Assignment 4: SURE, Systems of equations and panel data

Barcelona Graduate School of Economics
Msc in Competition and Market Regulation

Course: Econometrics
Prof.: Javier Gómez Biscarri

Bendegúz István Biksi, Xia Mengzhen

November 25, 2015

1) SURE example.

Estimation results from the 25 separate OLS regressions.

Variables	(1) agr	(2) psg	(3) hrr	(4) pop	(5) gas
market	1.579***	1.293***	0.647***	0.906***	0.933***
	(0.389)	(0.220)	(0.237)	(0.112)	(0.181)
Constant	-0.0270	-0.00688	0.0104	0.00951	0.0141
	(0.0207)	(0.0117)	(0.0126)	(0.00598)	(0.00964)
Observations	60	60	60	60	60
R-squared	0.221	0.374	0.113	0.529	0.314

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Variables	(6) ura	(7) azk	(8) sch	(9) vid	(10) upl
market	1.735***	1.668***	1.012***	1.095***	1.272***
	(0.219)	(0.298)	(0.137)	(0.283)	(0.207)
Constant	-0.0182	0.0144	0.000633	0.0138	-0.0150
	(0.0117)	(0.0159)	(0.00727)	(0.0151)	(0.0110)
Observations	60	60	60	60	60
R-squared	0.519	0.351	0.486	0.205	0.394

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

	(11)	(12)	(13)	(14)	(15)
Variables	cgi	alb	eur	ibe	tud
market	0.823***	1.013***	0.755***	1.025***	1.012***
	(0.227)	(0.166)	(0.128)	(0.108)	(0.281)
Constant	-0.00541	-0.00101	0.00850	0.00323	-0.00798
	(0.0121)	(0.00882)	(0.00682)	(0.00573)	(0.0150)
Observations	60	60	60	60	60
R-squared	0.185	0.392	0.374	0.610	0.182

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

	(16)	(17)	(18)	(19)	(20)
Variables	caf	bkt	urb	fnz	ele
market	0.658***	1.190***	1.355***	0.696**	0.993***
	(0.241)	(0.102)	(0.246)	(0.268)	(0.109)
Constant	0.00412	-0.000167	-0.0174	-0.00818	0.0106*
	(0.0128)	(0.00540)	(0.0131)	(0.0143)	(0.00581)
Observations	60	60	60	60	60
R-squared	0.114	0.703	0.344	0.104	0.588

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

	(21)	(22)	(23)	(24)	(25)
Variables	tef	nan	ffr	byb	mvc
market	0.929***	0.826***	0.573**	0.274	1.253***
	(0.125)	(0.229)	(0.239)	(0.243)	(0.153)
Constant	0.00526	-0.00706	-0.0270**	-0.00111	-0.0132
	(0.00668)	(0.0122)	(0.0127)	(0.0129)	(0.00814)
Observations	60	60	60	60	60
R-squared	0.486	0.184	0.090	0.021	0.537

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Estimation results from the Seemingly Unrelated Regression estimation.

	(1)	(2)	(3)	(4)	(5)
Variables	agr	psg	hrr	pop	gas
market	1.579***	1.293***	0.647***	0.906***	0.933***
	(0.383)	(0.216)	(0.233)	(0.111)	(0.178)
Constant	-0.0270	-0.00688	0.0104	0.00951	0.0141
	(0.0204)	(0.0115)	(0.0124)	(0.00588)	(0.00947)
	(6)	(7)	(8)	(9)	(10)
Variables	ura	azk	sch	vid	upl
market	1.735***	1.668***	1.012***	1.095***	1.272***
	(0.216)	(0.293)	(0.134)	(0.278)	(0.204)
Constant	-0.0182	0.0144	0.000633	0.0138	-0.0150
	(0.0115)	(0.0156)	(0.00715)	(0.0148)	(0.0108)
	(11)	(12)	(13)	(14)	(15)
Variables	cgi	alb	eur	ibe	tud
market	0.823***	1.013***	0.755***	1.025***	1.012***
	(0.223)	(0.163)	(0.126)	(0.106)	(0.277)
Constant	-0.00541	-0.00101	0.00850	0.00323	-0.00798
	(0.0119)	(0.00867)	(0.00671)	(0.00563)	(0.0147)
	(16)	(17)	(18)	(19)	(20)
Variables	caf	bkt	urb	fnz	ele
market	0.658***	1.190***	1.355***	0.696***	0.993***
	(0.237)	(0.0998)	(0.242)	(0.264)	(0.107)
Constant	0.00412	-0.000167	-0.0174	-0.00818	0.0106*
	(0.0126)	(0.00531)	(0.0129)	(0.0140)	(0.00571)
	(21)	(22)	(23)	(24)	(25)
Variables	tef	nan	ffr	byb	mvc
market	0.929***	0.826***	0.573**	0.274	1.253***
	(0.123)	(0.225)	(0.235)	(0.239)	(0.150)
Constant	0.00526	-0.00706	-0.0270**	-0.00111	-0.0132
	(0.00657)	(0.0120)	(0.0125)	(0.0127)	(0.00801)

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The estimated sensitivities (coefficient of the market variable) and their standard errors are pretty much the same in the above two method of estimation (OLS: equation by equation and

SUR). So the gain by using SUR is minor. Therefore we can say that the disturbances of the different stocks' excess return are weakly correlated. The estimated sensitivities are close to one. 13 stock has higher, and 12 stock has lower sensitivity than 1. The avarage sensitivity estimated by SUR is 1.021. According to the SUR estimate for testing the null hyptohesis that all the constants are 0's at the same time, we ran the appropiate command and obatined the following results: chi2(25) = 45.69 with Prob > chi2 = 0.0070. That is to say, one is able to reject the null hypothesis, since the p-value of the test-statistics is lower, than .05, so the coefficients of the constants across all the equations are not 0's at the same time. In other words the CAPM model does not make much sense, since it implies that, the intercepts should be 0.

2) Systems of equations.

Demand: $Y_t = \beta_0 + \beta_1 * P_t + \beta_2 * cpi_t + \beta_3 * income_t + e_{1,t}$

Supply: $P_t = \alpha_0 + \alpha_1 * Y_t + \alpha_2 * l_t + \alpha_3 * nptcost_t + e_{2,t}$

Where:
Y – total output
P – price of output
nptcost – input cost index
cpi – aggregate price index
l - land
income – a measure of consumer income
t- runs from 1960-1986

a.

Study the identification of the parameters in the two equations of the system.

In the demand equation there is one included endogenous variable (P) and two excluded exogenous variables (land and inputcost). Since the number of the excluded exogenous variables, 2 is higher than the number of the included endogenous variables 1, the order condition is satisfied in this case. Now the rank condition is also satisfied since the matrix of the parameters $[\alpha_2 \quad \alpha_3]$ has a rank 1, that is (higher or) equal to the number of the included endogenous variables, that is also 1.

Now let us consider the supply equation. In this equation there is one included endogenous variable (Y) and two excluded exogenous variables (cpi and income). So here the order condition is satisfied, because again the number of the excluded exogenous variables, 2 is higher than the number of the included endogenous variables 1. Now the matrix of $[\beta_2 \quad \beta_3]$ has a rank equal to 1, that is (higher or) equal to the number of the included endogenous variables, that is also 1, so the rank condition is also satisfied. Since both equations satisfy both the order and the rank conditions, the whole system of the two equations are identifiable, and all the parameters can be recovered in theory.

b.

Decide whether you want to estimate the system in levels or in logs. Then estimate the system via 2SLS and 3SLS. Are the estimates different?

We decided to use the log-log scenario, since with this assumption, we have constant elasticity across the whole demand/supply function, plus the coefficients of the estimations are directly elasticities. First, we estimated the system via 2sls.

Figure 1: 2sls with logs

```
. reg3 (demand: l_y = l_p l_cpi l_income) (supply: l_p = l_y l_l l_nptcost), 2sls
```

Two-stage least-squares regression

Equation	Obs	Parms	RMSE	"R-sq"	F-Stat	P
demand	27	3	.0655471	0.8748	53.90	0.0000
supply	27	3	.1282061	0.9074	76.54	0.0000

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
demand						
l_p	.094473	.0884378	1.07	0.291	-.083543	.2724891
l_cpi	.1436712	.0902859	1.59	0.118	-.038065	.3254073
l_income	.3583597	.1831878	1.96	0.057	-.0103782	.7270975
_cons	.1114996	1.333797	0.08	0.934	-2.573294	2.796293
supply						
l_y	-1.13194	1.351815	-0.84	0.407	-3.853003	1.589122
l_l	.2092298	.3299442	0.63	0.529	-.4549135	.873373
l_nptcost	.8832858	.3260891	2.71	0.009	.2269026	1.539669
_cons	4.733115	4.576647	1.03	0.306	-4.479196	13.94543

Endogenous variables: l_y l_p
Exogenous variables: l_cpi l_income l_l l_nptcost

Second, we estimated the system via 3sls.

Figure 2: 3sls with logs

```
. reg3 (demand: l_y = l_p l_cpi l_income) (supply: l_p = l_y l_l l_nptcost)

Three-stage least-squares regression
```

Equation	Obs	Parms	RMSE	"R-sq"	chi2	P
demand	27	3	.0612986	0.8714	191.32	0.0000
supply	27	3	.1110637	0.9184	268.71	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
demand						
l_p	-.0140798	.0773194	-0.18	0.856	-.1656231	.1374635
l_cpi	.2516406	.0791644	3.18	0.001	.0964813	.4067999
l_income	.3032877	.168553	1.80	0.072	-.0270701	.6336456
_cons	.5408966	1.226682	0.44	0.659	-1.863356	2.94515
supply						
l_y	-.7704187	1.219009	-0.63	0.527	-3.159633	1.618796
l_l	.0673837	.2860986	0.24	0.814	-.4933592	.6281266
l_nptcost	.9567521	.2960769	3.23	0.001	.3764521	1.537052
_cons	3.37533	4.104321	0.82	0.411	-4.668991	11.41965

```
Endogenous variables:  l_y l_p
Exogenous variables:   l_cpi l_income l_l l_nptcost
```

One is able to see, that in both estimations (2sls and 3sls) we have more insignificant explanatory variable than significant. What is more the majority of the independent variables in both estimations are insignificant. However the 3sls estimation produced more significant variables than the 2sls. The estimated coefficients are different across the two method. There is only one reasonable variable in the 2sls case, that is the logarithm of the input costs, and 2 in the 3sls case, that are the logarithm of the consumer price index and the logarithm of the input costs. According to the coefficient of the inputcosts, it is similar in both estimations, but again the whole structure looks like unreasonable, although the 3sls produces more significant explanatory power, compared to the 2sls process.

c.

What are the implied price elasticities of supply[1] and demand? And the income elasticity of demand? And the cross-price elasticity?

[1] This is calculated by $1/\alpha_1$, since the supply equation should be inverted to calculate the own price elasticity.

Elasticity of demand	2sls	3sls
Own Price	.094	-.014
Income	.358	.303
Cross Price	.144	.252
Elasticity of supply		
Own Price	-.883	-1.298

d.

Do you think that the data are correctly identifying the demand and supply equations?

Unfortunately not, presumable because the sample size is only 27. IV estimations and here 2sls/3sls would need a higher sample size in order to get significant estimates of the parameters. Note that the table above contains the estimated elasticities. However only one of them is significant in 5% significance level. That is the cross price elasticity of other goods with the 3sls method. Since it is greater than zero, one is able to say, that the agricultural goods and the "other" goods are substitutes, since if the price of other goods increases by 1 percent, the quantity demanded for agricultural goods is also increasing by .3 percent, but again this is almost the only useable result of the whole estimation process.

3) Panel Data Example.

We will examine a simple model of total cost of production, of the form:

$$lnc_{i,t} = \beta_0 + \beta_1 * lnq_{i,t} + \beta_2 * lnpf_{i,t} + \beta_3 * lf_{i,t} + e_{i,t}$$

where:

i – Airline identifier

t – Time period identifier

lnq – log of the output, in revenue passenger miles (index number)

lc – log of total cost, in $1000

lpf – log of fuel price

lf – load factor, average capacity utilization of the fleet.

a.

Estimate the equation by pooled OLS. Knowing that the economies of scale can be measured by 1/eqc-1, where eqc is the output elasticity of cost, comment on what the OLS results imply about the existence of economies of scale. Comment also on the other two coefficients.

Figure 3: Pooled OLS

```
. reg l_c l_q l_pf lf
```

Source	SS	df	MS
Model	112.705452	3	37.5684839
Residual	1.33544153	86	.01552839
Total	114.040893	89	1.28135835

Number of obs =	90
F(3, 86) =	2419.34
Prob > F =	0.0000
R-squared =	0.9883
Adj R-squared =	0.9879
Root MSE =	.12461

l_c	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
l_q	.8827385	.0132545	66.60	0.000	.8563895	.9090876
l_pf	.453977	.0203042	22.36	0.000	.4136136	.4943404
lf	-1.62751	.345302	-4.71	0.000	-2.313948	-.9410727
_cons	9.516923	.2292445	41.51	0.000	9.0612	9.972645

According to the definition of economic scale, we can calculate that $1/e_{qc} - 1$ equal to 0.1328, that is positive. In another words, it is increasing economics of scale. When we increase the output, the avarage cost will decrease. If the air company sell more tickets, namely there are less vacancies for a flight, the company can reduce its fix cost (e.g. the investment in

purchasing and fixing airplanes) and variable cost (e.g. the salary of workers, the cost on fuel, etc.). Therefore, with the increasing of output, the company will be benefit more.
As for the load factor it has a significant impact on the cost. It's Coef. (-1.62751) is negative, that is to say, when the average capacity increases, namely more and more vacancies are used, the average cost will decrease to a large degree compared to other factors, because it takes advantage of space of the airplane efficiently to reduce the fixed cost and variable cost.

The fuel price (Coef. 0.4597) has positive impact on the cost. In other words, when the price of fuel increases, the cost of running an airplane will also increases.

b.

Estimate now the model again as pooled OLS but where you:

1. Include firm specific intercepts (firm dummies) and test for the joint significance of the effects. Remember that these estimates are inconsistent, but we do not want to worry much about that.

Figure 4: Pooled OLS with firm dummies

```
. xi: regress l_c l_q l_pf lf i.i
i.i               _Ii_1-6             (naturally coded; _Ii_1 omitted)

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F(  8,    81) = 3935.79
       Model |   113.74827     8  14.2185338           Prob > F      =  0.0000
    Residual |  .292622872    81  .003612628           R-squared     =  0.9974
-------------+------------------------------           Adj R-squared =  0.9972
       Total |  114.040893    89  1.28135835           Root MSE      =  .06011

------------------------------------------------------------------------------
         l_c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         l_q |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        l_pf |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
          lf |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _Ii_2 |  -.0412359   .0251839    -1.64   0.105    -.0913441    .0088722
       _Ii_3 |  -.2089211   .0427986    -4.88   0.000    -.2940769   -.1237652
       _Ii_4 |   .1845557   .0607527     3.04   0.003     .0636769    .3054345
       _Ii_5 |   .0240547   .0799041     0.30   0.764    -.1349293    .1830387
       _Ii_6 |   .0870617   .0841995     1.03   0.304     -.080469    .2545924
       _cons |   9.705942    .193124    50.26   0.000     9.321686     10.0902
------------------------------------------------------------------------------
```

We can reject the null hypothesis, since the value of the test-statistics is lower then .05. We reject that, firms have the same unobserved time invariant characteristics.

Figure 5: Test for the joint significance of the firm dummies

```
. testparm _Ii*

 ( 1)  _Ii_2 = 0
 ( 2)  _Ii_3 = 0
 ( 3)  _Ii_4 = 0
 ( 4)  _Ii_5 = 0
 ( 5)  _Ii_6 = 0

       F(  5,    81) =   57.73
            Prob > F =    0.0000
```

2. Include time effects but not firm effects (i.e. include a dummy for each period of time). Test for the joint significance of the time effects.

Test for the joint significance of the 14 year dummies. Here we can not reject the null hypothesis, since the value of the test-statistics is higher than .05. So we can not reject, that the time effect is constant over time.

Figure 6: Pooled OLS with year dummies

```
. xi: regress l_c l_q l_pf lf i.t
i.t               _It_1-15            (naturally coded; _It_1 omitted)

      Source |       SS       df       MS              Number of obs =      90
-------------+------------------------------           F( 17,     72) =  439.62
       Model |  112.952703    17  6.64427664           Prob > F      =  0.0000
    Residual |  1.08819022    72  .015113753           R-squared     =  0.9905
-------------+------------------------------           Adj R-squared =  0.9882
       Total |  114.040893    89  1.28135835           Root MSE      =  .12294

------------------------------------------------------------------------------
         l_c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         l_q |   .8677268   .0154082    56.32   0.000     .8370111    .8984424
        l_pf |  -.4844835   .3641085    -1.33   0.188    -1.210321    .2413535
          lf |  -1.954404   .4423777    -4.42   0.000    -2.836268    -1.07254
       _It_2 |   .0822341   .0720337     1.14   0.257    -.0613624    .2258306
       _It_3 |   .1599262   .0745989     2.14   0.035      .011216    .3086365
       _It_4 |   .2449538   .0813503     3.01   0.004     .0827849    .4071227
       _It_5 |   .7040272   .2425407     2.90   0.005     .2205312    1.187523
       _It_6 |    .915816   .3369745     2.72   0.008     .2440696    1.587562
       _It_7 |   1.007546   .3690931     2.73   0.008     .2717724     1.74332
       _It_8 |   1.158223   .4197076     2.76   0.007     .3215517    1.994895
       _It_9 |   1.333766   .4541888     2.94   0.004     .4283574    2.239175
      _It_10 |   1.617998   .5881371     2.75   0.008     .4455679    2.790428
      _It_11 |   1.969523   .7439351     2.65   0.010     .4865151     3.45253
      _It_12 |   2.155532   .8023798     2.69   0.009     .5560174    3.755047
      _It_13 |    2.12075   .7800367     2.72   0.008     .5657757    3.675725
      _It_14 |   2.056424   .7500037     2.74   0.008     .5613184    3.551529
      _It_15 |   2.040961   .7346895     2.78   0.007     .5763841    3.505538
       _cons |    20.4958   4.209528     4.87   0.000     12.10426    28.88735
------------------------------------------------------------------------------
```

Figure 7: Test for the joint significance of the year dummies

```
. testparm _It*

 ( 1)  _It_2 = 0
 ( 2)  _It_3 = 0
 ( 3)  _It_4 = 0
 ( 4)  _It_5 = 0
 ( 5)  _It_6 = 0
 ( 6)  _It_7 = 0
 ( 7)  _It_8 = 0
 ( 8)  _It_9 = 0
 ( 9)  _It_10 = 0
 (10)  _It_11 = 0
 (11)  _It_12 = 0
 (12)  _It_13 = 0
 (13)  _It_14 = 0
 (14)  _It_15 = 0

       F( 14,    72) =    1.17
            Prob > F =    0.3178
```

c.

Estimate now the model as a random-effects panel and a fixed-effects panel, with no time effects. Check that the estimates of $\beta 1$, β_2 and β_3 in fixed effects are the same as those in b-1.

Figure 8: Random effects

```
. xtreg l_c l_q l_pf lf, re

Random-effects GLS regression                   Number of obs      =        90
Group variable: i                               Number of groups   =         6

R-sq:  within  = 0.9925                         Obs per group: min =        15
       between = 0.9856                                        avg =      15.0
       overall = 0.9876                                        max =        15

                                                Wald chi2(3)       =  11091.33
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
         l_c |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         l_q |   .9066805    .025625    35.38   0.000     .8564565    .9569045
        l_pf |   .4227784   .0140248    30.15   0.000     .3952904    .4502665
          lf |  -1.064499   .2000703    -5.32   0.000    -1.456629    -.672368
       _cons |   9.627909    .210164    45.81   0.000     9.215995    10.03982
-------------+----------------------------------------------------------------
     sigma_u |  .12488859
     sigma_e |  .06010514
         rho |  .81193816   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Figure 9: Fixed effects

```
. xtreg l_c l_q l_pf lf, fe

Fixed-effects (within) regression               Number of obs      =        90
Group variable: i                               Number of groups   =         6

R-sq:  within  = 0.9926                         Obs per group: min =        15
       between = 0.9856                                        avg =      15.0
       overall = 0.9873                                        max =        15

                                                F(3,81)            =   3604.80
corr(u_i, Xb)  = -0.3475                        Prob > F           =    0.0000

------------------------------------------------------------------------------
         l_c |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         l_q |   .9192846   .0298901    30.76   0.000     .8598126    .9787565
        l_pf |   .4174918   .0151991    27.47   0.000     .3872503    .4477333
          lf |  -1.070396     .20169    -5.31   0.000    -1.471696   -.6690963
       _cons |   9.713528    .229641    42.30   0.000     9.256614    10.17044
-------------+----------------------------------------------------------------
     sigma_u |   .1320775
     sigma_e |  .06010514
         rho |  .82843653   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(5, 81) =    57.73               Prob > F = 0.0000
```

The estimated results of $\beta 1$, β_2 and β_3 are the same with fixed effects and the pooled OLS that contains the firm specific dummies.

d.

Perform the Breusch-Pagan test for random effects versus no random effects. What is the conclusion of the test?

Figure 10: Breusch-Pagan test for random effects

```
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

        l_c[i,t] = Xb + u[i] + e[i,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                     l_c |   1.281358       1.131971
                       e |   .0036126       .0601051
                       u |   .0155972       .1248886

        Test:   Var(u) = 0
                             chibar2(01) =   334.85
                          Prob > chibar2 =   0.0000
```

We can reject the null hypothesis, since the value of the test-statistics is higher than .05. That means we reject that, the unobserved heterogenity is the same for all firms. Therefore one should not use the simple pooled OLS method. Fixed effect or random effect estimation should be used in this case.

e.

Do the Hausman test of fixed effects (always consistent) versus random effects (efficient under random individual heterogeneity, inconsistent otherwise). Interpret the outcome of the test.

Figure 11: Hausman test

```
. hausman fixed_eff random_eff

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |   fixed_eff    random_eff     Difference          S.E.
-------------+----------------------------------------------------------------
         l_q |    .9192846     .9066805        .0126041        .0153877
        l_pf |    .4174918     .4227784       -.0052867        .0058583
          lf |   -1.070396    -1.064499       -.0058974        .0255088
------------------------------------------------------------------------------
                 b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

                chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                        =        2.12
              Prob>chi2 =      0.5469
              (V_b-V_B is not positive definite)
```

Since the matrix of $V_b - V_\beta$ is not positive definite, it can not be inverted, therefore this test-statistics is not useable.