Geographical Economics

Appendix: Toolkit

Rosella Nicolini

UAB

October 2014

"Mickey Mouse" toolkits: econometrics

- Panel data: macro and micro series.

"Mickey Mouse" toolkits: econometrics

- Panel data: macro and micro series.
- Fixed effects

"Mickey Mouse" toolkits: example

$$w_{it} = \alpha + \beta x_{it} + \varepsilon_{it}; \qquad i = 3; t = 2012; 2013$$

$$\begin{pmatrix} w_{1,2012} \\ w_{2,2012} \\ w_{3,2012} \\ w_{1,2013} \\ w_{2,2013} \\ w_{3,2013} \end{pmatrix} = \overline{\alpha} + \overline{\beta} \begin{pmatrix} x_{1,2012} \\ x_{2,2012} \\ x_{3,2012} \\ x_{1,2013} \\ x_{2,2013} \\ x_{3,2013} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,2012} \\ \varepsilon_{2,2012} \\ \varepsilon_{3,2012} \\ \varepsilon_{1,2013} \\ \varepsilon_{2,2013} \\ \varepsilon_{3,2013} \end{pmatrix}$$

Panel data (also known as longitudinal or cross-sectional time-series data) is a dataset in which the behavior of entities are observed across time.

These entities could be states, companies, individuals, countries, etc.

Panel data looks like this

country	year	Y	X1	X2	X3
1	2000	6.0	7.8	5.8	1.3
1	2001	4.6	0.6	7.9	7.8
1	2002	9.4	2.1	5.4	1.1
2	2000	9.1	1.3	6.7	4.1
2	2001	8.3	0.9	6.6	5.0
2	2002	0.6	9.8	0.4	7.2
3	2000	9.1	0.2	2.6	6.4
3	2001	4.8	5.9	3.2	6.4
3	2002	9.1	5.2	6.9	2.1

Fixed effects

The equation for the fixed effects model becomes:

$$Y_{it} = \beta_1 X_{it} + \alpha_i + u_{it} \qquad \text{[eq.1]}$$

Where

- α_i (i=1….n) is the unknown intercept for each entity (*n* entity-specific intercepts).
- Y_{it} is the dependent variable (DV) where i = entity and t = time.
- X_{it} represents one independent variable (IV),
- β_1 is the coefficient for that IV,
- u_{it} is the error term

Fixed effects

Another way to see the fixed effects model is by using binary variables. So the equation for the fixed effects model becomes:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \gamma_2 E_2 + \ldots + \gamma_n E_n + u_{it} \qquad \text{[eq.2]}$$

Where

–Y_{it} is the dependent variable (DV) where i = entity and t = time.

–$X_{k,it}$ represents independent variables (IV),

–β_k is the coefficient for the IVs,

–u_{it} is the error term

–E_n is the entity n. Since they are binary (dummies) you have n-1 entities included in the model.

–γ_2 Is the coefficient for the binary repressors (entities)

Fixed effects

You could add time effects to the entity effects model to have a *time and entity fixed effects regression model*:

$$Y_{it} = \beta_0 + \beta_1 X_{1,it} + \ldots + \beta_k X_{k,it} + \gamma_2 E_2 + \ldots + \gamma_n E_n + \delta_2 T_2 + \ldots + \delta_t T_t + u_{it} \quad \text{[eq.3]}$$

Where

- –Y_{it} is the dependent variable (DV) where i = entity and t = time.
- –$X_{k,it}$ represents independent variables (IV),
- –β_k is the coefficient for the IVs,
- –u_{it} is the error term
- –E_n is the entity n. Since they are binary (dummies) you have n-1 entities included in the model.
- –γ_2 is the coefficient for the binary regressors (entities).
- –T_t is time as binary variable (dummy), so we have t-1 time periods.
- –δ_t is the coefficient for the binary time regressors .

Control for time effects whenever unexpected variation or special events my affect the outcome variable.

From: O. Torres-Reyna, Princeton University

Fixed effects: *n* entity-specific intercepts

$Y_{it} = \beta_1 X_{it} + \ldots + \beta_k X_{kt} + \alpha_i + e_{it}$ [see eq.1]

NOTE: Add the option 'robust' to control for heteroskedasticity

Outcome variable

Predictor variable(s)

Fixed effects option

Total number of cases (rows)

Total number of groups (entities)

```
. xtreg  y  x1, fe

Fixed-effects (within) regression               Number of obs      =        70
Group variable: country                         Number of groups   =         7

R-sq:  within  = 0.0747                         Obs per group: min =        10
       between = 0.0763                                        avg =      10.0
       overall = 0.0059                                        max =        10

                                                F(1,62)            =      5.00
corr(u_i, Xb)  = -0.5468                        Prob > F           =    0.0289

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          x1 |   2.48e+09   1.11e+09     2.24   0.029     2.63e+08    4.69e+09
       _cons |   2.41e+08   7.91e+08     0.30   0.762    -1.34e+09    1.82e+09
-------------+----------------------------------------------------------------
     sigma_u |  1.818e+09
     sigma_e |  2.796e+09
         rho |  .29726926   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The errors u_i are correlated with the regressors in the fixed effects model

If this number is < 0.05 then your model is ok. This is a test (F) to see whether all the coefficients in the model are different than zero.

Coefficients of the regressors. Indicate how much *Y* changes when *X* increases by one unit.

Two-tail p-values test the hypothesis that each coefficient is different from 0. To reject this, the p-value has to be lower than 0.05 (95%, you could choose also an alpha of 0.10), if this is the case then you can say that the variable has a significant influence on your dependent variable (y)

29.7% of the variance is due to differences across panels.

'rho' is known as the intraclass correlation

$$rho = \frac{(sigma\_u)^2}{(sigma\_u)^2 + (sigma\_e)^2}$$

sigma_u = sd of residuals within groups u_i

sigma_e = sd of residuals (overall error term) e_i

t-values test the hypothesis that each coefficient is different from 0. To reject this, the t-value has to be higher than 1.96 (for a 95% confidence). If this is the case then you can say that the variable has a significant influence on your dependent variable (y). The higher the t-value the higher the relevance of the variable.

For more info see Hamilton, Lawrence, *Statistics with STATA*.

"Mickey Mouse" toolkits: interpretation results

- Time dummies; collinearity;

"Mickey Mouse" toolkits: interpretation results

- Time dummies; collinearity;
- Coeff test,

"Mickey Mouse" toolkits: interpretation results

- Time dummies; collinearity;
- Coeff test,
- F-test,

"Mickey Mouse" toolkits: interpretation results

- Time dummies; collinearity;
- Coeff test,
- F-test,
- Cluster correction.