ML Estimation of State Space Models

March 17, 2017

Recap from last time:

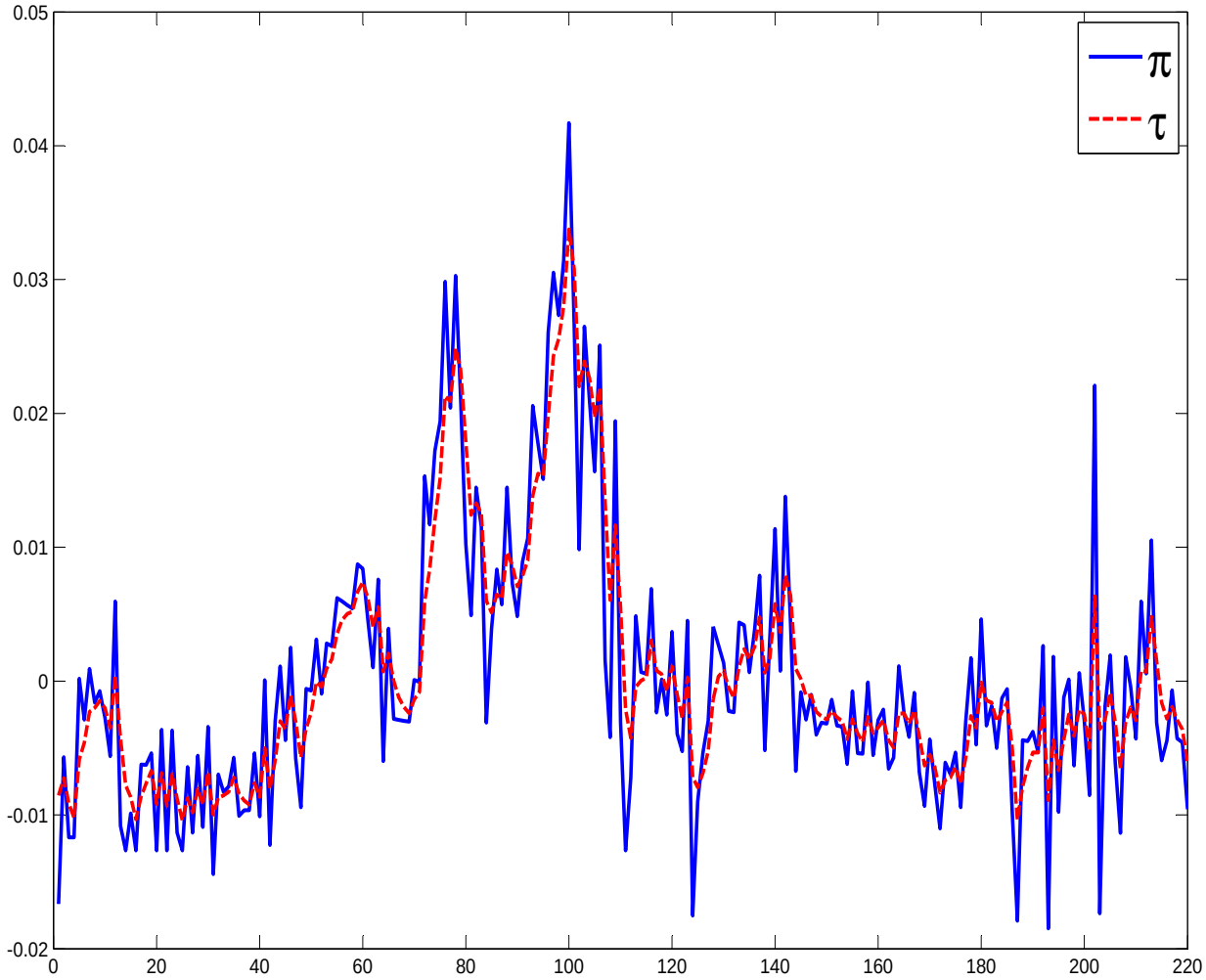
Unobserved Component model of Inflation

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

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

1. Estimated the parameters of the system, i.e. σ_η^2 and σ_ε^2 using grid search
   - $\widehat{\sigma}_\varepsilon^2 = 0.0028$
   - $\widehat{\sigma}_\eta^2 = 0.0051$
2. Computed both real time and smoothed estimates of the permanent component τ_t at different points in time

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 different 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

Numerical maximization of likelihood functions

Today:
Numerical maximization

- (Very brief) review of grid search, steepest ascent and Newton-Raphson algorithms
- Simulated annealing

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

Work through an example:

- Simple DSGE model

Grid search

Divide range of parameters into grid and evaluate all possible combinations

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

Steepest Ascent method

A blind man climbing a mountain.
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.

A minimalistic DSGE model

The linearized structural system

After linearizing, the main model equations are given by

$$\begin{aligned} \pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - \overline{y}_t) \\ y_t &= E_t y_{t+1} - \sigma\left(i_t - E_t \pi_{t+1}\right) \\ i_t &= \phi \pi_t \\ x_t &= \rho x_{t-1} + u_t^x : u_t^x \sim N(0, \sigma_u^2) \end{aligned}$$

where π_t, y_t, y_t, i_t are inflation, output, potential output and nominal interest rate respectively.

Solving the model

3 ways to solve a linear model

Solving a model using full information rational expectations as the equilibrium concept involves integrating out expectations terms from the structural equations of the model by replacing agents' expectations with the mathematical expectation, conditional on the state of the model.

Three different ways of doing this.

1. Method of undetermined coefficients, can be very quick when feasible and illustrates the fixed point nature of the rational expectations solution.
2. Replacing expectations with linear projections onto observable variables
3. Decouple the stable and unstable dynamics of the model and set the unstable part to zero.

The 3 equation NK model

As a vehicle to demonstrate the different solution methods, we will use a simple New-Keynesian model

$$\begin{aligned} \pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - \overline{y}_t) \\ y_t &= E_t y_{t+1} - \sigma\left(i_t - E_t \pi_{t+1}\right) \\ i_t &= \phi \pi_t \\ x_t &= \rho x_{t-1} + u_t : u_t \sim N(0, \sigma_u^2) \end{aligned}$$

where π_t, y_t, y_t, i_t are inflation, output, potential output and nominal interest rate respectively.

- Single variable, potential output x_t, as the state.

Method I:
Method of undetermined coefficients

Pros

- Method is quick when feasible
- Illustrates well the fixed point nature of rational expectations equilibria.

Cons

- Difficult to implement in larger models

Method of undetermined coefficients

$$
\begin{aligned}
\pi_t &= \beta E_t \pi_{t+1} + \kappa(y_t - x_t) \\
y_t &= E_t y_{t+1} - \sigma\left(i_t - E_t \pi_{t+1}\right) \\
i_t &= \phi \pi_t \\
x_t &= \rho x_{t-1} + u_t : u_t \sim N(0, \sigma_u^2)
\end{aligned}
$$

Start by substituting in the interest rate in the Euler equation

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

Solving model using method of undetermined coefficients

Conjecture that model can be put in the form

$$\begin{aligned} x_t &= \rho x_{t-1} + u_t^x \\ y_t &= a x_t \\ \pi_t &= b x_t \end{aligned}$$

Why is this a good guess?

Solving model using method of undetermined coefficients

Substitute in conjectured form of solution into structural equation

$$\begin{aligned} ax_t &= a\rho x_t - \frac{1}{\gamma}\left[\phi_\pi b x_t - b\rho x_t\right] \\ bx_t &= b\rho x_t + \kappa\left[ax_t - x_t\right] \end{aligned}$$

where we used that $x_t = \rho x_{t-1} + u_t^x$ implies that $\mathrm{E}[x_{t+1} \mid x_t] = \rho x_t$

Solving model using method of undetermined coefficients

Equate coefficients on right and left hand side

$$\begin{aligned} a &= a\rho - \frac{1}{\gamma}\phi_\pi b + \frac{1}{\gamma} b\rho \\ b &= b\rho + \kappa\left[a - 1\right] \end{aligned}$$

or

$$\begin{bmatrix} (1-\rho) & \frac{1}{\gamma}(\phi_\pi - \rho) \\ -\kappa & (1-\rho) \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ -\kappa \end{bmatrix}$$

Solving model using method of undetermined coefficients

Solve for a and b

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} (1-\rho) & \frac{1}{\gamma}(\phi_\pi - \rho) \\ -\kappa & (1-\rho) \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ -\kappa \end{bmatrix}$$

or

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -\kappa \frac{\phi-\rho}{-c} \\ \kappa\gamma \frac{1-\rho}{-c} \end{bmatrix}$$

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

The solved model

The solved model is of the form

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

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

Method II: Replacing expectations with linear projections

The second method uses that projections of the future values of variables on observables gives optimal expectations (in the sense of minimum error variance) if the observables span the space of the state.

How does it work?

- Replace $E_t\pi_{t+1}$ and $E_t y_{t+1}$ with linear projections of these variables on current inflation.
- There is nothing special about inflation. Projecting onto current output would also work.

Least squares estimation via the projection theorem

To find the estimate $\widehat{x}$ as a linear function of y simply use that

$$\begin{aligned} \langle x - \beta y, y \rangle &= E\left[(x - \beta y)\, y'\right] \\ &= 0 \end{aligned}$$

and solve for β

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

The advantage of this approach is that once you have made sure that the variables y and x are in a well defined inner product space, there is no need to minimize the variance directly. The projection theorem ensures that an estimate with orthogonal errors is the (linear) minimum variance estimate.

Two useful properties of linear projections

1. If two random variables X and Y are Gaussian, then the projection of Y onto X coincides withe the conditional expectation $E(Y \mid X)$.
2. If X and Y are not Gaussian, the linear projection of Y onto X is the minimum variance linear prediction of Y given X.

Replacing expectations with linear projections

We will use that in equilibrium

$$
\begin{aligned}
E\left(\pi_{t+1} \mid \pi_t\right) &= \frac{cov(\pi_t, \pi_{t+1})}{var\left(\pi_t\right)}\pi_t \\
E\left(y_{t+1} \mid \pi_t\right) &= \frac{cov(\pi_t, y_{t+1})}{var\left(\pi_t\right)}\pi_t
\end{aligned}
$$

if the innovations u_t to x_t are Gaussian.

Replacing expectations with linear projections

Let

$$\begin{aligned} c_0\pi_t &= E^*\left(\pi_{t+1} \mid \pi_t\right) \\ d_0\pi_t &= E^*\left(y_{t+1} \mid \pi_t\right) \end{aligned}$$

denote initial candidate projections of expected inflation and output on current inflation. We can then write the structural equations as

$$\begin{aligned} \pi_t &= \beta c_0\pi_t + \kappa(y_t - x_t) \\ y_t &= d_0\pi_t - \sigma\left(\phi\pi_t - c_0\pi_t\right) \end{aligned}$$

Replacing expectations with linear projections

Put the whole system in matrix form

$$\begin{bmatrix} x_t \\ \pi_t \\ y_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ \kappa & 1-\beta c_0 & -\kappa \\ 0 & -d_0+\sigma\phi-\sigma c_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \pi_{t-1} \\ y_{t-1} \end{bmatrix}$$
$$+ \begin{bmatrix} 1 & 0 & 0 \\ \kappa & 1-\beta c_o & -\kappa \\ 0 & -d_0+\sigma\phi-\sigma c_0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_t$$

or

$$X_t = AX_{t-1} + Cu_t$$

Solution algorithm

1. Make an initial guess of c_0 and d_0 in (1)
2. Compute the implied covariances of current inflation and future inflation and output using

$$\begin{aligned} E\left[X_t X_t'\right] &= \Sigma_{XX} \\ \Sigma_{XX} &= A\Sigma_{XX}A' + CC' \end{aligned}$$

and

$$E\left[X_{t+1} X_t'\right] = A\Sigma_{XX}$$

3. Replace the c_s and d_s with the c_{s+1} and d_{s+1} in ((1))

$$\begin{aligned} c_{s+1} &= \frac{cov(\pi_t, \pi_{t+1})}{var\left(\pi_t\right)} \\ d_{s+1} &= \frac{cov(\pi_t, y_{t+1})}{var\left(\pi_t\right)} \end{aligned}$$

using the covariances from Step 2.

4. Repeat Step 2-3 until c_s and d_s converges.

Method III: Stable/unstable decoupling

Originally due to Blanchard and Kahn (1980)

- Computational aspects of the method has been further developed by others, for instance Klein (2000).
- The most accessible reference is probably Soderlind (1999), who also has code posted on his web site.

The method has several advantages:

- Fast
- Provides conditions for when a solution exists
- Provides conditions for when the solution is unique.

Method III: Stable/unstable decoupling

Start by putting the model into matrix form

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & \beta & 0 \\ 0 & \sigma & 1 \end{bmatrix} \begin{bmatrix} x_{t+1} \\ E_t\pi_{t+1} \\ E_t y_{t+1} \end{bmatrix}$$

$$= \begin{bmatrix} \rho & 0 & 0 \\ \kappa & 1 & -\kappa \\ 0 & \sigma\phi & 1 \end{bmatrix} \begin{bmatrix} x_{t-1} \\ \pi_t \\ y_t \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} u_{t+1}$$

or

$$A_0 \begin{bmatrix} x^1_{t+1} \\ E_t x^2_{t+1} \end{bmatrix} = A_1 \begin{bmatrix} x^1_t \\ x^2_t \end{bmatrix} + C_1 u_{t+1}$$

- x^1_t is vector containing the pre-determined and/or exogenous variables (i.e. x_t)
- x^2_t a vector containing the forward looking ("jump") variables (i.e. $E_t y_{t+1}$ and $E_t\pi_{t+1}$).

Method III: Stable/unstable decoupling

Pre-multiply both sides of

$$A_0 \begin{bmatrix} x_{t+1}^1 \\ E_t x_{t+1}^2 \end{bmatrix} = A_1 \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + C_1 u_{t+1}$$

by A_0^{-1} to get

$$\begin{bmatrix} x_{t+1}^1 \\ E_t x_{t+1}^2 \end{bmatrix} = A \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + C u_{t+1}$$

where $A = A_0^{-1} A_1$ and $C = A_0^{-1} C_1$.

For the model to have unique stable solution the number of stable eigenvalues of A must be equal to the number of exogenous/pre-determined variables.

Method III: Stable/unstable decoupling

Use a Schur decomposition to get

$$A = ZTZ^H$$

where T is upper block triangular

$$T = \begin{bmatrix} T_{11} & T_{12} \\ \mathbf{0} & T_{22} \end{bmatrix}$$

and Z is a unitary matrix so that $Z^H Z = ZZ^H = I$
($\implies Z^H = Z^{-1}$).

- For any square matrix W, $W^{-1}AW$ is a so called similarity transformation of A.
- Similarity transformations do not change the eigenvalues of a matrix
- We can always choose Z and T so that the unstable eigenvalues of A are shared with T_{22}

Method III: Stable/unstable decoupling

Define the auxiliary variables

$$\begin{bmatrix} \theta_t \\ \delta_t \end{bmatrix} = Z^H \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}$$

We can then rewrite the system (33) as

$$Z^H \begin{bmatrix} x_{t+1}^1 \\ E_t x_{t+1}^2 \end{bmatrix} = Z^H Z T Z^H \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix}$$

or equivalently

$$E \begin{bmatrix} \theta_{t+1} \\ \delta_{t+1} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ \mathbf{0} & T_{22} \end{bmatrix} \begin{bmatrix} \theta_t \\ \delta_t \end{bmatrix}$$

since $Z^H Z = I$.

Method III: Stable/unstable decoupling

For this system to be stable, the auxiliary variables associated with the unstable roots in T_{22} must be zero for all t. (WHY?)

Imposing $\delta_t = 0 \forall t$ reduces the relevant state dynamics to

$$\theta_t = T_{11}\theta_{t-1}$$

To get back the original variables we simply use that

$$\begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} Z_{11} \\ Z_{21} \end{bmatrix} \theta_t$$

or

$$\begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} = \begin{bmatrix} Z_{11} \\ Z_{21} \end{bmatrix} Z_{11}^{-1} x_t^1$$

which is the solution to the model. It is in the form

$$\begin{aligned} x_t^1 &= M x_{t-1}^1 + \varepsilon_t \\ x_t^2 &= G x_t^1 \end{aligned}$$

where $M = Z_{11} T_{11} Z_{11}^{-1}$ ($= \rho$ in our example) and $G = Z_{21} Z_{11}^{-1}$.

Estimating a DSGE model using Simulated Annealing

Remember our benchmark NK model:

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

To estimate the model using three time series, we need to add more shocks

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

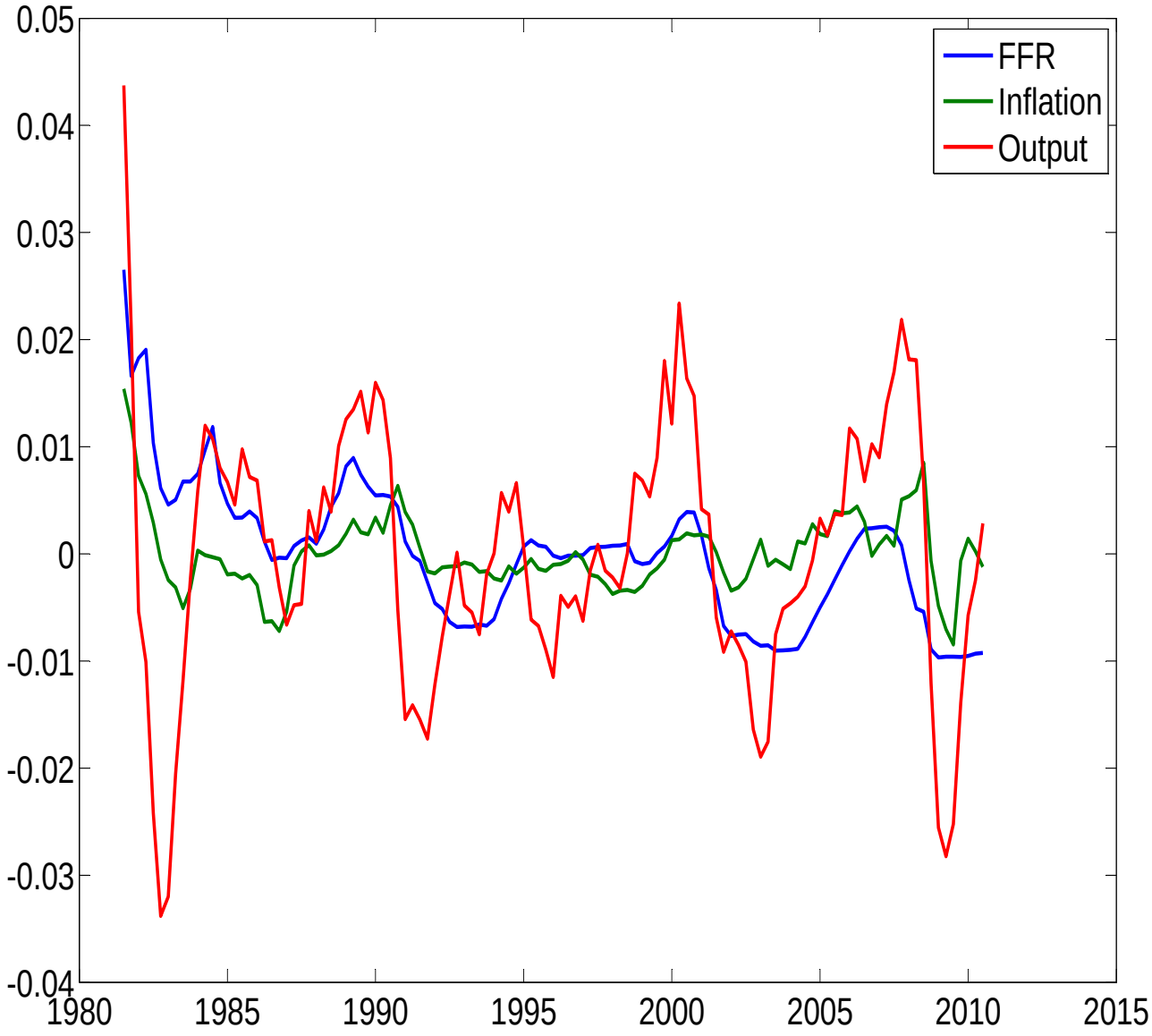
For a given state space system

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

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


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

```
% 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);
```

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`

initial loss function value:
-706.3706

No. of evaluations
46

current temperature
5

current optimum function value
-840.2525

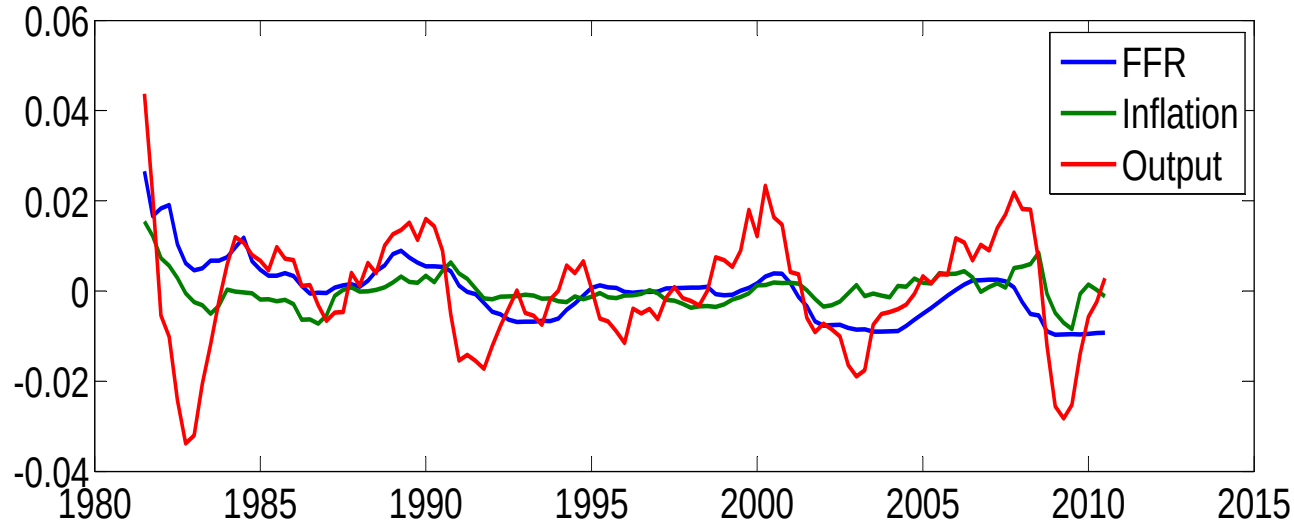
No. of downhill steps
13

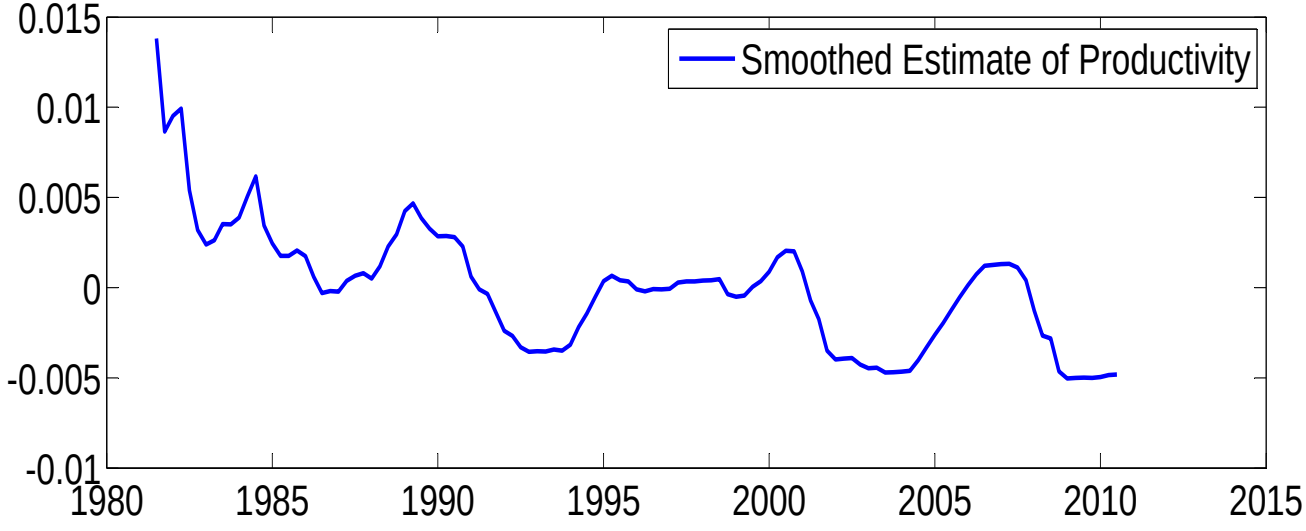
No. of accepted uphill steps
10

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

Components for ML estimation of State Space Models

1. A function that maps parameters Θ into a SSS
2. A function that evaluates the likelihood function $\mathcal{L}(Z \mid \Theta)$
3. A maximizer that takes 1 and 2 and starting values for Θ as inputs

That's all you need.

No-arbitrage term structure models

State Space models and Principal Components

State Space models and Principal Components

Previously, we used principal components to find the factors of a system that was in state space form

- What is the relation ship between the state X_t and the factors F_t
  - Can we find one from the other?
- When is state space models and filtering preferable to principal components?

Principal component as factors

Recall:

$$\begin{array}{rcl} \underset{(r\times 1)}{F_t} & = & \Phi F_{t-1} + \mathbf{u}_t^F \\ \underset{(N\times 1)}{Y_t} & = & \underset{(N\times r)}{W}\ \underset{(r\times 1)}{F_t} + v_t \end{array}$$

- W contains the eigenvectors of $EY_tY_t' = W\Lambda W'$ and $WW' = I$ so that $F_t = W'Y_t$
- Λ is a diagonal matrix containing the ordered eigenvalues

This looks like a special case of a state space system

Can we find a mapping from X_t to F_t?

Yes:

- When N is large or Σ_{vv} is small the following holds:

$$\begin{aligned} F_t &= W' Y_t \\ &= W' D X_t \end{aligned}$$

If D is of rank n (where $n =$ dimension of X), then $(W'D)^{-1}$ exists so the mapping works in both directions.

How can we find W for a state space model?

$$\begin{aligned} EZ_t Z_t' &= D\Sigma_{xx} D' + \Sigma_{vv} \\ &= W \Lambda W' \end{aligned}$$

where Σ_{xx} solves

$$\Sigma_{xx} = A\Sigma_{xx} A' + CC'$$

Doing the eigenvector/value decomposition of $EZ_t Z_t'$ thus gives us W so that the factors can be computed as $F_t = W' D X_t$

How about the dynamics of the factors?

We have:

$$\begin{aligned} F_t &= \Phi F_{t-1} + \mathbf{u}_t^F \\ X_t &= AX_{t-1} + C\mathbf{u}_t \end{aligned}$$

To find Φ, use that $F_t = W'DX_t$ and $X_t = (W'D)^{-1} F_t$

$$\begin{aligned} F_t &= W'DAX_{t-1} + W'DC\mathbf{u}_t \\ &= W'DA\left(W'D\right)^{-1} F_{t-1} + W'DC\mathbf{u}_t \end{aligned}$$

to get

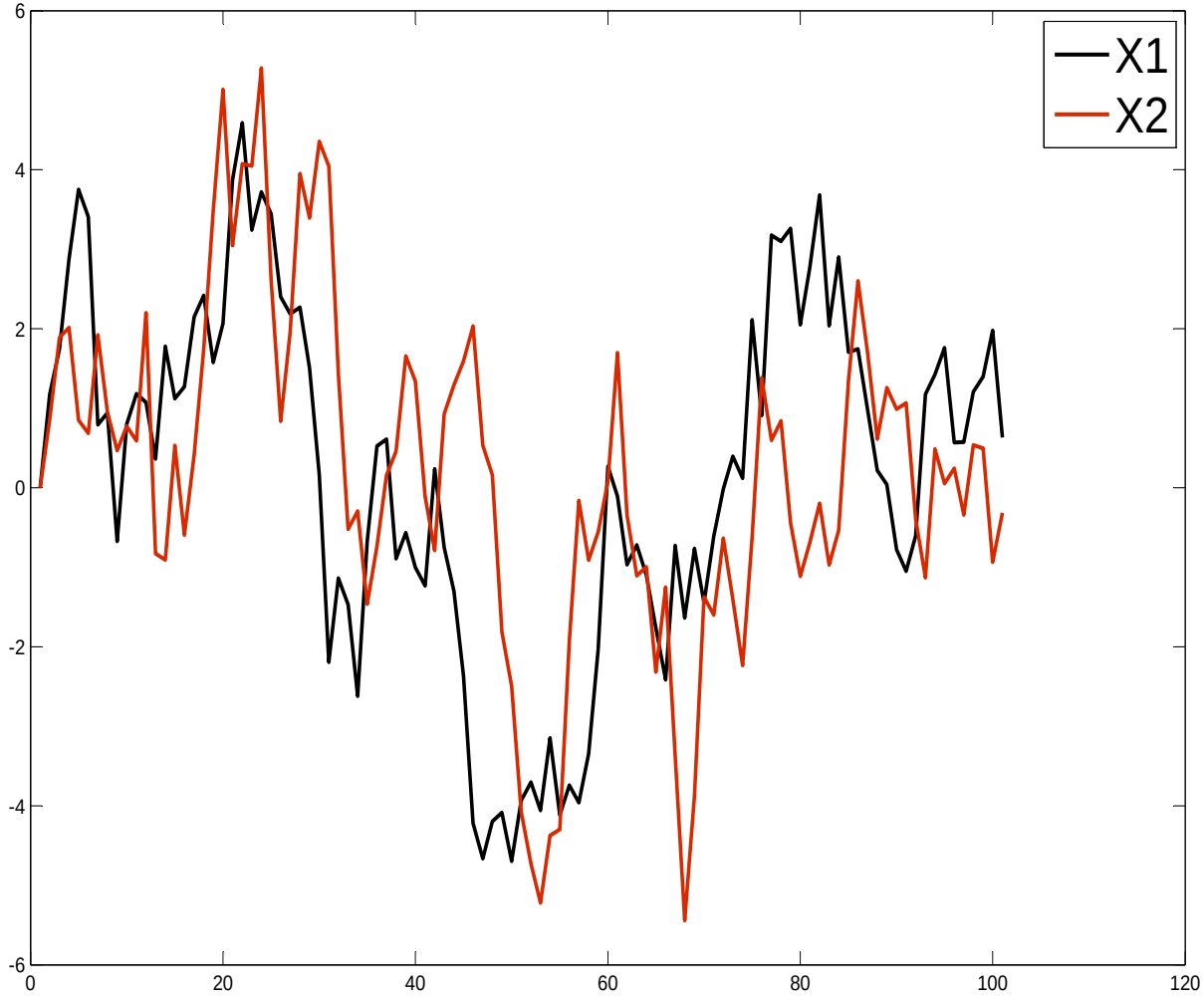
$$\Phi = W'DA\left(W'D\right)^{-1}, E\mathbf{u}_t^F\mathbf{u}_t'^F = W'DC\left(W'DC\right)'$$

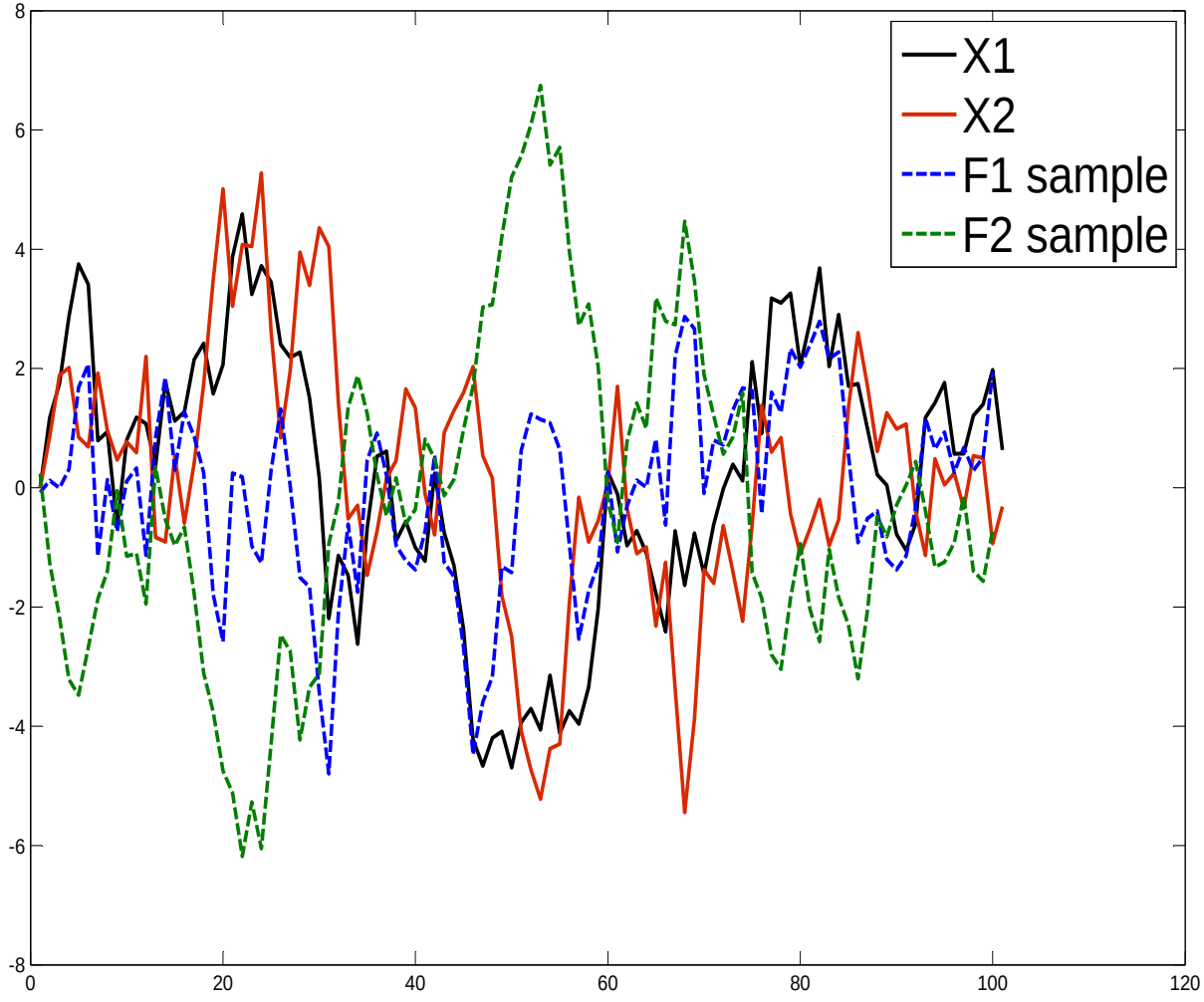
Example

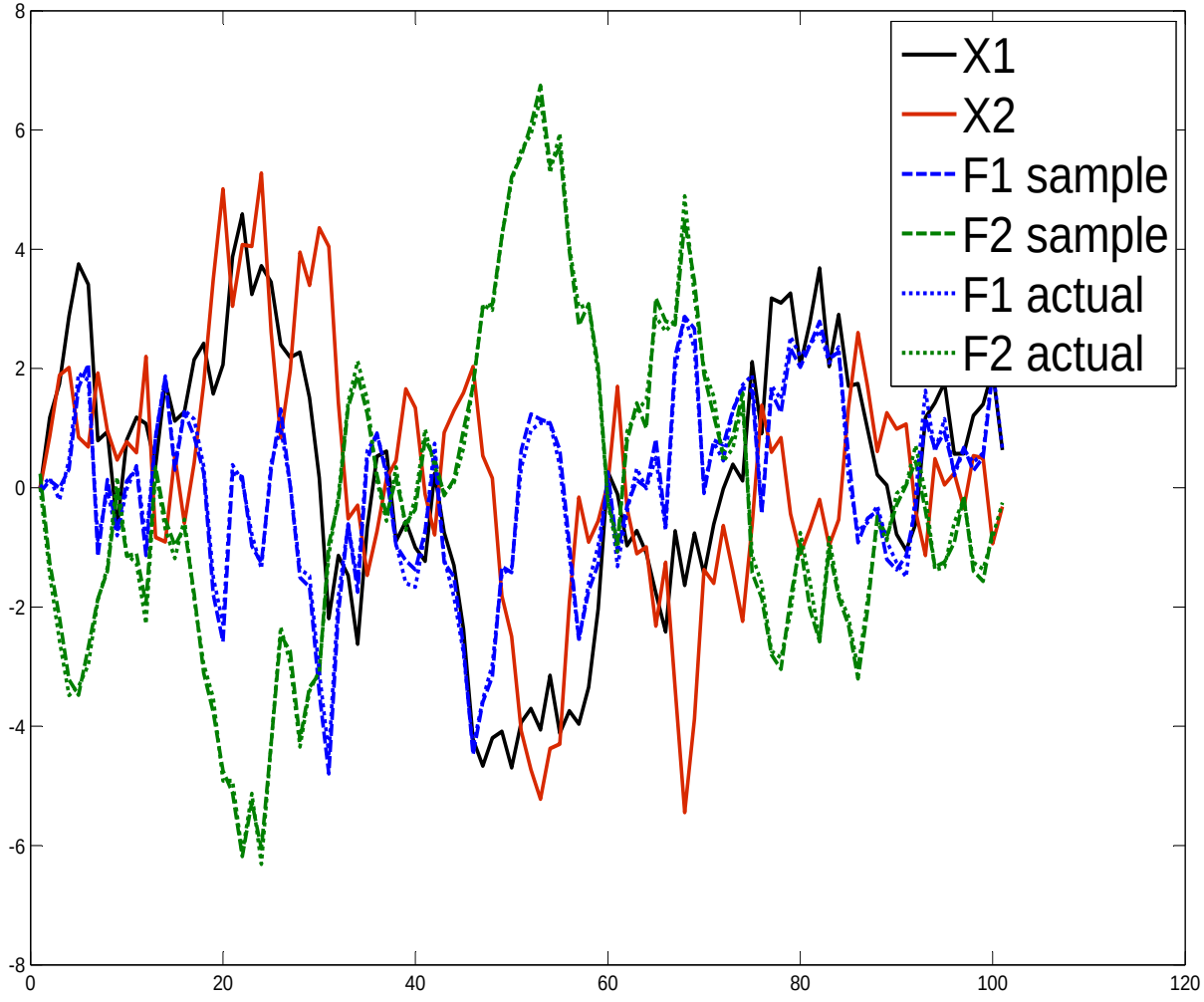
Simulate data using

$$\begin{aligned} A &= \begin{bmatrix} 0.9 & 0 \\ 0.2 & 0.7 \end{bmatrix} \\ C &= I, D = I, \Sigma_{vv} = 0.1 \times I \end{aligned}$$

with $T = 100$.


X1
X2
6
4
2
0
-2
-4
-6
0
20
40
60
80
100
120


X1
X2
F1 sample
F2 sample
8
6
4
2
0
-2
-4
-6
-8
0
20
40
60
80
100
120


X1
X2
F1 sample
F2 sample
F1 actual
F2 actual
8
6
4
2
0
-2
-4
-6
-8
0
20
40
60
80
100
120

T=100 sample and theoretical Ψ

Using $\Phi = W'DA\left(W'D\right)^{-1}$

$$\Phi = \begin{bmatrix} 0.69 & 0.008 \\ -0.19 & 0.91 \end{bmatrix}$$

$$\widehat{\Phi} = \begin{bmatrix} 0.71 & 0.025 \\ -0.19 & 0.93 \end{bmatrix}$$

Differences are due to small sample and non-zero measurement errors

Take homes from PCs and State Space models

State space systems that have the same implications for observables are not unique

- "Rotations" of state variables in X_t can give different interpretations
- Different rotations span the same space, so no difference in predictive content

Is this important?

- Depends on the question.
  - In factor models of the term structure, PC imply that the factors will have the level, slope and curvature interpretation.
  - In macro models, usually too few degrees of freedom to do any rotations since number of deep parameters is lower than free parameters in the state space system.

What is better when?

State space model and Kalman filter is better...

- ...if number of different observable time series is small...
- ...and measurement errors are large.

Time dimension is important when law of large numbers do not work in the cross-section

- If noise is small or number of time series very large, PC and SSM give the same results
- Choose whatever is more convenient