Bayesian Methods for DSGE models
Lecture 2
State Space Models and the Kalman Filter

Kristoffer Nimark
Cornell University

July 5, 2016

State Space Models

The most general form to write linear models is as state space systems

$$
\begin{aligned}
X_t &= A_t X_{t-1} + C_t \mathbf{u}_t : \mathbf{u}_t \sim N(0, I) \text{ (state equation)} \\
Z_t &= D_t X_t + \mathbf{v}_t : \mathbf{v}_t \sim N(0, \Sigma_v) \text{ (measurement equation)}
\end{aligned}
$$

Nests "observable" VAR(p), MA(p) and VARMA(p,q) processes as well as systems with latent variables.

State Space Models: Examples

The $VAR(p)$ model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-1} + u_t$$

can be written as

$$\begin{aligned} X_t &= A_t X_{t-1} + C_t \mathbf{u}_t \\ Z_t &= D_t X_t + \mathbf{v}_t \end{aligned}$$

where

$$\begin{aligned} A &= \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_p \\ I & 0 & & 0 \\ 0 & \ddots & & \ddots \\ 0 & 0 & I & 0 \end{bmatrix}, C = \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix} u_t \\ D &= \begin{bmatrix} I & 0 & \cdots & 0 \end{bmatrix}, \Sigma_{vv} = 0 \end{aligned}$$

MA(1) in State Space Form

The MA(1) process

$$y_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

can be written as

$$\begin{aligned}
\begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix} &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1} \\ \varepsilon_{t-2} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \varepsilon_t \\
y_t &= \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \varepsilon_{t-1} \end{bmatrix}
\end{aligned}$$

which is also of the form

$$\begin{aligned}
X_t &= A_t X_{t-1} + C_t \mathbf{u}_t \\
Z_t &= D_t X_t + \mathbf{v}_t
\end{aligned}$$

A simple DSGE model as a State Space System

Our benchmark 3 equation New Keynesian model:

$$\begin{aligned}
x_t &= \rho x_{t-1} + u_t^x \\
y_t &= E_t(y_{t+1}) - \frac{1}{\gamma}\left[r_t - E_t\left(\pi_{t+1}\right)\right] + u_t^y \\
\pi_t &= E_t\left(\pi_{t+1}\right) + \kappa\left[y_t - x_t\right] + u_t^\pi \\
r_t &= \phi \pi_t + u_t^r
\end{aligned}$$

Same as yesterday but with more shocks.

We want to estimate the parameters
$\theta = \{\rho, \gamma, \kappa, \phi, \sigma_x, \sigma_y, \sigma_\pi, \sigma_r, \}$

Benchmark 3 equation New Keynesian model

The solved model

$$
\begin{aligned}
x_t &= \rho x_{t-1} + u_t^x \\
y_t &= -\kappa \frac{\rho - \phi}{c} x_t + u_t^y + \frac{1}{\gamma} u_t^r \\
\pi_t &= \kappa\gamma \frac{\rho - 1}{c} x_t + u_t^\pi
\end{aligned}
$$

where $c = \gamma - \kappa\rho - 2\gamma\rho + \kappa\phi + \gamma\rho^2 < 0$

Benchmark 3 equation New Keynesian model

The solved model can be put in state space form

$$\begin{aligned} X_t &= AX_{t-1} + Cu_t \\ Z_t &= DX_t + v_t \end{aligned}$$

where

$$\begin{aligned} X_t &= x_t, A = \rho, Cu_t = u_t^x \\ Z_t &= \begin{bmatrix} r_t \\ \pi_t \\ y_t \end{bmatrix}, D = \begin{bmatrix} \phi\kappa\gamma\frac{1-\rho}{-c} \\ \kappa\gamma\frac{1-\rho}{-c} \\ -\kappa\frac{\phi-\rho}{-c} \end{bmatrix}, v_t = R \begin{bmatrix} u_t^r \\ u_t^\pi \\ u_t^y \end{bmatrix} \end{aligned}$$

The Kalman Filter

The Kalman Filter

The Kalman filter is mainly used for two purposes:

1. Form an estimate of the unobservable state X_t
2. To evaluate the likelihood function associated with a state space model

The Kalman Filter

For state space systems of the form

$$\begin{aligned} X_t &= A_t X_{t-1} + C_t \mathbf{u}_t \\ Z_t &= D_t X_t + \mathbf{v}_t \end{aligned}$$

the Kalman filter recursively computes estimates of X_t conditional on the history of observations $Z_t, Z_{t-1}, ...Z_0$ and an initial estimate (or prior) $X_{0|0}$ with variance $P_{0|0}$.
The form of the filter is

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t X_{t|t-1} \right)$$

and the task is thus to find the Kalman gain K_t so that the estimates $X_{t|t}$ are in some sense "optimal".

Notation

Define

$$X_{t|t-s} \equiv E[X_t \mid Z^{t-s}]$$

and

$$P_{t|t-s} \equiv E(X_t - X_{t|t-s})(X_t - X_{t|t-s})'$$

A Simple Example

A Simple Example

Let's say that we have a noisy measures z^1 of the unobservable process x so that

$$\begin{aligned} z_1 &= x + v_1 \\ v_1 &\sim N(0, \sigma_1^2) \end{aligned}$$

Since the signal is unbiased, the minimum variance estimate $E\left[x \mid z^1\right] \equiv \widehat{x}$ of x is simply given by

$$\widehat{x} = z_1$$

and its variance is equal to the variance of the noise

$$E\left[\widehat{x} - x\right]^2 = \sigma_1^2$$

Introducing a second signal

Now, let's say we have an second measure z_2 of x so that

$$\begin{aligned} z_2 &= x + v_2 \\ v_2 &\sim N(0, \sigma_2^2) \end{aligned}$$

How can we combine the information in the two signals to find the a minimum variance estimate of x?

If we restrict ourselves to linear estimators of the form

$$\widehat{x} = (1 - a)\, z_1 + a z_2$$

we can simply minimize

$$E\left[(1 - a)\, z_1 + a z_2 - x\right]^2$$

with respect to a.

Minimizing the variance

Rewrite expression for variance as

$$\begin{aligned} & E\left[(1-a)(x+v_1)+a(x+v_2)-x\right]^2 \\ = \;& E\left[(1-a)v_1+av_2\right]^2 \\ = \;& \sigma_1^2-2a\sigma_1^2+a^2\sigma_1^2+a^2\sigma_2^2 \end{aligned}$$

where the third line follows from the fact that v^1 and v^2 are uncorrelated so all expected cross terms are zero. Differentiate w.r.t. a and set equal to zero

$$-2\sigma_1^2+2a\sigma_1^2+2a\sigma_2^2=0$$

and solve for a

$$a=\sigma_1^2/(\sigma_1^2+\sigma_2^2)$$

The minimum variance estimate of x

The minimum variance estimate of x is thus given by

$$\widehat{x} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} z_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} z_2$$

with conditional variance

$$E\left[\widehat{x} - x\right]^2 = \left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1}$$

For $\sigma_2^2 < \infty$ we have that

$$\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)^{-1} < \sigma_1^2$$

so we get a better estimate with two signals.

The Scalar Filter

The Scalar Filter

Consider the process

$$\begin{array}{rcl} x_t & = & \rho x_{t-1} + u_t \\ z_t & = & x_t + v_t \\ \left[\begin{array}{c} u_t \\ v_t \end{array} \right] & \sim & N\left(0, \left[\begin{array}{cc} \sigma_u^2 & 0 \\ 0 & \sigma_v^2 \end{array} \right] \right) \end{array}$$

We want to form an estimate of x_t conditional on $z^t = \{z_t, z_{t-1,\ldots,}z_1\}$.

In addition to the knowledge of the state space system above we have a "prior" knowledge about the initial value of the state x_0 so that

$$\begin{array}{rcl} x_{0|0} & = & \overline{x}_0 \\ E\left(\overline{x}_0 - x_0\right)^2 & = & p_0 \end{array}$$

With this information we can form a prior about x_1.

The scalar filter cont'd.

Using the state transition equation we get

$$x_{1|0} \equiv E\left[x_1 \mid x_{0|0}\right] = \rho x_{0|0}$$

The variance of the prior estimate then is

$$E\left(x_{1|0} - x_1\right)^2 = \rho^2 p_0 + \sigma_u^2$$

- $\rho^2 p_0$ is the uncertainty from period 0 carried over to period 1
- σ_u^2 is the uncertainty in period 0 about the period 1 innovation to x_t

Denote prior variance as

$$p_{1|0} = \rho^2 p_0 + \sigma_u^2$$

The scalar filter cont'd.

The information in the signal z_1 can be combined with the information in the prior in exactly the same way as we combined the two signals in the previous section.

The optimal weight k_1 in

$$x_{1|1} = (1 - k_1)x_{1|0} + k_1 z_1$$

is thus given by

$$k_1 = \frac{p_{1|0}}{p_{1|0} + \sigma_v^2}$$

and the period 1 posterior error covariance $p_{1|1}$ then is

$$p_{1|1} = \left(\frac{1}{p_{1|0}} + \frac{1}{\sigma_v^2} \right)^{-1}$$

or equivalently

$$p_{1|1} = p_{1|0} - p_{1|0}^2 (p_{1|0} + \sigma_v^2)^{-1}$$

The Scalar Filter Cont'd.

We can again propagate the posterior error variance $p_{1|1}$ one step forward to get the next period prior variance $p_{2|1}$

$$p_{2|1} = \rho^2 p_{1|1} + \sigma_u^2$$

or

$$p_{2|1} = \rho^2 \left(p_{1|0} - p_{1|0}^2 (p_{1|0} + \sigma_v^2)^{-1} \right) + \sigma_u^2$$

By an induction type argument, we can find a general difference equation for the evolution of prior error variances

$$p_{t|t-1} = \rho^2 \left(p_{t-1|t-2} - p_{t-1|t-2}^2 (p_{t-1|t-2} + \sigma_v^2)^{-1} \right) + \sigma_u^2$$

The associated period t Kalman gain is then given by

$$k_t = p_{t|t-1}(p_{t|t-1} + \sigma_v^2)^{-1}$$

which allows us to compute

$$x_{t|t} = (1 - k_t) x_{t|t-1} + k_t z_t$$

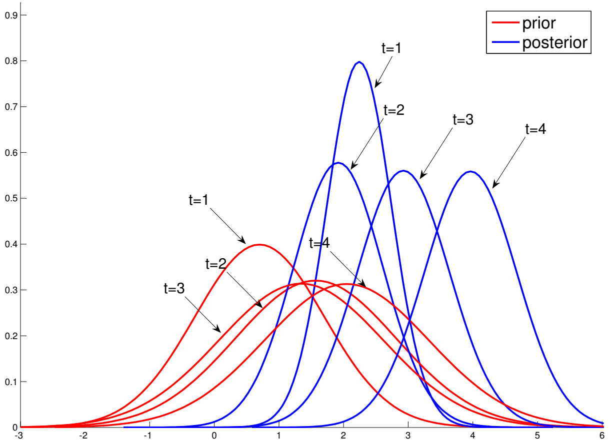
The scalar filter

$$\begin{aligned} x_t &= \rho x_{t-1} + u_t : u_t \sim N(0, \sigma_u^2) \text{ (state equation)} \\ z_t &= x_t + v_t : v_t \sim N(0, \sigma_v^2) \text{ (measurement equation)} \end{aligned}$$

gives the Kalman update equations

$$\begin{aligned} x_{t|t} &= \rho x_{t-1|t-1} + k_t \left(z_1 - \rho x_{t-1|t-1} \right) \\ k_t &= p_{t|t-1}(p_{t|t-1} + \sigma_v^2)^{-1} \\ p_{t|t-1} &= \rho^2 \underbrace{\left(p_{t-1|t-2} - p_{t-1|t-2}^2 (p_{t-1|t-2} + \sigma_v^2)^{-1} \right)}_{p_{t-1|t-1}} + \sigma_u^2 \end{aligned}$$

Propagation of the filter



Properties

There are two things worth noting about the difference equation for the prior error variances:

1. The prior error variance is bounded both from above and below so that

$$\sigma_u^2 \leq p_{t|t-1} \leq \frac{1}{(1-\rho^2})\sigma_u^2$$

2. For $0 \leq |\rho| < 1$ the iteration is a contraction

The upper bound in (1) is given by the optimality of the filter: we cannot do worse than making the unconditional mean our estimate of x_t for all t.

The lower bound is given by that the future is inherently uncertain as long as there are innovations in the x_t process, so even with a perfect estimate of x_{t-1}, x_t will still not be known with certainty.

The scalar filter

$$
\begin{aligned}
x_t &= \rho x_{t-1} + u_t : u_t \sim N(0, \sigma_u^2) \text{ (state equation)} \\
z_t &= x_t + v_t : v_t \sim N(0, \sigma_v^2) \text{ (measurement equation)}
\end{aligned}
$$

gives the Kalman update equations

$$
\begin{aligned}
x_{t|t} &= \rho x_{t-1|t-1} + k_t \left(z_1 - \rho x_{t-1|t-1} \right) \\
k_t &= p_{t|t-1}(p_{t|t-1} + \sigma_v^2)^{-1} \\
p_{t|t-1} &= \rho^2 \underbrace{\left(p_{t-1|t-2} - p_{t-1|t-2}^2 (p_{t-1|t-2} + \sigma_v^2)^{-1} \right)}_{p_{t-1|t-1}} + \sigma_u^2
\end{aligned}
$$

What determines the Kalman gain k_t?

Kalman filter optimally combine information in prior $\rho x_{t-1|t-1}$ and signal z_t to form posterior estimate $x_{t|t}$ with covariance $p_{t|t}$

$$x_{t|t} = (1 - k_t)\rho x_{t-1|t-1} + k_t z_1$$

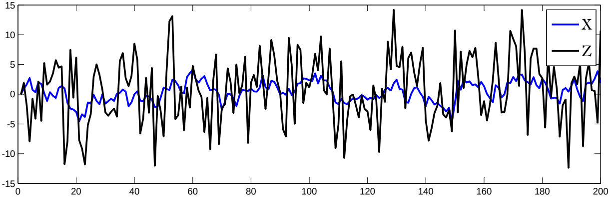
- More weight on signal (large kalman gain k_t) if prior variance is large or if signal is very precise
- Prior variance can be large either because previous state estimate was imprecise (i.e. $p_{t-1|t-1}$ is large) or because variance of state innovations is large (i.e. σ_u^2 is large)

Example 1

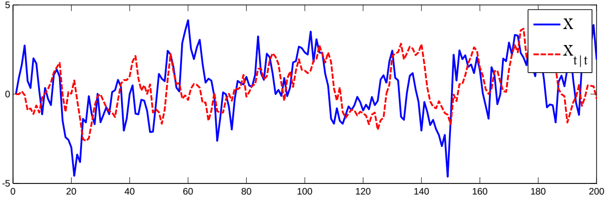
Set

- $\rho = 0.9$
- $\sigma_u^2 = 1$
- $\sigma_v^2 = 5$

Example 1



15
10
5
0
-5
-10
-15
0
20
40
60
80
100
120
140
160
180
200
X
Z



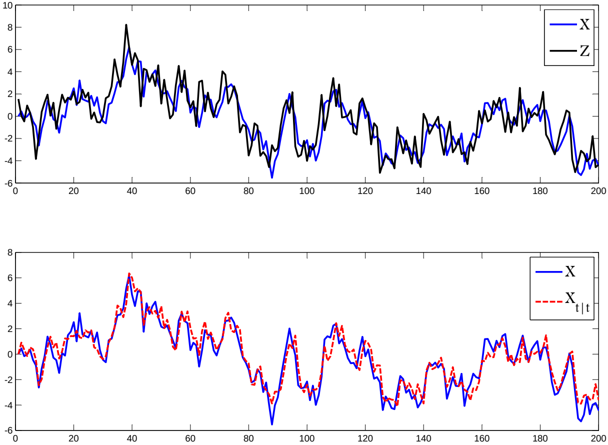
5
0
-5
0
20
40
60
80
100
120
140
160
180
200
X
X
t|t

Example 2

Set

- $\rho = 0.9$
- $\sigma_u^2 = 1$
- $\sigma_v^2 = 1$

Example 2: Smaller measurement error variance



Convergence to time invariant filter

If $\rho < 1$ and if ρ, σ_u^2 and σ^2 are constant, the prior variance of the state estimate

$$p_{t|t-1} = \rho^2 \left(p_{t-1|t-2} - p_{t-1|t-2}^2 (p_{t-1|t-2} + \sigma_v^2)^{-1} \right) + \sigma_u^2$$

will converge to

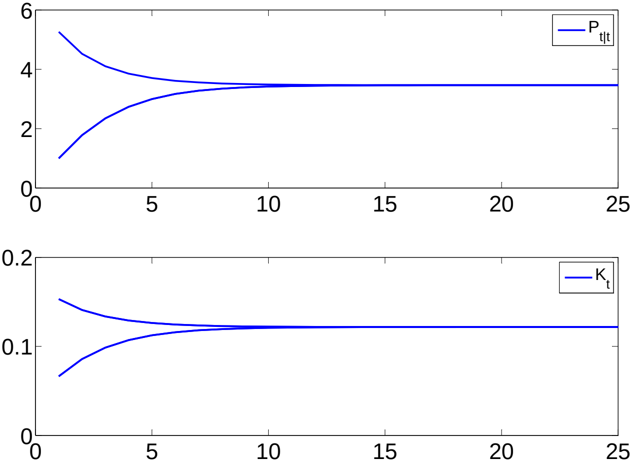
$$p = \rho^2 \left(p - p^2 (p + \sigma_v^2)^{-1} \right) + \sigma_u^2$$

The Kalman gain will then also converge:

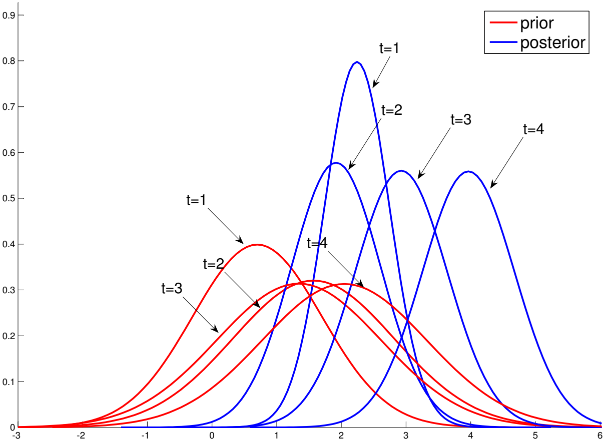
$$k = p(p + \sigma_v^2)^{-1}$$

We can illustrate this by starting from the boundaries of possible values for $p_{1|0}$

Convergence to time invariant filter



Convergence to time invariant filter



The Multivariate Filter

The Kalman Filter

For state space systems of the form

$$\begin{aligned} X_t &= A_t X_{t-1} + C_t \mathbf{u}_t \\ Z_t &= D_t X_t + \mathbf{v}_t \end{aligned}$$

the Kalman filter recursively computes estimates of X_t conditional on the history of observations $Z_t, Z_{t-1}, ...Z_0$ and an initial estimate (or prior) $X_{0|0}$ with variance $P_{0|0}$.
The form of the filter is

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t X_{t|t-1} \right)$$

and the task is thus to find the Kalman gain K_t so that the estimates $X_{t|t}$ are in some sense "optimal".

We further assume that $X_{0|0} - X_0$ is uncorrelated with the shock processes $\{\mathbf{u}_t\}$ and $\{\mathbf{v}_t\}$.

A Brute Force Linear Minimum Variance Estimator

The general period t problem:

$$\min_{\alpha} E\left[X_t - \sum_{j=0}^{t}\alpha_j Z_{t-j}\right]\left[X_t - \sum_{j=0}^{t}\alpha_j Z_{t-j}\right]'$$

We want to find the linear projection of X_t on the history of observables $Z_t, Z_{t-1}, ...Z_1$. From the projection theorem, the linear combination $\sum_{j=1}^{t}\alpha_j Z_{t-j+1}$ should imply errors that are orthogonal to $Z_t, Z_{t-1}, ...Z_1$ so that

$$\left(X_t - \sum_{j=0}^{t}\alpha_j Z_{t-j}\right) \perp \{Z_j\}_{j=1}^{t}$$

holds.

A Brute Force Linear Minimum Variance Estimator

We could compute the αs directly as

$$P\left(X_t \mid Z_t, Z_{t-1}, \ldots Z_1\right) = E\left(X_t \left[Z_t' \; Z_{t-1}' Z_1'\right]'\right) \times$$
$$\left(E\left[Z_t' \; Z_{t-1}' \ldots Z_1'\right]\left[Z_t' \; Z_{t-1}' \ldots Z_1'\right]'\right)^{-1} \times \left[Z_t' \; Z_{t-1}' \ldots Z_1' \quad\right]'$$

but that is not particularly convenient as $t \to \infty$.

2 tricks to find recursive formulation

1. Gram-Schmidt Orthogonalization
2. Exploit a convenient property of projections onto mutually orthogonal variables

Gram-Schmidt Orthogonalization in $\mathbb{R}^m$

Let the matrix $Y \;\; (m \times n)$ have columns $\mathbf{y}_1, \mathbf{y}_2, \ldots.\mathbf{y}_n$.

$$Y = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_n \end{bmatrix}$$

- The first column can be chosen arbitrarily so we might as well keep the first column of Y as it is.
- The second column should be orthogonal to the first. Subtract the projection of $\mathbf{y}_2$ on $\mathbf{y}_1$ from $\mathbf{y}_2$ and define a new column vector $\widetilde{\mathbf{y}}_2$

  $$\widetilde{\mathbf{y}}_2 = \mathbf{y}_2 - \mathbf{y}_1 \left(\mathbf{y}_1' \mathbf{y}_1\right)^{-1} \mathbf{y}_1' \mathbf{y}_2$$

  or

  $$\widetilde{\mathbf{y}}_2 = \left(I - \mathcal{P}_{y_1}\right) \mathbf{y}_2$$

  and then subtract the projection of $\mathbf{y}_3$ on $[\mathbf{y}_1 \;\; \mathbf{y}_2]$ from $\mathbf{y}_3$ to construct $\widetilde{\mathbf{y}}_3$ and so on.

Projections onto uncorrelated variables

Let Z and Y be two uncorrelated mean zero variables so that

$$E\left[ZY'\right] = 0$$

then

$$E\left[X \mid Z, Y\right] = E\left[X \mid Z\right] + E\left[X \mid Y\right]$$

To see why, just write out the projection formula. If the variables that we project on are orthogonal, the inverse will be taken of a diagonal matrix.

Finding the Kalman gain K_t

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t X_{t|t-1} \right)$$

Finding the Kalman gain K_1

Start from the first period problem of how to optimally combine the information in the prior $X_{0|0}$ and the signal Z_1 : Use that

$$Z_1 = D_1 A_0 X_0 + D_1 C \mathbf{u}_1 + \mathbf{v}_1$$

and that we know that $\mathbf{u}_t$ and $\mathbf{v}_t$ are orthogonal to $X_{0|0}$ to first find the optimal projection of Z_1 on $X_{0|0}$

$$Z_{1|0} = D_1 A_0 X_{0|0}$$

We can then define the period 1 innovation $\widetilde{Z}_1$ in Z_1 as

$$\widetilde{Z}_1 = Z_1 - Z_{1|0}$$

We know that

$$E\left(X_1 \mid \widetilde{Z}_1, X_{0|0}\right) = E\left(X_1 \mid \widetilde{Z}_1\right) + E\left(X_1 \mid X_{0|0}\right)$$

since $\widetilde{Z}_1 \perp X_{0|0}$ and $E\left(Z_1 \mid X_{0|0}\right) = D_1 A_0 X_{0|0}$.

Finding K_1

From the projection theorem, we know that we should look for a K_1 such that the inner product of the projection error and $\widetilde{Z}_1$ are zero

$$\left\langle X_1 - K_1 \widetilde{Z}_1, \widetilde{Z}_1 \right\rangle = 0$$

Defining the inner product $\langle X, Y \rangle$ as $E\left(XY'\right)$ we get

$$\begin{aligned} E\left[\left(X_1 - K_1 \widetilde{Z}_1\right) \widetilde{Z}_1'\right] &= 0 \\ E\left[X_1 \widetilde{Z}_1'\right] - K_1 E\left[\widetilde{Z}_1 \widetilde{Z}_1'\right] &= 0 \\ K_1 &= E\left[X_1 \widetilde{Z}_1'\right] \left(E\left[\widetilde{Z}_1 \widetilde{Z}_1'\right]\right)^{-1} \end{aligned}$$

We thus need to evaluate the two expectational expressions above.

Finding $E\left[X_1 \widetilde{Z}_1'\right]$

Before doing so it helps to define the state innovation

$$\widetilde{X}_1 = X_1 - X_{1|0}$$

that is, $\widetilde{X}_1$ is the one period error. The first expectation factor of K_1 in (43) can now be manipulated in the following way

$$\begin{aligned} E\left[X_1 \widetilde{Z}_1'\right] &= E\left(\widetilde{X}_1 + X_{1|0}\right)\widetilde{Z}_1' \\ &= E\widetilde{X}_1 \widetilde{Z}_1' \\ &= E\widetilde{X}_1\left(\widetilde{X}_1' D' + \mathbf{v}_1'\right) \\ &= P_{1|0} D' \end{aligned}$$

Evaluating $E\left[\widetilde{Z}_1\widetilde{Z}_1'\right]$

Evaluating the second expectation factor

$$\begin{aligned} E\left[\widetilde{Z}_1\widetilde{Z}_1'\right] &= E\left[\left(D_1\widetilde{X}_1 + \mathbf{v}_t\right)\left(D_1\widetilde{X}_1 + \mathbf{v}_t\right)'\right] \\ &= D_1 P_{1|0} D_1' + \Sigma_{vv} \end{aligned}$$

gives us the last component needed for the formula for K_1

$$K_1 = P_{1|0} D_1' \left(D_1 P_{1|0} D_1' + \Sigma_{vv}\right)^{-1}$$

where we know that $P_{1|0} = A_0 P_{0|0} A_0' + C_0 C_0'$.

The period 1 estimate of X

We can add the projections of X_1 on $\widetilde{Z}_1$ and $X_{0|0}$ to get our linear minimum variance estimate $X_{1|1}$

$$
\begin{aligned}
X_{1|1} &= E\left(X_1 \mid X_{0|0}\right) + E\left(X_t \mid \widetilde{Z}_1\right) \\
&= A_0 X_{0|0} + K_1 \widetilde{Z}_1
\end{aligned}
$$

Finding the covariance $P_{t|t-1}$

We also need to find an expression for $P_{t|t}$.

We can rewrite

$$X_{t|t} = K_t \widetilde{Z}_t + X_{t|t-1}$$

as

$$X_t - X_{t|t} + K_t \widetilde{Z}_t = X_t - X_{t|t-1}$$

by adding X_t to both sides and rearranging. Since the period t error $X_t - X_{t|t}$ is orthogonal to $\widetilde{Z}_t$ the variance of the right hand side must be equal to the sum of the variances of the terms on the left hand side. We thus have

$$P_{t|t} + K_t \left(D P_{t|t-1} D' + \Sigma_{vv} \right) K_t' = P_{t|t-1}$$

Finding the covariance $P_{t|t-1}$ cont'd.

We thus have

$$P_{t|t} + K_t \left(DP_{t|t-1}D' + \Sigma_{vv} \right) K_t' = P_{t|t-1}$$

or by rearranging

$$\begin{aligned} P_{t|t} &= P_{t|t-1} - K_t \left(DP_{t|t-1}D' + \Sigma_{vv} \right) K_t' \\ &= P_{t|t-1} - P_{t|t-1}D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} D_t P_{t|t-1} \end{aligned}$$

It is then straightforward to show that

$$\begin{aligned} P_{t+1|t} &= A_t P_{t|t} A_t' + CC' \\ &= A_t' \left(P_{t|t-1} - P_{t|t-1}D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv} \right)^{-1} D_t P_{t|t-1} \right) A_t' \\ &\quad + CC' \end{aligned}$$

Summing up the Kalman Filter

For the state space system

$$
\begin{aligned}
X_t &= A_t X_{t-1} + C_t \mathbf{u}_t \\
Z_t &= D_t X_t + \mathbf{v}_t \\
\begin{bmatrix} \mathbf{u}_t \\ \mathbf{v}_t \end{bmatrix} &\sim N\left(\mathbf{0}, \begin{bmatrix} I_n & \mathbf{0}_{n\times l} \\ \mathbf{0}_{l\times n} & \Sigma_{vv} \end{bmatrix}\right)
\end{aligned}
$$

we get the state estimate update equation

$$X_{t|t} = A_t X_{t-1|t-1} + K_t \left(Z_t - D_t X_{t|t-1}\right)$$

$$K_t = P_{t|t-1} D_t' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv}\right)^{-1}$$

$$
\begin{aligned}
P_{t+1|t} &= A_t \left(P_{t|t-1} - P_{t|t-1} D_{t1}' \left(D_t P_{t|t-1} D_t' + \Sigma_{vv}\right)^{-1} D_t P_{t|t-1}\right) A_t' \\
&\quad + C_{t+1} C_{t+1}'
\end{aligned}
$$

The innovation sequence can be computed recursively from the innovation representation

$$\widetilde{Z}_t = Z_t - D_t X_{t|t-1}, \quad X_{t+1|t} = A_{t-1} X_{t|t-1} + A_{t-1} K_t \widetilde{Z}_t$$

Estimating the parameters in a State Space System

Estimating the parameters in a State Space System

For a given state space system

$$\begin{aligned} X_t &= AX_{t-1} + C\mathbf{u}_t : \mathbf{u}_t \sim N(0, I) \\ Z_t &= DX_t + \mathbf{v}_t : \mathbf{v}_t \sim N(0, \Sigma_{vv}) \end{aligned}$$

How can we find the A, C, D and Σ_v that best fits the data?

The Likelihood Function of a State Space model

We can use that the innovations $\widetilde{Z}_t$ are conditionally independent Gaussian random vectors to write down the log likelihood function as

$$L(Z \mid \theta) = (-T/2)\log(2\pi) - \frac{T}{2}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T}\widetilde{Z}_t'\Omega_t^{-1}\widetilde{Z}_t$$

where

$$\begin{aligned} \widetilde{Z}_t &= Z_t - DAX_{t-1|t-1} \\ X_{t|t} &= AX_{t-1|t-1} + K_t\left(Z_t - DAX_{t-1|t-1}\right) \\ \Omega_t &= DP_{t|t-1}D' + \Sigma_{vv} \end{aligned}$$

We can start the Kalman filter recursions from the unconditional mean and variance.

But how do we find the MLE?

Numerical maximization of likelihood functions

Numerical maximization

- Grid search
- Steepest ascent
- Newton-Raphson algorithms
- Simulated annealing

Based on selected parts of Ch 5 of Hamilton and articles by Goffe, Ferrier and Rogers (1994).

Two examples:

- Unobserved components model (Grid search)
- New Keynesian DSGE (Simulated Annealing)

The basic idea

How can we estimate parameters when we cannot maximize likelihood analytically?

We need to

- Be be able to evaluate the likelihood function for a given set of parameters
- Find a way to evaluate a sequence of likelihoods conditional on different parameter vectors so that we can feel confident that we have found the parameter vector that maximizes the likelihood

Maximum Likelihood and Unobserved Components Models

Unobserved Component model of inflation

$$
\begin{aligned}
\pi_t &= \tau_t + \eta_t \\
\tau_t &= \tau_{t-1} + \varepsilon_t
\end{aligned}
$$

Decomposes inflation into permanent (τ) and transitory (η) component

- Fits the data well
  - But we may be concerned about having an actual unit root root in inflation on theoretical grounds
- Based on simplified (constant parameters) version of Stock and Watson (JMCB 2007)

The basic formulas

We want to:

1. Estimate the parameters of the system, i.e. estimate σ_η^2 and σ_ε^2
   1.1 Parameter vector is given by $\Theta = \left\{\sigma_\eta^2, \sigma_\varepsilon^2\right\}$
   1.2 $\widehat{\Theta} = \arg\max_{\theta \in \Theta} L(\pi^t \mid \mathbf{\Theta})$
2. Find an estimate of the permanent component τ_t at different points in time

The Likelihood function

We have the state space system

$$\begin{aligned} \pi_t &= \tau_t + \eta_t \text{ (measurement equation)} \\ \tau_t &= \tau_{t-1} + \varepsilon_t \text{ (state equation)} \end{aligned}$$

implying that $A = 1, D = 1, C = \sqrt{\sigma_\varepsilon^2}, \Sigma_v = \sigma_\eta^2$. The likelihood function for a state space system is (as always) given by

$$L(Z \mid \mathbf{\Theta}) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T}\widetilde{Z}_t'\Omega_t^{-1}\widetilde{Z}_t$$

where

$$\begin{aligned} \widetilde{Z}_t &= Z_t - DAX_{t-1|t-1} \\ \Omega_t &= DP_{t|t-1}D' + \Sigma_{vv} \end{aligned}$$

and n is the number of observable variables, i.e. the dimension of Z_t.

Starting the Kalman recursions

How can we choose initial values for the Kalman recursions?

- Unconditional variance is infinite because of unit root in permanent component
- A good choice is to choose "neutral" values, i.e. something akin to uninformative priors
  - One such choice is $X_{0|0} = \pi_1$ and $P_{0|0}$ very large (but finite) and constant

$$L(Z \mid \mathbf{\Theta}) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T}\widetilde{Z}_t'\Omega_t^{-1}\widetilde{Z}_t$$

Maximizing the Likelihood function

How can we find $\widehat{\Theta} = \arg\max_{\theta \in \Theta} L(\pi^t \mid \mathbf{\Theta})$?

- The dimension of the parameter vector is low so we can use grid search

Define grid for variances σ_ε^2 and σ_η^2

$$\begin{aligned}
\sigma_\varepsilon^2 &= \{0, 0.001, 0.002, ..., \sigma_{\varepsilon,\max}^2\} \\
\sigma_\eta^2 &= \{0, 0.001, 0.002, ..., \sigma_{\eta,\max}^2\}
\end{aligned}$$

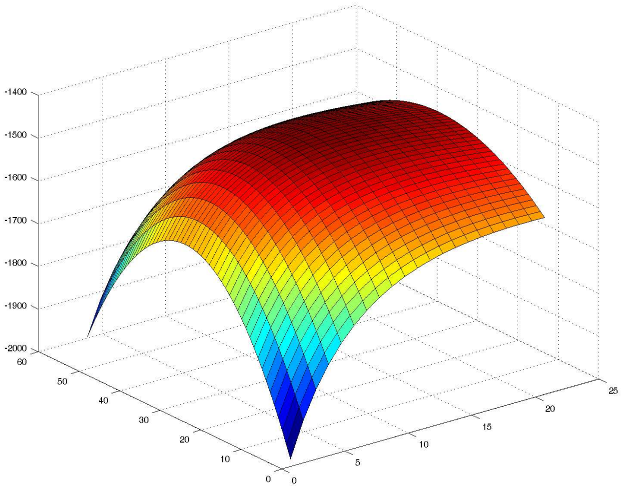
and evaluate likelihood function for all combinations.
How do we choose boundaries of grid?

- Variances are non-negative
- Both $\widehat{\sigma}_\varepsilon^2$ and $\widehat{\sigma}_\eta^2$ should be smaller than or equal to the sample variance of inflation so we can set $\sigma_{\varepsilon,\max}^2 = \sigma_{\eta,\max}^2 = \frac{1}{T}\sum \pi_t^2$

Grid Search: Fill out the x's

$\sigma_\epsilon^2 \backslash \sigma_\eta^2$	0	0.5	1	1.5	2	2.5
-1	x	x	x	x	x	x
-0.5	x	x	x	x	x	x
0	x	x	x	x	x	x
0.5	x	x	x	x	x	x
1	x	x	x	x	x	x

Maximizing the Likelihood function



Grid search

Pros:

- With a fine enough grid, grid search always finds the global maximum (if parameter space is bounded)

Cons:

- Computationally infeasible for models with large number of parameters

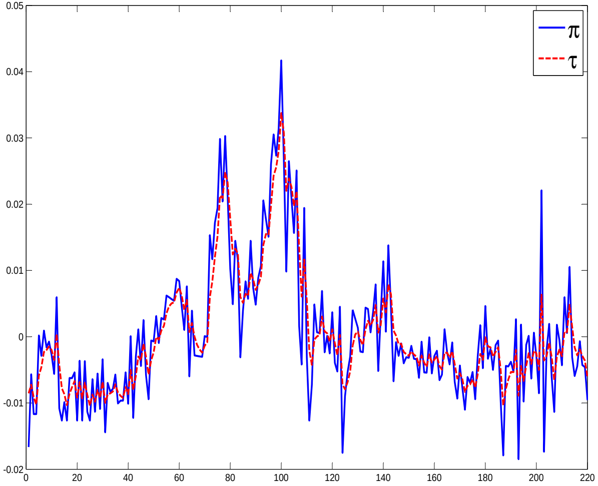
Maximizing the Likelihood function

Estimated parameter values:

- $\widehat{\sigma}_{\varepsilon}^{2} = 0.0028$
- $\widehat{\sigma}_{\eta}^{2} = 0.0051$

We can also estimate the permanent component

Actual Inflation and filtered permanent component



Maximizing the likelihood for larger models

How can we estimate parameters when we cannot maximize likelihood analytically and when grid search is not feasible?

We need to

- Be be able to evaluate the likelihood function for a given set of parameters
- Find a way to evaluate a sequence of likelihoods conditional on difference parameter vectors so that we can feel confident that we have found the parameter vector that maximizes the likelihood

Numerical maximization of likelihood functions

Estimating richer state space models

- Likelihood surface may not be well behaved

We will need more sophisticated maximization routines

Steepest Ascent method

1. Make initial guess of $\Theta = \Theta^{(0)}$
2. Find direction of "steepest ascent" by computing the gradient

$$\mathbf{g}(\Theta) \equiv \frac{\partial \mathcal{L}(Z \mid \Theta)}{\partial \Theta}$$

   which is a vector which can be approximated element by element

$$\frac{\partial \mathcal{L}(Z \mid \Theta^{(0)})}{\partial \theta_i}$$
$$\approx \quad \frac{\mathcal{L}(Z \mid \theta_j = \theta_j^{(0)} + \varepsilon : j = i; \theta_j = \theta_j^{(0)} \text{ otherwise}) - \mathcal{L}(Z \mid \Theta^{(0)})}{\varepsilon}$$

   for each θ_j in $\Theta = \{\theta_1, \theta_2, ...\theta_J\}$.

3. Take step proportional to gradient, i.e. in the direction of "steepest ascent" by setting new value of parameter vector as $\Theta^{(1)} = \Theta^{(0)} + s\mathbf{g}(\Theta)$

4. Repeat Steps 2 and 3 until convergence.

Steepest Ascent method

Pros:

- Feasible for models with a large number of parameters

Cons:

- Can be hard to calibrate even for simple models to achieve the right rate of convergence
  - Too small steps and "convergence" is achieved to soon
  - Too large step and parameters may be sent off into orbit.
- Can converge on local maximum. (How could a blind man on $K2$ find his way to Mt Everest?)

Newton-Raphson

Newton-Raphson is similar to steepest ascent, but also computes the step size

- Step size depends on second derivative
- May converge faster than steepest ascent
- Requires concavity, so is less robust when shape of likelihood function is unknown

Simulated Annealing Goffe et al (1994)

- Language is from thermodynamics
- Combines elements of grid search with (strategically chosen) random movements in the parameter space
- Has a good record in practice, but cannot be proven to reach global max quicker than grid search.

Simulated Annealing: The Algorithm

Main inputs: $\Theta^{(0)}$, temperature T, boundaries of Θ, temperature reduction parameter r_T (and the function to be max/minimized $f(\Theta)$).

1. $\theta_j' = \theta_j^{(0)} + r \cdot v_j$ where $r \sim U[-1,1]$ and v_i is an element of the step size vector V.
2. Evaluate $f(\Theta')$ and compare with $f(\Theta^{(0)})$. If $f(\Theta') > f(\Theta^{(0)})$ set $\Theta^{(1)} = \Theta'$. If $f(\Theta') < f(\Theta^{(0)})$ set $\Theta^{(1)} = \Theta'$ with probability $e^{(f(\Theta')-f(\Theta^{(0)})/T}$ and $\Theta^{(1)} = \Theta^{(0)}$ with probability $1- e^{(f(\Theta')-f(\Theta^{(0)})/T}$.
3. After N_s loops through 1 and 2 step length vector V is adjusted in direction so that approx 50% of all moves are accepted.
4. After N_T loops through 1 and 3 temperature is reduced so that $T' = r_T \cdot T$ so that fewer downhill steps are accepted.

Estimating a DSGE model using Simulated Annealing

Estimating a DSGE model using Simulated Annealing

Remember our benchmark NK model:

$$\begin{aligned}
x_t &= \rho x_{t-1} + u_t^x \\
y_t &= E_t(y_{t+1}) - \frac{1}{\gamma}\left[r_t - E_t\left(\pi_{t+1}\right)\right] + u_t^y \\
\pi_t &= E_t\left(\pi_{t+1}\right) + \kappa\left[y_t - x_t\right] + u_t^\pi \\
r_t &= \phi \pi_t + u_t^r
\end{aligned}$$

Estimating a DSGE model using Simulated Annealing

The solved model can be put in state space form

$$\begin{aligned} X_t &= AX_{t-1} + Cu_t \\ Z_t &= DX_t + v_t \end{aligned}$$

where

$$\begin{aligned} X_t &= x_t, A = \rho, Cu_t = u_t^x \\ Z_t &= \begin{bmatrix} r_t \\ \pi_t \\ y_t \end{bmatrix}, D = \begin{bmatrix} \phi\kappa\gamma\frac{1-\rho}{-c} \\ \kappa\gamma\frac{1-\rho}{-c} \\ -\kappa\frac{\phi-\rho}{-c} \end{bmatrix}, v_t = R \begin{bmatrix} u_t^r \\ u_t^\pi \\ u_t^y \end{bmatrix} \end{aligned}$$

where $c = \gamma - \kappa\rho - 2\gamma\rho + \kappa\phi + \gamma\rho^2 < 0$

We want to estimate the parameters $\theta = \{\rho, \gamma, \kappa, \phi, \sigma_x, \sigma_y, \sigma_\pi, \sigma_r\}$

The log likelihood function of a state space system

For a given state space system

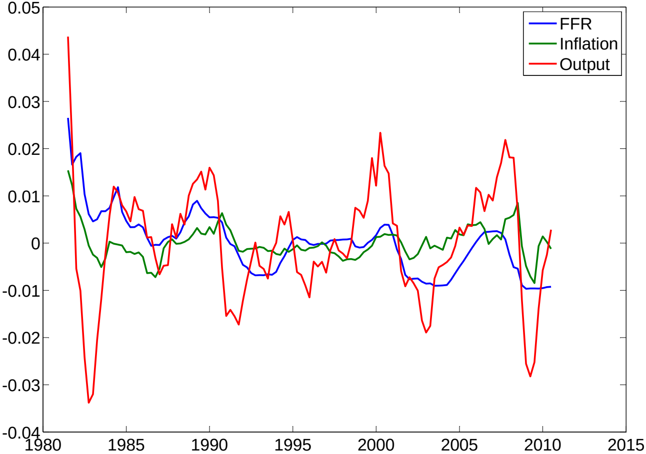
$$\begin{aligned} X_t &= AX_{t-1} + C\mathbf{u}_t \\ \underset{(p\times 1)}{Z_t} &= DX_t + \mathbf{v}_t \end{aligned}$$

we can evaluate the log likelihood by computing

$$\mathcal{L}(Z \mid \Theta) = -.5 \sum_{t=0}^{T} \left[p \ln(2\pi) + \ln |\Omega_t| + \widetilde{Z}_t' \Omega_t^{-1} \widetilde{Z}_t \right]$$

where $\widetilde{Z}_t$ are the innovation from the Kalman filter

The data



FFR
Inflation
Output
0.05
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
1980
1985
1990
1995
2000
2005
2010
2015

Code has three components

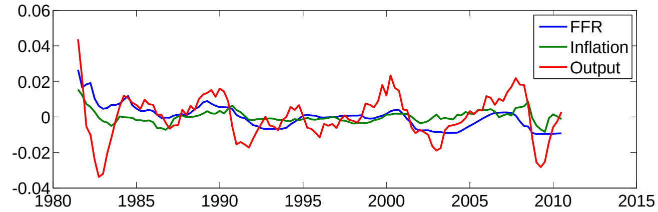
1. The main program that defines starting values for simulated annealing algorithm etc
2. A function that translates Θ into a state space system
3. A function that evaluates $\mathcal{L}(Z \mid \Theta)$

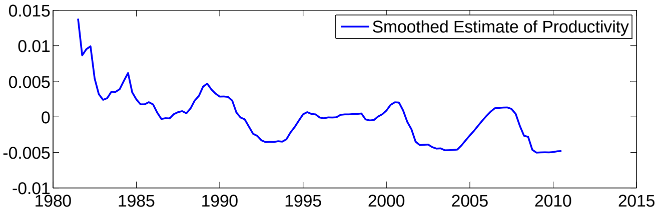
Point 2 and 3 are both done by `LLDSGE.m`

```
% Set up and estimate miniature DSGE model
clc
clear all
close all
global Z
load('Z');

r=0.95; %productivity persistence
g=5; %relative risk aversion
d=0.75; %Calvo parameter
b=0.99;   %discount factor
k=((1-d)*(1-d*b))/d; %slope of Phillips curve
f=1.5;% coefficient on inflation in Taylor rule
sigx=0.1;% s.d. prod shock
sigy=0.11;% s.d. demand shock
sigp=0.1;% s.d. cost push shock
sigr=0.1;% s.d. monetary policy shock

theta=[r,g,d,b,f,sigx,sigy,sigp,sigr]';%Starting value for paramter vector
LB=[0,1,0,0,1,zeros(1,4);]'; UB=[1,10,1,1,5,1*ones(1,4);]';
x=theta;

sa_t= 5; sa_rt=.3; sa_nt=5; sa_ns=5;

[xhat]=simannb( 'LLDSGE', x, LB, UB, sa_t, sa_rt, sa_nt, sa_ns, 1);
```

```
initial loss function value:
 -706.3706

No. of evaluations
    46

  current temperature
     5

current optimum function value
 -840.2525

No. of downhill steps
    13

No. of accepted uphill steps
    10

No. of rejections
    22

current optimum vector
    0.1338
    8.9575
    0.5270
    0.2829
    1.5000
```

```
Elapsed time is 15.624657 seconds.
No. of evaluations
        3376
  current temperature
  2.3915e-007
current optimum function value
 -1.7554e+003
No. of downhill steps
    67
No. of accepted uphill steps
    32
No. of rejections
   126
current optimum vector
    0.8964
    1.5185
    0.9399
    0.8396
    1.9204
    0.0009
    0.0031
    0.0000
    0.0123
Elapsed time is 15.835753 seconds.
simulated annealing achieved termination after 3376 evals
optimum function value
 -1.7554e+003
```


0.06
0.04
0.02
0
-0.02
-0.04
1980
1985
1990
1995
2000
2005
2010
2015
FFR
Inflation
Output


0.015
0.01
0.005
0
-0.005
-0.01
1980
1985
1990
1995
2000
2005
2010
2015
Smoothed Estimate of Productivity

Summing up

We can view any DSGE model as a function

- Input: Vector of parameters θ
- Output: A state space system

The Kalman filter can be used to

- Estimate latent variables in state space system
- Evaluate the likelihood function for given parameterized state space system