et} \\ \epsilon_{It} \end{bmatrix}$$

- State-transition equation:

$$GDP_t = \mu(1-\rho) + \rho GDP_{t-1} + \epsilon_{Gt}.$$

- We can also allow for correlation between measurement errors and state-transition innovations:

$$(\epsilon_{Gt}, \epsilon_{Et}, \epsilon_{It})' \sim iid\ N(\underline{0}, \Sigma), \quad \text{where} \quad \Sigma = \begin{bmatrix} \sigma^2_{GG} & 0 & 0 \\ 0 & \sigma^2_{EE} & \sigma^2_{EI} \\ 0 & \sigma^2_{IE} & \sigma^2_{II} \end{bmatrix}.$$

Filtering - General Idea

- State-space model:

$$
\begin{aligned}
y_t &= \Psi s_t + u_t \\
s_t &= \Phi s_{t-1} + \epsilon_t
\end{aligned}
$$

- Likelihood function:

$$
p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1}, \theta)
$$

- A filter generates a sequence of conditional distributions $s_t|Y_{1:t}$.
- Iterations:
  - Initialization at time $t-1$: $p(s_{t-1}|Y_{1:t-1}, \theta)$
  - Forecasting t given $t-1$:
    1. Transition equation: $p(s_t|Y_{1:t-1}, \theta) = \int p(s_t|s_{t-1}, Y_{1:t-1}, \theta) p(s_{t-1}|Y_{1:t-1}, \theta) ds_{t-1}$
    2. Measurement equation: $p(y_t|Y_{1:t-1}, \theta) = \int p(y_t|s_t, Y_{1:t-1}, \theta) p(s_t|Y_{1:t-1}, \theta) ds_t$
  - Updating with Bayes theorem. Once y_t becomes available:

$$
p(s_t|Y_{1:t}, \theta) = p(s_t|y_t, Y_{1:t-1}, \theta) = \frac{p(y_t|s_t, Y_{1:t-1}, \theta) p(s_t|Y_{1:t-1}, \theta)}{p(y_t|Y_{1:t-1}, \theta)}
$$

Kalman Filter

- Our model:

$$\begin{aligned} y_t &= \Psi s_t + u_t \\ s_t &= \Phi s_{t-1} + \epsilon_t \end{aligned}$$

  where $\epsilon_t \sim iidN(0, \Sigma)$ and $u_t \sim iidN(0, H)$.

- In linear Gaussian state-space model all distributions that appear in the filter are Gaussian. Thus, Kalman filter only tracks means and covariance matrices.

Kalman Filter – Initialization

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- Write $\mathbb{E}[s_0] = \hat{s}_{0|0}$ and $\mathbb{V}[s_0] = P_{0|0}$.
- Prior distribution for initial state s_0: $s_0 \sim N(\bar{s}_{0|0}, P_{0|0})$.
- Initialization: if
  1. s_t is stationary we can initialize the filter with the unconditional distribution of s_t. Covariance matrix:
     $$\mathbb{E}[s_t s_t'] = \Phi \mathbb{E}[s_t s_t'] \Phi' + \Sigma;$$
  2. otherwise, could assume that $s_t = 0$ for $t = -\tau$ or treat s_0 as parameter.

Kalman Filter – Forecasting

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- At $(t-1)^+$, that is, after observing y_{t-1}, the belief about the state vector has the form $s_{t-1}|Y_{1:t-1} \sim N(\bar{s}_{t-1|t-1}, P_{t-1|t-1})$.
- "Posterior" from period $t-1$ turns into a prior for $(t-1)^+$.
- Since s_{t-1} and ϵ_t are independent multivariate normal random variables, it follows that

  $$s_t|Y_{1:t-1} \sim N(\bar{s}_{t|t-1}, P_{t|t-1})$$

  where

  $$\begin{aligned} \bar{s}_{t|t-1} &= \Phi \bar{s}_{t-1|t-1} \\ P_{t|t-1} &= \Phi P_{t-1|t-1} \Phi' + \Sigma \end{aligned}$$

Kalman Filter – Forecasting and Likelihood Function

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- The conditional distribution of $y_t | s_t, Y_{1:t-1}$ is of the form

$$y_t | s_t, Y_{1:t-1} \sim N(\Psi s_t, H)$$

- Since $s_t | Y_{1:t-1} \sim N(\bar{s}_{t|t-1}, P_{t|t-1})$, we can deduce that the marginal distribution of y_t conditional on $Y_{1:t-1}$ is of the form

$$y_t | Y_{1:t-1} \sim N(\bar{y}_{t|t-1}, F_{t|t-1})$$

where $\bar{y}_{t|t-1} = \Psi \bar{s}_{t|t-1}$ and $F_{t|t-1} = \Psi P_{t|t-1} \Psi' + H$.

- Likelihood Function:

$$p(Y_{1:T} | \Psi, \Phi, \Sigma, H)$$
$$= (2\pi)^{-nT/2} \left(\prod_{t=1}^{T} |F_{t|t-1}| \right)^{-1/2} \exp \left\{ -\frac{1}{2} \sum_{t=1}^{T} (y_t - \bar{y}_{t|t-1})' F_{t|t-1}^{-1} (y_t - \bar{y}_{t|t-1}) \right\}$$

Kalman Filter – Updating

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- To obtain the posterior distribution of $s_t | y_t, Y_{1:t-1}$ note that

$$\begin{aligned} s_t &= \bar{s}_{t|t-1} + \big(s_t - \bar{s}_{t|t-1}\big) \\ y_t &= \bar{y}_{t|t-1} + \Psi\big(s_t - \bar{s}_{t|t-1}\big) + u_t \end{aligned}$$

- and the joint distribution of s_t and y_t is given by

$$\left[\begin{array}{c} s_t \\ y_t \end{array} \right] \Big| Y^{t-1} \sim N\left(\left[\begin{array}{c} \bar{s}_{t|t-1} \\ \bar{y}_{t|t-1} \end{array} \right], \left[\begin{array}{cc} P_{t|t-1} & P_{t|t-1}\Psi' \\ \Psi P'_{t|t-1} & F_{t|t-1} \end{array} \right] \right)$$

Kalman Filter – Updating

$y_t = \Psi s_t + u_t$, $s_t = \Phi s_{t-1} + \epsilon_t$ where $\epsilon_t \sim N(0, \Sigma)$ and $u_t \sim N(0, H)$.

- Applying Bayes theorem, i.e., calculating a conditional distribution based on a joint...

$$s_t | y_t, Y_{1:t-1} \sim N(\bar{s}_{t|t}, P_{t|t})$$

where

$$\begin{aligned} \bar{s}_{t|t} &= \bar{s}_{t|t-1} + P_{t|t-1}\Psi' F^{-1}_{t|t-1}(y_t - \bar{y}_{t|t-1}) \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}\Psi' F^{-1}_{t|t-1}\Psi P_{t|t-1} \end{aligned}$$

- The conditional mean and variance $\bar{y}_{t|t-1}$ and $F_{t|t-1}$ were given above.
- This completes one iteration of the algorithm. The posterior $s_t | Y_{1:t}$ is the prior for the next iteration.

Summary: Conditional Distributions for Kalman Filter

	Distribution	Mean and Variance
$s_{t-1}\|(Y_{1:t-1},\theta)$	$N\big(\bar{s}_{t-1\|t-1}, P_{t-1\|t-1}\big)$	Given from Iteration $t-1$
$s_t\|(Y_{1:t-1},\theta)$	$N\big(\bar{s}_{t\|t-1}, P_{t\|t-1}\big)$	$\bar{s}_{t\|t-1} = \Phi_1 \bar{s}_{t-1\|t-1}$ $P_{t\|t-1} = \Phi_1 P_{t-1\|t-1}\Phi_1' + \Phi_\epsilon \Sigma_\epsilon \Phi_\epsilon'$
$y_t\|(Y_{1:t-1},\theta)$	$N\big(\bar{y}_{t\|t-1}, F_{t\|t-1}\big)$	$\bar{y}_{t\|t-1} = \Psi_0 + \Psi_1 t + \Psi_2 \bar{s}_{t\|t-1}$ $F_{t\|t-1} = \Psi_2 P_{t\|t-1}\Psi_2' + \Sigma_u$
$s_t\|(Y_{1:t},\theta)$	$N\big(\bar{s}_{t\|t}, P_{t\|t}\big)$	$\bar{s}_{t\|t} = \bar{s}_{t\|t-1} + P_{t\|t-1}\Psi_2' F_{t\|t-1}^{-1}(y_t - \bar{y}_{t\|t-1})$ $P_{t\|t} = P_{t\|t-1} - P_{t\|t-1}\Psi_2' F_{t\|t-1}^{-1}\Psi_2 P_{t\|t-1}$

Simulation Smoother

- A filter generates the sequence of densities $p(s_t|Y_{1:t})$.
- However, often we are interested in $p(s_t|Y_{1:T})$.
- We will derive a simulation smoother that samples from $p(S_{1:T}|Y_{1:T})$:
  - It turns out that we can draw the states sequentially, starting from $s_T|Y_{1:T}$, which is obtained in the T'th iteration of the filter.
  - We then continue with

$$p(s_t|S_{t+1:T}, Y_{1:T}) \propto p(s_t, S_{t+1:T}, Y_{1:T})$$

An Important Result

Consider the following factorization

$$
\begin{aligned}
& p(s_t, S_{t+1:T}, Y_{1:T}) \\
&= \int p(S_{1:T}, Y_{1:T}) dS_{1:t-1} \\
&= \int p(S_{1:t}, Y_{1:t}) \cdot \big[p(s_{t+1}|s_t)p(y_{t+1}|s_{t+1})\big] \cdot \big[p(s_{t+2}|s_{t+1})p(y_{t+2}|s_{t+2})\big] \\
& \quad \ldots \big[p(s_T|s_{T-1})p(y_T|s_T)\big] dS_{1:t-1} \\
&= p(s_t, Y_{1:t}) \cdot \big[p(s_{t+1}|s_t)p(y_{t+1}|s_{t+1})\big] \cdot \text{terms without } s_t.
\end{aligned}
$$

We deduce

$$
\begin{aligned}
p(s_t|S_{t+1:T}, Y_{1:T}) &\propto p(s_t, Y_{1:t})p(s_{t+1}|s_t) \\
&\propto p(s_t, s_{t+1}, Y_{1:t}) \\
&= p(s_t|s_{t+1}, Y_{1:t})
\end{aligned}
$$

Summary: Simulation Smoother

- We now can write

$$p(S_{1:T}|Y_{1:T}) = p(s_T|Y_{1:T}) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, Y_{1:T}) = p(s_T|Y_{1:T}) \prod_{t=1}^{T-1} p(s_t|s_{t+1}, Y_{1:t}),$$

  where $p(s_t|s_{t+1}, Y_{1:t}) \propto p(s_t|Y_{1:t})p(s_{t+1}|s_t)$.

- Draw $s_t^{(i)} \sim p(s_t \mid s_{t+1}^{(i)}, Y_{1:T})$ for $t = T, \ldots, 1$.
  (a) Run the Kalman filter to get $\{\hat{s}_{t|t}, P_{t|t}\}_{t=1}^{T}$ where $s_t|Y_{1:t} \sim N(\hat{s}_{t|t}, P_{t|t})$;
  (b) Draw $s_T^{(i)} \sim N(\hat{s}_{T|T}, P_{T|T})$;
  (c) For $t = T-1, \ldots, 1$,

$$s_t^{(i)} \sim N\left(\hat{s}_{t|t+1},\ P_{t|t+1}\right)$$

  where

$$\hat{s}_{t|t+1} = \hat{s}_{t|t} + P_{t|t}\Phi' P_{t+1|t}^{-1}(s_{t+1}^{(i)} - \Phi\hat{s}_{t|t})$$
$$P_{t|t+1} = P_{t|t} - P_{t|t}\Phi' P_{t+1|t}^{-1}\Phi P_{t|t}$$

Application 4: Stylized DSGE Models

State-space representation:

$$
\begin{aligned}
y_t &= \Psi_0(\theta) + \Psi_1(\theta)s_t \\
s_t &= \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t
\end{aligned}
$$

System matrices:

$$
\Psi_0(\theta) = M_y' \begin{bmatrix} \log\gamma \\ \log(lsh) \\ \log\pi^* \\ \log(\pi^*\gamma/\beta) \end{bmatrix}, \quad x_\phi = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\phi}, \quad x_\lambda = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\lambda}, \quad x_z = \frac{\rho_z\psi_p}{1-\psi_p\rho_z}, \quad x_{\epsilon_R} = -\psi_p\sigma_R
$$

$$
\Psi_1(\theta) = M_y' \begin{bmatrix} x_\phi & x_\lambda & x_z+1 & x_{\epsilon_R} & -1 \\ 1+(1+\nu)x_\phi & (1+\nu)x_\lambda & (1+\nu)x_z & (1+\nu)x_{\epsilon_R} & 0 \\ \frac{\kappa_p}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \frac{\kappa_p}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \frac{\kappa_p}{1-\beta\rho_z}(1+\nu)x_z & +\kappa_p(1+\nu)x_{\epsilon_R} & 0 \\ \frac{\kappa_p/\beta}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \frac{\kappa_p/\beta}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \frac{\kappa_p/\beta}{1-\beta\rho_z}(1+\nu)x_z & (\kappa_p(1+\nu)x_{\epsilon_R}/\beta+\sigma_R) & 0 \end{bmatrix}
$$

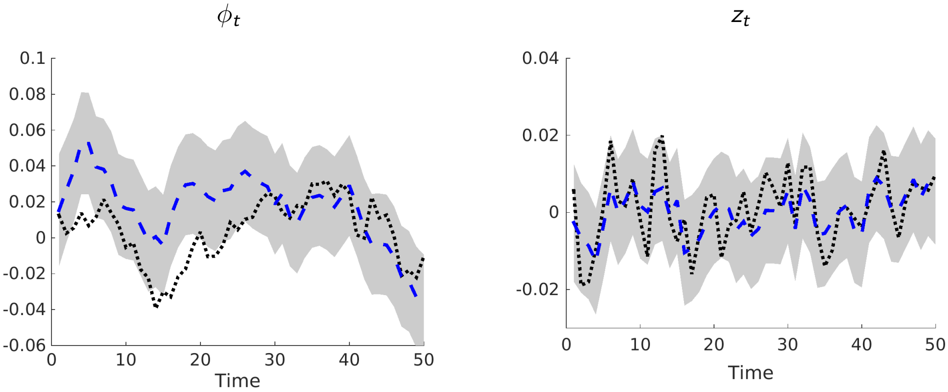
$$
\Phi_1(\theta) = \begin{bmatrix} \rho_\phi & 0 & 0 & 0 & 0 \\ 0 & \rho_\lambda & 0 & 0 & 0 \\ 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_\phi & x_\lambda & x_z & x_{\epsilon_R} & 0 \end{bmatrix}, \quad \Phi_\epsilon(\theta) = \begin{bmatrix} \sigma_\phi & 0 & 0 & 0 \\ 0 & \sigma_\lambda & 0 & 0 \\ 0 & 0 & \sigma_z & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}
$$

M_y' is an $n_y \times 4$ selection matrix that selects rows of Ψ_0 and Ψ_1.

Filtered States from DSGE Model

- Simulate $T = 50$ observations from the stylized DSGE model.
- Plot filtered and true latent shock processes ϕ_t and z_t: $\mathbb{E}[s_t | Y_{1:t}]$.
- First set of figures: only use output growth as observable.
- Second set of figures: add inflation and labor share as observable.
- "Error" bands: 90% credible intervals which are centered around the filtered estimates and based on the standard deviations $\sqrt{\mathbb{V}[s_t | Y_{1:t}]}$.

Filtering Based on $y_t = \log(X_t/X_{t-1})$



Filtering Based on $y_t = [\log(X_t/X_{t-1}), lsh_t, \pi_t]'$