9. Instrumental variable methods (IVM)

- review on the least-squares method
- description of IV methods
- choice of instruments
- extended IV methods

Revisit the LS method

using linear regression in dynamic models (SISO)

$$A(q^{-1})y(t) = B(q^{-1})u(t) + \nu(t)$$

where $\nu(t)$ denotes the equation error

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a}, \quad B(q^{-1}) = b_1 q^{-1} + \cdots + b_{n_b} q^{-n_b}$$

we can write the dynamic as

$$y(t) = H(t)\theta + \nu(t)$$

where

$$\begin{aligned} H(t) &= \begin{bmatrix} -y(t-1) & \cdots & -y(t-n_a) & u(t-1) & \cdots & u(t-n_b) \end{bmatrix} \\ \theta &= \begin{bmatrix} a_1 & \cdots & a_{n_a} & b_1 & \cdots & b_{n_b} \end{bmatrix} \end{aligned}$$

the least-squares solution is the value of $\hat{\theta}$ that minimizes

$$\frac{1}{N}\sum_{t=1}^{N}\|\nu(t)\|^2$$

and is given by

$$\hat{\theta}_{\rm ls} = \left(\frac{1}{N}\sum_{t=1}^{N} H(t)^T H(t)\right)^{-1} \left(\frac{1}{N}\sum_{t=1}^{N} H(t)^T y(t)\right)$$

to examine if $\hat{\theta}$ is consistent ($\hat{\theta} \to \theta$ as $N \to \infty$), note that

$$\begin{aligned}
\hat{\theta}_{\rm ls} - \theta &= \left(\frac{1}{N}\sum_{t=1}^{N} H(t)^T H(t)\right)^{-1} \left\{\frac{1}{N}\sum_{t=1}^{N} H(t)^T y(t) - \left(\frac{1}{N}\sum_{t=1}^{N} H(t)^T H(t)\right)\theta\right\} \\
&= \left(\frac{1}{N}\sum_{t=1}^{N} H(t)^T H(t)\right)^{-1} \left(\frac{1}{N}\sum_{t=1}^{N} H(t)^T \nu(t)\right)
\end{aligned}$$

hence, $\hat{\theta}_{\rm ls}$ is consistent if

- $\mathbf{E}[H(t)^T H(t)]$ is nonsingular
  satisfied in most cases, except u is not persistently exciting of order n
- $\mathbf{E}[H(t)^T \nu(t)] = 0$
  not satisfied in most cases, except $\nu(t)$ is white noise

summary:

- LS method for dynamical models is still certainly simple to use
- consistency is not readily obtained since the information matrix (H) is no longer deterministic
- it gives consistent estimates under restrictive conditions

to obtain consistency of the estimates, we modify the normal equation so that the output and the disturbance become uncorrelated

Solutions:

- PEM (Prediction error methods)
  - model the noise
  - applicable to general model structures
  - generally very good properties of the estimates
  - computationally quite demanding
- IVM (Instrumental variable methods)
  - do not model the noise
  - retain the simple LS structure
  - simple and computationally efficient approach
  - consistent for correlated noise
  - less robust and statistically less effective than PEM

Description of IVM

define $Z(t) \in \mathbf{R}^{n_\theta}$ with entries uncorrelated with $\nu(t)$

$$\frac{1}{N}\sum_{t=1}^{N} Z(t)^T \nu(t) = \frac{1}{N}\sum_{t=1}^{N} Z^T(t)[y(t) - H(t)\theta] = 0$$

The basic IV estimate of θ is given by

$$\hat{\theta} = \left(\frac{1}{N}\sum_{t=1}^{N} Z(t)^T H(t)\right)^{-1} \left(\frac{1}{N}\sum_{t=1}^{N} Z(t)^T y(t)\right)$$

provided that the inverse exists

- $Z(t)$ is called **the instrument** and is up to user's choice
- if $Z(t) = H(t)$, the IV estimate reduces to the LS estimate

Choice of instruments

the instruments $Z(t)$ have to be chosen such that

- $Z(t)$ is uncorrelated with noise $\nu(t)$

$$\mathbf{E}Z(t)^T\nu(t) = 0$$

- the matrix

$$\frac{1}{N}\sum_{t=1}^{N} Z(t)^T H(t) \rightarrow \mathbf{E}Z(t)^T H(t)$$

  has full rank

in other words, $Z(t)$ and $H(t)$ are correlated one possibility is to choose

$$Z(t) = \begin{bmatrix} -\eta(t-1) & \dots & -\eta(t-n_a) & u(t-1) & \dots & u(t-n_b) \end{bmatrix}$$

where the signal $\eta(t)$ is obtained by filtering the input,

$$C(q^{-1})\eta(t) = D(q^{-1})u(t)$$

Special choices:

- let C, D be *a prior* estimates of A and B
- simple choice: pick $C(q^{-1}) = 1$, $D(q^{-1}) = -q^{-n_b}$

$$Z(t) = \begin{bmatrix} u(t-1) & \dots & u(t-n_a-n_b) \end{bmatrix}$$

(with a reordering of $Z(t)$)

note that $u(t)$ and the noise $\nu(t)$ are assumed to be independent

Example via Yule-Walker equations

consider a scalar ARMA process:

$$A(q^{-1})y(t) = C(q^{-1})e(t)$$

$$y(t) + a_1 y(t-1) + \ldots + a_p y(t-p) = e(t) + c_1 e(t-1) + \ldots + c_r e(t-r)$$

where $e(t)$ is white noise with zero mean and variance λ^2

define $R_k = \mathbf{E} y(t) y(t-k)^T$, we obtain

$$R_k + a_1 R_{k-1} + \ldots + a_p R_{k-p} = 0, \quad k = r+1, r+2, \ldots$$

where we have used $\mathbf{E} C(q^{-1}) e(t) y(t-k)^T = 0, \quad k > r$

this is referred to as **Yule-Walker equations** enumerate from $k = r+1, \ldots, r+m$, where $m \geq p$,

the Yule-Walker equations can be fit into a matrix form

$$\begin{bmatrix} R_r & R_{r-1} & \ldots & R_{r+1-p} \\ R_{r+1} & R_r & \ldots & R_{r+2-p} \\ \vdots & \vdots & \vdots & \\ R_{r+m-1} & R_{r+m-2} & \ldots & R_{r+m-p} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{r+1} \\ R_{r+2} \\ \vdots \\ R_{r+m} \end{bmatrix} \triangleq \mathbf{R}\theta = -r$$

$\mathbf{R}$ and r are typically replaced by their sample esimates:

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} y(t-r-1) \\ \vdots \\ y(t-r-m) \end{bmatrix} \begin{bmatrix} y(t-1) & \ldots & y(t-p) \end{bmatrix}$$

$$\hat{r} = \frac{1}{N} \sum_{t=1}^{N} \begin{bmatrix} y(t-r-1) \\ \vdots \\ y(t-r-m) \end{bmatrix} y(t)$$

hence $\hat{\mathbf{R}}\hat{\theta} = -\hat{r}$ is equivalent to

$$\frac{1}{N}\sum_{t=1}^{N} \underbrace{\begin{bmatrix} y(t-r-1) \\ \vdots \\ y(t-r-m) \end{bmatrix}}_{Z(t)^T} \underbrace{\begin{bmatrix} -y(t-1) & \ldots & -y(t-p) \end{bmatrix}}_{H(t)}$$

$$= \frac{1}{N}\sum_{t=1}^{N} \begin{bmatrix} y(t-r-1) \\ \vdots \\ y(t-r-m) \end{bmatrix} y(t)$$

this is the relationship in basic IVM

$$\frac{1}{N}\sum_{t=1}^{N} Z(t)^T H(t)\theta = \frac{1}{N}\sum_{t=1}^{N} Z(t)^T y(t)$$

where we use the delayed output as an instrument

$$Z(t) = \begin{bmatrix} -y(t-r-1) & y(t-r-2) & \ldots & y(t-r-m) \end{bmatrix}^T$$

Extended IV methods

The *extended* IV method is to generalize the basic IV in two directions:

- allow $Z(t)$ to have more elements than θ ($n_z \geq n_\theta$)
- use prefiltered data

and the extended IV estimate of θ is obtained by

$$\min_\theta \left\| \sum_{t=1}^{N} Z(t)^T F(q^{-1})(y(t) - H(t)\theta) \right\|_W^2$$

where $\|x\|_W^2 = x^T W x$ and $W \succ 0$ is given

when $F(q^{-1}) = I, n_z = n_\theta, W = I$, we obtain the basic IV estimate

Define

$$A_N = \frac{1}{N}\sum_{t=1}^{N} Z(t)^T F(q^{-1})H(t)$$
$$b_N = \frac{1}{N}\sum_{t=1}^{N} Z(t)^T F(q^{-1})y(t)$$

then $\hat{\theta}$ is obtained by

$$\hat{\theta} = \operatorname*{argmin}_{\theta} \|b_N - A_N\theta\|_W^2$$

this is a weighted least-squares problem

the solution is given by

$$\hat{\theta} = (A_N^T W A_N)^{-1} A_N^T W b_N$$

note that this expression is only of theoretical interest

Theoretical analysis

Assumptions:

1. the system is strictly causal and asymptotically stable
2. the input $u(t)$ is persistently exciting of a sufficiently high order
3. the disturbance $\nu(t)$ is a stationary stochastic process with rational spectral density,
$$\nu(t) = G(q^{-1})e(t), \quad \mathbf{E}e(t)^2 = \lambda^2$$
4. the input and the disturbance are independent
5. the model and the true system have the same transfer function if and only if $\hat{\theta} = \theta$ (uniqueness)
6. the instruments and the disturbances are uncorrelated

from the system description

$$y(t) = H(t)\theta + \nu(t)$$

we have

$$\begin{aligned}
b_N &= \frac{1}{N}\sum_{t=1}^{N} Z(t)^T F(q^{-1}) y(t) \\
&= \frac{1}{N}\sum_{t=1}^{N} Z(t)^T F(q^{-1}) H(t)\theta + \frac{1}{N}\sum_{t=1}^{N} Z(t)^T F(q^{-1}) \nu(t) \\
&\triangleq A_N\theta + q_N
\end{aligned}$$

thus,

$$\hat{\theta} - \theta = (A_N^T W A_N)^{-1} A_N^T W b_N - \theta = (A_N^T W A_N)^{-1} A_N^T W q_N$$

as $N \to \infty$,

$$(A_N^T W A_N)^{-1} A_N^T W q_N \to (A^T W A)^{-1} A^T W q$$

where

$$\begin{aligned} A &\triangleq \lim_{N\to} A_N = \mathbf{E}[Z(t)^T F(q^{-1}) H(t)] \\ q &\triangleq \lim_{N\to} q_N = \mathbf{E}[Z(t)^T F(q^{-1}) \nu(t)] \end{aligned}$$

hence, the IV estimate is consistent ($\lim_{N\to\infty} \hat{\theta} = \theta$) if

- A has full rank
- $\mathbf{E}[Z(t)^T F(q^{-1}) \nu(t)] = 0$

Numerical example

the true system is given by

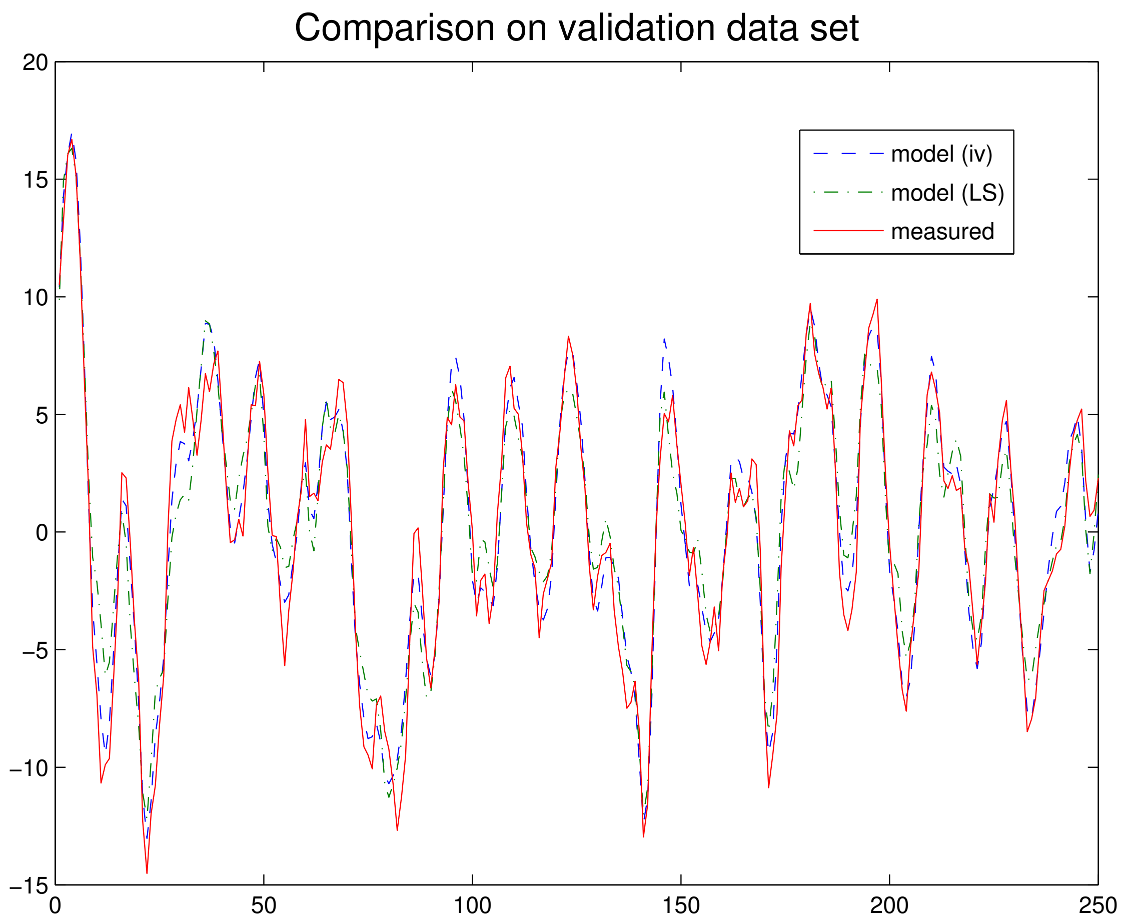
$$(1 - 1.5q^{-1} + 0.7q^{-2})y(t) = (1.0q^{-1} + 0.5q^{-2})u(t) + (1 - 1.0q^{-1} + 0.2q^{-2})e(t)$$

- ARMAX model
- $u(t)$ is from an ARMA process, independent of $e(t)$
- $e(t)$ is white noise withzero mean and variance 1
- $N = 250$ (number of data points)

estimation

- use ARX model and assume $n_a = 2, n_b = 2$
- compare the LS method with IVM

fit $\triangleq 100(1 - \|y - \hat{y}\| / \|y - \bar{y}\|)$



LS fit $= 66.97\%$, IV fit $= 77.50\%$

Example of MATLAB codes

```
%% Generate the data
close all; clear all;
N = 250; Ts = 1;
a = [1 -1.5 0.7]; b = [0 1 .5]; c = [1 -1 0.2];
Au = [1 -0.1 -0.12]; Bu = [0 1 0.2]; Mu = idpoly(Au,Bu,Ts);
u = sim(Mu,randn(2*N,1));   % u is ARMA process
noise_var = 1; e = randn(2*N,1);
M = idpoly(a,b,c,1,1,noise_var,Ts);
y = sim(M,[u e]);
uv = u(N+1:end); ev = e(N+1:end); yv = y(N+1:end);
u = u(1:N); e = e(1:N); y = y(1:N);
DATe = iddata(y,u,Ts); DATv = iddata(yv,uv,Ts);

%% Identification
na = 2; nb = 2; nc = 2;
theta_iv = iv4(DATe,[na nb 1]);    % ARX using iv4
theta_ls = arx(DATe,[na nb 1]);    % ARX using LS
```

```
%% Compare the measured output and the model output
[yhat2,fit2] = compare(DATv,theta_iv);
[yhat4,fit4] = compare(DATv,theta_ls);

figure;t = 1:N;
plot(t,yhat2{1}.y(t),'--',t,yhat4{1}.y(t),'-.',t,yv(t));
legend('model (iv)','model (LS)','measured')
title('Comparison on validation data set','FontSize',16);
```

References

Chapter 8 in
T. Söderström and P. Stoica, *System Identification*, Prentice Hall, 1989

Lecture on
Instrumental variable methods, System Identification (1TT875), Uppsala University, `http://www.it.uu.se/edu/course/homepage/systemid/vt05`