An introduction to econometrics of panel data

F. Legendre

Université Paris-Est

Seminars given at UKMA departement of finance

Application - II

Data from French national statistic office

2.202 Valeur ajoutée par branche en volume aux prix de l'année...					
		1949	1950	1951	...
DA	Agriculture, sylviculture, pêche	9.11260	10.63951	10.07511	...
DB	Industrie (= EB à EG)	31.11276	34.12560	37.52775	...
EB	Industries agricoles et alimentaires	9.17839	10.06890	11.31772	...
FB1	Industries de la viande et du lait				...
...	...	...	...	...	...

Application - III

Data needed by E-Views

	QDA	QFB1	QFB2	···
1979	23.1605	6.50860	15.4239	···
1980	22.9535	7.14595	15.8512	···
1981	22.8904	7.54329	16.2384	···
1982	···	···	···	···

as I want to run the program

```
WORKFILE A 1979 2008

POOL MyPool DA FB1 FB2 FC1 FC2 FC3 FC4 ED FE1 FE2 FE3
    FF1 FF2 FF3 FF4 FF5 FF6 FG1 FG2 FH1 FH2 FJ1 FJ2 FJ3
    EK FL1 FL2 FM1 FM2 FN1 FN2 FN3 FN4 FP1 FP2 FP3 FQ1
    FQ2 FQ3 FR1 FR2

MyPool.READ(A2) "...\Q.xls" QDA QFB1 QFB2 QFC1 QFC2 QFC3
     QFC4 QED QFE1 QFE2 QFE3 QFF1 QFF2 QFF3 QFF4 QFF5
    QFF6 QFG1 QFG2 QFH1 QFH2 QFJ1 QFJ2 QFJ3 QEK QFL1 QFL2
     QFM1 QFM2 QFN1 QFN2 QFN3 QFN4 QFP1 QFP2 QFP3 QFQ1
    QFQ2 QFQ3 QFR1 QFR2
```

Application - IV

The following program gets wrong results :(

```
WORKFILE A 1979 2008

POOL MyPool DA FB1 FB2 FC1 FC2 FC3 FC4 ED FE1 FE2 FE3
    FF1 FF2 FF3 FF4 FF5 FF6 FG1 FG2 FH1 FH2 FJ1 FJ2 FJ3
    EK FL1 FL2 FM1 FM2 FN1 FN2 FN3 FN4 FP1 FP2 FP3 FQ1
    FQ2 FQ3 FR1 FR2

MyPool.READ(A2) "...\Q.xls" Q?
```

I do not find the bug

Application - V – Data management (a)

Data management flow

1. With Excel, save the `xls` sheet file as 2003 `xml` data
2. Build an C++ program to read the `xml` data and to generate the right file
3. Use the command line “ssconvert” (an open source tool) to generate the `xls` right sheet file

Application - V – Data management (b)

C++ program (do not be afraid)

```
int main() {

  const char * names [] = { "t_2202", "t_2205", "t_2604", } ;
  const char * vars [] = { "Q", "L", "K", } ;

  const char * sectors [] = { "DA", ... "FR2", } ;

  const auto year_beg = 1978 ;
  const auto year_end = 2008 ;

  for ( size_t i = 0 ; i < NR_ELEMENTS(names) ; ++ i ) {

    vector< vector<double> >
       table( year_end-year_beg+1, vector<double>(NR_ELEMENTS(sectors)) ) ;
    Sheet sheet(names[i]+string(".xml")) ;

    size_t col = 0 ;
    ...

  return 0 ;
}
```

Application - VI – Overall dimension (a)

```
MyPool.GENR lq? = LOG(Q?)
MyPool.DELETE Q?
MyPool.GENR lk? = LOG(K?)
MyPool.DELETE K?
MyPool.GENR ll? = LOG(L?)
MyPool.DELETE L?

GENR t = @year - 1979

MyPool.LS lq? lk? ll? t
```

Application - VI – Overall dimension (b)

```
Dependent Variable: LQ?
Method: Pooled Least Squares
Date: 11/03/12   Time: 13:20
Sample: 1979 2008
Included observations: 30
Number of cross-sections used: 41
Total panel (balanced) observations: 1230

Variable Coefficient Std. Error t-Statistic Prob.

C        -1.325172    0.081599   -16.24000  0.0000
LK?       0.400369    0.010774    37.16046  0.0000
LL?       0.481730    0.013936    34.56617  0.0000
T         0.010148    0.001485    6.835410  0.0000

R-squared            0.765435  Mean dependent var  2.893650
Adjusted R-squared   0.764861  S.D. dependent var  0.918047
S.E. of regression   0.445171  Sum squared resid   242.9652
F-statistic          1333.567  Durbin-Watson stat  0.013086
Prob(F-statistic)    0.000000
```

Full information maximum likelihood estimator - I

Linear model as usual plus random effect hypothesis

$$\underline{y} = X\underline{a} + \underline{u} \quad \text{and} \quad u_{it} = \mu_i + v_{it} \quad \text{so} \quad \mathrm{V}(\underline{u}) = \Omega \neq \sigma^2 I_{NT}$$

The likelihood is

$$\mathcal{L}(\underline{a}, \sigma_\mu^2, \sigma_v^2) = (2\pi)^{-NT/2} |\Omega|^{-1/2} \exp\left[-\frac{1}{2} (\underline{y} - X\underline{a})' \Omega^{-1} (\underline{y} - X\underline{a}) \right]$$

where

$$\Omega = \lambda_w W + \lambda_b B \quad \text{with} \quad \lambda_w = \sigma_v^2 \ \text{ and } \ \lambda_b = (T\sigma_\mu^2 + \sigma_v^2)$$

One can show

$$|\Omega| = \lambda_w^{N(T-1)} \lambda_b^N \ \text{ and } \ \Omega^{-1} = \lambda_w^{-1} W + \lambda_b^{-1} B$$

Full information maximum likelihood estimator - II

So log-likelihood is

$$-\frac{1}{2}N(T-1)\log\lambda_w - \frac{1}{2}N\log\lambda_b - \frac{1}{2\lambda_w}(\underline{y}-X\underline{a})'W(\underline{y}-X\underline{a}) - \frac{1}{2\lambda_b}(\underline{y}-X\underline{a})'B(\underline{y}-X\underline{a})$$

First order conditions (after tedious manipulations!)

$$\hat{\underline{a}}_{ML} = \left(\frac{1}{\hat{\lambda}_{w\,ML}}X'WX + \frac{1}{\hat{\lambda}_{b\,ML}}X'BX\right)^{-1}\left(\frac{1}{\hat{\lambda}_{w\,ML}}X'W\underline{y} + \frac{1}{\hat{\lambda}_{b\,ML}}X'B\underline{y}\right)$$

$$\hat{\lambda}_{w\,ML} = \frac{(\underline{y}-X\hat{\underline{a}}_{ML})'W(\underline{y}-X\hat{\underline{a}}_{ML})}{N(T-1)}$$

$$\hat{\lambda}_{b\,ML} = \frac{(\underline{y}-X\hat{\underline{a}}_{ML})'B(\underline{y}-X\hat{\underline{a}}_{ML})}{N}$$

Iterative evaluation may be a device to numerically solve this system (with an initial guess)

Generalised least squares estimator

If Ω is know (very strange assumption), then

$$\underline{\hat{a}}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}\underline{y}$$

We already show

$$\Omega^{-1} = \lambda_w^{-1}W + \lambda_b^{-1}B$$

So the GLS estimator can be written as

$$\underline{\hat{a}}_{GLS} = [X'(W+\theta B)X]^{-1}X'(W+\theta B)\underline{y} \quad \text{with} \quad \theta = \lambda_w/\lambda_b$$

Remember that $W^2 = W$, $B^2 = B$, and $W'B = 0$

$$\underline{\hat{a}}_{GLS} = \left\{\left[\left(W+\sqrt{\theta}B\right)X\right]'\left[\left(W+\sqrt{\theta}B\right)X\right]\right\}^{-1}\left[\left(W+\sqrt{\theta}B\right)X\right]'\left[\left(W+\sqrt{\theta}B\right)\underline{y}\right]$$

Then, after the transformation

$$z_{it}^{\star} = (z_{it}-\bar{z}_i)+\sqrt{\theta}\bar{z}_i = z_{it}-(1-\sqrt{\theta})\bar{z}_i$$

you can use OLS formulas

$$\underline{\hat{a}}_{GLS} = (X^{\star\prime}X^{\star})^{-1}X^{\star\prime}\underline{y}^{\star}$$

Feasible generalised least squares estimator

Two steps

- Get a convergent estimation of θ

$$\hat{\theta} = \hat{\lambda}_w / \hat{\lambda}_b = \frac{SSR_W/(NT-N-k)}{SSR_B/(N-k)}$$

- Use OLS formulas on transformed data

$$\hat{z}^{\star}_{it} = (z_{it} - \overline{z}_i) + \sqrt{\hat{\theta}}\overline{z}_i = z_{it} - (1-\sqrt{\hat{\theta}})\overline{z}_i$$

$$\underline{\hat{a}}_{FGLS} = (\widehat{X}^{\star\prime}\widehat{X}^{\star})^{-1}\widehat{X}^{\star\prime}\underline{\hat{y}}^{\star}$$

Application - VII – Random effect hypothesis (a)

```
MyPool.GENR lq? = LOG(Q?)
MyPool.DELETE Q?
MyPool.GENR lk? = LOG(K?)
MyPool.DELETE K?
MyPool.GENR ll? = LOG(L?)
MyPool.DELETE L?

GENR t = @year - 1979

MyPool.LS(R) lq? lk? ll? t
```

Application - VII – Random effect hypothesis (b)

```
Dependent Variable: LQ?
Method: GLS (Variance Components)
Date: 11/03/12   Time: 13:24
Sample: 1979 2008
Included observations: 30
Number of cross-sections used: 41
Total panel (balanced) observations: 1230

Variable Coefficient Std. Error t-Statistic Prob.

C        -0.400375    0.162875   -2.458167  0.0141
LK?       0.048407    0.033859    1.429654  0.1531
LL?       0.490450    0.032325    15.17248  0.0000
T         0.017539    0.000943    18.60080  0.0000

Random Effects
DA--C-0.192887
...
FR2--C-1.069435

GLS Transformed Regression

R-squared            0.961699    Mean dependent var   2.893650
Adjusted R-squared   0.961606    S.D. dependent var   0.918047
S.E. of regression   0.179887    Sum squared resid    39.67238
Durbin-Watson stat   0.075329

Unweighted Statistics including Random Effects

R-squared            0.964129    Mean dependent var   2.893650
Adjusted R-squared   0.964041    S.D. dependent var   0.918047
S.E. of regression   0.174087    Sum squared resid    37.15537
Durbin-Watson stat   0.080432
```