Structural Macroeconometrics

Chapter III
Kalman Filter

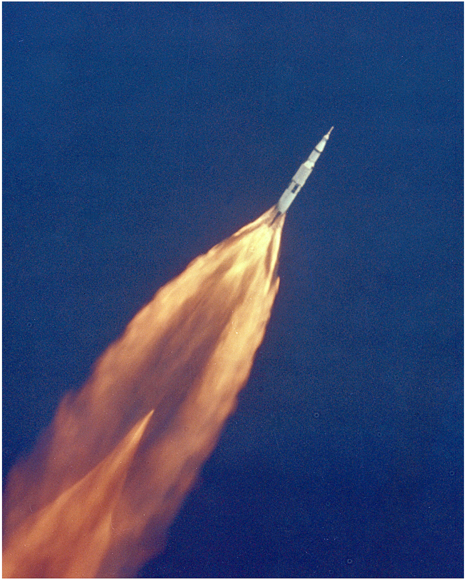
Johannes Pfeifer

Kobe University

Summer 2018

Outline

Cool Stuff



The Basic Concept

- named after seminal work of Kálmán (1960) and Kálmán and Bucy (1961)

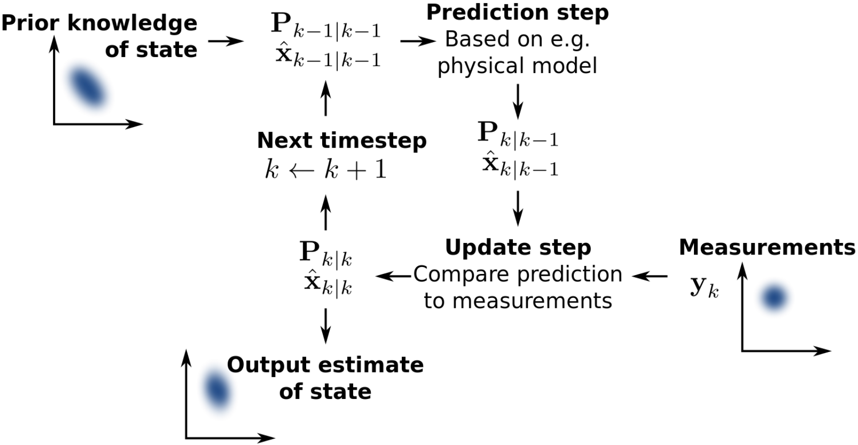


Figure 1: Source: "Basic concept of Kalman filtering" by Petteri Aimonen

References

- Handout online
- Ljungqvist and Sargent (2012, Chapter 2.5)
- Hamilton (1994, Chapter 13)
- Durbin and Koopman (2012)
- Canova (2007, Chapter 6)

Gaussian State Space

- Solution to linearized RBC model took **state space form**, which can be written as

$$x_{t+1} = Fx_t + w_{t+1},\ w_{t+1} \overset{iid}{\sim} \mathcal{N}(0, Q) \tag{1}$$

$$y_t = Hx_t + \nu_t,\ \nu_t \overset{iid}{\sim} \mathcal{N}(0, R) \tag{2}$$

- F, H, Q, R are functions of the model parameters θ
- x_t is an $n_x \times 1$ vector of states
- y_t is an $n_y \times 1$ vector of observables
- w_t is a $p \times 1$ vector of structural errors
- ν_t a vector of measurement errors
- Assumption: w_t and ν_t are orthogonal

$$E_t\left(w_{t+1}\nu_s\right) = 0 \quad \forall\ t+1 \text{ and } s \geq 0$$

Fundamental Problem: Unobserved States

- This implies that
$$y_t = H\left(Fx_{t-1} + w_t\right) + \nu_t \tag{3}$$
- Thus, y_t is normally distributed:
$$y_t \sim \mathcal{N}(HFx_{t-1}, HQH' + R) \tag{4}$$
- If all states were observed, we could directly construct the likelihood $f(y_T, \ldots, y_1|\theta)$
- We could then run optimizer over our estimated parameter set $\tilde{\theta} \subseteq \theta$ to get ML estimate of $\tilde{\theta}$
- Problem: we have **unobserved states** and cannot use equation (4)
- Solution: turn to Kalman filter to back out states from the observed data $\rightarrow$ **Filtering problem**

Initial Values

- initial value x_0 distributed as

$$x_0 \sim \mathcal{N}\left(\hat{x}_{0|-1}, \Sigma_{0|-1}\right) \tag{5}$$

- subscript -1 denotes information at the beginning of times
- particular observation: y_t
- complete history of observations up to a certain point in time: $y^t = \{y_t, \ldots, y_0\}$
- What does this imply for $y_0 = Hx_0 + \nu_0$?

Initial Values

- Let's define

$$\hat{y}_{0|-1} \equiv E(y_0|y^{-1}) = E(Hx_0 + \nu_0|y^{-1}) = H\hat{x}_{0|-1} \tag{6}$$

- **Mean Squared Error** of y_0 given information up to $t = -1$

$$\begin{aligned}
&E\left[\left(y_0 - \hat{y}_{0|-1}\right)\left(y_0 - \hat{y}_{0|-1}\right)' |y^{-1}\right] \\
&= E\left[\left(Hx_0 + \nu_0 - H\hat{x}_{0|-1}\right)\left(Hx_0 + \nu_0 - H\hat{x}_{0|-1}\right)' |y^{-1}\right] \\
&= E\left[H(x_0 - \hat{x}_{0|-1})(x_0 - \hat{x}_{0|-1})'H' + \nu_0\nu_0'|y^{-1}\right] \\
&= H\Sigma_{0|-1}H' + R
\end{aligned} \tag{7}$$

- Hence:

$$y_0 \sim N\left(H\hat{x}_{0|-1}, H\Sigma_{0|-1}H' + R\right) \tag{8}$$

What's the Goal?

- Economist only observes the observables: $y^t = \{y_t, \ldots, y_0\}$
- Wants to infer the **unobserved state variables** $x_t, \ldots, x_0$
- Assumption: Economist knows state space structure, i.e. H, F, R, Q
- Aim: find **recursive** formulas for the state forecast

$$\hat{x}_{t|t-1} \equiv E[x_t | y^{t-1}] \tag{9}$$

  and the Mean Squared Error/covariance matrices of the forecast error (FE)

$$\Sigma_{t|t-1} \equiv E\left[\left(x_t - \hat{x}_{t|t-1}\right)\left(x_t - \hat{x}_{t|t-1}\right)' | y^{t-1}\right] \tag{10}$$

- Recursiveness allows for online tracking

Idea

- regression of the (unknown) state FE on observation FE

$$x_t - \hat{x}_{t|t-1} = L_t \left(y_t - \hat{y}_{t|t-1} \right) + \eta_t \tag{11}$$

L_t: regression coefficient; η_t: regression residual

- Forecast $\hat{y}_{t|t-1}$ of y_t given by

$$\hat{y}_{t|t-1} = H\hat{x}_{t|t-1} \tag{12}$$

- regression equation can then be written as

$$x_t - \hat{x}_{t|t-1} = L_t \left(y_t - H\hat{x}_{t|t-1} \right) + \eta_t \tag{13}$$

- Problem: don't know the left-hand side $\Rightarrow$ can't compute L_t
- But: if we somehow knew L_t, could form a forecast of our state FE

Computing L_t

- Goal: compute L_t at time t
- General formula for a **population regression**

$$\beta = E\left(YX'\right)\left[E\left(XX'\right)\right]^{-1}$$

- In our case

$$\begin{aligned} L_t =& E\left[\left(x_t - \hat{x}_{t|t-1}\right)\left(y_t - H\hat{x}_{t|t-1}\right)'\right] \\ & \times \left(E\left[\left(y_t - H\hat{x}_{t|t-1}\right)\left(y_t - H\hat{x}_{t|t-1}\right)'\right]\right)^{-1} \\ =& \Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1} \end{aligned} \tag{14}$$

Stochastic Singularity

- Computation requires $\left(H\Sigma_{t|t-1}H' + R\right)^{-1}$
- For the inversion, state-space model must not feature **stochastic singularity**, i.e. the forecast error matrix of the observables must have full rank
- Typical requirement: as least as many shocks as observables
- Having more shocks than observables is not a problem
- Also requires that there is no collinearity between observables

$\rightarrow$ Problem if all components of budget constraint are observed

- Way out: add measurement error (see Altug 1989; Ireland 2004; Sargent 1989)
- May also help with model misspecification (Del Negro and Schorfheide 2009)

Forecasting x_1

- Rewrite state transition equation

$$\begin{aligned} x_1 &= F x_0 + w_1 \\ &= F\hat{x}_{0|-1} + F\left(x_0 - \hat{x}_{0|-1}\right) + w_1 \\ &\overset{(13)}{=} F\hat{x}_{0|-1} + F\left(L_0\left(y_0 - H\hat{x}_{0|-1}\right) + \eta_0\right) + w_1 \end{aligned} \tag{15}$$

- Forecast tomorrow's state just based on yesterday's forecast and today's observation

$$\begin{aligned} \hat{x}_{1|0} = E\left[x_1|y^0\right] &= F\hat{x}_{0|-1} + \underbrace{FL_0}_{K_0}\left(y_0 - H\hat{x}_{0|-1}\right) \\ &= F\hat{x}_{0|-1} + K_0\left(y_0 - H\hat{x}_{0|-1}\right) \end{aligned} \tag{16}$$

- **Kalman Gain**

$$K_0 = FL_0 = F\Sigma_{0|-1}H'\left(H\Sigma_{0|-1}H' + R\right)^{-1} \tag{17}$$

$\rightarrow$ how much is state estimate updated based on previous FE

x_2, x_3,...?

- We want to use population regression

$$L_1 = \Sigma_{1|0} H' \left(H \Sigma_{1|0} H' + R \right)^{-1}$$

- For this, we need covariance/mean squared error matrix $\Sigma_{1|0}$

$$\begin{aligned} \Sigma_{1|0} =& E\left[\left(x_1 - \hat{x}_{1|0} \right) \left(x_1 - \hat{x}_{1|0} \right)' |y^0 \right] \\ =& \left(F - K_0 H \right) \Sigma_{0|-1} \left(F - K_0 H \right)' + Q + K_0 R K_0{}' \end{aligned} \tag{18}$$

- From this, we now know that

$$x_1 \sim \mathcal{N} \left(\hat{x}_{1|0}, \Sigma_{1|0} \right) \tag{19}$$

- We can now start over again, delivering the recursion we were looking for

Summary

At time t, given $\hat{x}_{t|t-1}, \Sigma_{t|t-1}$ and observing y_t

1. Compute the forecast error in the observations using

$$a_t = y_t - H\hat{x}_{t|t-1} \tag{20}$$

2. Compute the **Kalman Gain** K_t using

$$K_t = F\Sigma_{t|t-1}H'\left(H\Sigma_{t|t-1}H' + R\right)^{-1} \tag{21}$$

3. Compute the state forecast for next period given today's information

$$\hat{x}_{t+1|t} = F\hat{x}_{t|t-1} + K_t\left(y_t - H\hat{x}_{t|t-1}\right) = F\hat{x}_{t|t-1} + K_t a_t \tag{22}$$

4. Update the covariance matrix

$$\Sigma_{t+1|t} = (F - K_t H)\,\Sigma_{t|t-1}\,(F - K_t H)' + Q + K_t R K_t' \tag{23}$$

Initialization

- How to initialize filter at $t = 0$ where no observations are available?

  $\rightarrow$ start with **unconditional** mean $E(x)$ and Variance Σ

- Given covariance stationarity, the unconditional mean is

$$E(x) = Ex_{t+1} = E(Fx_t + w_{t+1}) = FE(x) \Rightarrow (I - F)E(x) = 0$$

  hence, $E(x) = 0$

- For the covariance matrix, we have

$$\begin{aligned} \Sigma &= E\left[(Fx_t + w_t)\,(Fx_t + w_t)'\right] \\ &= E\left[Fx_t{x_t}'F' + w_t{w_t}'\right] \\ &= F\Sigma F' + Q \end{aligned} \tag{24}$$

  $\rightarrow$ so-called **Lyapunov-equation**

Initialization

- If A, B, C are conformable matrices

$$vec(ABC) = (C' \otimes A)vec(B)$$

- Hence

$$vec(\Sigma) = (F \otimes F)vec(\Sigma) + vec(Q)$$

- This has solution

$$vec(\Sigma) = \left(I_{n_x^2} - (F \otimes F)\right)^{-1} vec(Q) \tag{25}$$

- Problem: Inversion of n_x^2 matrix $\rightarrow$ slooooow
- Alternative: so-called **doubling algorithm** or other algorithms Dynare offers in `lyapunov_symm`

Initialization: Diffuse Kalman Filter

- Consider model with unit root: does not violate Blanchard and Kahn (1980) conditions
- Problem: unconditional variance does not exist for some variables, only the conditional one

→ cannot use unconditional one for initialization

- Alternative: **diffuse Kalman filter** (De Jong 1991; Koopman and Durbin 2003) that considers initial state as diffuse vector (infinite variance)

  → requires specifying the `diffuse_filter`-option of Dynare

Kalman Smoothing and other Outputs

- The Kalman filter recursions provide us with **filtered variables** $E_t x_{t+1}$, i.e. the best prediction for tomorrow's state given information up to today
- We can also compute **updated variables** $E_t x_t$, i.e. our best estimate of the state today given information up to today (remember that x_t is not in the information set here!)
- Note: the terminology here follows Dynare as there seems to be no standard for naming these two objects
- More importantly, we are regularly interested in our best estimates of shocks and states given the full observed data up to time T
- The **Kalman Smoother** provides recursions for obtaining these estimates $E_T(x_t)$ and $E_T(w_t)$ by working backwards in time (for details, see the handout)

From Filtering to Estimation

- Typically we are not interested in filtering per se, but rather in estimating a model
- Start with complete history of observables up to time t: $y^t = \{y_t, \ldots, y_0\}$
- Likelihood function $f(y_T, \ldots, y_0)$ can be factored as

$$f(y_T, \ldots, y_0) = f\left(y_T | y^{T-1}\right) \times f\left(y_{T-1} | y^{T-2}\right) \times \ldots \times f\left(y_1 | y^0\right) \times f(y_0)$$

- Earlier we derived

$$y_t \sim \mathcal{N}\left(H\hat{x}_{t|t-1}, \underbrace{H\Sigma_{t|t-1}H' + R}_{\equiv \Omega_t}\right)$$

Log-likelihood

- Hence, the probability density of observing y_t given y^{t-1} is given by

$$f\left(y_t|y^{t-1}\right) = \frac{1}{\sqrt{(2\pi)^{n_y}\det\left(\Omega_t\right)}} e^{-\frac{1}{2}\left(y_t - H\hat{x}_{t|t-1}\right)'\Omega_t^{-1}\left(y_t - H\hat{x}_{t|t-1}\right)},$$

- Taking logs leads to the **log-likelihood function** for each observation:

$$\log\left(f\left(y_t|y^{t-1}\right)\right) = -\frac{n_y}{2}\log\left(2\pi\right) - \frac{1}{2}\log\left(\det\left(\Omega_t\right)\right) - \frac{1}{2}a_t{}'\Omega_t^{-1}a_t \tag{26}$$

$\rightarrow$ Kalman filter delivers everything needed to compute the likelihood

- Can be easily integrated into our VAR-ML-code

Optimality

- Kalman filter is basically a least squares estimator
- If initial conditions and innovations are normal, it is the best predictor! You cannot get better!
- Otherwise, it is the best linear predictor
- Can only be used to construct likelihood function for linear (asymptotically) Gaussian state-space systems

$\rightarrow$ works perfectly for first-order approximated DSGE models

$\rightarrow$ does not work with higher order approximations/non-linearities

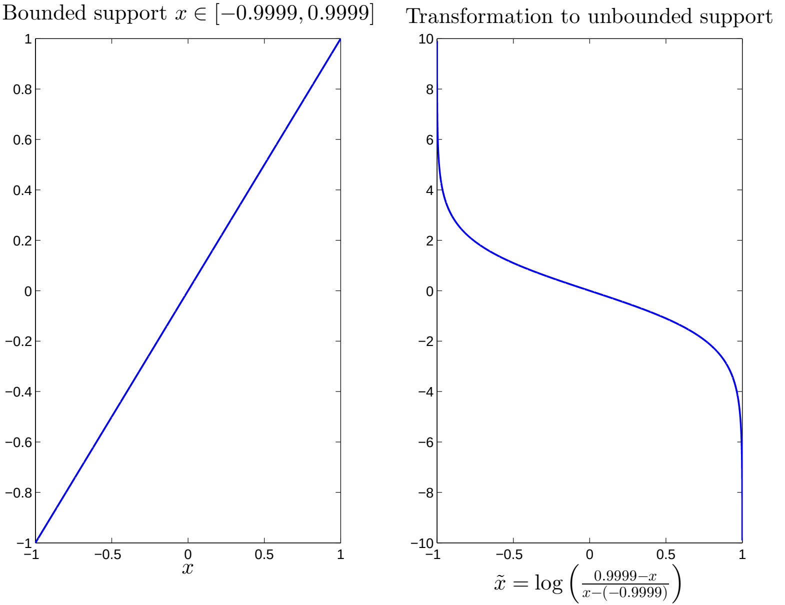
Finding the Global Maximum

- Finding a **global maximum** is hard!
- Newton-type optimizers are inherently **local**
- One way out: try numerous starting points
- Alternative: use **global optimizers** like simulated annealing (e.g. Corona, Marchesi, Martini, and Ridella 1987; Goffe, Ferrier, and Rogers 1994) or covariance matrix adaptive evolutionary strategy (CMA-ES) (Hansen, Müller, and Koumoutsakos 2003)
- Particularly CMA-ES (`mode_compute=9` in Dynare) seems to perform well in practice (Andreasen 2010)
- Dynare supports various different optimizers, calling them sequentially seems to work pretty well in hard cases

Constrained Numerical Optimization

- Most numerical optimizers are **unconstrained optimizers**
- Problem: economic parameters are typically **bounded**; e.g. $\beta \in [0, 1]$
- One way out: directly rely on constrained optimizers (e.g. `fmincon`)
- Or: use one-to-one transformation of parameter x into unbounded parameter $\tilde{x}$ before passing it to optimizer
- Then retransform $\tilde{x}$ back to x before using it in economic model inside of optimizer:
- $x \in [LB, Inf)$:
  $\tilde{x} = \log(x - LB) \Leftrightarrow x = LB + exp(\tilde{x})$
- $x \in (-Inf, UB]$:
  $\tilde{x} = \log(-x + UB) \Leftrightarrow x = UB - exp(\tilde{x})$
- $x \in [LB, UB]$:
  $\tilde{x} = \log\left(\frac{UB-x}{x-LB}\right) \Leftrightarrow x = \frac{UB+exp(\tilde{x})\times LB}{1+exp(\tilde{x})}$

Logistic Transformation for Autocorrelation Parameter



Steady State Kalman Filter

- It can be shown that the matrix recursions in the Kalman filter converge (See Hamilton 1994, Chapter 13.5)
- Kalman gain and forecast error variance matrix will settle on final values
- Continuing with recursions after they have settled within reasonable tolerance is computationally expensive

$\rightarrow$ check for convergence and then continue with fixed matrices

- Dynare by default uses this **steady state Kalman filter**

Numerical Issues and Invalid Parameter Draws

- There will generally be parameter draws for which the model cannot be solved (no valid steady state, BK conditions not satisfied)
- We could set their likelihood to 0, but many optimizers have issues with the arising cliffs
- Better: add **penalty function** that penalizes deviations from desired target so as to provide algorithms with a direction
- Also: when conducting Kalman filter recursions in the computer, various numerical problems can appear
- E.g. assuring perfect symmetry of covariance matrices in each step can improve the performance considerably
- It is often better to use existing routines than to reinvent the wheel

Observation Equations

- Up to this point, we have been silent on how to map the observed data to the model variables
- This is the topic of specifying the **observation equations**
- The biggest problem is stationarity in the model vs. trends in the data
- The guiding principle is: **the data variable needs to perfectly correspond to a variable defined in the model**
- A lengthy treatment of many possible cases is available in Pfeifer (2013)
- We will most of the time consider demeaned growth rates of variables together with a loglinear model
- This gives rise to a straightforward observation equation of the form: `y_obs=y-y(-1);`

Bibliography I

Altug, Sumru (1989). "Time-to-build and aggregate fluctuations: some new evidence". *International Economic Review 30* (4), 889–920.

Andreasen, Martin M. (2010). "How to maximize the likelihood function for a DSGE model". *Computational Economics 35* (2), 127–154.

Blanchard, Olivier J. and Charles M. Kahn (1980). "The solution of linear difference models under rational expectations". *Econometrica 48* (5), 1305–11.

Canova, Fabio (2007). *Methods for applied macroeconomic research*. Princeton University Press.

Corona, Angelo, M. Marchesi, Claudio Martini, and Sandro Ridella (1987). "Minimizing multimodal functions of continuous variables with the "simulated annealing" algorithm". *ACM Transactions on Mathematical Software 13* (3), 262–280.

De Jong, Piet (1991). "The diffuse Kalman filter". *The Annals of Statistics 19* (2), 1073–1083.

Bibliography II

Del Negro, Marco and Frank Schorfheide (2009). "Monetary policy analysis with potentially misspecified models". *American Economic Review 99* (4), 1415–50.

Durbin, James and Siem J. Koopman (2012). *Time series analysis by state space methods*. Second Revised Edition. Oxford: Oxford University Press.

Goffe, William L., Gary D. Ferrier, and John Rogers (1994). "Global optimization of statistical functions with simulated annealing". *Journal of Econometrics 60* (1/2), 65–100.

Hamilton, James D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Hansen, Nikolaus, Sybille D. Müller, and Petros Koumoutsakos (2003). "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)". *Evolutionary Computation 11* (1), 1–18.

Bibliography III

Ireland, Peter N. (2004). “A Method for Taking Models to the Data”. *Journal of Economic Dynamics and Control 28* (6), 1205–1226.

Kálmán, Rudolf E. (1960). “A new approach to linear filtering and prediction problems”. *Journal of Basic Engineering 82* (1).

Kálmán, Rudolf E. and Richard S. Bucy (1961). “New results in linear filtering and prediction theory”. *Journal of Basic Engineering 83*, 95–108.

Koopman, Siem J. and James Durbin (2003). “Filtering and smoothing of state vector for diffuse state-space models”. *Journal of Time Series Analysis 24* (1), 85–98.

Ljungqvist, Lars and Thomas J. Sargent (2012). *Recursive macroeconomic theory*. 3rd ed. Cambridge, MA: MIT Press.

Pfeifer, Johannes (2013). “A guide to specifying observation equations for the estimation of DSGE models”. Mimeo. University of Cologne.

Sargent, Thomas J. (1989). “Two models of measurements and the investment accelerator”. *Journal of Political Economy 97* (2), 251–87.