How to use MKFM6

Zhiyong Zhang

1-30-2005

What is MKFM6

- A free program by Conor V. Dolan for estimating state-space model by MLE via Kalman Filter
- Any model can be estimated only if it can be written as its state-space representation

State-Space Model /Time Series Model

- MKFM estimates the following time-series model:

$$\begin{aligned}\mathbf{y}_t &= \mathbf{S}\mathbf{a}_t + \mathbf{d} + \mathbf{e}_t + \mathbf{Z}\mathbf{x}_t \\ \mathbf{a}_{t+1} &= \mathbf{H}\mathbf{a}_t + \mathbf{c} + \mathbf{G}\mathbf{z}_{t+1}\end{aligned}$$

where $t = 1, \dots,$ T and
$\mathbf{y}_t$ observed random variable (NY×1)
$\mathbf{a}_t$ latent random variable (NE×1)
$\mathbf{e}_t$ latent (residual) random (NY×1)
$\mathbf{z}_t$ latent (residual) random (NE×1)
$\mathbf{x}_t$ observed fixed exogenous regressors (NX×1)

Parameters in the state-space model

-

$$\begin{aligned}\mathbf{y}_t &= \mathbf{S}\mathbf{a}_t + \mathbf{d} + \mathbf{e}_t + \mathbf{Z}\mathbf{x}_t \\ \mathbf{a}_{t+1} &= \mathbf{H}\mathbf{a}_t + \mathbf{c} + \mathbf{G}\mathbf{z}_{t+1}\end{aligned}$$

$\mathbf{S}$ NY×NE matrix of regression pars
$\mathbf{d}$ NY×1 vector of intercepts
$\mathbf{H}$ NE×NE matrix of regression pars
$\mathbf{c}$ NE×1 vector of intercepts
$\mathbf{R}$ NY×NY matrix of covariance matrix for $\mathbf{e}$

$\mathbf{Q}$ NE×NE matrix of covariance matrix for $\mathbf{z}$
$\mathbf{Z}$ NY×NX matrix of regression pars
$\mathbf{G}$ NE×NE matrix of regression pars

Moving Average Model

- A qth order moving average process, denoted *MA*(q), is characterized by

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q}.$$

- The first-order model is simplified as

$$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

- The state-space representation of the *AR*(1) model is

$$\begin{aligned} Y_t &= \mu + (\theta \; 1) \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_t \end{pmatrix} \\ \begin{pmatrix} \varepsilon_{t-1} \\ \varepsilon_t \end{pmatrix} &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \varepsilon_{t-2} \\ \varepsilon_{t-1} \end{pmatrix} + \begin{pmatrix} 0 \\ \varepsilon_t \end{pmatrix} \end{aligned}$$

Autoregressive Model

- The pth-order autoregressive process, denoted *AR*(p), satisfies

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-1} + \ldots + \phi_p Y_{t-p} + \epsilon_t.$$

- The first-oder model is simplified as

$$Y_t = c + \phi_1 Y_{t-1} + \epsilon_t.$$

- The state-space representation for AR(1) model is

$$\begin{aligned} Y_t &= c + a_t \\ a_t &= \phi_t a_{t-1} + \epsilon_t \end{aligned}$$

Dynamic Factor Model - White Noise Factor Score Model

- The model is

$$\mathbf{y}_t = \sum_{l=0}^{L} \mathbf{\Lambda}_l \mathbf{f}_{t-l} + \mathbf{e}_t.$$

- Its state-space representation is

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{\Lambda}_0 & \mathbf{\Lambda}_1 & \dots & \mathbf{\Lambda}_L \end{pmatrix} \begin{pmatrix} vec(\mathbf{f}_t) \\ vec(\mathbf{f}_{t-1}) \\ \vdots \\ vec(\mathbf{f}_{t-L}) \end{pmatrix} + \mathbf{e}_t$$

$$\begin{pmatrix} vec(\mathbf{f}_t) \\ vec(\mathbf{f}_{t-1}) \\ \vdots \\ vec(\mathbf{f}_{t-L}) \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix} \begin{pmatrix} vec(\mathbf{f}_{t-1}) \\ vec(\mathbf{f}_{t-2}) \\ \vdots \\ vec(\mathbf{f}_{t-L-1}) \end{pmatrix} + \begin{pmatrix} vec(\mathbf{f}_t) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

- For a model with four variables, one factor, and lag one model

$$\mathbf{y}_t = \mu + \mathbf{\Lambda}_0 f_t + \mathbf{\Lambda}_1 f_{t-1} + \mathbf{e}_t.$$

- Its state-space form is

$$\mathbf{y}_t = \mu + \begin{pmatrix} \mathbf{\Lambda}_0 & \mathbf{\Lambda}_1 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_t \end{pmatrix} + \mathbf{e}_t$$

$$\begin{pmatrix} f_{t-1} \\ f_t \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} f_{t-1} \\ f_t \end{pmatrix} + \begin{pmatrix} 0 \\ f_t \end{pmatrix}$$

DFM - Direct Autoregressive Factor Score Model

- The DAFS model is

$$\mathbf{y}_t = \mathbf{\Lambda} \cdot \mathbf{f}_t + \mathbf{u}_t$$

$$\mathbf{f}_t = \sum_{l=1}^{\mathrm{L}} \mathbf{B}_l \cdot \mathbf{f}_{t-l} + \mathbf{v}_t$$

- Its state-space form is

$$\mathbf{y}_t = (\mathbf{\Lambda 0 0} \cdots \mathbf{0}) \begin{pmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \\ \vdots \\ \mathbf{f}_{t-L+1} \end{pmatrix} + \mathbf{e}_t$$

$$\begin{pmatrix} \mathbf{f}_t \\ \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \\ \vdots \\ \mathbf{f}_{t-L+1} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_1 & \mathbf{B}_2 & \cdots & \mathbf{B}_{L-1} & \mathbf{B}_L \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{f}_{t-1} \\ \mathbf{f}_{t-2} \\ \mathbf{f}_{t-3} \\ \vdots \\ \mathbf{f}_{t-L} \end{pmatrix} + \begin{pmatrix} \mathbf{v}_t \\ \mathbf{0} \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{pmatrix}$$

Scripts for one DAFS model

title example simulated !the title of the analysis

!nm: # of models; se:calculate the standard errors.

nm=1 se=yes

!mo: model index; ny: # of y in the model; ne: number of variables in $\mathbf{a}_t$; nx: # of variables in $\mathbf{x}$; it: # of iterations

mo=1 ny=6 ne=2 nx=0 it=1000

!df: specify data file, save the script and the data file and the program in the same folder. rf:a file to save the estimates for the states; ns: # of subjects; mi:missing data

df=datakf.dat rf=states ns=1 mi=-999.

!Tell the program which matrix in the full state-space model is present in the current model

S=1 R=1 H=1 Q=1 d=0 c=0 P=1 Z=0 G=1

S fi

0 0

0 0

0 0

0 0

0 0

0 0

S fr

1 0

2 0

3 0

0 4

0 5

0 6

Q fi sy

1

0 1

Q fr sy

0

7 0

G fi di

1 1

G fr

0 0

H fi fu

0 0

0 0

H fr fu

8 0

0 9

R fi di

0 0 0 0 0 0

R fr di

10 11 12 13 14 15

P fi sy

50

0 50

P fr sy

0

0 0

st

.1 -.1 -.1 -.1

.6 .1 .6

.1 .1

.1 .1 .1 .1 .1 .1 .5 .5

lb

-10 -10 -10 -10 -10 -10

-1 0 0

.0001 .00001 .0001 .0001 .0001 .0001

ub

10 10 10 10 10 10

1 1 1

10 10 10 10 10 10

Data File

50

0 0

0.530079877 -0.087123175 0.555901635 -0.015002067 0.961175609 -0.246165734

0.588365226 0.201891482 -0.078339517 -0.017637439 -0.199191429 -0.501216579

- The first line gives the sample size

- The second line gives the initial values for the states
- From the third line, every line represents the data for one subject
- **NOTE: There must have a blank line at the end of the data file.**

Run the Program

- It is suggested to put the script file, data file, and the program in the same folder.
- Only the single program file mkfm6.exe is needed.
- This program is run in a DOS window.
- Command: mkfm5 < scriptfile > outfile

Output

max nm= 5 nt=5000 ns= 10 ny=30 nx= 5 ne=30 npar=400

Read from input file

title Simulation study DFMs by Johnny Zhang

nm=1 se=yes it=1000

mo=1 ny=6 ne=2 nx=0

df=mkfm.txt rf=states ns=1 mi=-999

S=1 R=1 H=1 Q=1 Z=0 d=0 c=0 G=1 P=1

===================

MKFv1 Nov. 2002

===================

title Simulation study DFMs by Johnny Zhang

Model 1 of 1

S fr parameters (nonzero)

0 0

1 0

2 0

0 0

0 3

0 4

R fr parameters (nonzero) - diagonal

10 11 12 13 14 15

H fr parameters (nonzero)

8 0

0 9

Q fr parameters (nonzero)

5

6 7

DATA SUMMARY

MODEL 1 of 1 NY= 6 NX= 0 NE= 2 Ncases= 1 START= 1 END= 1

CASE 1 T= 50 N of T missing= 0 datafile mkfm.txt

State_0 0.00 0.00

var 1 2 3 4 5 6

%miss 0.0000 0.0000 0.0000 0.0000 0.0000 0.0000

mean 0.05 0.06 0.04 -0.09 -0.08 -0.15

var 0.37 0.41 0.27 0.67 0.56 1.14

std 0.61 0.64 0.52 0.82 0.75 1.07

min -1.17 -1.24 -1.08 -1.77 -1.79 -2.51

max 1.23 1.42 0.99 2.72 2.15 2.28

N of penalized error - during npsol

— P matrix not posdef in KF 26 times

— Density less/equal zero 12 times

ML parameter estimates

nr 1 0.48509 g 0.000002 se 0.0577 t 8.40

nr 2 0.38860 g -0.000006 se 0.0460 t 8.44

nr 3 0.61405 g 0.000007 se 0.0736 t 8.35

nr 4 0.93233 g -0.000007 se 0.0974 t 9.57

nr 5 1.12023 g -0.000002 se 0.2763 t 4.05

nr 6 0.59347 g 0.000002 se 0.1865 t 3.18

nr 7 0.88435 g -0.000001 se 0.2335 t 3.79

nr 8 0.27206 g 0.000001 se 0.1370 t 1.99

nr 9 0.40902 g 0.000000 se 0.1319 t 3.10

nr 10 0.05361 g -0.000006 se 0.0233 t 2.31

nr 11 0.11432 g 0.000001 se 0.0309 t 3.70

nr 12 0.07506 g -0.000006 se 0.0198 t 3.79

nr 13 0.13851 g 0.000000 se 0.0409 t 3.38

nr 14 0.14962 g 0.000019 se 0.0401 t 3.73

nr 15 0.19398 g -0.000002 se 0.0669 t 2.90

Logl 22.997 -2xLogL -45.995 Inform(NPSOL) 0 title Simulation study DFMs by Johnny Zhang

Model 1 of 1

S parameters

0.500 0.000

0.485 0.000

0.389 0.000

0.000 0.700

0.000 0.614

0.000 0.932

R parameters - diagonal

0.054 0.114 0.075 0.139 0.150 0.194

H parameters

0.272 0.000

0.000 0.409

Q parameters

1.120

0.593 0.884

P parameters

50.000

0.000 50.000

G parameters - diagonal

1.000 1.000

P(t|t) error cov

0.100

0.007 0.082

start: 15 11 56 end: 15 12 1 date: 23 5 2005

start getse: 15 12 0 end getse: 15 12 1

References

1. Dolan, C. V. (2005). MKFM6: Multi-group, multi-subject stationary timeseries modeling based on the Kalman filter. (This program can be downloaded from http://users.fmg.uva.nl/cdolan/.)

2. Hamaker, E.L.. Kalman filter and state-space representations.

3. Hamilton, J. D. (1994). *Time series analysis*. Princeton, N.J. : Princeton University Press.