

Technical tips on time series with Stata

Gustavo Sánchez

Senior Statistician
StataCorp

2011 Mexican Stata Users Group Meeting 2011

Outline

- Tip 1: Specifying the time structure
 - `tsset`
 - Date formats
- Tip 2: Why some predictions with `-arima-` do not match my manual computations - Kalman Filter recursions
- Tip 3: What is the initial shock for impulse response functions after `-var-`
- Tip 4: How do I fit my unobserved component model with `-sspace-`
 - linear regression and random walk
- Tip 5: How do I specify restrictions on the long-run cointegrating relationship in the VEC model

TIP 1: Specifying the time structure

- `tsset timevar [, options]`
 - Date frequency (daily, weekly, monthly,...)
 - Clocktime (hours, minutes, seconds,..., milliseconds)
 - Generic
 - `delta()`

Example:

```
tsset timevar,daily delta(7)  
lags in terms of seven days
```

TIP 1: Specifying the time structure

- Date formats
 - Example – Daily format

```
clear
input      str12   date
           "1/01/2008"
           "1/02/2008"
           "1/03/2008"
           "1/04/2008"
           "1/05/2008"

end

generate mydate1=date(date,"DMY")
format   mydate1 %td
generate mydate2=date(date,"DMY")
format   mydate2 %tdmon-DD,_CCYY
```


TIP 1: Specifying the time structure

- Date formats
 - Example – Daily format

```
. list date mydate1 mydate2
```

```
+-----+  
|          date          mydate1          mydate2          |  
+-----+  
1. | 1/01/2008    01jan2008    jan-01, 2008 |  
2. | 1/02/2008    01feb2008    feb-01, 2008 |  
3. | 1/03/2008    01mar2008    mar-01, 2008 |  
4. | 1/04/2008    01apr2008    apr-01, 2008 |  
5. | 1/05/2008    01may2008    may-01, 2008 |  
+-----+
```

TIP 1: Specifying the time structure

- Date formats
 - Example – Daily format

```
. tsset mydate1
   time variable: mydate1, 01jan2008 to 01may2008,
                   but with gaps
                   delta: 1 day

. list mydate1 if tin(01feb2008,01apr2008)
```

```
  +-----+
  |   mydate1   |
  +-----+
  2. | 01feb2008 |
  3. | 01mar2008 |
  4. | 01apr2008 |
  +-----+
```

TIP 1: Specifying the time structure

- Date formats
 - Example – Clock format

```
clear
Input          str20      etime          y
              "06feb2010 12:40:00"      2
              "06feb2010 12:42:00"      5
              "06feb2010 12:44:00"      7
              "06feb2010 12:46:00"      6
              "06feb2010 12:48:00"      9

end

generate double mytime = clock(etime, "DMY hms")
format mytime %tc DMYHH:MM:SS
```

TIP 1: Specifying the time structure

- Date formats
 - Example – Clock format

```
. tsset mytime,delta(2 minute)
   time variable: mytime, 06feb2010 12:40:00 to 06feb2010 12:48:00
   delta: 2 minutes

. generate my_ly=1.y
(1 missing value generated)

. list mytime y ly my_ly
```

```
-----+-----
|               mytime    y    my_ly |
|-----+-----|
1. | 06feb2010 12:40:00    2      . |
2. | 06feb2010 12:42:00    5      2 |
3. | 06feb2010 12:44:00    7      5 |
4. | 06feb2010 12:46:00    6      7 |
5. | 06feb2010 12:48:00    9      6 |
-----+-----
```


TIP 2: Predictions with -arima- Kalman Filter recursions

- Let's consider the following moving average (MA1) model:

$$y_t = \alpha + \theta \varepsilon_{t-1} + \varepsilon_t \quad ; \quad \varepsilon_t \sim i.i.d.N(0, \sigma^2)$$

Command line

to fit the model:

```
arima y, ma(1)
```

And we get the

predictions with:

```
predict double y_hat
```

TIP 2: Predictions with -arima- Kalman Filter recursions

- Users try to manually reproduce the predictions with:

$$\hat{Y}_t = \hat{\alpha} + \hat{\theta} \times \hat{\epsilon}_{t-1}$$

$$\hat{\epsilon}_{t-1} = Y_{t-1} - \hat{Y}_{t-1}$$

- However, the results do not match the predictions obtained with:

```
predict double y_hat
```

WHY?

TIP 2: Predictions with -arima- Kalman Filter recursions

- Code for manual predictions

```
use http://www.stata-press.com/data/r11/lutkepohl, clear
arima dlinvestment, ma(1)
predict double yhat
scalar b0 = _b[_cons]
scalar t1 = [ARMA]_b[L1.ma]
gen double my_yhat = b0
gen double myehat = dlinvestment - b0 in 2
forvalues i = 3/91 {
    qui replace my_yhat = my_yhat ///
        + t1*L.myehat in `i'
    qui replace myehat = dlinvestment - my_yhat in `i'
}
```

TIP 2: Predictions with -arima- Kalman Filter recursions

- List first 12 predictions

```
. list qtr yhat my_yhat in 1/13, sep(11)
```

```
+-----+
|      qtr      yhat      my_yhat |
+-----+
1. | 1960q1      .01686688      .01686688 |
2. | 1960q2      .01686688      .01686688 |
3. | 1960q3      .02052151      .02062398 |
4. | 1960q4      .01478403      .0147996 |
5. | 1961q1      .01312365      .01312617 |
6. | 1961q2      .00326376      .00326418 |
7. | 1961q3      .02471242      .02471249 |
8. | 1961q4      .01691061      .01691062 |
9. | 1962q1      .01412974      .01412975 |
10. | 1962q2      .00643301      .00643301 |
11. | 1962q3      .01940009      .0194001 |
+-----+
12. | 1962q4      .01649863      .01649863 |
13. | 1963q1      .01749646      .01749646 |
+-----+
```


TIP 2: Predictions with -arima- Kalman Filter recursions

- Stata uses the recursive formula for the Kalman filter prediction based on:

$$\hat{\epsilon}_{t-1} = \left[\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\theta}^2 \times p_{t-1}} \right] \times (Y_{t-1} - \hat{Y}_{t-1}) \quad \hat{\epsilon}_0 = 0$$

- Where:

$$p_t = \left[\frac{\hat{\sigma}^2 \times \hat{\theta}^2 \times p_{t-1}}{\hat{\sigma}^2 + \hat{\theta}^2 \times p_{t-1}} \right] \quad p_1 = \hat{\sigma}^2$$

$\hat{\sigma}^2$ estimated variance of the white noise disturbance

TIP 2: Predictions with -arima- Kalman Filter recursions

```
use http://www.stata-press.com/data/r11/lutkepohl,clear
arima dlinvestment, ma(1)
predict double yhat
```

** Coefficient estimates and σ^2 from ereturn list **

```
scalar b0 = _b[_cons]
scalar t1 = [ARMA]_b[L1.ma]
scalar sigma2 = e(sigma)^2
```

** pt and shrinking factor for the first two observations**

```
gen double pt=sigma2 in 1/2
gen double myratio=(sigma2)/(sigma2+t1^2*pt) in 2
```

** Predicted series and errors for the first two observations **

```
gen double my_yhat = b0
generate double myehat = myratio*(dlinvestment - my_yhat) in 2
```

** Predictions with the Kalman filter recursions **

```
forvalues i = 3/91 {
    qui replace my_yhat = my_yhat + t1*_l.myehat in `i'
    qui replace pt= (sigma2)*(t1^2)*(L.pt)/ (sigma2+t1^2*L.pt) in `i'
    qui replace myratio=(sigma2)/(sigma2+t1^2*pt) in `i'
    qui replace myehat=myratio*(dlinvestment - my_yhat) in `i'
}
```

TIP 2: Predictions with -arima- Kalman Filter recursions

- List first 10 predictions

```
. list qtr yhat my_yhat pt myratio in 1/10
```

```
+-----+
|      qtr      yhat      my_yhat      pt      myratio |
+-----+
1. | 1960q1      .01686688      .01686688      .00192542      . |
2. | 1960q2      .01686688      .01686688      .00192542      .97272668 |
3. | 1960q3      .02052151      .02052151      .00005251      .99923589 |
4. | 1960q4      .01478403      .01478403      1.471e-06      .99997858 |
5. | 1961q1      .01312365      .01312365      4.125e-08      .9999994 |
+-----+
6. | 1961q2      .00326376      .00326376      1.157e-09      .99999998 |
7. | 1961q3      .02471242      .02471242      3.243e-11      1 |
8. | 1961q4      .01691061      .01691061      9.092e-13      1 |
9. | 1962q1      .01412974      .01412974      2.549e-14      1 |
10. | 1962q2      .00643301      .00643301      7.147e-16      1 |
+-----+
```


TIP 3: Initial shock for Impulse response functions (IRF) after -var-

VAR model

$$\Delta Y_t = \alpha + \beta_1 \Delta Y_{t-1} + \beta_2 \Delta Y_{t-2} + \dots + \beta_p \Delta Y_{t-p} + \varepsilon_t$$

Where:

$Y_t = (Y_{1t}, Y_{2t}, \dots, Y_{kt})$: I(1) Endogenous variables

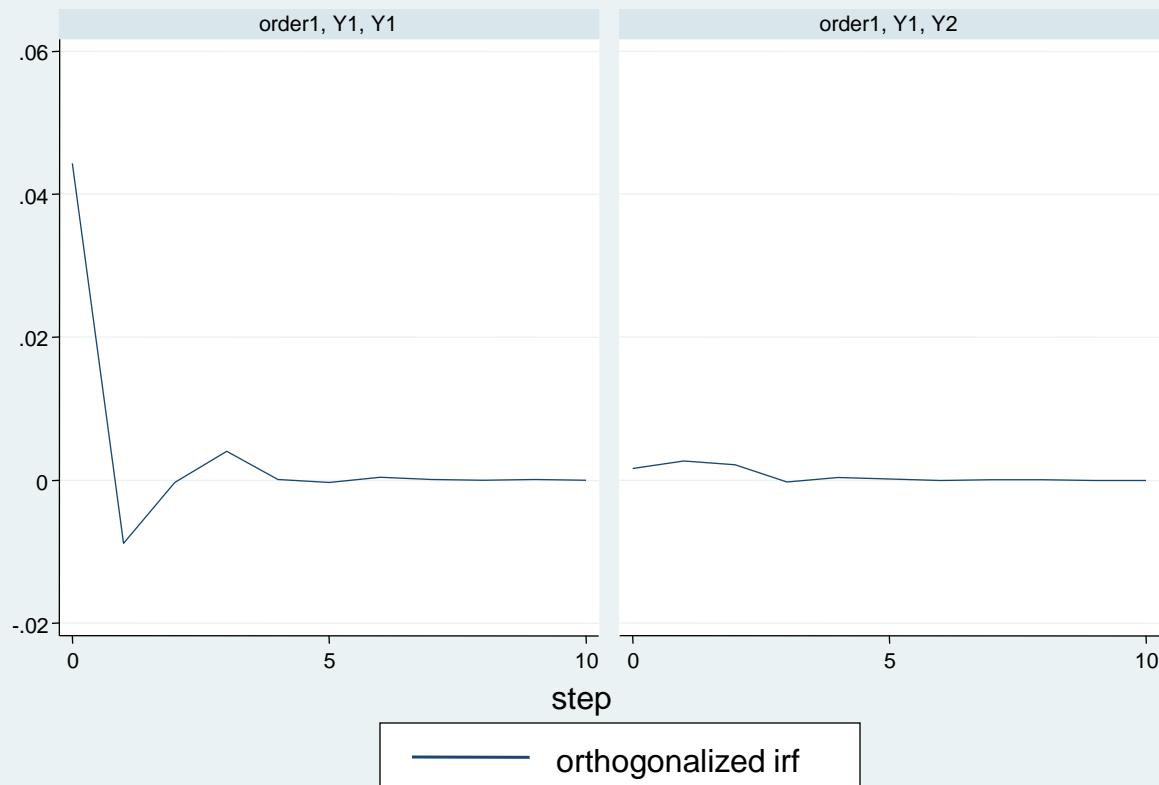
β_i : Matrix with coefficients associated to lag i

α : Vectors with coefficients associated to the intercepts

ε_t : Vector with innovations

TIP 3: Initial shock for Impulse response functions (IRF) after -var-

- Orthogonalized IRF functions for a shock in Y1

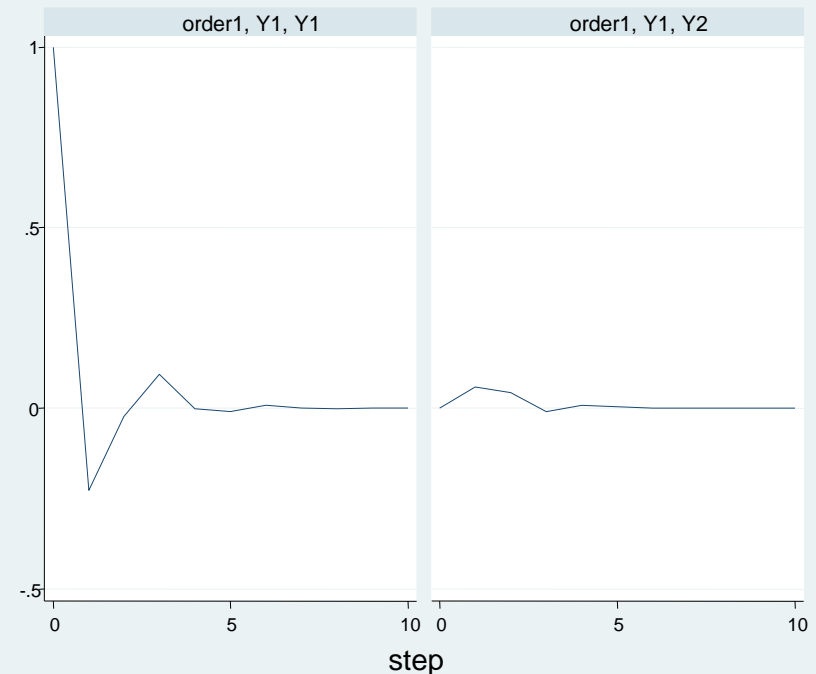


Graphs by irfname, impulse variable, and response variable

TIP 3: Initial shock for Impulse response functions after -var-

What is the magnitude of the shock in the IRF graph?

- `-irf graph,irf-`: simple IRF
 - correspond to one-time unit increase
 - the effects do not have a causal interpretation

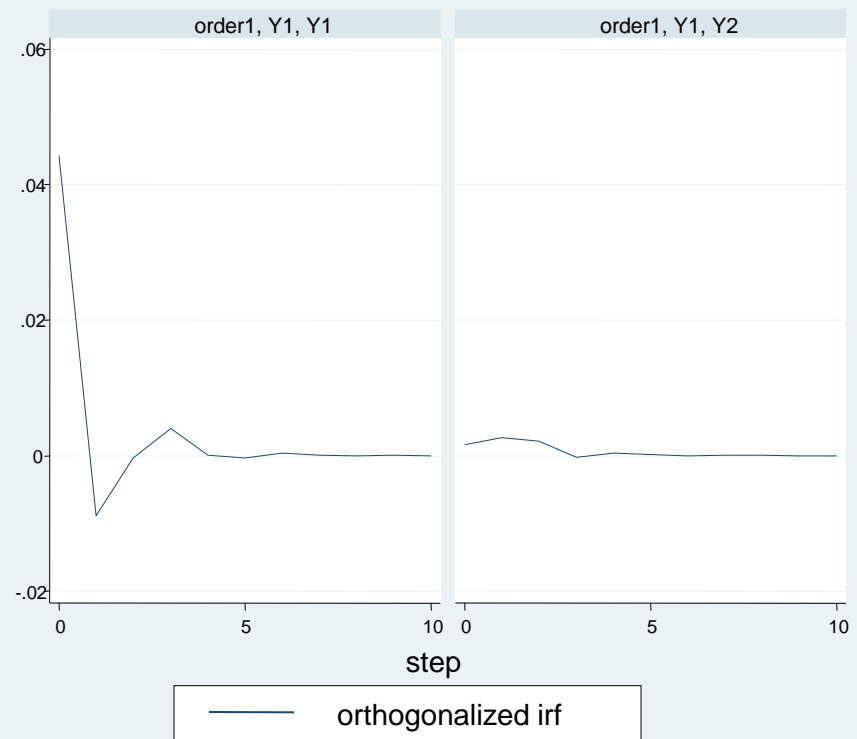


Graphs by irfname, impulse variable, and response variable

TIP 3: Initial shock for Impulse response functions after -var-

What is the magnitude of the shock in the IRF graph?

- `-irf graph,oirf-`: orthogonal IRF
 - orthogonalization is produced via the Cholesky decomposition
 - the magnitude of the shock corresponds to one unit standard deviation
- `-irf graph,sirf-` structural IRF
 - `-irf graph,sirf-` IRF functions are derived from the constraints imposed on the SVAR
 - the magnitude of the shock corresponds to one unit standard deviation



Graphs by irfname, impulse variable, and response variable

TIP 3: Initial shock for Impulse response functions after -var-

- Let's fit a VAR model:

use <http://www.stata-press.com/data/r11/lutkepohl>

```
var dlinvestment dlincome, lags(1/2) dfk
```


TIP 3: Initial shock for Impulse response functions after -var-

```
. var dlinvestment dlincome,lags(1/2) dfk
```

Vector autoregression

Equation	Parms	RMSE	R-sq	chi2	P>chi2		
dlinvestment	5	.04424	0.0856	8.32989	0.0802		
dlincome	5	.011403	0.1027	10.1916	0.0373		
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]		
dlinvestment	-----						
dlinvestment	-----						
L1.	-.2274192	.1053092	-2.16	0.031	-.4338214	-.021017	
L2.	-.1159636	.1057698	-1.10	0.273	-.3232686	.0913415	
dlincome	-----						
L1.	.7103053	.3948248	1.80	0.072	-.0635372	1.484148	
L2.	.5149489	.3935121	1.31	0.191	-.2563206	1.286218	
_cons	-.0012273	.0111362	-0.11	0.912	-.0230539	.0205993	
dlincome	-----						
dlinvestment	-----						
L1.	.0597466	.0271441	2.20	0.028	.0065451	.1129481	
L2.	.0563513	.0272629	2.07	0.039	.002917	.1097855	
dlincome	-----						
L1.	.0209461	.1017687	0.21	0.837	-.1785169	.220409	
L2.	.0833252	.1014303	0.82	0.411	-.1154745	.2821249	
_cons	.0150368	.0028704	5.24	0.000	.0094108	.0206627	

TIP 3: Initial shock for Impulse response functions after -var-

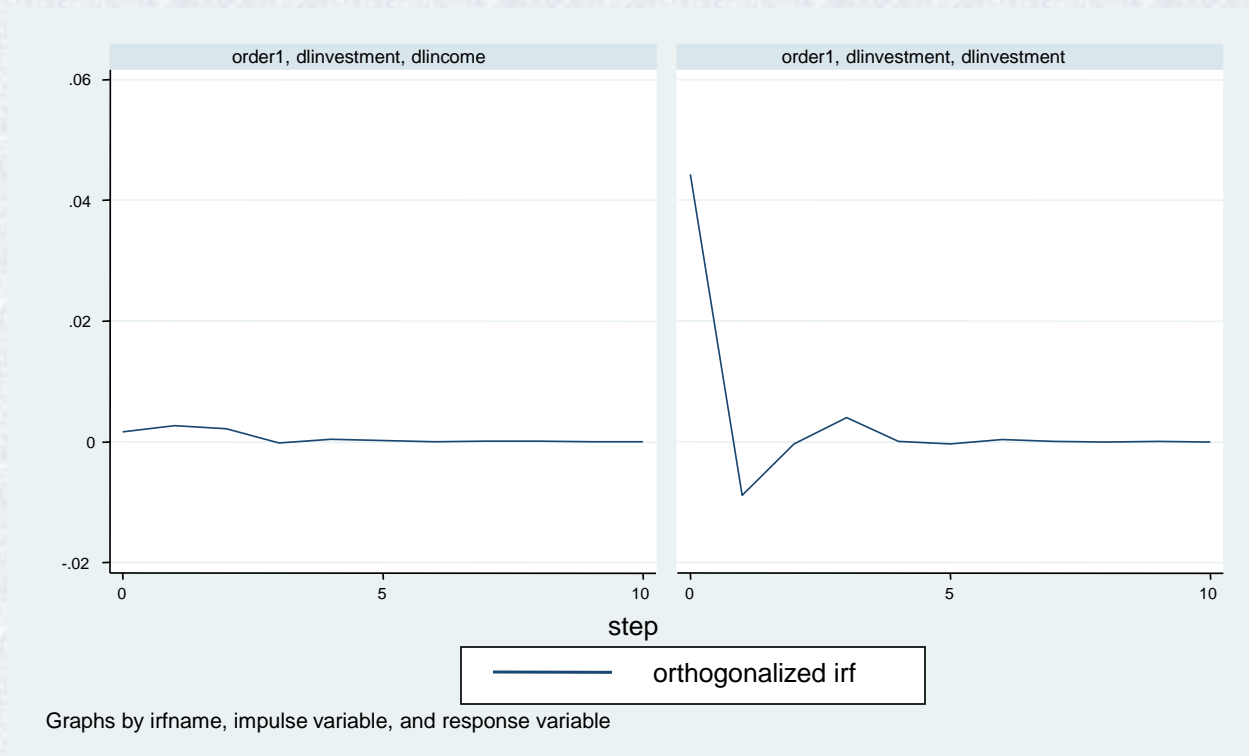
- Plot the IRF function for a shock in dlinvestment

```
use http://www.stata-press.com/data/r11/lutkepohl
var dlinvestment dlincome, lags(1/2) dfk
irf create order1, step(10) set(myirf1,replace)
irf graph oirf, impulse(dlinvestment) ///
      response(dlinvestment dlincome)
```

TIP 3: Initial shock for Impulse response functions after -var-

- Plot the IRF function for a shock in dlinvestment

```
irf graph oirf, impulse(dlinvestment) ///  
response(dlinvestment dlincome)
```



TIP 3: Initial shock for Impulse response functions after -var-

- Table for the OIRF function for a shock in dlinvestment

```
. irf table oirf, irf(order1) impulse(dlinvestment) ///
> response(dlincome dlinvestment)
```

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)
step	oirf	Lower	Upper	oirf	Lower	Upper
0	.001641	-.000715	.003998	.04424	.037741	.050739
1	.002678	.000241	.005114	-.008895	-.018459	.000669
2	.002154	-.000283	.004592	-.00036	-.009879	.009159
3	-.000255	-.001358	.000849	.004022	-.001068	.009113
4	.000394	-.000347	.001136	.000056	-.002258	.00237
5	.000217	-.000207	.000641	-.00033	-.002245	.001585
6	.000021	-.000237	.000279	.000426	-.000457	.001309
7	.000025	-.000101	.000152	.000068	-.000353	.000488
8	.00003	-.000055	.000116	-.000036	-.000356	.000284
9	4.4e-06	-.000034	.000043	.000035	-.000074	.000143
10	2.7e-06	-.000022	.000027	.000015	-.000063	.000093

95% lower and upper bounds reported

(1) irfname = order1, impulse = dlinvestment, and response = dlincome

(2) irfname = order1, impulse = dlinvestment, and response = dlinvestment

TIP 3: Initial shock for Impulse response functions after -var-

- Table for the IRF function for a shock in dlinvestment

```
. irf table irf, irf(order1) impulse(dlinvestment) ///
> response(dlincome dlinvestment)
```

Results from order1

	(1)	(1)	(1)	(2)	(2)	(2)
step	irf	Lower	Upper	irf	Lower	Upper
0	0	0	0	1	1	1
1	.059747	.004985	.114509	-.227419	-.439876	-.014963
2	.044015	-.010388	.098419	-.021806	-.237257	.193646
3	-.008218	-.032283	.015847	.093362	-.027102	.213826
4	.007845	-.007056	.022745	-.001875	-.054015	.050264
5	.004629	-.004709	.013967	-.00906	-.054602	.036483
6	.000104	-.005125	.005332	.009605	-.010735	.029945
7	.000451	-.002119	.003022	.001323	-.00833	.010977
8	.000638	-.001136	.002413	-.001041	-.008544	.006462
9	.000063	-.000688	.000814	.000769	-.001641	.003179
10	.000042	-.000454	.000538	.00032	-.001466	.002105

95% lower and upper bounds reported

(1) irfname = order1, impulse = dlinvestment, and response = dlincome

(2) irfname = order1, impulse = dlinvestment, and response = dlinvestment

TIP 4: How do I fit my unobserved component model with `-sspace-`

- State Space representation

$$z_t = Az_{t-1} + Bx_t + C\varepsilon_t$$

$$y_t = Dz_t + Fw_t + Gu_t$$

Where:

z_t : is an $m \times 1$ vector of unobserved state variables;

x_t : is a $k_x \times 1$ vector of exogenous variables;

ε_t : is a $q \times 1$ vector of state-error terms, ($q \leq m$);

y_t : is an $n \times 1$ vector of observed endogenous variables;

w_t : is a $k_w \times 1$ vector of exogenous variables;

u_t : is an $r \times 1$ vector of observation-error terms, ($r \leq n$); and

A, B, C, D, F, and G are parameter matrices.

TIP 4: How do I fit my unobserved component model with `–sspace–`

- State Space representation for linear regression

$$z_t = \varepsilon_t$$

$$y_t = z_t + \alpha + \beta w_t$$

Command specification

```
constraint 1 [z]L.z = 0
```

```
constraint 2 [y]z = 1
```

```
sspace (z L.z, state noconstant) ///  
      (y w z,noerror ), constraints(1/2)
```


TIP 4:

■ State Space estimation for linear regression

use <http://www.stata-press.com/data/r11/lutkepohl,clear>

constraint 1 [z]L.z = 0

constraint 2 [dlinvestment]z = 1

sspace (z L.z, state noconstant) (dlinvestment dlincome z,noerror), constraints(1/2) nolog

State-space model

Sample: 1960q2 - 1982q4

Number of obs = 91

Wald chi2(1) = 0.88

Log likelihood = 154.44197

Prob > chi2 = 0.3487

(1) [z]L.z = 0

(2) [dlinvestment]z = 1

```
-----+-----
```

		OIM				
dlinvestment	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
z						
	L1.	(omitted)				
-----+-----						
dlinvestment						
	z	1
dlincome		.3668678	.3914794	0.94	0.349	-.4004178 1.134153
_cons		.0096556	.008925	1.08	0.279	-.007837 .0271483
-----+-----						
var(z)		.0019651	.0002913	6.75	0.000	.0013941 .0025361
-----+-----						

Note: Tests of variances against zero are conservative and are provided only for reference.

TIP 4: How do I fit my unobserved component model with `-sspace-`

- Random Walk $y_t = y_{t-1} + \varepsilon$

- State Space representation

$$z_t = z_{t-1} + \varepsilon_t$$

$$y_t = z_t$$

- Command specification

```
constraint 1 [z]L.z = 1
```

```
constraint 2 [y]z = 1
```

```
sspace (z L.z, state noconstant) ///
       (y z,noerror noconstant), constraints(1/2)
```

TIP 4:

■ State Space estimation for Random Walk

use <http://www.stata-press.com/data/r11/lutkepohl,clear>

constraint 1 [z]L.z = 1

constraint 2 [dlinvestment]z = 1

sspace (z L.z, state noconstant) (dlinvestment z,noerror noconstant), constraints(1/2) nolog

State-space model

Sample: 1960q2 - 1982q4

Number of obs = 91

Log likelihood = 112.76541

(1) [z]L.z = 1

(2) [dlinvestment]z = 1

```
-----+-----
```

		OIM				
		Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
dlinvestment						
	z					
	L1.	1
dlinvestment						
	z	1
var(z)		.0046812	.0006978	6.71	0.000	.0033135 .006049

```
-----+-----
```

Note: Model is not stationary.

Note: Tests of variances against zero are conservative and are provided only for reference.

TIP 5:VEC – Johansen identification

Reduced form for a VEC model

$$\Delta z_t = a + bt - \alpha\beta z_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta z_{t-i} + v_t$$

Where:

z_t : I(1) Endogenous variables

$\alpha\beta$: Matrices containing the long-run adjustment coefficients and coefficients for the cointegrating relationships

Γ_i : Matrix with coefficients associated to short-run dynamic effects

a, b : Vectors with coefficients associated to the intercepts and trends

v_t : Vector with innovations

TIP 5:

Example: VEC with three endogenous variables

$$\begin{bmatrix} \Delta z_{1t} \\ \Delta z_{2t} \\ \Delta z_{3t} \end{bmatrix} = a + bt - \alpha\beta \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \\ z_{3t-1} \end{bmatrix} + \sum_{i=1}^{p-1} \Gamma_i \begin{bmatrix} \Delta z_{1t-i} \\ \Delta z_{2t-i} \\ \Delta z_{3t-i} \end{bmatrix} + v_t$$

Where:

$$\alpha\beta z_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \\ z_{3t-1} \end{bmatrix}$$

- Identifying α and β requires r^2 restrictions (r : number of cointegrating vectors).
- Johansen FIML estimation identifies α and β by imposing r^2 atheoretical restrictions.

$$\alpha\beta z_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} 1 & 0 & \beta_{13} \\ 0 & 1 & \beta_{23} \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \\ z_{3t-1} \end{bmatrix}$$

TIP 5:

- Restrictions based on Johansen normalization (Default)

use <http://www.stata-press.com/data/r11/lutkepohl>,clear

vec linvestment lincome lconsumption, rank(2) lags(2) noetable trend(none)

Vector error-correction model

```

      .
      .
      .
      .
      .
  
```

Identification: beta is exactly identified

Johansen normalization restrictions imposed

beta	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	

_ce1						
linvestment	1
lincome	(omitted)					
lconsumption	-.7943718	.0125908	-63.09	0.000	-.8190493	-.7696942

_ce2						
linvestment	(omitted)					
lincome	1
lconsumption	-1.013321	.0013846	-731.87	0.000	-1.016035	-1.010608

TIP 5:

Instead of the Johansen atheoretical restrictions we could use economic theory to impose restrictions to identify $\alpha\beta$.

For example, let's assume the following cointegrating equations:

$$z_{1t-1} = 0.75z_{2t-1} + \beta_{13}z_{3t-1}$$

$$z_{3t-1} = .85z_{2t-1} + \beta_{21}z_{1t-1}$$

Which implies

$$\alpha\beta z_{t-1} = \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \\ \alpha_{31} & \alpha_{32} \end{bmatrix} \begin{bmatrix} 1 & -.75 & \beta_{13} \\ \beta_{21} & -.85 & 1 \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \\ z_{3t-1} \end{bmatrix}$$

TIP 5:

- Restrictions specified by the user

```

constraint define 1 [_ce1]linvestment=1
constraint define 2 [_ce1]lincome=-.75
constraint define 3 [_ce2]lconsumption=1
constraint define 4 [_ce2]lincome=-.85
vec investment lincome lconsumption, rank(2) lags(2) noetable trend(none) bconstraints(1/4)
    
```

Identification: beta is exactly identified

- (1) [_ce1]linvestment = 1
- (2) [_ce1]lincome = -.75
- (3) [_ce2]lconsumption = 1
- (4) [_ce2]lincome = -.85

beta	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
-----+-----						
_ce1						
linvestment	1
lincome	-.75
lconsumption	-.0343804	.0122816	-2.80	0.005	-.0584519	-.010309
-----+-----						
_ce2						
linvestment	-.1745742	.00322	-54.22	0.000	-.1808852	-.1682632
lincome	-.85
lconsumption	1
-----+-----						

Summary

- Tip 1: Specifying the time structure
- Tip 2: Predictions with `–arima-`. Kalman Filter recursions
- Tip 3: Initial shock for Impulse response functions after `-var-`
- Tip 4: Unobserved component models with `–sspace-`
- Tip 5: Restrictions on cointegrating relationship for VEC models

Technical tips on time series with Stata

Gustavo Sánchez

Senior Statistician
StataCorp

2011 Mexican Stata Users Group Meeting 2011