ES1004 Econometrics by Example

Lecture 12: Panel Data Models

Dr. Hany Abdel-Latif

Swansea University, UK

Gujarati textbook, second edition [chapter 17]

11th February 2017

The Economic Society

Time Series Data

Table: An example

country	year	x	y
Egypt	2000	0.28	0.46
Egypt	2001	0.24	0.51
Egypt	2002	0.26	0.23
Egypt	2003	0.55	0.65
Egypt	2004	0.63	0.95
Egypt	2005	0.25	0.19
Egypt	2006	0.20	0.28
Egypt	2007	0.56	0.78
Egypt	2008	0.50	0.52
Egypt	2009	0.86	0.86
Egypt	2010	0.11	0.20

Cross-Sectional Data

Table: An Example

country	year	x	y
Egypt	2000	0.31	0.09
USA	2000	0.30	0.10
Italy	2000	0.30	0.72
Greece	2000	0.17	0.34
Lebanon	2000	0.87	0.28
Algeria	2000	0.24	0.99
Tunisia	2000	0.27	0.51
Kuwait	2000	0.66	0.15
Germany	2000	0.73	0.60
France	2000	0.94	0.88
UK	2000	0.20	0.02

Panel Data

Table: An Example

country	year	x	y
Egypt	2000	0.06	0.15
Egypt	2001	0.36	0.91
Egypt	2002	0.02	0.28
Egypt	2003	0.84	0.44
Egypt	2004	0.48	0.24
USA	2000	0.55	0.01
USA	2001	0.15	0.54
USA	2002	0.05	0.19
USA	2003	0.53	0.80
USA	2004	0.19	0.54
USA	2005	0.45	0.26

The Importance of Panel Data

- can take heterogeneity explicitly into account by allowing for subject-specific variables
- more informative, more variability less collinearity, more degrees of freedom
- allows for the study of the dynamics of change

Example: Charitable Giving

- table17.1 data on charitable giving
  - 47 individuals over 1979-1988
  - charity: sum of cash and other property contributions
  - income: gross income
  - price: one minus the marginal tax rate
  - age: a dummy equal to 1 if the taxpayer is over 64
  - MS: a dummy equal to 1 if married
  - DEPS: number of dependants
- find out the effect of the marginal tax rate on charitable giving [charity function]

Example: Data

subject	time	charity	income	price	age	ms	deps
1	1979	6.44572	10.1781	-0.3285	0	1	2
1	1980	6.13123	10.1857	-0.3285	0	1	1
1	1981	6.32794	10.3542	-0.38566	0	1	1
1	1982	6.41346	10.4694	-0.40048	0	1	1
1	1983	6.49224	10.5216	-0.35667	0	1	1
1	1984	6.56526	10.5662	-0.3285	0	1	1
1	1985	6.59987	10.6428	-0.40048	0	1	1
1	1986	6.66568	10.7012	-0.40048	0	1	1
1	1987	2.30259	10.7091	-1.60944	0	1	1
1	1988	2.30259	10.7716	-0.3285	0	1	1
2	1979	7.58274	10.0121	-0.30111	0	0	1
2	1980	6.24611	10.0543	-0.30111	0	0	1
2	1981	7.57147	9.9993	-0.27444	0	0	1
2	1982	7.18159	10.2367	-0.3285	1	0	1
2	1983	7.91132	10.1542	-0.23572	1	0	0
2	1984	7.55433	10.3258	-0.30111	1	0	0
2	1985	6.99393	10.3189	-0.3285	1	0	0
2	1986	7.49499	10.274	-0.30111	1	0	0
2	1987	8.12445	10.523	-0.69315	1	0	0
2	1988	7.90249	10.678	-0.3285	0	0	0
3	1979	6.24998	10.306	-0.4943	0	0	0
3	1980	6.41346	10.332	-0.4943	0	0	0

Example: Balanced & Short Panel

- balanced panel: the number of time observation is the same of each individual [equals 10 in this example]
- unbalanced panel:
- short panel: the number of cross-sectional or individual units N is greater than the number of time periods
  [in this example N=47 & T=10]
- long panel: T is greater than N

The Economic Society

Estimation: Five Options

1. individual times series of charity functions
   - very few degrees of freedom; meaningless statistical analysis
   - neglect information about other individuals' charity contributions; they all operate in the same regulatory environment
2. cross-sectional charity functions
   - neglect the dynamic aspect of charitable giving

Estimation: Five Options

3. pooled OLS charity function
   - pool all 470 observations [47×10]
   - neglect the dual nature of time series and cross-sectional data
   - assumes that the coefficients remain constant across time and cross-section
   - *aka* constant coefficient model

Estimation: Five Options

4. fixed effects least squares dummy variable (LSDV) model
    - pool all observations but allow each individual unit to have its individual intercept dummy
    - this is different from the within estimator which we will discuss shortly

Estimation: Five Options

5. random effects model
    - assumes that the intercepts values of the 47 individuals are random drawings from a much larger population of individuals

Example: pooled OLS

$$C_{it} = \beta_1 + \beta_2 Age_{it} + \beta_3 Income_{it} + \beta_4 Price_{it} + \beta_5 DEPS_{it} + \beta_6 MS_{it} + u_{it}$$

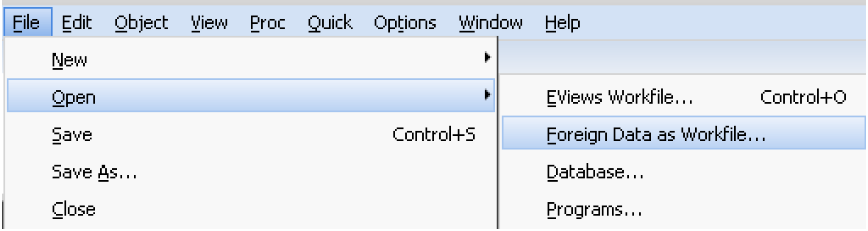
$i = 1, 2, \ldots, 47;\ t = 1, 2, \ldots, 10$

- notice we have two subscripts ($i\&t$)
- nonstochastic regressors, or if stochastic, uncorrelated with the error term
- the error terms satisfies the usual classical assumptions

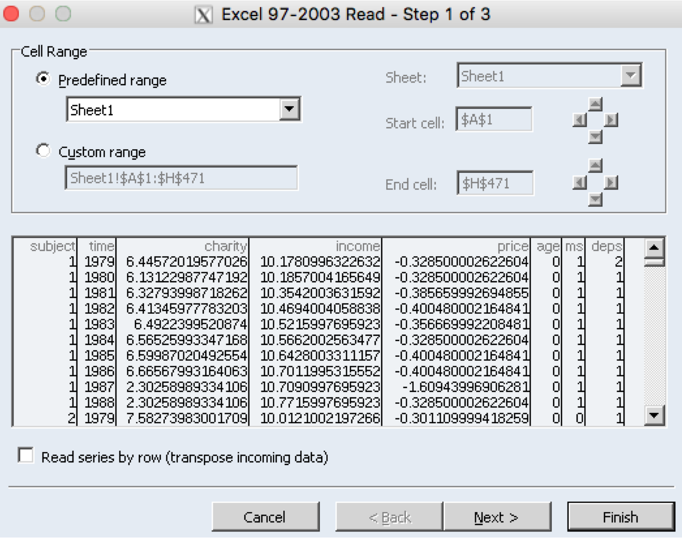
Example: a priori

- we would expect
  - age, income, price and marital status to have a positive impact on charitable giving
  - the number of dependants to have a negative impact
- the price variable represents the opportunity cost of giving charitable contributions
  - the higher the marginal tax, the lower the opportunity cost

EViews: import data

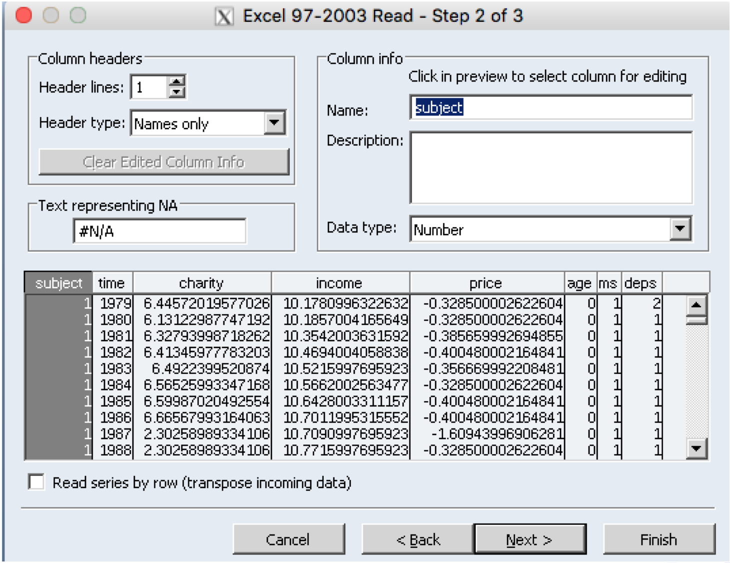


EViews: import data

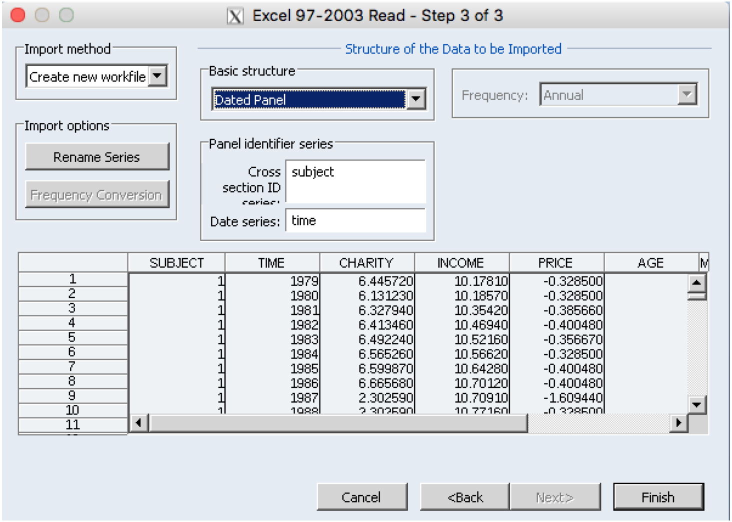


subject	time	charity	income	price	age	ms	deps
1	1979	6.44572019577026	10.1780996322632	-0.328500002622604	0	1	2
1	1980	6.13122987747192	10.1857004165649	-0.328500002622604	0	1	1
1	1981	6.32793998718262	10.3542003631592	-0.385659992694855	0	1	1
1	1982	6.41345977783203	10.4694004058838	-0.400480002164841	0	1	1
1	1983	6.4922399520874	10.5215997695923	-0.356669992208481	0	1	1
1	1984	6.56525993347168	10.5662002563477	-0.328500002622604	0	1	1
1	1985	6.59987020492554	10.6428003311157	-0.400480002164841	0	1	1
1	1986	6.66567993164063	10.7011995315552	-0.400480002164841	0	1	1
1	1987	2.30258989334106	10.7090997695923	-1.60943996906281	0	1	1
1	1988	2.30258989334106	10.7715997695923	-0.328500002622604	0	1	1
2	1979	7.58273983001709	10.0121002197266	-0.301109999418259	0	0	1

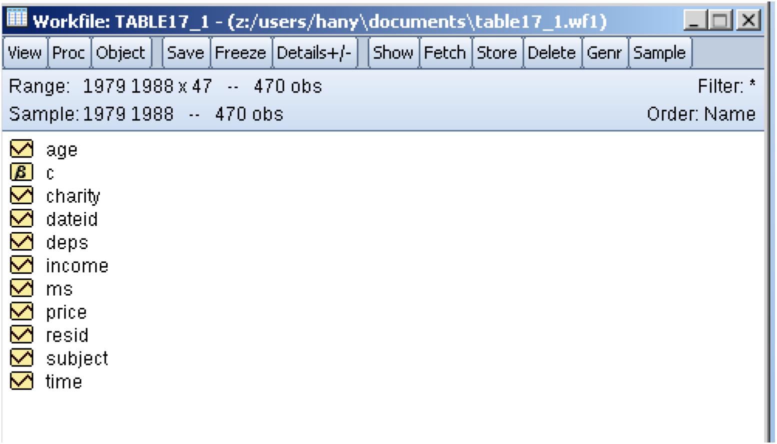
EViews: import data



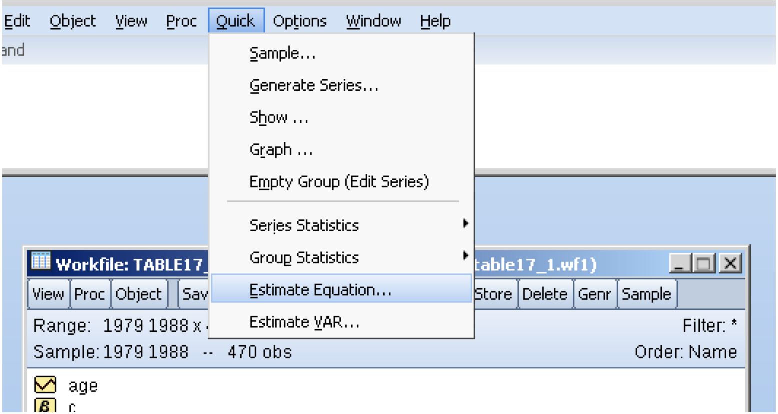
EViews: import data



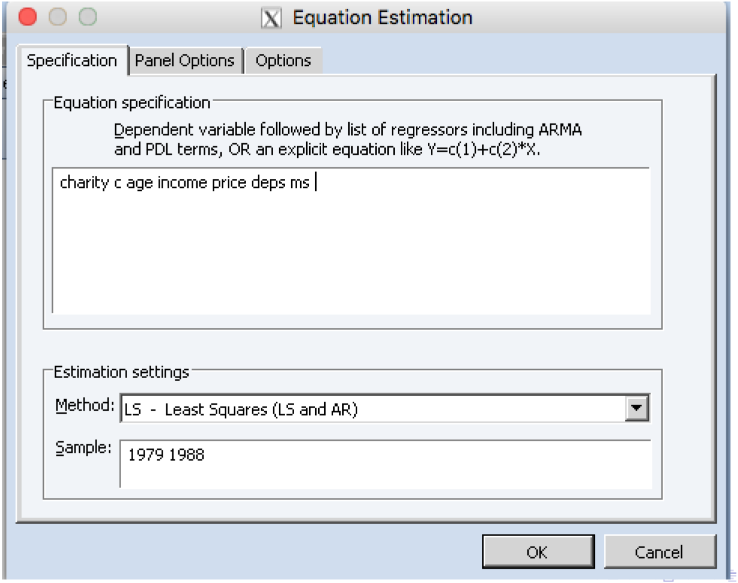
EViews: import data



Estimation: pooled OLS



Estimation: pooled OLS



Estimation: pooled OLS

Dependent Variable: CHARITY
Method: Least Squares
Date: 02/11/17 Time: 10:39
Sample: 1 470
Included observations: 470

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.674220	1.298134	-3.600722	0.0004
AGE	1.547275	0.216955	7.131788	0.0000
INCOME	1.035779	0.128944	8.032767	0.0000
PRICE	0.483092	0.207703	2.325875	0.0205
DEPS	0.175368	0.042642	4.112556	0.0000
MS	-0.008036	0.184849	-0.043475	0.9653

R-squared	0.224488	Mean dependent var	6.577150
Adjusted R-squared	0.216131	S.D. dependent var	1.313659
S.E. of regression	1.163067	Akaike info criterion	3.152681
Sum squared resid	627.6639	Schwarz criterion	3.205695
Log likelihood	-734.8801	Hannan-Quinn criter.	3.173538
F-statistic	26.86281	Durbin-Watson stat	0.701077
Prob(F-statistic)	0.000000		

Fixed Effects Least Squares Dummy Variables LSDV

- one way to take into the heterogeneity among 47 individuals is to allow each individual to have his or her own intercept

$$C_{it} = \beta_{1i} + \beta_2 Age_{it} + \beta_3 Income_{it} + \beta_4 Price_{it} + \beta_5 DEPS_{it} + \beta_6 MS_{it} + u_{it}$$

$i = 1, 2, \ldots, 47;\ t = 1, 2, \ldots, 10$

- notice we have added the subscript i to the intercept
- differences may be due to special features of each individual, such as education or religion

Fixed Effects Regression Model LSDV

- the term 'fixed effects'
  - each taxpayers' intercept is different from the intercepts of other taxpayers
  - however individual intercepts do not very over time [time invariant]

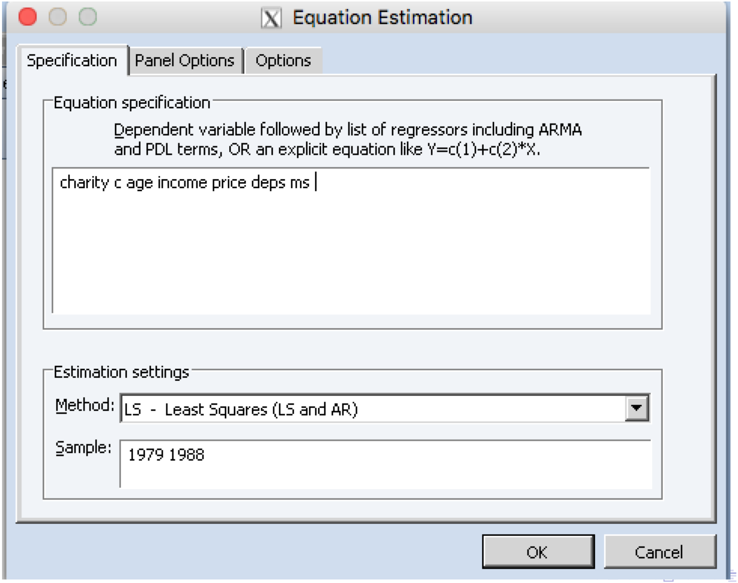
Fixed Effects Regression Model LSDV

- done easily by introducing differential intercept dummies

$$C_{it} = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \cdots + \beta_4 6 D_{46i} + \beta_{47} Age_{it} + \beta_{48} Income_{it} + \beta_{49} Price_{it} + \beta_{50} DEPS_{it} + \beta_{51} MS_{it} + u_{it}$$

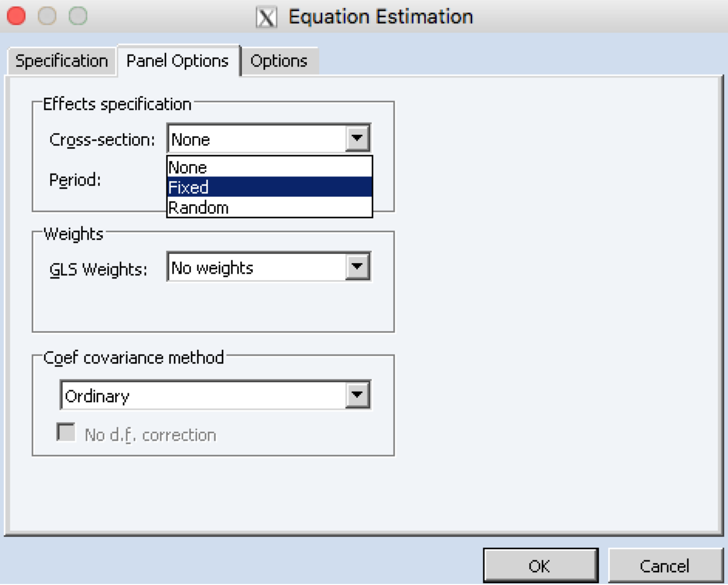
- 46 dummies represents 47 individuals to avoid dummy variable trap (perfect multicollinearity)
- individual 1 is our benchmark or reference category; any individual can be chosen for that purpose though

Estimation: LSDV



The Economic Society

Estimation: LSDV



Estimation: LSDV

Dependent Variable: CHARITY
Method: Panel Least Squares
Date: 02/11/17 Time: 12:08
Sample: 1979 1988
Periods included: 10
Cross-sections included: 47
Total panel (balanced) observations: 470

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.089972	1.131118	-1.847705	0.0654
AGE	0.102249	0.208039	0.491490	0.6233
INCOME	0.838810	0.111267	7.538727	0.0000
PRICE	0.366080	0.124294	2.945266	0.0034
DEPS	-0.086352	0.053483	-1.614588	0.1072
MS	0.199833	0.263890	0.757257	0.4493

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.763177	Mean dependent var	6.577150
Adjusted R-squared	0.734282	S.D. dependent var	1.313659
S.E. of regression	0.677163	Akaike info criterion	2.162215
Sum squared resid	191.6735	Schwarz criterion	2.621666
Log likelihood	-456.1204	Hannan-Quinn criter.	2.342975

Pooled OLS vs LSDV

- a test to find out if the fixed effects model is better than the OLS pooled model
- the pooled model is restricted
  - neglects the heterogeneity effects that are explicitly taken into account in the fixed effects model
- we can use the restricted F test

$$F = \frac{(R^2_{ur} - R^2_r)/m}{(1 - R^2_{ur})/(n-k)}$$

m number of parameters excluded (46 here)
k number of parameters estimated in the unrestricted model (52 here)

Pooled OLS vs LSDV

$$F = \frac{(0.7632 - 0.2245)/46}{(1 - 0.7632)/(418)} = 20.672$$

- for 46 df in the numerator and 418 in the dominator, this F is highly significant
- confirms that the fixed effect model is superior to the pooled regression model

Limitations to the Fixed Effects LSDV model

- every additional dummy variable will cost an additional degree of freedom (very costly in small samples)
- too many additive and multiplicative dummies may lead to multicollinearity

Limitations to the Fixed Effects LSDV model

- we assume that $u_{it} \sim N(0, \sigma^2)$ but since we have i and t we may have to modify this classical assumption
- there are several possibilities
  1. assume the error variance is the same for all cross-sectional unit or it is heteroscedastic
  2. for each subject, we can assume there is no autocorrelation over time, or we can assume autocorrelation of the AR(1) type
  3. at any given time, we can allow the error term of individual 1 to be non-correlated with the error for say individual 2, or we can assume that there is such correlation

The Economic Society

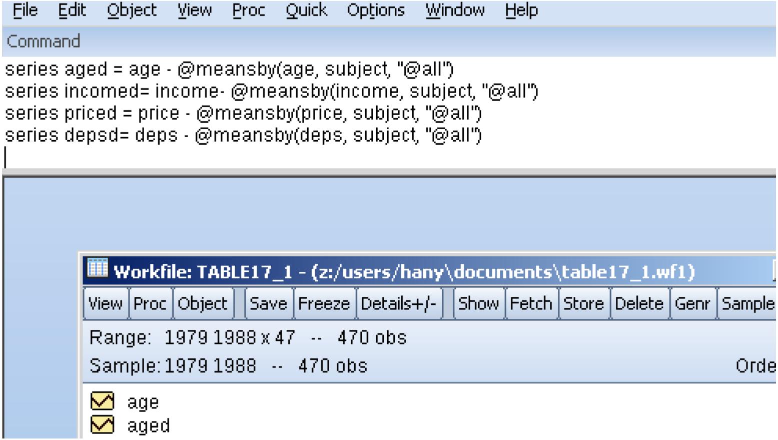
Fixed Effects Within Group Estimator

- we can eliminate the fixed effect term by expressing both the regressand and regressors as deviations for their respective (group) mean values
- run the regression on the mean corrected variables

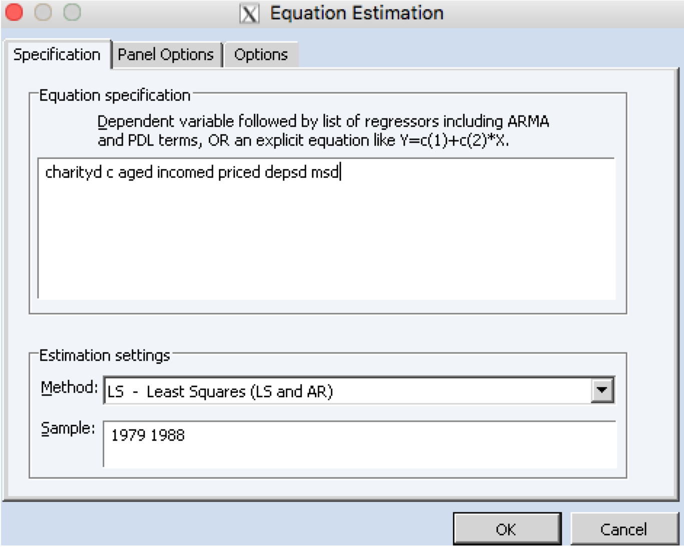
$$C_{it} - \overline{C}_i = B_2(Age_{it} - \overline{Age}_i) + \beta_3(Income_{it} - \overline{income}_i)$$
$$+\beta_4(Price_{it} - \overline{Price}_i) + \beta_5(Deps_{it} - \overline{Deps}_i) +$$
$$\beta_6(MS_{it} - \overline{MS}_i) + u_{it}$$

- notice how fixed effects or individual effect intercept term β_{1i} drops out

Estimation: Demeaned Variables



Estimation: WG Estimator



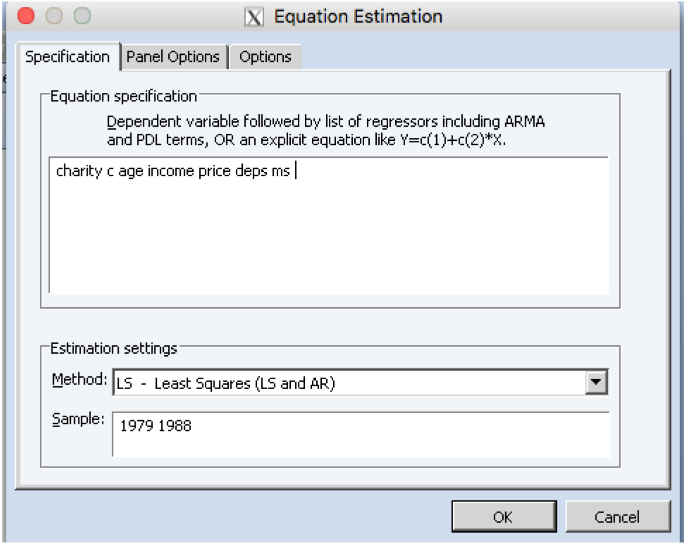
Estimation: WG Estimator

Dependent Variable: CHARITYD
Method: Panel Least Squares
Date: 02/11/17 Time: 13:39
Sample: 1979 1988
Periods included: 10
Cross-sections included: 47
Total panel (balanced) observations: 470

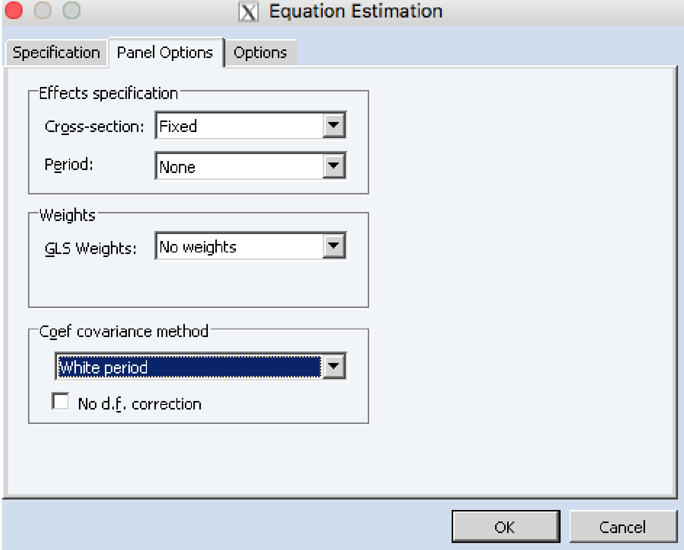
Variable	Coefficient	Std. Error	t-Statistic	Prob.
AGED	0.102249	0.197458	0.517828	0.6048
INCOMED	0.838810	0.105608	7.942712	0.0000
PRICED	0.366080	0.117973	3.103097	0.0020
DEPSD	-0.086352	0.050762	-1.701111	0.0896
MSD	0.199833	0.250468	0.797837	0.4254
C	-1.43E-16	0.029646	-4.82E-15	1.0000

R-squared	0.134974	Mean dependent var	-3.59E-17
Adjusted R-squared	0.125653	S.D. dependent var	0.687353
S.E. of regression	0.642720	Akaike info criterion	1.966470
Sum squared resid	191.6735	Schwarz criterion	2.019483
Log likelihood	-456.1204	Hannan-Quinn criter.	1.987327
F-statistic	14.48002	Durbin-Watson stat	1.234015
Prob(F-statistic)	0.000000		

FEM Robust Standard Errors



FEM Robust Standard Errors



FEM Robust Standard Errors

Dependent Variable: CHARITY
Method: Panel Least Squares
Date: 02/11/17 Time: 13:50
Sample: 1979 1988
Periods included: 10
Cross-sections included: 47
Total panel (balanced) observations: 470
White period standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.674220	2.295730	-2.036049	0.0423
AGE	1.547275	0.324334	4.770620	0.0000
INCOME	1.035779	0.217883	4.753819	0.0000
PRICE	0.483092	0.186064	2.596376	0.0097
DEPS	0.175368	0.087411	2.006254	0.0454
MS	-0.008036	0.352005	-0.022830	0.9818

R-squared	0.224488	Mean dependent var	6.577150
Adjusted R-squared	0.216131	S.D. dependent var	1.313659
S.E. of regression	1.163067	Akaike info criterion	3.152681
Sum squared resid	627.6639	Schwarz criterion	3.205695
Log likelihood	-734.8801	Hannan-Quinn criter.	3.173538
F-statistic	26.86281	Durbin-Watson stat	0.449191
Prob(F-statistic)	0.000000		

Random Effects Model REM

- FEM assumes individual specific coefficient β_{1i} is fixed for each subject [i.e., time-invariant]
- REM β_{1i} is a random variable with a mean value of β_1 (no i subscript here) and the intercept of any cross-section unit is expressed as

$$\beta_{1i} = \beta_1 + \epsilon_i$$

where ϵ_i is a random error term with mean 0 and variance σ_ϵ^2

Random Effects Model REM

$$C_{it} = \beta_1 + \beta_2 Age_{it} + \beta_3 Income_{it} + \beta_4 Price_{it} + \beta_5 DEPS_{it} + \beta_6 MS_{it} + w_{it}$$

where

$$w_{it} = \epsilon_i + u_{it}$$

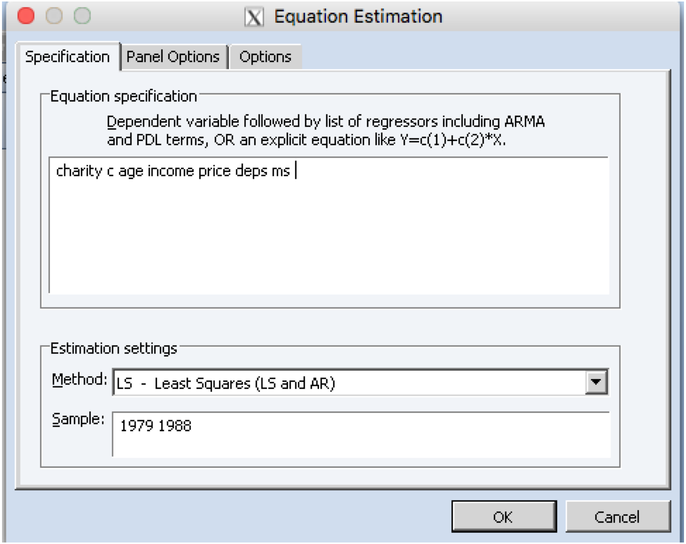
- the 47 individuals are drawn from a much larger universe and have a common mean value for the intercept $(= \beta_1)$
- differences in the individual values of the intercept for each donor are reflected in the error term ϵ_i

Random Effects Model REM

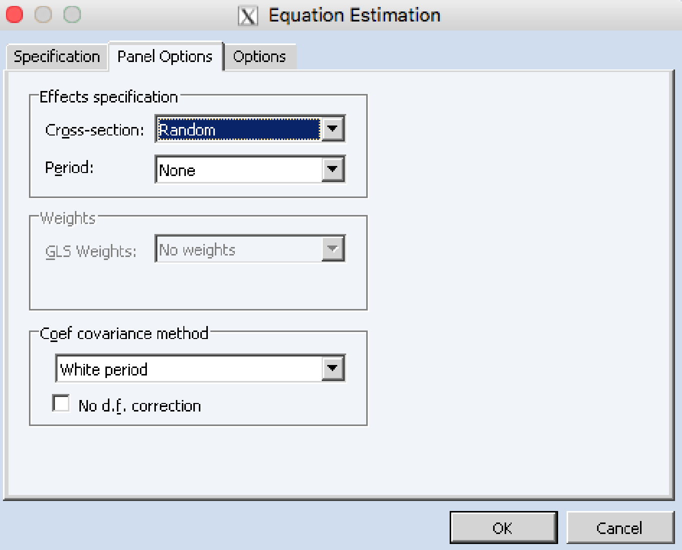
$$w_{it} = \epsilon_i + u_{it}$$

- the composite error term w_{it} has two components
  - ϵ_i the cross-section or individual-specific error component, and
  - u_{it} the combined time series and cross-section error component [aka idiosyncratic term because it varies over cross-section i.e., individual as well as time]

Estimation: REM



Estimation: REM



Estimation: REM

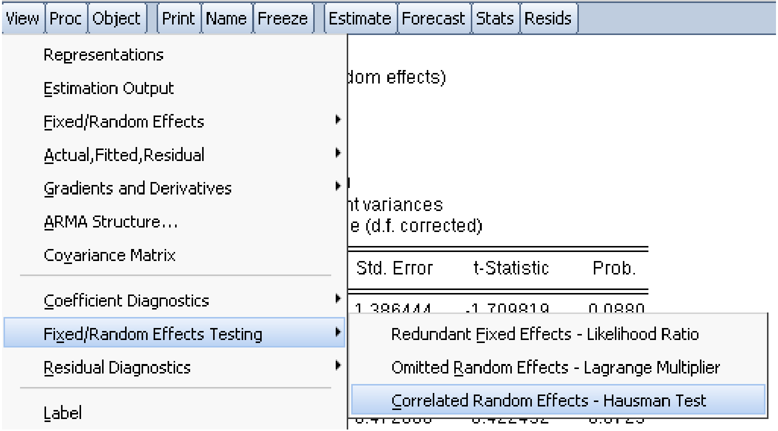
Dependent Variable: CHARITY
Method: Panel EGLS (Cross-section random effects)
Date: 02/11/17 Time: 15:08
Sample: 1979 1988
Periods included: 10
Cross-sections included: 47
Total panel (balanced) observations: 470
Swamy and Arora estimator of component variances
White period standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-2.370568	1.386444	-1.709819	0.0880
AGE	0.277063	0.127176	2.178576	0.0299
INCOME	0.852997	0.126574	6.739101	0.0000
PRICE	0.370199	0.140054	2.643254	0.0085
DEPS	-0.036254	0.064181	-0.564874	0.5724
MS	0.199669	0.472666	0.422432	0.6729

Effects Specification

	S.D.	Rho
Cross-section random	0.930938	0.6540
Idiosyncratic random	0.677163	0.3460

Hausman Test



Hausman Test

Correlated Random Effects - Hausman Test
Equation: Untitled
Test cross-section random effects

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	15.964267	5	0.0069

Cross-section random effects test comparisons:

Variable	Fixed	Random	Var(Diff.)	Prob.
AGE	0.102249	0.277063	0.003539	0.0033
INCOME	0.838810	0.852997	0.000830	0.6224
PRICE	0.366080	0.370199	0.000087	0.6595
DEPS	-0.086352	-0.036254	0.000487	0.0232
MS	0.199833	0.199669	0.016167	0.9990