Linear state-space models, the Kalman filter, and MLE estimation of linearized DSGE models

Constantino Hevia
UTDT

August 22, 2024

Linear state space model

$$
\begin{aligned}
x_{t+1} &= Ax_t + Cw_{t+1} \qquad (1)\\
y_t &= Gx_t + v_t \qquad (2)
\end{aligned}
$$

- (1) is the state equation and (2) is the observation equation.
- x_t is an $n \times 1$ vector of state variables (possibly unobserved).
- y_t is an $m \times 1$ vector of observed variables that depend on the state and random noise v_t.
- $w_{t+1} \sim N(0, I)$ is a $p \times 1$ vector of i.i.d. random variables
- $v_t \sim N(0, R)$ is a $m \times 1$ vector of i.i.d. random variables. We can interpret v_t as "measurement" error. In other applications, v_t are expectational errors.
- shocks w_{t+1} and v_t are orthogonal an satisfy $E[v_t v_{t-j}] = 0$ and $E[w_t w_{t-j}] = 0$ for all $j \neq 0$.
- A is $n \times n$; C is $n \times p$; G is $m \times n$ and R is $m \times m$.

Example 1: MA(1) process in state-space form

$$x_{t+1} = Ax_t + Cw_{t+1}$$
$$y_t = Gx_t + v_t$$

MA(1) stochastic process:

$$z_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

- State variables:

$$x_t = \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix}$$

- Observed variable:

$$y_t = z_t - \mu = \epsilon_t + \theta\epsilon_{t+1}$$

- State equation:

$$\begin{bmatrix} \epsilon_{t+1} \\ \epsilon_t \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_{t+1} \\ 0 \end{bmatrix}$$

- Observation equation

$$y_t = \begin{bmatrix} 1 & \theta \end{bmatrix} \begin{bmatrix} \epsilon_t \\ \epsilon_{t-1} \end{bmatrix}$$

Example 2: ARMA(1,1) in state-space form

$$x_{t+1} = Ax_t + Cw_{t+1}$$
$$y_t = Gx_t + v_t$$

ARMA(1,1) stochastic process:

$$z_{t+1} - \mu = \rho(z_t - \mu) + \epsilon_{t+1} + \theta\epsilon_t$$

- State variables:

$$x_t = \begin{bmatrix} z_t - \mu \\ \epsilon_t \end{bmatrix}$$

- Observed variable:

$$y_t = z_t - \mu$$

- State equation:

$$\begin{bmatrix} z_{t+1} - \mu \\ \epsilon_{t+1} \end{bmatrix} = \begin{bmatrix} \rho & \theta \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_t - \mu \\ \epsilon_t \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \end{bmatrix} \epsilon_{t+1}$$

- Observation equation

$$y_t = \begin{bmatrix} 1 & 0 \end{bmatrix} x_t$$

Example 3: Linearized DSGE model

- Linearized DSGE models have a state-space representation

$$
\begin{aligned}
x_{t+1} =& A(\theta)x_t + C(\theta)w_{t+1} \\
y_t =& G(\theta)x_t
\end{aligned}
$$

  the matrices $A(\theta)$, $C(\theta)$ and $G(\theta)$ depend on a $k \times 1$ vector θ of deep parameters of the model.

- We often add measurement errors to avoid stochastic singularity (more on this later)

$$
\begin{aligned}
x_{t+1} =& A(\theta)x_t + C(\theta)w_{t+1} \\
y_t =& G(\theta)x_t + v_t
\end{aligned}
$$

  where $v_t \sim N(0, R)$ is a measurement error.

Example 4: Non-zero mean process

Stochastic process is

$$
\begin{aligned}
x_{t+1} &= d + Ax_t + Cw_{t+1} \\
y_t &= e + Gx_t + v_t
\end{aligned}
$$

where d and e are constant vectors.

- Suppose that x_t is covariance stationary.

- We can write this system as in (1) and (2) as follows.
  1. Note that $\bar{x} = E(x_t) = (I - A)^{-1}d$ and $\bar{y} = e + G\bar{x}$.
  2. Define the demeaned variables $\tilde{x}_t = x_t - \bar{x}$ and $\tilde{y}_t = y_t - \bar{y}$.
  3. Then,

$$
\begin{aligned}
\tilde{x}_{t+1} &= A\tilde{x}_t + Cw_{t+1} \\
\tilde{y}_t &= G\tilde{x}_t + v_t.
\end{aligned}
$$

  which has the same form as (1) and (2).

Example 5: Time-varying parameters

The Kalman filter can be adapted to more general models of the form

$$\begin{aligned} x_{t+1} &= A(z_t)x_t + C(z_t)w_{t+1} \\ y_t &= e(z_t) + G(z_t)'x_t + v_t \end{aligned}$$

where

$$v_t \sim N(0, R(z_t)).$$

and z_t may contain:

- exogenous variables
- lagged dependent variables $y_{t-1}, y_{t-2}\ldots$

Example 6: Estimate the ex-ante real interest rate (Fama and Gibbons)

Want to estimate the ex-ante (unobserved) real interest rate

$$r_t^e = i_t - \pi_t^e$$

r_t^e is the expected real interest rate between $t-1$ and t, i_t is the nominal rate between $t-1$ and t (known at $t-1$) and π_t^e is the expected inflation between $t-1$ and t.

- **Unobserved state:** demeaned ex-ante real rate

$$x_t = i_t - \pi_t^e - \mu$$

  μ is the average real interest rate.

- **Evolution of the state x_t:** ex-ante real rate follows the AR(1) process

$$x_{t+1} = \phi x_t + w_{t+1}$$

Example 6: Estimate the ex-ante real interest rate (Fama and Gibbons)

- We observe the ex-post real interest rate $r_t = i_t - \pi_t$, where π_t is realized inflation
- Ex-post real rate can be written as

$$i_t - \pi_t = (i_t - \pi_t^e) + (\pi_t^e - \pi_t) = \mu + x_t + v_t$$

  where $v_t = \pi_t^e - \pi_t$ is the inflation forecast error.
- Rational expectations: v_t is white noise and orthogonal to any variable known at $t-1$.
- **Observation equation:** demeaned ex-post real rate:

$$y_t = i_t - \pi_t - \mu.$$

- **State-space model**:

$$\begin{aligned} x_{t+1} &= \phi x_t + w_{t+1} \\ y_t &= x_t + v_t. \end{aligned}$$

Other examples

- **Common factors**: let x_t be a univariate process and $y_t \in R^m$ be an m vector of observed variables,

$$\begin{aligned} x_{t+1} =& \alpha x_t + w_{t+1} \\ y_t =& G x_t + v_t \\ v_t =& D v_{t-1} + \eta_t \end{aligned}$$

  $G \in R^m$ captures common components and v_t capture idiosyncratic dynamics.

- **Potential output and output gap**: Let x_t be the unobserved potential output and y_t be observed output. Then

$$\begin{aligned} x_{t+1} =& \alpha x_t + w_{t+1} \\ y_t =& x_t + v_t \end{aligned}$$

  v_t is the "output gap".

- **VARs with time-varying coefficients**: Kalman filter can be applied to models of the form

$$\begin{aligned} A_t =& A_{t-1} + w_t \\ y_t =& A_t y_{t-1} + v_t \end{aligned}$$

  This is a special case of Example 5: model with time-varying parameters.

Filering problem

State space model

$$x_{t+1} = Ax_t + Cw_{t+1}$$
$$y_t = Gx_t + v_t.$$

- **Problem**: Estimate the probability distribution of the unobserved state x_t conditional on the history of observations $Y^t = \{y_1, y_2, \ldots y_t\}$.
- **Solution**: Use the Kalman filter
  - Recursive algorithm for estimating the unobserved state.
  - Kalman filter is composed of two steps:
    - Prediction step
    - Updating step
- As a byproduct, the Kalman filter delivers the likelihood of the data

$$L(\theta; Y^T) = f(y_0, y_1, y_2, ..., y_T | \theta)$$

- If shocks are non-normal, the Kalman filter is the *best linear predictor* and can be used to perform quasi-maximum likelihood estimation.

Notation

- **Filering:**

$$x_{t|t} = E\left[x_t \,\middle|\, Y^t\right]$$

$$P_{t|t} = E\left[\left(x_t - x_{t|t}\right)\left(x_t - x_{t|t}\right)' \,\middle|\, Y^t\right]$$

- **Predicting:**

$$x_{t+1|t} = E\left[x_{t+1} \,\middle|\, Y^t\right]$$

$$P_{t+1|t} = E\left[\left(x_{t+1} - x_{t+1|t}\right)\left(x_{t+1} - x_{t+1|t}\right)' \,\middle|\, Y^t\right]$$

$$y_{t+1|t} = E\left[y_{t+1} \,\middle|\, Y^t\right]$$

$$\Omega_{t+1|t} = E\left[\left(y_{t+1} - y_{t+1|t}\right)\left(y_{t+1} - y_{t+1|t}\right)' \,\middle|\, Y^t\right]$$

- **Smoothing:**

$$x_{t|T} = E\left[x_t \,\middle|\, Y^T\right]$$

$$P_{t|T} = E\left[\left(x_t - x_{t|T}\right)\left(x_t - x_{t|T}\right)' \,\middle|\, Y^T\right]$$

Recursive projection formula

We will use the formula for updating a linear projection when new information arrives (first set of notes)

- We have a linear projection of x on Y, $P[x|Y]$
- Information z arrives and we want to compute the projection of x on $\{Y, z\}$, $P[x|Y, z]$
- The projection on the larger information set satisfies the following formula

$$P[x|Y, z] = P[x|Y] + P[(x - P[x|Y]) \,|\, (z - P[z|Y])] \tag{3}$$

  where
  - $P[x|Y]$: Original projection.
  - $P[(x - P[x|Y]) \,|\, (z - P[z|Y])]$: Projection of forecast errors on forecast errors.

Kalman filter: Prediction step (I)

- We start with filtered estimates $x_{t-1|t-1}$ and $P_{t-1|t-1}$
- Objective is to obtain the prediction moments

$$
\begin{aligned}
x_{t|t-1} &= E\left[x_t \middle| Y^{t-1}\right] \\
P_{t|t-1} &= E\left[\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' \middle| Y^{t-1}\right] \\
y_{t|t-1} &= E\left[y_t \middle| Y^{t-1}\right] \\
\Omega_{t|t-1} &= E\left[\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)' \middle| Y^{t-1}\right]
\end{aligned}
$$

- Use the state equation

$$x_t = Ax_{t-1} + Cw_t$$

- Taking expectations conditional on Y^{t-1} gives

$$E\left[x_t \middle| Y^{t-1}\right] = AE\left[x_{t-1} \middle| Y^{t-1}\right] + CE\left[w_t \middle| Y^{t-1}\right]$$

or

$$x_{t|t-1} = Ax_{t-1|t-1}.$$

Kalman filter: Prediction step (II)

- Mean squared error of the prediction (e.g. covariance matrix) is

$$
\begin{aligned}
P_{t|t-1} =& E\left[\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' \middle| Y^{t-1}\right] \\
=& E\left[\left(x_t - Ax_{t-1|t-1}\right)\left(x_t - Ax_{t-1|t-1}\right)' \middle| Y^{t-1}\right] \\
=& E\left[\left(Ax_{t-1} + Cw_t - Ax_{t-1|t-1}\right)\left(Ax_{t-1} + Cw_t - Ax_{t-1|t-1}\right)' \middle| Y^{t-1}\right] \\
=& E\left[\left(A\left(x_{t-1} - x_{t-1|t-1}\right) + Cw_t\right)\left(A\left(x_{t-1} - x_{t-1|t-1}\right) + Cw_t\right)' \middle| Y^{t-1}\right] \\
=& AE\left[\left(x_{t-1} - x_{t-1|t-1}\right)\left(x_{t-1} - x_{t-1|t-1}\right)' \middle| Y^{t-1}\right] A' + CE\left[w_t w_t' \middle| Y^{t-1}\right] C'
\end{aligned}
$$

or

$$P_{t|t-1} = AP_{t-1|t-1}A' + CC'$$

Kalman filter: Prediction step (III)

- We now compute the prediction moments of the observable variable $y_{t|t-1}$ and $\Omega_{t|t-1}$.
- Using the observation equation (2),

$$y_t = Gx_t + v_t$$

gives

$$y_{t|t-1} = Gx_{t|t-1}$$

and

$$\begin{aligned}
\Omega_{t|t-1} =& E\left[\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)' \middle| Y^{t-1}\right] \\
=& E\left[\left(Gx_t + v_t - y_{t|t-1}\right)\left(Gx_t + v_t - y_{t|t-1}\right)' \middle| Y^{t-1}\right] \\
=& E\left[\left(G(x_t - x_{t|t-1}) + v_t\right)\left(G(x_t - x_{t|t-1}) + v_t\right)' \middle| Y^{t-1}\right] \\
=& GE\left[(x_t - x_{t|t-1})(x_t - x_{t|t-1})' \middle| Y^{t-1}\right] G' + E\left[v_t v_t' \middle| Y^{t-1}\right]
\end{aligned}$$

or

$$\Omega_{t|t-1} = GP_{t|t-1}G' + R$$

Kalman filter: Prediction step (IV)

Summarizing, given filtered estimates

$$x_{t-1|t-1} \text{ and } P_{t-1|t-1},$$

the prediction step is given by the following equations

- **State equation**:

$$x_{t|t-1} = Ax_{t-1|t-1} \tag{4}$$
$$P_{t|t-1} = AP_{t-1|t-1}A' + CC' \tag{5}$$

- **Observation equation**:

$$y_{t|t-1} = Gx_{t-1|t-1} \tag{6}$$
$$\Omega_{t|t-1} = GP_{t-1|t-1}G' + R \tag{7}$$

Kalman filter: Updating step (I)

- We have the prediction estimates $x_{t|t-1}$, $P_{t|t-1}$, $y_{t|t-1}$, and $\Omega_{t|t-1}$ and observe the (new) data y_t,
- Objective is to obtain the filtered (updated) moments

$$\begin{aligned} x_{t|t} =& E\left[x_t \middle| Y^t\right] \\ P_{t|t} =& E\left[\left(x_t - x_{t|t}\right)\left(x_t - x_{t|t}\right)' \middle| Y^t\right] \end{aligned}$$

- Linear model implies that $E\left[x_t \middle| Y^t\right]$ is the same as the linear projection of x_t on Y^t
- Using the recursive projection formula (3) with the information set $\left\{Y^{t-1}, y_t\right\}$ gives

$$x_{t|t} = x_{t|t-1} + P\left[x_t - x_{t|t-1} \middle| y_t - y_{t|t-1}\right]$$

- We thus need to compute the linear projection (the $n \times m$ matrix K_t)

$$P\left[x_t - x_{t|t-1} \middle| y_t - y_{t|t-1}\right] = K_t\left(y_t - y_{t|t-1}\right)$$

The matrix K_t is called the "Kalman gain".

Kalman filter: Updating step (II)

By the orthogonality principle , the residuals of the projection must be orthogonal to $\left(y_t - y_{t|t-1}\right)$,

$$
\begin{aligned}
\underset{(n\times m)}{0} =& E\left[\underbrace{\left(x_t - x_{t|t-1} - K_t\left(y_t - y_{t|t-1}\right)\right)}_{\text{Residual of projection } (n\times 1)} \times \underbrace{\left(y_t - y_{t|t-1}\right)'}_{\text{Explanatory vars } (1\times m)}\right] \\
=& E\left[\left(x_t - x_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right] - K_t \underbrace{E\left[\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right]}_{=\Omega_{t|t-1}} \\
=& E\left[\left(x_t - x_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right] - K_t\Omega_{t|t-1} \\
=& E\left[\left(x_t - x_{t|t-1}\right)\left(\left(x_t - x_{t|t-1}\right)' G' + v_t'\right)\right] - K_t\Omega_{t|t-1} \\
=& \underbrace{E\left[\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)'\right]}_{=P_{t|t-1}} G' - K_t\Omega_{t|t-1} \\
=& P_{t|t-1} G' - K_t\Omega_{t|t-1}
\end{aligned}
$$

so that

$$K_t = P_{t|t-1} G' \Omega_{t|t-1}^{-1}$$

Kalman filter: Updating step (III)

Thus, the updating step is

$$x_{t|t} = x_{t|t-1} + K_t \left(y_t - y_{t|t-1} \right)$$

The associated covariance matrix is

$$\begin{aligned}
P_{t|t} &= E\left[\left(x_t - x_{t|t} \right) \left(x_t - x_{t|t} \right)' \right] \\
&= E\left[\left(x_t - x_{t|t-1} - K_t \left(y_t - y_{t|t-1} \right) \right) \left(x_t - x_{t|t-1} - K_t \left(y_t - y_{t|t-1} \right) \right)' \right] \\
&= E\left[\left(x_t - x_{t|t-1} \right) \left(x_t - x_{t|t-1} \right)' \right] - E\left[\left(x_t - x_{t|t-1} \right) \left(y_t - y_{t|t-1} \right)' \right] K_t' \\
&\quad -K_t E\left[\left(y_t - y_{t|t-1} \right) \left(x_t - x_{t|t-1} \right)' \right] + K_t E\left[\left(y_t - y_{t|t-1} \right) \left(y_t - y_{t|t-1} \right)' \right] K_t' \\
&= P_{t|t-1} - K_t \Omega_{t|t-1} K_t' - K_t \Omega_{t|t-1} K_t' + K_t \Omega_{t|t-1} K_t' \\
&= P_{t|t-1} - K_t \Omega_{t|t-1} K_t' \\
&= P_{t|t-1} - K_t \Omega_{t|t-1} \Omega_{t|t-1}^{-1} G P_{t|t-1} \\
&= P_{t|t-1} - K_t G P_{t|t-1} \\
&= \left(I - K_t G \right) P_{t|t-1}
\end{aligned}$$

where, in the fourth equality we used the result from the previous slide (see third equality), and in the sixth equality we used that $K_t' = \Omega_{t|t-1}^{-1} G P_{t|t-1}$.

Kalman filter: Updating step (IV)

Summarizing, given the prediction estimates

$$x_{t|t-1},\ P_{t|t-1},\ y_{t|t-1},\ \Omega_{t|t-1}$$

the updating step is given by the following equations

$$x_{t|t} = x_{t|t-1} + K_t \left(y_t - y_{t|t-1} \right) \tag{8}$$

$$P_{t|t} = (I - K_t G)\, P_{t|t-1}. \tag{9}$$

where the Kalman gain K_t is given by

$$K_t = P_{t|t-1} G' \Omega_{t|t-1}^{-1}. \tag{10}$$

Kalman filter: Initialization of the filter (I)

- So far we derived an algorithm for computing updated and prediction estimates given previous estimates.
- We now discuss how to initialize the filter.
- More than one way to do it. Here is one when x_t is covariance stationary.
- We need starting values for $x_{1|0}$ and $P_{1|0}$:
- $x_{1|0}$ is the forecast of x_1 based on no observations on y_t or x_t.
- We just set $x_{1|0}$ to the unconditional mean of x_1:

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

with mean squared error

$$P_{1|0} = E\left[\left(x_1 - x_{1|0}\right)\left(x_1 - x_{1|0}\right)'\right]$$

Kalman filter: Initialization of the filter (II)

- The state equation is $x_{t+1} = Ax_t + Cw_{t+1}$. If A is a stable matrix, then

$$E\left[x_{t+1}\right] = AE\left[x_t\right] \Rightarrow (I - A)E\left[x_t\right] = 0$$

  Since 1 is not an eigenvalue of A, $I - A$ is non-singular and the equation has a unique solution

$$E\left[x_t\right] = 0 \Rightarrow x_{1|0} = E\left[x_1\right] = 0.$$

- We now want a starting value for $P_{1|0}$.

$$P_{1|0} = E\left[\left(x_1 - x_{1|0}\right)\left(x_1 - x_{1|0}\right)'\right] = E\left[x_1 x_1'\right]$$

- Using $x_1 = Ax_0 + Cw_1$ we have

$$\begin{aligned} E\left[x_1 x_1'\right] &= E\left[\left(Ax_0 + Cw_1\right)\left(Ax_0 + Cw_1\right)'\right] \\ &= AE\left[x_0 x_0'\right]A' + CE\left[w_1 w_1'\right]C' \end{aligned}$$

Kalman filter: Initialization of the filter (III)

- Using stationarity (so that $P = E\left[x_t x_t'\right]$ for all t) and $E\left[w_t w_t'\right] = I$ gives

$$P = APA' + CC'.$$

- Set $P_{1|0} = P$, where P solves the previous equation.
- The solution can be expressed as

$$vec(P) = \left[I_{n^2} - A \otimes A\right]^{-1} vec(CC'). \tag{11}$$

- Given the initial values $x_{1|0}$ and $P_{0|0} = P_{1|0} = P$, can compute the prediction moments

$$\begin{aligned} y_{1|0} &= Ax_{1|0} \\ \Omega_{1|0} &= GP_{0|0}G' + R. \end{aligned}$$

Summary of the Kalman filter

1. **Initialization:** Let P solve (11). Initialize the filter using

$$\begin{aligned} x_{1|0} &= \mathbf{0}_{n\times 1}; \quad P_{1|0} = P \\ y_{1|0} &= Ax_{1|0}; \quad \Omega_{1|0} = GPG' + R. \end{aligned}$$

2. **Filtering step.** For $t = 1, 2, ..., T$, compute the updated moments

$$x_{t|t} = x_{t|t-1} + K_t \left(y_t - y_{t|t-1} \right)$$

$$P_{t|t} = (I - K_t G)\, P_{t|t-1}.$$

   where

$$K_t = P_{t|t-1} G' \Omega_{t|t-1}^{-1}$$

3. **Prediction step.** Given filtered moments $x_{t|t}$ and $P_{t|t}$, compute the prediction moments

$$\begin{aligned} x_{t+1|t} &= Ax_{t|t} \\ P_{t+1|t} &= AP_{t|t}A' + CC' \\ y_{t+1|t} &= Gx_{t+1|t} \\ \Omega_{t+1|t} &= GP_{t+1|t}G' + R. \end{aligned}$$

   Set $t \rightarrow t+1$ and return to step 2.

More general state-space models

- The Kalman filter can be applied to more general state space models:
  - Models with time-varying parameters and exogenous variables

$$\begin{aligned} x_{t+1} &= A_t x_t + B_t z_t + C_t w_{t+1} \\ y_t &= G_t x_t + H_t z_t + v_t \end{aligned}$$

    where A_t, B_t, C_t, G_t, and H_t are known matrices function of time; and z_t are exogenous variables.
  - Models with correlated noises

$$E\left[w_{t+1} v_t'\right] = \Psi \neq 0$$

  - Models with serially correlated measurement errors

$$v_t = D v_{t-1} + \eta_t.$$

- It is possible to use the Kalman filter on those models using appropriate transformations (See Anderson and Moore, Hansen and Sargent, and Hamilton for details.)

Recursive evaluation of the likelihood function (I)

- The Kalman filter enables a recursive algorithm to evaluate a Gaussian likelihood function for the observations y_t, $t = 1, 2, ..., T$ on the $m \times 1$ vector y_t.
- Use the factorization (prediction-error decomposition) of the likelihood function

$$\begin{aligned} f(y_T, y_{T-1}, ..., y_1) \quad &= \quad f_T(y_T | y_{T-1}, y_{T-2}, ..., y_1) \times f_{T-1}(y_{T-1} | y_{T-2}, y_{T-3}, ..., y_1) \\ & \quad\quad \ldots f_1(y_2 | y_1) \times f(y_1). \end{aligned}$$

- The state-space model with normal errors (1) and (2) implies that the conditional density $f(y_t | y_{t-1}, y_{t-2}, ..., y_1)$ is normal with mean

$$E\left(y_t | Y^{t-1}\right) = y_{t|t-1}$$

and covariance matrix

$$E\left[\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right] = \Omega_{t|t-1}.$$

- The Kalman filter delivers the quantities $y_{t|t-1}$ and $\Omega_{t|t-1}$.

Recursive evaluation of the likelihood function (II)

- Then

$$f(y_t|Y^{t-1}) = (2\pi)^{-\frac{m}{2}} \det\left(\Omega_{t|t-1}\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(y_t - y_{t|t-1}\right)' \Omega_{t|t-1}^{-1}\left(y_t - y_{t|t-1}\right)\right).$$

- Letting θ summarize all the relevant parameters, the log-likelihood function is then given by

$$\mathcal{L}(\theta|Y^T) = \sum_{t=1}^{T}\left(-\frac{m}{2}\log(2\pi) - \frac{1}{2}\log\det(\Omega_{t|t-1}) - \frac{1}{2}\left(y_t - y_{t|t-1}\right)' \Omega_{t|t-1}^{-1}\left(y_t - y_{t|t-1}\right)\right)$$

or

$$\mathcal{L}(\theta|Y^T) = -\frac{1}{2}\sum_{t=1}^{T}\left(\log\det(\Omega_{t|t-1}) + \left(y_t - y_{t|t-1}\right)' \Omega_{t|t-1}^{-1}\left(y_t - y_{t|t-1}\right)\right) - \frac{1}{2}Tm\log(2\pi).$$

Maximization of the likelihood function

- Suppose the true parameter is θ_0. Pick an initial guess θ.
- Compute the log-likelihood function and maximize it with respect to the parameter vector θ.
    - Each time you try a new guess θ, you need to solve the model and run the Kalman filter to evaluate the likelihood function.
- Maximization step is done numerically. In some problems (DSGE) this step is tough.
- Obtain standard errors from the Hessian of $\mathcal{L}(\theta|Y^T)$ using the asymptotic result

$$\hat{\theta} \approx N\left(\theta_0, -\left[\frac{\partial^2 \mathcal{L}(\theta)}{\partial\theta\partial\theta'}\bigg|_{\theta=\hat{\theta}}\right]^{-1}\right).$$

- Standard errors are the square root of the diagonal elements of the negative of the Hessian.
- Most numerical software performs minimization rather than maximization. We thus minimize the negative of the log-likelihood function.

Maximization of the likelihood function: standard errors

- Several ways of computing standard errors, all asymptotically equivalent if the model is correctly specified:
  - Inverse of Hessian of log-likelihood (what we did above):

$$Var(\hat{\theta}) = \left(-E\left[\frac{\partial^2 \mathcal{L}}{\partial\theta\partial\theta'}\right]\right)^{-1};$$

  - Outer product of the gradient of the log-likelihood:

$$Var(\hat{\theta}) = \left(E\left[\frac{\partial \mathcal{L}}{\partial\theta}\frac{\partial \mathcal{L}'}{\partial\theta}\right]\right)^{-1};$$

  - A combination of both:

$$Var(\hat{\theta}) = E\left[\left[\frac{-\partial^2 \mathcal{L}}{\partial\theta\partial\theta'}\right]^{-1}\frac{\partial \mathcal{L}}{\partial\theta}\frac{\partial \mathcal{L}'}{\partial\theta}\left[\frac{-\partial^2 \mathcal{L}}{\partial\theta\partial\theta'}\right]^{-1}\right]$$

- Practical issue: sometimes the inverse of the Hessian gives an imaginary number. In that case use the outer product of the gradient.

Issues in the maximization step

- Which optimizer to use?
  - Most (all) optimizers are only able to find local optima. Some are more robust than others for finding the global optimum.

- Typical local methods.
  1. Newton-Raphson and Quasi Newton-Raphson. Quadratic approximation of the likelihood function. Need to compute (numerical) derivatives and is good near the optimum. In Matlab, performed by the function `fminunc.m`.
  2. Methods without derivatives. Simplex method (Nelder-Mead). Robust method but slow to converge. Also, since it does not use derivatives, the gradient of the likelihood need not be close to zero.

- Methods that (try to) avoid getting stuck in local optima
  - Simulated annealing: random search algorithm. More robust than local methods but very slow to converge near the optimum.
  - Genetic algorithms.

Issues in the maximization step

- It is good practice to try several optimizers, particularly when there are many parameters to estimate.

- One possibility (my approach):
  - If we don't have a good initial guess, start with simulated annealing from different initial guesses. Keep best estimate to use as starting value for other optimization methods.
  - Switch to a faster, local search algorithm.
    - If problem is well behaved, use a Newton Raphson method.
    - If not well behaved, stick to simplex until objective function does not increase.
    - After that run a Newton-Rapshon method.

- If hard to solve, mix all optimization methods always keeping the best estimate so far as initial guess.

Constraints on the parameters

- Sometimes we need to impose constraints on the parameters. For example, coefficient of risk aversion cannot be negative.
- Could use constrained optimization routines. But it is usual to stick to unconstrained optimization which are more robust.
- In many cases, can perform a simple transformation of the parameter to make sure that the constraint is always satisfied.
- Suppose that the parameter θ (a scalar) is constrained in some form.
- Often we can define a new variable ϕ such that ϕ is unconstrained but $\theta = g(\phi)$ is a transformation of ϕ that satisfies the constraint.

Constraints on the parameters: examples

- In all the cases that follow θ satisfies some constraint and ϕ is an unconstrained number.

- The following are useful transformations:

Constraint	Transformation	Constraint	Transformation
$\theta > 0$	$\theta = e^{\phi}$	$\theta \in (a, b)$	$\theta = \frac{b+ae^{-\phi}}{1+e^{-\phi}}$
$\theta > 0$	$\theta = \phi^2$	$\theta \in (-1, 1)$	$\theta = \frac{e^{\phi}-e^{-\phi}}{e^{\phi}+e^{-\phi}}$
$\theta < 0$	$\theta = -e^{\phi}; \quad \theta = -\phi^2$	$\theta \in (-1, 1)$	$\theta = \frac{\phi}{1+\lvert\phi\rvert}$
$\theta \in (0, 1)$	$\theta = \frac{1}{1+e^{-\phi}}$	$\theta \in (-1, 1)$	$\theta = 2\pi \tan^{-1}(\phi)$

- This simplifies a lot the optimization. But there is a cost:
  - The Hessian delivers standard errors for the unconstrained parameters ϕ.
  - But we are interested in the standard errors of the actual parameters θ.
  - How to fix this?

Delta method: standard errors of a transformation

- Suppose $\theta \in R^k$ is the variable of interest, but we make a transformation $\theta = g(\phi) : R^k \to R^k$ that makes the problem unconstrained in the variables $\phi \in R^k$
- By usual results, $\sqrt{T}(\phi - \hat{\phi}) \to N(0, Var(\phi))$. We are interested in the standard error of $\theta = g(\phi)$. Using a first order Taylor expansion,

$$g(\phi) = g(\hat{\phi}) + \frac{\partial g(\hat{\phi})}{\partial \phi'}(\phi - \hat{\phi}) + o_P(1)$$

  where $\frac{\partial g(\hat{\phi})}{\partial \phi}$ is the $k \times k$ Jacobian matrix of the transformation $g(\phi)$.

$$\sqrt{T}(\theta - \hat{\theta}) = \frac{\partial g(\hat{\phi})}{\partial \phi'}\sqrt{T}(\phi - \hat{\phi}) + o_P(1)$$

  which implies

$$Var(\theta) = Var(g(\phi)) = \frac{\partial g(\hat{\phi})}{\partial \phi'} Var(\phi) \frac{\partial g(\hat{\phi})}{\partial \phi}$$

- In simple cases, this is done by hand. With many constraints, perhaps easier to use numerical differentiation

Fama and Gibbons: estimating the ex-ante real interest rate

- **Observation equation:** demeaned ex-post real rate

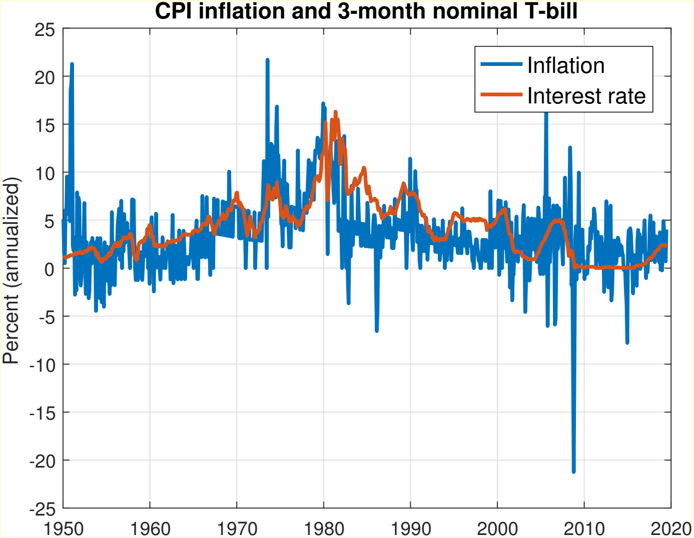
$$y_{t+1} = i_t - \pi_{t+1} - \mu.$$

- **State-space system**:

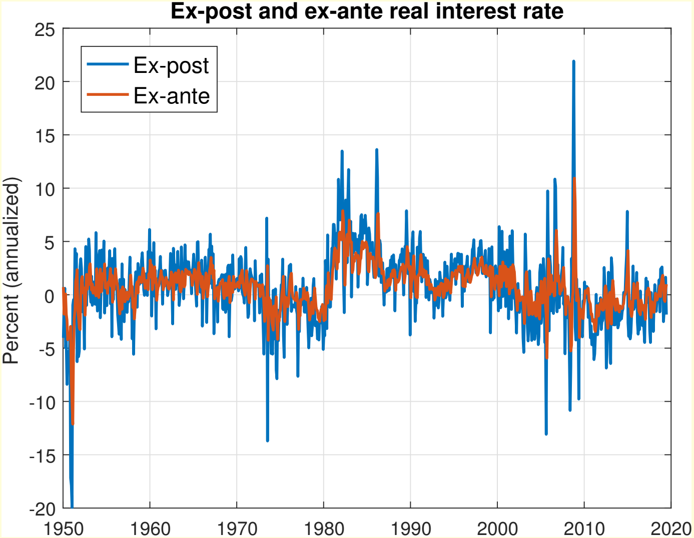
$$\begin{aligned} x_{t+1} &= \phi x_t + w_{t+1} \\ y_{t+1} &= x_t + v_{t+1} \end{aligned}$$

- US monthly data on 3-month treasury bills and CPI.

Fama and Gibbons: estimating the ex-ante real interest rate



Fama and Gibbons: estimating the ex-ante real interest rate



Dynamic Nelson and Siegel model of the yield curve

- Diebold and Li (2006) proposed the following parametric model of the yield curve (i.e. the return of nominal zero coupon bonds)

$$y_t(\tau) = L_t + S_t \left(\frac{1 - e^{\lambda\tau}}{\lambda\tau} \right) + C_t \left(\frac{1 - e^{\lambda\tau}}{\lambda\tau} - e^{\lambda\tau} \right).$$

- τ : is the maturity of the bond (in months)
- $y_t(\tau)$ is the yield (return) of a zero-coupon bond that matures in τ months
- λ is a positive parameter.
- (L_t, S_t, C_t) is a 3×1 vector of unobserved factors interpreted as "level," "slope," and "curvature," respectively.
- The unobserved factors evolve according to the state equation

$$\begin{bmatrix} L_t - \mu^L \\ S_t - \mu^S \\ C_t - \mu^C \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} L_{t-1} - \mu^L \\ S_{t-1} - \mu^S \\ C_{t-1} - \mu^C \end{bmatrix} + \begin{bmatrix} w_t^L \\ w_t^S \\ w_t^C \end{bmatrix} \tag{12}$$

where $w_t \sim N(0, Q)$.

Dynamic Nelson and Siegel model of the yield curve

- The observed yields (observation equations) are then given by

$$\begin{bmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \vdots \\ y_t(\tau_N) \end{bmatrix} = \begin{bmatrix} 1 & \left(\frac{1-e^{\lambda\tau_1}}{\lambda\tau_1}\right) & \left(\frac{1-e^{\lambda\tau_1}}{\lambda\tau_1} - e^{\lambda\tau_1}\right) \\ 1 & \left(\frac{1-e^{\lambda\tau_2}}{\lambda\tau_2}\right) & \left(\frac{1-e^{\lambda\tau_2}}{\lambda\tau_2} - e^{\lambda\tau_2}\right) \\ \vdots & \vdots & \vdots \\ 1 & \left(\frac{1-e^{\lambda\tau_N}}{\lambda\tau_N}\right) & \left(\frac{1-e^{\lambda\tau_N}}{\lambda\tau_N} - e^{\lambda\tau_N}\right) \end{bmatrix} \begin{bmatrix} L_t \\ S_t \\ C_t \end{bmatrix} + \begin{bmatrix} v_t^1 \\ v_t^2 \\ \vdots \\ v_t^N \end{bmatrix} \tag{13}$$

  where τ_1, τ_2, ..., τ_N are N observed maturities and v_t^1, v_t^2,..., v_t^N are measurement errors distributed as $v_t \sim N(0, R)$.
- If more than three observed yields, we need to add measurements errors for otherwise the model is stochastically singular.
- This model has a state space representation and, therefore can be estimated using Maximum Likelihood as described.
- See the accompanying Matlab programs.

MLE estimation of (linearized) DSGE MODELS (I)

- Express the solution of the model in state-space form.

- Write the linearized equilibrium conditions of the model as

$$A_0(\theta) E_t \begin{bmatrix} x_{t+1} \\ u_{t+1} \end{bmatrix} = B_0(\theta) \begin{bmatrix} x_t \\ u_t \end{bmatrix}$$

  x_t : state variables; u_t : control variables; and θ is a vector of structural parameters.

- The matrices $A_0(\theta)$ and $B_0(\theta)$ are functions of the structural parameters of the model.

- We usually order state variables so that endogenous state variables (e.g. capital) appear first and exogenous state variables (e.g. productivity shocks) appear second:

$$x_t = \begin{bmatrix} \text{endogenous state variables} \\ \text{exogenous state variables} \end{bmatrix}$$

- Exogenous state variables are assumed to evolve according to

$$x^2_{t+1} = \Lambda x^2_t + \eta \epsilon_{t+1}; \; \epsilon_{t+1} \sim N(0, I).$$

MLE estimation of (linearized) DSGE MODELS (II)

- The QZ decomposition gives the equilibrium of the model in the form

$$\begin{aligned} E_t\left[x_{t+1}\right] =& F(\theta)x_t \\ u_t =& P(\theta)x_t \end{aligned}$$

- Thus, the state-space representation of the DSGE model is

$$\begin{aligned} x_{t+1} =& F(\theta)x_t + \begin{bmatrix} 0 \\ \eta \end{bmatrix} \epsilon_{t+1} \\ u_t =& P(\theta)x_t. \end{aligned}$$

- First, in terms of the state-space model (1) and (2) we have

$$A(\theta) = F(\theta);\; C(\theta) = \begin{bmatrix} 0 \\ \eta \end{bmatrix}$$

so that

$$x_{t+1} = A(\theta)x_t + C(\theta)\epsilon_{t+1}.$$

MLE estimation of (linearized) DSGE MODELS (III)

- Second, to estimate the model we need a set of observable variables $y_t \in R^m$.
- Observable variables could be a subset of the control and/or state variables. In particular,

$$y_t = D(\theta)x_t,$$

  where $D(\theta)$ picks the relevant row of the matrix $P(\theta)$ for the control variables or is an indicator vector in the case of state variables.
- Example:
  - Suppose that the first observable in y_t is the *second state variable*. Then, the first row of $D(\theta)$ is $\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \end{bmatrix}$ with a 1 in the second column.
  - Suppose the third element of y_t is a control variable, say the fourth element of u_t. Then the third row of $D(\theta)$ is the fourth row of the matrix $P(\theta)$.

Stochastic singularity of DSGE models

- **Definition:** The state space representation of a DSGE model is stochastically singular if the number of exogenous shocks is smaller than the number of observable variables.

- Stochastic singularity arises because DSGE models typically use a small number of structural shocks to generate predictions about a large number of observable variables.

- Linearized models with less shocks than observable variables predict that certain linear combinations of observable variables should hold deterministically (i.e. without noise).

- It is impossible to estimate by MLE a model that is stochastically singular.

Stochastic singularity of DSGE models (II)

Example: RBC model with only productivity shocks (all variables expressed as deviations from steady state).

- The state variables are (k_t, z_t), the stock of capital and the productivity shock.
  1. Only one exogenous state variable: z_t.
  2. k_t is an endogenous state variable.

- Suppose that the observable variables are output, labor, and consumption (y_t, n_t, c_t).

- The solution of the observable variables take the form

$$\begin{bmatrix} y_t \\ n_t \\ c_t \end{bmatrix} = \begin{bmatrix} \phi_{yk} & \phi_{yz} \\ \phi_{nk} & \phi_{nz} \\ \phi_{ck} & \phi_{cz} \end{bmatrix} \begin{bmatrix} k_t \\ z_t \end{bmatrix}.$$

- Use the observation of output and labor to solve for capital and the technology shock:

$$\begin{bmatrix} k_t \\ z_t \end{bmatrix} = \begin{bmatrix} \phi_{yk} & \phi_{yz} \\ \phi_{nk} & \phi_{nz} \end{bmatrix}^{-1} \begin{bmatrix} y_t \\ n_t \end{bmatrix}$$

Stochastic singularity of DSGE models (III)

- Therefore, the row for consumption will be given by

$$c_t = \begin{bmatrix} \phi_{ck} & \phi_{cz} \end{bmatrix} \begin{bmatrix} k_t \\ z_t \end{bmatrix} = \begin{bmatrix} \phi_{ck} & \phi_{cz} \end{bmatrix} \begin{bmatrix} \phi_{yk} & \phi_{yz} \\ \phi_{nk} & \phi_{nz} \end{bmatrix}^{-1} \begin{bmatrix} y_t \\ n_t \end{bmatrix}$$

  Rearranging gives

$$(\phi_{yk}\phi_{cz} - \phi_{yz}\phi_{ck})n_t + (\phi_{nz}\phi_{ck} - \phi_{nk}\phi_{cz})y_t + (\phi_{yz}\phi_{nk} - \phi_{nz}\phi_{yk})c_t = 0$$

- There is a linear combination of the observable variables that hold without noise.

- That is, the model predicts a singular covariance matrix of (y_t, n_t, c_t) for any sample size and parameter values.

- Needless to say, this does not hold in the data.

- The likelihood function cannot be evaluated.

Stochastic singularity of DSGE models (IV)

- We now discuss the general problem of stochastic singularity and show that the likelihood function cannot be evaluated.

- Let $x_t \in R^n$ be the set of state variables. Suppose that $\tilde{n} \leq n$ of the elements of x_t correspond to exogenous (stochastic) state variables (there are $\tilde{n}$ exogenous shocks) and the remaning $n - \tilde{n}$ are state variables whose time t values are known with certainty at time $t - 1$.

- Let $y_t \in R^m$ be the set of observable variables that satisfy

$$y_t = Dx_t$$

  where D is an $m \times n$ matrix.

Stochastic singularity of DSGE models (V)

- We apply the Kalman filter and compute the covariances

$$
\begin{aligned}
P_{t|t-1} &= E\left[\left(x_t - x_{t|t-1}\right)\left(x_t - x_{t|t-1}\right)' \middle| Y^{t-1}\right] \\
\Omega_{t|t-1} &= E\left[\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)' \middle| Y^{t-1}\right]
\end{aligned}
$$

  where $P_{t|t-1}$ is $n \times n$ and $\Omega_{t|t-1}$ is $m \times m$.

- Because $n - \tilde{n}$ state variables are known with certainty one period in advance, $P_{t|t-1}$ is of rank $\tilde{n}$.

- Using $y_t - y_{t|t-1} = D(x_t - x_{t|t-1})$, it follows that

$$
\Omega_{t|t-1} = DP_{t|t-1}D'.
$$

- If the number of observables is greater than the number of exogenous shocks, $m > \tilde{n}$, the $m \times m$ matrix $\Omega_{t|t-1}$ has rank $\tilde{n} < m$.

Stochastic singularity of DSGE models (VI)

- Thus, $\Omega_{t|t-1}$ is singular—this is the equivalent to saying that there is a linear combination of elements in $y_t t$ that hold with certainty.

- Therefore, the inverse of $\Omega_{t|t-1}$ does not exist and the likelihood function cannot be evaluated because we need to compute

$$\left(y_t - y_{t|t-1}\right)' \Omega_{t|t-1}^{-1} \left(y_t - y_{t|t-1}\right).$$

Solutions to the stochastic singularity problem

1. Drop observables until number of observables = number of exogenous variables.
   - Problem: throwing away information.

2. Extend the model adding more shocks.
   - Problem: you are changing the model.

3. Add measurement errors. Assume that the econometrician observes y_t with noise:

$$y_t = D(\theta)x_t + v_t;\ E[v_t] = 0, E[v_t v_t'] = R.$$

   - Add measurement errors until the $\#$ of measurement errors + $\#$ of exogenous exogenous = $\#$ of observable variables. Problem: it might not be obvious to which variables we should add the measurement errors.
   - Add one measurement error for each observable variable. Problem: you may be adding to many additional parameters to estimate, namely, the covariance matrix of the measurement errors. But you can also assume uncorrelated measurement errors, so that the matrix R is diagonal.

Example RBC model

$$\max E_0 \left[\sum_{t=0}^{\infty} \beta^t \left(\log c_t - \eta \frac{l_t^{1+\nu}}{1+\nu} \right) \right] \text{ subject to } c_t + k_{t+1} = A_t k_t^{\alpha} l_t^{1-\alpha} + (1-\delta) k_t.$$

Equilibrium conditions:

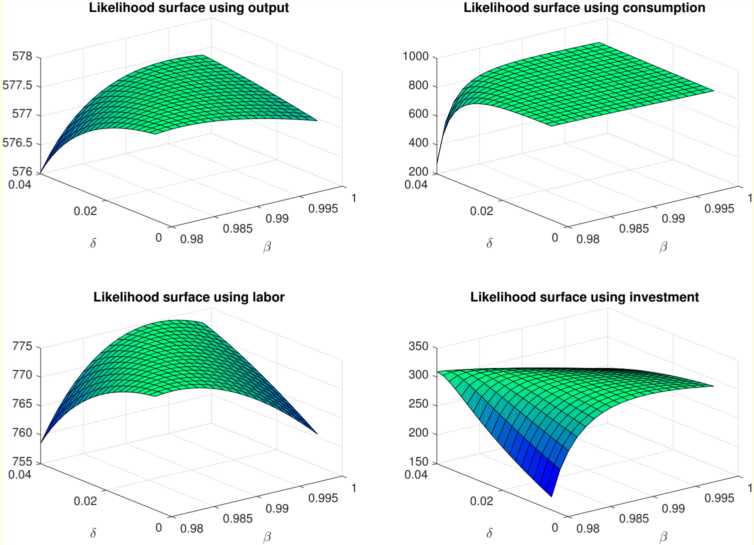
$$\begin{aligned}
\frac{1}{c_t} =& \lambda_t \\
\eta l_t^{1/\eta} =& \lambda_t (1-\alpha) y_t / l_t \\
\lambda_t =& \beta E_t \left[\lambda_{t+1} \left(\alpha \frac{y_{t+1}}{k_{t+1}} + 1 - \delta \right) \right] \\
y_t =& c_t + x_t \\
y_t =& A_t k_t^{\alpha} l_t^{1-\alpha} \\
k_{t+1} =& x_t + (1-\delta) k_t \\
\log A_{t+1} =& \rho \log A_t + \sigma \epsilon_{t+1}.
\end{aligned}$$

Log-linear RBC model

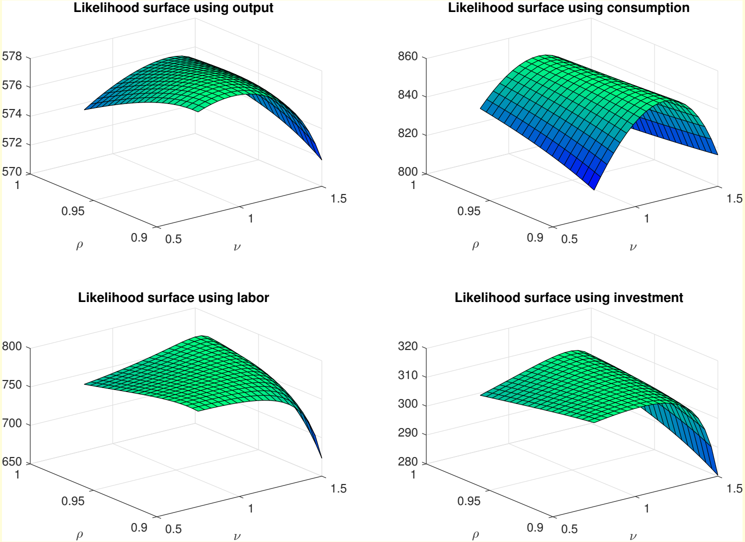
$$
\begin{aligned}
0 =& \hat{c}_t + \hat{\lambda}_t \\
0 =& (1 + 1/\nu)\hat{l}_t - \hat{\lambda}_t - \hat{y}_t \\
0 =& \hat{y}_t - \hat{A}_t - \alpha\hat{k}_t - (1-\alpha)\hat{l}_t \\
0 =& \bar{y}\hat{y}_t - \bar{c}\hat{c}_t - \bar{x}\hat{x}_t \\
E_t[k_{t+1}] =& (1-\delta)\hat{k}_t + \delta\hat{x}_t \\
E_t[\hat{\lambda_{t+1}} + \beta\alpha(\bar{y}/\bar{k})(\hat{y_{t+1}} - \hat{k_{t+1}})] =& \hat{\lambda}_t \\
E_t[\log A_{t+1}] =& \rho \log A_t.
\end{aligned}
$$

- Parameters: $\theta = [\beta, \eta, \nu, \delta, \alpha, \rho, \sigma] = [0.99, 7.59, 1, 0.017, 0.33, 0.95, 1.019]$.
- Simulate model for $T = 200$ periods given parameter values.
- Compute log-likelihood using 1 observable variable at a time: y_t, c_t, n_t, x_t.
- Plot surface of log-likelihood moving 2 parameters at a time fixing the others at their true values.

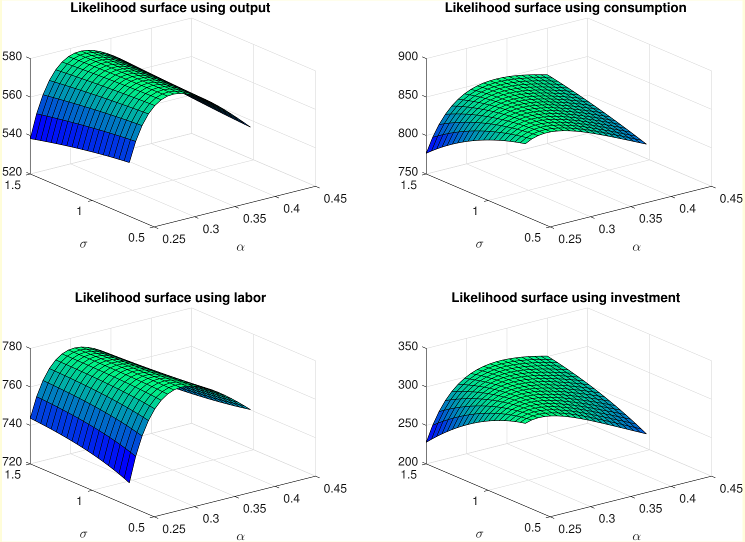
RBC log-likelihood surface moving (β, δ)



RBC log-likelihood surface moving (ρ, ν)



RBC log-likelihood surface moving (σ, α)



RBC log-likelihood surface moving (η, ν)