13.1) Seemingly Unrelated Regressions (SUR)

Vitor Kamada

August 2019

Reference

Wooldridge (2010). **Econometric Analysis of Cross Section and Panel Data.** Ch 7.7

https://ebookcentral.proquest.com/lib/wayne/detail.action?docID=3339196&

System of Equations

$$housing = \beta_{10} + \beta_{11} houseprc + \beta_{12} foodprc + \beta_{13} clothprc + \beta_{14} income + \beta_{15} size + \beta_{16} age + u_1.$$

$$food = \beta_{20} + \beta_{21} houseprc + \beta_{22} foodprc + \beta_{23} clothprc + \beta_{24} income + \beta_{25} size + \beta_{26} age + u_2.$$

$$clothing = \beta_{30} + \beta_{31} houseprc + \beta_{32} foodprc + \beta_{33} clothprc + \beta_{34} income + \beta_{35} size + \beta_{36} age + u_3.$$

$$E(u_g | x_1, x_2, ..., x_G) = 0$$

$$g = 1, ..., G$$

Generalized Least Squares (GLS)

$$E(\epsilon\epsilon'|X) = \sigma^2 \underset{(n \times n)}{V(X)}$$

$$V^{-1} = C'C$$

$$Cy = CX\beta + C\epsilon$$

$$\hat{\beta}_{GLS} = (X'V^{-1}X)^{-1}X'V^{-1}y$$

GLS - Two Equations

$$y_g = X_g\beta_g + u_g, \qquad g = 1, 2$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

$$\Sigma \otimes I_N = \begin{bmatrix} \sigma_{11} I_N & \sigma_{12} I_N \\ \sigma_{21} I_N & \sigma_{22} I_N \end{bmatrix}$$

$$\hat{\beta}_{GLS} = \begin{bmatrix} \sigma_{11} X_1' X_1 & \sigma_{12} X_1' X_2 \\ \sigma_{21} X_2' X_1 & \sigma_{22} X_2' X_2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{11} X_1' y_1 + \sigma_{12} X_1' y_2 \\ \sigma_{21} X_2' y_1 + \sigma_{22} X_2' y_2 \end{bmatrix}$$

Seemingly Unrelated Regressions (SUR)

$$\begin{bmatrix} y_{i1} \\ \vdots \\ y_{iG} \end{bmatrix} = \begin{bmatrix} x'_{i1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x'_{iG} \end{bmatrix} \begin{bmatrix} \beta_{i1} \\ \vdots \\ \beta_{iG} \end{bmatrix} + \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iG} \end{bmatrix}$$

$$\hat{\beta}_{GLS} = \{X'(\Sigma^{-1} \otimes I_N)X\}^{-1}\{X'(\Sigma^{-1} \otimes I_N)y\}$$

$$Var(\hat{\beta}) = \{X'(\Sigma^{-1} \otimes I_N)X\}^{-1}$$

SUR Estimator = FGLS Estimator

1) Estimate each equation by OLS

2) Estimate Σ, using:

$$\hat{u}_j = y_j - X_j\hat{\beta}_j \text{ and } \hat{\sigma}_{jj'} = \hat{u}_j'\hat{u}_{j'}/N$$

3) Use $\hat{\Sigma}$ to obtain $\hat{\beta}_{FGLS}$

SUR Model for Hourly Wages and Hourly Benefits

$$r_{12} = .32$$

Explanatory Variables	*hrearn*	*hrbens*
educ	.459 (.069)	.077 (.008)
exper	−.076 (.057)	.023 (.007)
exper^2	.0040 (.0012)	−.0005 (.0001)
tenure	.110 (.084)	.054 (.010)
tenure^2	−.0051 (.0033)	−.0012 (.0004)
union	.808 (.408)	.366 (.049)
south	−.457 (.552)	−.023 (.066)
nrtheast	−1.151 (0.606)	−.057 (.072)
nrthcen	−.636 (.556)	−.038 (.066)
married	.642 (.418)	.058 (.050)
white	1.141 (0.612)	.090 (.073)
male	1.785 (0.398)	.268 (.048)
intercept	−2.632 (1.228)	−.890 (.147)

Wooldridge (2010)

Medical Expenditure Panel Survey (MEPS)

- Medicare-eligible population
- Aged > 65 years
- Medicare does not cover all medical expenses
- People usually buy private insurance

ldrugexp: log of expenditure on prescribed drugs

ldrugexp: log of expenditure on all categories of medical services other than drugs

actlim: activity limitation

```
X1 = sm.add_constant(df[['age', 'age2',
 'actlim', 'totchr', 'medicaid', 'private']])
X2 = sm.add_constant(df[['age', 'age2',
 'actlim','totchr', 'educyr', 'private']])
```

```
OLS1 = sm.OLS(df.ldrugexp, X1,
 missing='drop').fit(cov_type='HC1')
```

```
                 coef    std err          z      P>|z|
------------------------------------------------------
const         -4.4022      2.972     -1.481      0.139
age            0.2764      0.079      3.484      0.000
age2          -0.0018      0.001     -3.475      0.001
actlim         0.3574      0.046      7.854      0.000
totchr         0.4035      0.016     24.768      0.000
medicaid       0.0893      0.062      1.435      0.151
private        0.0775      0.044      1.750      0.080
```

```python
OLS2 = sm.OLS(df.ltotothr, X2,
 missing='drop').fit(cov_type='HC1')
```

	coef	std err	z	P>|z|
const	-6.1414	3.853	-1.594	0.111
age	0.3174	0.103	3.081	0.002
age2	-0.0021	0.001	-3.047	0.002
actlim	0.7421	0.064	11.664	0.000
totchr	0.2960	0.020	14.460	0.000
educyr	0.0650	0.008	8.531	0.000
private	0.2590	0.054	4.773	0.000

$r_{12} = \hat{\sigma}_{12} / \sqrt{\hat{\sigma}_{11}\hat{\sigma}_{22}} = 0.17$

```
from linearmodels.system import SUR
from collections import OrderedDict
Equation = OrderedDict()
Equation['ldrugexp'] = {'dependent': df.ldrugexp, 'exog': X1}
Equation['ltotothr'] = {'dependent': df.ltotothr, 'exog': X2}
SUR_Reg = SUR(Equation).fit()
```

```
res1 = SUR_Reg.equations['ldrugexp'].resids
res2 = SUR_Reg.equations['ltotothr'].resids
np.corrcoef(res1, res2)
```

```
array([[1.        , 0.17427516],
       [0.17427516, 1.        ]])
```

Breusch–Pagan or Lagrange Multiplier (LM) Test

$$Nr_{12}^2 = 3251 \times 0.1741^2 = 98.54$$

SUR Results

	Parameter	Std. Err.	T-stat	P-value
const	-3.8913	2.6290	-1.4802	0.1388
age	0.2630	0.0702	3.7478	0.0002
age2	-0.0017	0.0005	-3.7387	0.0002
actlim	0.3547	0.0400	8.8715	0.0000
totchr	0.4005	0.0144	27.828	0.0000
medicaid	0.1068	0.0539	1.9798	0.0477
private	0.0810	0.0390	2.0788	0.0376

Equation: ltotothr, Dependent Variable:

	Parameter	Std. Err.	T-stat	P-value
const	-5.1983	2.6469	-1.9640	0.0495
age	0.2928	0.0707	4.1399	0.0000
age2	-0.0019	0.0005	-4.0930	0.0000
actlim	0.7387	0.0431	17.143	0.0000
totchr	0.2874	0.0141	20.318	0.0000
educyr	0.0653	0.0053	12.238	0.0000
private	0.2689	0.0373	7.2116	0.0000