2. Pooled Cross Sections and Panels

2.1 Pooled Cross Sections versus Panel Data

Pooled Cross Sections are obtained by collecting random samples from a large polulation independently of each other at different points in time. The fact that the random samples are collected independently of each other implies that they need not be of equal size and will usually contain different statistical units at different points in time.

Consequently, serial correlation of residuals is not an issue, when regression analysis is applied. The data can be pretty much analyzed like ordinary cross-sectional data, except that we must use dummies in order to account for shifts in the distribution between different points in time.

Panel Data or *longitudinal data* consists of time series for each statistical unit in the cross section. In other words, we randomly select our cross section only once, and once that is done, we follow each statistical unit within this cross section over time. Thus all cross sections are equally large and consist of the same statistical units.

For panel data we cannot assume that the observations are independently distributed across time and serial correlation of regression residuals becomes an issue. We must be prepared that unobserved factors, while acting differently on different cross-sectional units, may have a lasting effect upon the same statistical unit when followed through time. This makes the statistical analysis of panel data more difficult.

2.2 Independent Pooled Cross Sections

Example 1: Women's fertility over time. Regressing the number of children born per woman upon year dummies and controls such as education, age etc. yields information about the development of fertility unexplained by the controls. (Base year 1972)

Dependent Variable: KIDS
Method: Least Squares
Date: 11/23/12 Time: 08:21
Sample: 1 1129
Included observations: 1129

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-7.742457	3.051767	-2.537040	0.0113
EDUC	-0.128427	0.018349	-6.999272	0.0000
AGE	0.532135	0.138386	3.845283	0.0001
AGESQ	-0.005804	0.001564	-3.710324	0.0002
BLACK	1.075658	0.173536	6.198484	0.0000
EAST	0.217324	0.132788	1.636626	0.1020
NORTHCEN	0.363114	0.120897	3.003501	0.0027
WEST	0.197603	0.166913	1.183867	0.2367
FARM	-0.052557	0.147190	-0.357072	0.7211
OTHRURAL	-0.162854	0.175442	-0.928248	0.3535
TOWN	0.084353	0.124531	0.677367	0.4983
SMCITY	0.211879	0.160296	1.321799	0.1865
Y74	0.268183	0.172716	1.552737	0.1208
Y76	-0.097379	0.179046	-0.543881	0.5866
Y78	-0.068666	0.181684	-0.377945	0.7055
Y80	-0.071305	0.182771	-0.390136	0.6965
Y82	-0.522484	0.172436	-3.030016	0.0025
Y84	-0.545166	0.174516	-3.123871	0.0018

R-squared	0.129512	Mean dependent var	2.743136
Adjusted R-squared	0.116192	S.D. dependent var	1.653899
S.E. of regression	1.554847	Akaike info criterion	3.736447
Sum squared resid	2685.898	Schwarz criterion	3.816627
Log likelihood	-2091.224	Hannan-Quinn criter.	3.766741
F-statistic	9.723282	Durbin-Watson stat	2.010694
Prob(F-statistic)	0.000000		

We can also interact a time dummy with key explanatory variables to see if the effect of that variable has changed over time.

Example 2:

Changes in the return to education and the gender wage gap between 1978 and 1985.

Dependent Variable: LWAGE
Method: Least Squares
Date: 11/27/12 Time: 17:25
Sample: 1 1084
Included observations: 1084

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.458933	0.093449	4.911078	0.0000
EXPER	0.029584	0.003567	8.293165	0.0000
EXPER^2	-0.000399	7.75E-05	-5.151307	0.0000
UNION	0.202132	0.030294	6.672233	0.0000
EDUC	0.074721	0.006676	11.19174	0.0000
FEMALE	-0.316709	0.036621	-8.648173	0.0000
Y85	0.117806	0.123782	0.951725	0.3415
Y85*EDUC	0.018461	0.009354	1.973509	0.0487
Y85*FEMALE	0.085052	0.051309	1.657644	0.0977

R-squared	0.426186	Mean dependent var	1.867301
Adjusted R-squared	0.421915	S.D. dependent var	0.542804
S.E. of regression	0.412704	Akaike info criterion	1.076097
Sum squared resid	183.0991	Schwarz criterion	1.117513
Log likelihood	-574.2443	Hannan-Quinn criter.	1.091776
F-statistic	99.80353	Durbin-Watson stat	1.918367
Prob(F-statistic)	0.000000		

The return to education has risen by about 1.85% and the gender wage gap narrowed by about 8.5% between 1978 and 1985, other factors being equal.

Policy Analysis with Pooled Cross Sections

Example 3: How does a garbage incinerator's location affect housing prices? (Kiel and McClain 1995). We use data on housing prices from 1978, before any planning of an incinarator and from 1981, when construction work began. Naively, one might be tempted to use only 1981 data and to estimate a model like

$$\text{(1)} \qquad \texttt{rprice} = \gamma_0 + \gamma_1 \texttt{nearinc} + u,$$

where **rprice** is the housing price in 1978 dollars and **nearinc** is a dummy variable equal to one if the house is near the incinerator. Estimation yields

$$\text{(2)} \qquad \widehat{\texttt{rprice}} = \underset{(t=32.75)}{101\,307.5} - \underset{(t=-5.266)}{30\,688.27}\ \texttt{nearinc},$$

consistent with the notion that a location near a garbage incinerator depresses housing prices. However, another possible interpretation is that incinerators are built in areas with low housing prices. Indeed, estimating (1) on 1978 data yields

$$\text{(3)} \qquad \widehat{\texttt{rprice}} = \underset{(t=31.09)}{82\,517.23} - \underset{(t=-3.968)}{18\,824.37}\ \texttt{nearinc}.$$

To find the price impact of the incinerator, calculate the so called *difference-in-differences estimator*

$$\hat{\delta}_1 = -30\,688.27 - (-18\,824.37) = -11\,863.9.$$

So in this sample, vincinity of an incinerator depresses housing prices by almost \$12 000 on average, but we don't know yet whether the effect is statistically significant.

The previous example is called a *natural experiment* (or a *quasi-experiment*). It occurs when some external event (often a policy change) affects some group, called the *treatment group*, but leaves another group, called the *control group*, unaffected. It differs from a true experiment in that these groups are not randomly and explicitely chosen.

Let D_T be a dummy variable indicating whether an observation is from the treatment group and D_{after} be a dummy variable indicating whether the observation is from after the exogeneous event. Then the impact of the external event on y is given by δ_1 in the model

$$(4) \quad y = \beta_0 + \delta_0 D_{after} + \beta_1 D_T + \delta_1 D_{after} \cdot D_T \quad (+\text{other factors}).$$

If no other factors are included, $\hat{\delta}_1$ will be the difference-in-differences estimator

$$(5) \quad \hat{\delta}_1 = (\bar{y}_{after,T} - \bar{y}_{bef.,T}) - (\bar{y}_{after,C} - \bar{y}_{bef.,C}),$$

where T and C denote the treatment and control group, respectively.

Example 3: (continued.)

Estimating (4) yields

$$\widehat{\texttt{rprice}} = \underset{(t=30.26)}{82\,517} + \underset{(t=4.64)}{18\,790}\,\texttt{y81} \underset{(t=-3.86)}{-18\,824}\,\texttt{nearinc} - \underset{(t=-1.59)}{11\,863}\,\texttt{y81}\cdot\texttt{nearinc}$$

The p-value on the interaction term is 0.1126, so it is not yet significant. This changes however, once additional controls enter (4):

Dependent Variable: RPRICE
Method: Least Squares
Date: 11/29/12 Time: 08:28
Sample: 1 321
Included observations: 321

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13807.67	11166.59	1.236515	0.2172
Y81	13928.48	2798.747	4.976683	0.0000
NEARINC	3780.337	4453.415	0.848862	0.3966
Y81*NEARINC	-14177.93	4987.267	-2.842827	0.0048
AGE	-739.4510	131.1272	-5.639189	0.0000
AGE^2	3.452740	0.812821	4.247845	0.0000
INTST	-0.538635	0.196336	-2.743437	0.0064
LAND	0.141420	0.031078	4.550529	0.0000
AREA	18.08621	2.306064	7.842891	0.0000
ROOMS	3304.227	1661.248	1.989003	0.0476
BATHS	6977.317	2581.321	2.703002	0.0073

R-squared	0.660048	Mean dependent var	83721.36
Adjusted R-squared	0.649082	S.D. dependent var	33118.79
S.E. of regression	19619.02	Akaike info criterion	22.64005
Sum squared resid	1.19E+11	Schwarz criterion	22.76929
Log likelihood	-3622.729	Hannan-Quinn criter.	22.69166
F-statistic	60.18938	Durbin-Watson stat	1.677080
Prob(F-statistic)	0.000000		

We conclude that vincinity of a garbage incinerator depresses housing prices by about \$14,178 (at 1978 value), when controlling for other valuation-relevant properties of the house ($p = 0.0048$).

2.3 First-Difference Estimation in Panels

Recall from Econometrics 1 that omitting important variables in the model may induce severe bias to all parameter estimates. This was called the *omitted variable bias*. Panel data allows to mitigate, if not eliminate, this problem.

Example 4.

Crime and unemployment rates for 46 cities in 1982 and 1987. Regressing the crimerate (crimes per 1 000 people) `crmte` upon the unemployment rate `unem` (in percent) yields for the 1987 cross section

$$\widehat{\texttt{crmrte}} = \underset{(t=6.18)}{128.38} - \underset{(t=-1.22)}{4.16}\texttt{unem}.$$

Even though unemployment is nonsignificant ($p = 0.23$), a causal interpretation would imply that an increase in unemployment lowers the crime rate, which is hard to believe. The model probably suffers from omitted variable bias.

With panel data we view the unobserved factors affecting the dependent variable as consisting of two types: those that are constant and those that vary over time. Letting i denote the cross-sectional unit and t time:

$$y_{it} = \beta_0 + \delta_0 D_{2,t} + \beta_1 x_{it} + a_i + u_{it}. \quad t = 1, 2 \tag{6}$$

The dummy variable $D_{2,t}$ is zero for $t = 1$ and one for $t = 2$. It models the time-varying part of the unobserved factors. The variable a_i captures all unobserved, time-constant factors that affect y_{it}. a_i is generally called an *unobserved* or *fixed effect*. u_{it} is called the *idiosyncratic error*. A model of the form (6) is called an *unobserved effects model* or *fixed effects model*.

Example 4 (continued). A fixed effects model for city crime rates in 1982 and 1987 is

$$\mathtt{crmrte}_{\mathtt{it}} = \beta_0 + \delta_0 \mathtt{D}_{\mathtt{87,t}} + \beta_1 \mathtt{unem}_{\mathtt{it}} + \mathtt{a}_{\mathtt{i}} + \mathtt{u}_{\mathtt{it}}, \tag{7}$$

where D_{87} is dummy variable for 1987.

Naively, we might go and estimate a fixed effects model by pooled OLS. That is, we write (6) in the form

$$(8)\quad y_{it} = \beta_0 + \delta_0 D_{2,t} + \beta_1 x_{it} + \nu_{it}, \qquad t = 1, 2,$$

and apply OLS, where $\nu_{it} = a_i + u_{it}$ is called the *composite error*.

Such an approach is problematic for two reasons. As a minor complication it turns out that $\mathbb{C}\mathrm{ov}(\nu_{i1}, \nu_{i2}) = V(a_i)$ even though a_i and u_{it} are pairwise uncorrelated, such that the composite errors become positively correlated over time. This problem is minor because it can be solved by using standard errors which are robust to serial correlation in the residuals (HAC (Newey-West) resp. White period robust standard errors in EViews).

The main problem with applying pooled OLS is that we did very little to solve the omitted variable bias problem. Only the time-varying part (assumed to be common for all cross-sesctional units) has been taken out by introducing the time dummy. The fixed effect a_i, however, is still there; it has just been hidden in the composite error ν_{it}, and is therefore not modeled. That is, the parameter estimates are still biased, unless a_i is uncorrelated with x_{it}.

Example 4 (continued).

Pooled OLS on the crime rate data yields

$$\widehat{\texttt{crmrte}} = 93.42 + 7.94\texttt{D87} + 0.427\texttt{unem}. \tag{9}$$

The (wrong) p-value using OLS standard errors is 0.721, and applying Newey and West (1987) HAC standard errors $p = 0.693$. Thus, while the unemployment rate has now the expected sign, it is still deemed nonsignificant.

The main reason for collecting panel data is to allow for a_i to be correlated with the explanatory variables. This can be achieved by first writing down (6) explicitely for both time points:

$$\begin{aligned} y_{i1} &= \beta_0 + \beta_1 x_{i1} + a_i + u_{i1} && (t=1) \\ y_{i2} &= (\beta_0 + \delta_0) + \beta_1 x_{i2} + a_i + u_{i2} && (t=2). \end{aligned}$$

Subtract the first equation from the second:

$$(y_{i2} - y_{i1}) = \delta_0 + \beta_1 (x_{i2} - x_{i1}) + (u_{i2} - u_{i1}),$$

or

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i. \tag{10}$$

This is called the *first differenced equation*. Note that a_i has been "differenced away", which implies that estimation of (10) does not in any way depend upon whether a_i is correlated with x_{it} or not. When we obtain the OLS estimator of β_1 from (10), we call it the *first-difference estimator* (FD for short).

The parameters of (10) can be consistently estimated by OLS when the classical assumptions for regression analysis hold.

In particular, Δu_i must be uncorrelated with Δx_i, which holds if u_{it} is uncorrelated with x_{it} in <u>both</u> time periods. That is, we need *strict exogeneity*. In particular, this rules out including lagged dependent variables such as $y_{i,t-1}$ as explanatory variables. Lagged independent variables such as $x_{i,t-1}$ may be included without problems.

Another crucial assumption is that there must be variation in Δx_i. This rules out independent variables which do not change over time or change by the same amount for all cross-sectional units.

<u>Example 4 (continued).</u>

Estimation of (10) yields

$$\Delta\widehat{\texttt{crmrte}} = \underset{(t=3.28)}{15.40} + \underset{(t=2.52)}{2.22}\Delta\texttt{unem},$$

which now gives a positive, statistically significant relationship ($p = 0.015$) between unemployment and crime rates.

Policy Analysis with Two-Period Panel Data

Let y_{it} denote an outcome variable and let $\texttt{prog}_{it}$ be a program participation dummy variable. A simple unobserved effects model is

$$y_{it} = \beta_0 + \delta_0 D_{2,t} + \beta_1 \texttt{prog}_{it} + a_i + u_{it}. \tag{11}$$

Differencing yields

$$\Delta y_i = \delta_0 + \beta_1 \Delta\texttt{prog}_i + \Delta u_i. \tag{12}$$

In the special case that program participation occured only in the second period, $\Delta\texttt{prog}_i = \texttt{prog}_{i2}$, and the OLS estimator of β_1 has the simple interpretation

$$\hat{\beta}_1 = \overline{\Delta y}_{\texttt{Treatment}} - \overline{\Delta y}_{\texttt{Control}}, \tag{13}$$

which is the panel data version of the difference-in-differences estimator (5) in pooled cross sections.

The advantage of using panel data as opposed to pooled cross sections is that there is no need to include further variables to control for unit specific characteristics, since by using the same units at both times, these are automatically controlled for.

Example 5.
Job training program on worker productivity.

Let scrap_{it} denote the scrap rate of firm i during year t, and let grant_{it} be a dummy equal to one if firm i received a job training grant in year t. Pooled OLS yields using data from the years 1987 and 1988

$$\log(\texttt{scrap}_{it}) = \underset{(p=0.005)}{0.5974} - \underset{(p=0.566)}{0.1889}D_{88} + \underset{(p=0.896)}{0.0566}\texttt{grant},$$

suggesting that grants increase scrap rates. The preceding model suffers most likely from omitted variables bias. Estimating the first differenced equation (12) instead yields

$$\Delta \widehat{log(\texttt{scrap})} = \underset{(p=0.557)}{-0.057} - \underset{(p=0.059)}{0.317}\Delta\texttt{grant}.$$

Having a job training grant is estimated to lower the scrap rate by about 27.2%, since $\exp(-0.317) - 1 \approx -0.272$. The effect is significant at 10% but not at 5%. The large difference between $\widehat{\beta}_1$ obtained from pooled OLS and applying the first differenced estimator suggests that grants were mainly placed to firms which produce poorer quality.
No further variables (controls) with possible impact upon scrap rates need to be included in the model.

Differencing with More Than 2 Periods

The fixed effects model in the general case with k regressors and T time periods is

$$
\begin{aligned}
(14) \qquad y_{it} =& \delta_1 + \delta_2 D_{2,t} + \cdots + \delta_T D_{T,t} \\
&+ \sum_{j=1}^{k} \beta_j x_{tij} + a_i + u_{it},
\end{aligned}
$$

The key assumption is that the idiosyncratic errors are uncorrelated with the explanatory variables at all times (strict exogeneity):

$$
(15) \qquad \mathbb{C}\text{ov}(x_{itj}, u_{is}) = 0 \text{ for all } t, s \text{ and } j,
$$

which rules out using lagged dependent variables as regressors. Differencing(14) yields

$$
\begin{aligned}
(16) \qquad \Delta y_{it} =& \delta_2 \Delta D_{2,t} + \cdots + \delta_T \Delta D_{T,t} \\
&+ \sum_{j=1}^{k} \beta_j \Delta x_{tij} + \Delta u_{it}
\end{aligned}
$$

for $t = 2, \ldots, T$. Note that both the intercept δ_1 and the unobservable effect a_i have disappeared. This implies that while possible correlations between a_i and any of the explanatory variables causes omitted variables bias in (14), it causes no problem in estimating the first differenced equation (16).

Example 6.
Enterprise Zones and Unemployment Claims

Unemployment claims $\texttt{uclms}$ in 22 cities from 1980 to 1988 as a function of whether the city has an enterprise zone ($\texttt{ez} = 1$) or not:

$$\log(\texttt{uclms}_{it}) = \theta_t + \beta_1 \texttt{ez}_{it} + a_i + u_{it},$$

where θ_t shifts the intercept with appropriate year dummies. FD estimation output:

Dependent Variable: D(LUCLMS)
Method: Panel Least Squares
Date: 11/29/12 Time: 16:45
Sample (adjusted): 1981 1988
Periods included: 8
Cross-sections included: 22
Total panel (balanced) observations: 176

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D81	-0.321632	0.046064	-6.982279	0.0000
D82	0.457128	0.046064	9.923744	0.0000
D83	-0.354751	0.046064	-7.701262	0.0000
D84	-0.338770	0.050760	-6.673948	0.0000
D85	0.001449	0.048208	0.030058	0.9761
D86	-0.029478	0.046064	-0.639934	0.5231
D87	-0.267684	0.046064	-5.811126	0.0000
D88	-0.338684	0.046064	-7.352471	0.0000
D(EZ)	-0.181878	0.078186	-2.326211	0.0212

R-squared	0.622997	Mean dependent var	-0.159387
Adjusted R-squared	0.604937	S.D. dependent var	0.343748
S.E. of regression	0.216059	Akaike info criterion	-0.176744
Sum squared resid	7.795839	Schwarz criterion	-0.014617
Log likelihood	24.55348	Hannan-Quinn criter.	-0.110986
Durbin-Watson stat	2.441511		

The presence of an enterprise zone appears to reduce unemployment claims by about 18% ($p = 0.0212$).

Note that we have replaced the change in year dummies ΔD in (16) with the year dummies themselves. This can be shown to have no effect on the other parameter estimates (here D(EZ)).

2.4 Dummy Variable Regression in Panels

Another way to eliminate possible correlations with the unobservable factors a_i in (14) is to model them explicitly as dummy variables, where each cross-sectional unit gets its own dummy. This may be written as

$$\text{(17)} \qquad \boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\mu} + \boldsymbol{u}, \text{ where}$$

for N cross sections and T time periods:
$\boldsymbol{y}$ is a $(NT \times 1)$ vector of observations on y_{it},
$\boldsymbol{X}$ is a $(NT \times k)$ matrix of regressors x_{itj},
$\boldsymbol{\beta}$ is a $(k \times 1)$ vector of slope parameters β_j,
$\boldsymbol{Z}$ is a $(NT \times N)$ matrix of dummies,
$\boldsymbol{\mu}$ is a $(N \times 1)$ vector of unobservables a_i, and
$\boldsymbol{u}$ is a $(NT \times 1)$ vector of error terms u_{it}.

It is customs to stack $\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}$ and $\boldsymbol{u}$ such that the slower index is over cross sections i, and the faster index is over time points t, e.g.

$$\boldsymbol{y}' = (y_{11}, \ldots, y_{1T}, \;\cdots\;, y_{N1}, \ldots, y_{NT}).$$

Note that there is no constant in (17) in order to avoid exact multicollinearity (dummy variable trap). If you wish to include a constant, use only $N-1$ dummy variables for the N cross-sectional units.

Example 6. (continued)
Regressing log(uclms) on the year dummies, 22 dummies for the cities in sample and the enterprise zone dummy ez yields

Dependent Variable: LUCLMS
Method: Panel Least Squares
Date: 12/04/12 Time: 10:39
Sample: 1980 1988
Periods included: 9
Cross-sections included: 22
Total panel (balanced) observations: 198

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D81	-0.321632	0.060457	-5.319980	0.0000
D82	0.135496	0.060457	2.241179	0.0263
D83	-0.219255	0.060457	-3.626613	0.0004
D84	-0.579152	0.062318	-9.293490	0.0000
D85	-0.591787	0.065495	-9.035540	0.0000
D86	-0.621265	0.065495	-9.485616	0.0000
D87	-0.888949	0.065495	-13.57268	0.0000
D88	-1.227633	0.065495	-18.74379	0.0000
C1	11.67615	0.080079	145.8073	0.0000
C2	11.48266	0.079105	145.1574	0.0000
C3	11.29721	0.079105	142.8131	0.0000
C4	11.13498	0.079105	140.7621	0.0000
C5	11.68718	0.078930	148.0695	0.0000
C6	12.23073	0.080079	152.7326	0.0000
C7	12.42622	0.080079	155.1738	0.0000
C8	11.61739	0.078930	147.1852	0.0000
C9	12.02958	0.078930	152.4074	0.0000
C10	13.32116	0.079105	168.3987	0.0000
C11	11.54584	0.079105	145.9560	0.0000
C12	11.64117	0.079105	147.1612	0.0000
C13	10.84358	0.079105	137.0784	0.0000
C14	10.80252	0.078930	136.8613	0.0000
C15	11.44073	0.079105	144.6273	0.0000
C16	12.11190	0.079105	153.1118	0.0000
C17	11.23093	0.080079	140.2475	0.0000
C18	11.63326	0.079105	147.0611	0.0000
C19	11.76956	0.079105	148.7842	0.0000
C20	11.32518	0.080079	141.4244	0.0000
C21	12.13394	0.080079	151.5240	0.0000
C22	11.89479	0.079105	150.3673	0.0000
EZ	-0.104415	0.055419	-1.884091	0.0613

R-squared	0.933188	Mean dependent var	11.19078
Adjusted R-squared	0.921185	S.D. dependent var	0.714236
S.E. of regression	0.200514	Akaike info criterion	-0.233004
Sum squared resid	6.714401	Schwarz criterion	0.281826
Log likelihood	54.06741	Hannan-Quinn criter.	-0.024618
Durbin-Watson stat	1.306450		

(marginally significant decrease by 10%.)

2.5 Fixed Effects (FE) Estimation in Panels

Dummy variable regressions become impractical when the number of cross-sections gets large. An alternative method, which turns out to yield identical results, is called the *fixed effects* method.

As an example consider the simple model

$$y_{it} = \beta_1 x_{it} + a_i + u_{it}, \tag{18}$$

$$i = 1, \ldots, N, \quad t = 1, \ldots, T.$$

Thus there are altogether $N \times T$ observations.

Define means over the T time periods

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}, \quad \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}, \quad \bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}. \tag{19}$$

Then averaging over T yields

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{u}_i, \tag{20}$$

since

$$\frac{1}{T}\sum_{t=1}^{T} a_i = \frac{1}{T} T a_i = a_i.$$

Thus, subtracting (20) from (18) eliminates a_i and gives

$$y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i) \tag{21}$$

or

$$\dot{y}_{it} = \beta_1 \dot{x}_{it} + \dot{u}_{it}, \tag{22}$$

where e.g., $\dot{y}_{it} = y_{it} - \bar{y}_i$ is the time demeaned data on y.

This transformation is also called the *within transformation* and resulting (OLS) estimators of the regression parameters applied to (22) are called *fixed effect estimators* or *within estimators*. It generalizes to k regressors as

$$\dot{y}_{it} = \beta_1 \dot{x}_{it1} + \ldots + \beta_k \dot{x}_{itk} + \dot{u}_{it}. \tag{23}$$

Remark. The slope coefficient β_1 estimated from (20) (including a constant) is called the between estimator. $v_i = a_i + \bar{u}_i$ is the error term. This estimator is biased, however, if the unobserved component a_i is correlated with x.

Example 6. (continued)
Regressing the differences of log(uclms) from their means upon the differences of the year dummies from their means and the differences of the enterprize zone dummy ez from its means yields

Dependent Variable: LUCLMS-MLUCLMS
Method: Panel Least Squares
Date: 12/04/12 Time: 13:09
Sample: 1980 1988
Periods included: 9
Cross-sections included: 22
Total panel (balanced) observations: 198

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D81-MD81	-0.321632	0.056830	-5.659560	0.0000
D82-MD82	0.135496	0.056830	2.384236	0.0181
D83-MD83	-0.219255	0.056830	-3.858104	0.0002
D84-MD84	-0.579152	0.058579	-9.886703	0.0000
D85-MD85	-0.591787	0.061566	-9.612288	0.0000
D86-MD86	-0.621265	0.061566	-10.09109	0.0000
D87-MD87	-0.888949	0.061566	-14.43903	0.0000
D88-MD88	-1.227633	0.061566	-19.94022	0.0000
EZ-MEZ	-0.104415	0.052094	-2.004355	0.0465

R-squared	0.841596	Mean dependent var	-1.27E-16
Adjusted R-squared	0.834892	S.D. dependent var	0.463861
S.E. of regression	0.188483	Akaike info criterion	-0.455226
Sum squared resid	6.714401	Schwarz criterion	-0.305760
Log likelihood	54.06741	Hannan-Quinn criter.	-0.394727
Durbin-Watson stat	1.306450		

We recover the parameter estimates of the dummy variable regression, however not the standard errors. For example, the within estimator for the enterprise zone is -0.1044, the same as previously, but its standard error has decreased from 0.0554 to 0.0521 with a corresponding decrease in p-values from 0.0613 to 0.0465 now.

In order to understand the discrepancy in standard errors recall from STAT1010 (see also equations (18) and (22) of chapter 1) that the standard error of a slope coefficient is inverse proportional to the square root of the number of observations minus the number of regressors (including the constant).

In the dummy variable regression there are NT observations and $k+N$ regressors (k original regressors and N cross-sectional dummies). The degrees of freedom are therefore

$$df = NT - (k+N) = N(T-1) - k.$$

The demeaned regression sees only k regressors on the same NT observations, and therefore calculates the degrees of freedom (incorrectly, for our purpose) as $df_{\text{demeaned}} = NT - k$.

In order to correct for this, multiply the wrong standard errors of the demeaned regression by the square root of $NT - k$ and divide this with the square root of $N(T-1) - k$:

$$SE = \sqrt{\frac{NT-k}{N(T-1)-k}}\, SE_{\text{demeaned}}. \tag{24}$$

Example 6. (continued)

We have $N = 22$ cross-sectional units and $T = 9$ time periods for a total of $NT = 198$ observations. There is one dummy for the enterprize zone and eight year dummies for a total of $k = 9$ regressors. The correction factor for the standard errors is therefore

$$\sqrt{\frac{NT-k}{N(T-1)-k}} = \sqrt{\frac{22 \cdot 9 - 9}{22 \cdot 8 - 9}} = \sqrt{\frac{189}{167}} \approx 1.063831.$$

For example, multiplying the demeaned standard error of 0.052094 for the enterprise zone dummy with the correction factor yields

$$1.063831 \cdot 0.052094 = 0.055419,$$

which is the correct standard error that we found from the dummy regression earlier.

Taken together with its coefficient estimate of -0.1044 it will hence correctly reproduce the t-statistic of -1.884 with p-value 0.0613, however without the need to define 22 dummy variables!

EViews can do the degrees of freedom adjustment automatically, if you tell it that you have got panel data. In order to do that, choose

Structure/Resize Current Page...

from the Proc Menue. In the Workfile Structure Window, choose 'Dated Panel' and provide two identifyers: one for the cross section and one for time.

This will provide you with a 'Panel Options' tab in the estimation window. In order to apply the fixed effects estimator, (which, as we discussed, is equivalent to a dummy variable regression), change the effects specification for the cross-section into 'Fixed'.

Note that EViews reports a constant C, even though the demeaned regression shouldn't have any. C is to be interpreted as the average unobservable effect $\bar{a}_i$, or cross-sectional average intercept.

Example 6. (continued)

Applying the Fixed Effects option in Eviews yields

Dependent Variable: LUCLMS
Method: Panel Least Squares
Date: 12/05/12 Time: 11:47
Sample: 1980 1988
Periods included: 9
Cross-sections included: 22
Total panel (balanced) observations: 198

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	11.69439	0.042750	273.5544	0.0000
D81	-0.321632	0.060457	-5.319980	0.0000
D82	0.135496	0.060457	2.241179	0.0263
D83	-0.219255	0.060457	-3.626613	0.0004
D84	-0.579152	0.062318	-9.293490	0.0000
D85	-0.591787	0.065495	-9.035540	0.0000
D86	-0.621265	0.065495	-9.485616	0.0000
D87	-0.888949	0.065495	-13.57268	0.0000
D88	-1.227633	0.065495	-18.74379	0.0000
EZ	-0.104415	0.055419	-1.884091	0.0613

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.933188	Mean dependent var	11.19078
Adjusted R-squared	0.921185	S.D. dependent var	0.714236
S.E. of regression	0.200514	Akaike info criterion	-0.233004
Sum squared resid	6.714401	Schwarz criterion	0.281826
Log likelihood	54.06741	Hannan-Quinn criter.	-0.024618
F-statistic	77.75116	Durbin-Watson stat	1.306450
Prob(F-statistic)	0.000000		

The output coincides with that obtained from the dummy variable regression. C is the average of the cross-sectional city dummies C1 to C22.

R^2 in Fixed Effects Estimation

Note from the preceding example that while both the dummy regression and the fixed effects estimation yield an identical coefficient of determination of $R^2 = 0.933188$, it differs from $R^2 = 0.841596$, which we obtained when calculating the FE estimator by hand. Both ways of calculating R^2 are used.

The lower R^2 obtained from estimating (23) has the more intuitive interpretation as the amount of variation in y_{it} explained by the time variation in the explanatory variables.

The higher R^2 obtained in fixed effects estimation and dummy variable regressions should be used in F-tests when for example testing for joint significance of the unobservables a_i, that is the cross-sectional dummies in dummy variable regression.

Limitations

As with first differencing, the fact that we eliminated the unobservables a_i in estimation of (23) implies that any explanatory variable that is constant over time gets swept away by the fixed effects transformation. Therefore we cannot include dummies such as gender or race.

If we furthermore include a full set of time dummies, then, in order to avoid exact multicollinearity, we may neither include variables which change by a constant amount through time, such as working experience. Their effect will be absorbed by the year dummies in the same way as the effect of time-constant cross-sectional dummies is absorbed by the unobservables.

Example 7

Data set `wagepan.xls` (Wooldridge): $n = 545$, $T = 8$.

Is there a wage premium in belonging to labor union?

$$\begin{aligned}\log(\texttt{wage}_{it}) &= \beta_0 + \beta_1\texttt{educ}_{it} + \beta_3\texttt{expr}_{it} + \beta_4\texttt{expr}^2_{it} \\ &\quad +\beta_5\texttt{married}_{it} + \beta_6\texttt{union}_{it} + a_i + u_{it}\end{aligned}$$

Year (`d81` to `d87`) and race dummies (`black` and `hisp`) are also included. Pooled OLS with $\nu_{it} = a_i + u_{it}$ yields

Dependent Variable: LWAGE
Method: Panel Least Squares
Date: 12/11/12 Time: 12:32
Sample: 1980 1987
Periods included: 8
Cross-sections included: 545
Total panel (balanced) observations: 4360
White period standard errors & covariance (d.f. corrected)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.092056	0.160807	0.572460	0.5670
EDUC	0.091350	0.011073	8.249575	0.0000
BLACK	-0.139234	0.050483	-2.758032	0.0058
HISP	0.016020	0.039047	0.410265	0.6816
EXPER	0.067234	0.019580	3.433820	0.0006
EXPERSQ	-0.002412	0.001024	-2.354312	0.0186
MARRIED	0.108253	0.026013	4.161480	0.0000
UNION	0.182461	0.027421	6.653964	0.0000
D81	0.058320	0.028205	2.067692	0.0387
D82	0.062774	0.036944	1.699189	0.0894
D83	0.062012	0.046211	1.341930	0.1797
D84	0.090467	0.057941	1.561356	0.1185
D85	0.109246	0.066794	1.635577	0.1020
D86	0.141960	0.076174	1.863633	0.0624
D87	0.173833	0.085137	2.041805	0.0412

R-squared	0.189278	Mean dependent var	1.649147
Adjusted R-squared	0.186666	S.D. dependent var	0.532609
S.E. of regression	0.480334	Akaike info criterion	1.374764
Sum squared resid	1002.481	Schwarz criterion	1.396714
Log likelihood	-2981.986	Hannan-Quinn criter.	1.382511
F-statistic	72.45876	Durbin-Watson stat	0.864696
Prob(F-statistic)	0.000000		

The serial correlation in the residuals has been accounted for by using White period standard errors. But the parameter estimates are biased if a_i is correlated with any of the explanatory variables.

Example 7 (continued.)

Fixed Effects estimation yields

Dependent Variable: LWAGE
Method: Panel Least Squares
Date: 11/26/12 Time: 12:31
Sample: 1980 1987
Periods included: 8
Cross-sections included: 545
Total panel (balanced) observations: 4360

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.426019	0.018341	77.74835	0.0000
EXPERSQ	-0.005185	0.000704	-7.361196	0.0000
MARRIED	0.046680	0.018310	2.549385	0.0108
UNION	0.080002	0.019310	4.142962	0.0000
D81	0.151191	0.021949	6.888319	0.0000
D82	0.252971	0.024418	10.35982	0.0000
D83	0.354444	0.029242	12.12111	0.0000
D84	0.490115	0.036227	13.52914	0.0000
D85	0.617482	0.045244	13.64797	0.0000
D86	0.765497	0.056128	13.63847	0.0000
D87	0.925025	0.068773	13.45039	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.620912	Mean dependent var	1.649147
Adjusted R-squared	0.565718	S.D. dependent var	0.532609
S.E. of regression	0.350990	Akaike info criterion	0.862313
Sum squared resid	468.7531	Schwarz criterion	1.674475
Log likelihood	-1324.843	Hannan-Quinn criter.	1.148946
F-statistic	11.24956	Durbin-Watson stat	1.821184
Prob(F-statistic)	0.000000		

Note that we could not include the years of education and the race dummies, because they remain constant through time for each cross section. Likewise we could not include years of working experience, because they change by the same amount for all cross sections, and we included already a full set of year dummies.

The large changes in the premium estimates for marriage and union membership suggests that a_i is correlated with some of the explanatory variables.

Fixed effects or first differencing?

If the number of periods is 2 ($T = 2$) FE and FD give identical results.

When $T \geq 3$ the FE and FD are not the same.

Both are unbiased as well as consistent for fixed T as $N \to \infty$ under the assumptions FE.1-FE.4 below:

Assumptions:
FE.1: For each i, the model is

$$y_{it} = \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}, t = 1, \ldots T.$$

FE.2: We have a random sample.
FE.3: All explanatory variables change over time, and they are not perfectly collinear.
FE.4: $\mathbb{E}[u_{it}|\boldsymbol{X}_i, a_i] = 0$ <u>for all time periods</u> ($\boldsymbol{X}_i$ stands for all explanatory variables).

If we add the following two assumptions, FE is the best linear unbiased estimator:

FE.5: $\mathbb{V}\text{ar}[u_{it}|\boldsymbol{X}_i, a_i] = \sigma_u^2$ for all $t = 1, \ldots, T$.
FE.6: $\mathbb{C}\text{ov}[u_{it}, u_{is}|\boldsymbol{X}_i, a_i] = 0$ for all $t \neq s$.

In that case FD is worse than FE because FD is linear and unbiased under FE.1–FE.4.

While this looks like a clear case for FE, it is not, because often FE.6 is violated. If u_{it} is (highly) serially correlated, Δu_{it} may be less serially correlated, which may favor FD over FE. However, typically T is rather small, such that serial correlation is difficult to observe. Usually it is best to check both FE and FD.

If we add as a last assumption
FE.7: $u_{it}|\boldsymbol{X}_i, a_i \sim \text{NID}(0, \sigma_u^2)$,
then we may use exact t and F-statistics. Otherwise they hold only asymptotically for large N and T.

Balanced and unbalanced panels

A data set is called a **balanced panel** if the same number of time series observations are available for each cross section units. That is T is the same for all individuals. The total number of observations in a balanced panel is NT.

All the above examples are balanced panel data sets.

If some cross section units have missing observations, which implies that for an individual i there are available T_i time period observations $i = 1, \ldots, N$, $T_i \neq T_j$ for some i and j, we call the data set an **unbalanced panel**. The total number of observations in an unbalanced panel is $T_1 + \cdots + T_N$.

In most cases unbalanced panels do not cause major problems to fixed effect estimation.

Modern software packages make appropriate adjustments to estimation results.

2.6 Random effects models

Consider the simple unobserved effects model

(25) $$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it},$$

$i = 1, \ldots, n$, $t = 1, \ldots, T$.

Typically also time dummies are included in (25).

Using FD or FE eliminates the unobserved component a_i. As discussed earlier, the idea is to avoid omitted variable bias which arises necessarily as soon as a_i is correlated with x_{it}.

However, if a_i is uncorrelated with x_{it}, then using a transformation to eliminate a_i results in inefficient estimators. So called random effect (RE) estimators are more efficient in that case.

Generally, we call the model in equation (25) the **random effects model** if a_i is uncorrelated with all explanatory variables, i.e.,

$$\mathbb{C}\text{ov}[x_{it}, a_i] = 0, \ t = 1, \dots, T. \tag{26}$$

How to estimate β_1 efficiently?

If (26) holds, β_1 can be estimated consistently from a single cross section. So in principle, there is no need for panel data at all. But using a single cross section obviously discards a lot lot of useful information.

If the data set is simply pooled and the error term is denoted as $v_{it} = a_i + u_{it}$, we have the regression

$$y_{it} = \beta_0 + \beta_1 x_{it} + v_{it}. \tag{27}$$

Then $E[v_{it}^2] = \sigma_a^2 + \sigma_u^2$ and $E[v_{it}v_{is}] = \sigma_a^2$ for $t \neq s$, such that

$$\mathbb{C}\text{orr}[v_{it}, v_{is}] = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} \tag{28}$$

for $t \neq s$, where $\sigma_a^2 = \mathbb{V}\text{ar}[a_i]$ and $\sigma_u^2 = \mathbb{V}\text{ar}[u_{it}]$.

That is, the error terms v_{it} are (positively) autocorrelated, which biases the standard errors of the OLS $\hat{\beta}_1$.

If σ_a^2 and σ_u^2 were known, optimal estimators (BLUE) would be obtained by generalized least squares (GLS), which in this case would reduce to estimating the regression slope coefficients from the quasi demeaned equation (29)

$$y_{it} - \lambda\bar{y}_t = \beta_0(1-\lambda) + \beta_1(x_{it} - \lambda\bar{x}_i) + (v_{it} - \lambda\bar{v}_i),$$

where

$$\lambda = 1 - \left(\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}\right)^{\frac{1}{2}}. \tag{30}$$

In practice σ_u^2 and σ_a^2 are unknown, but they can be estimated for example as follows:

Estimate (27) from the pooled data set and use the OLS residuals $\widehat{v}_{it}$ to estimate σ_a^2 from the average covariance of $\widehat{v}_{it}$ and $\widehat{v}_{is}$ for $t \neq s$.

In the second step, estimate σ_u^2 from the variance of the OLS residuals $\widehat{v}_{it}$ as $\widehat{\sigma}_u^2 = \widehat{\sigma}_\nu^2 - \widehat{\sigma}_a^2$.

Finally plug these estimates of σ_a^2 and σ_u^2 into equation (30). Regression packages do this automatically.

The resulting GLS estimators for the regression slope coefficients are called **random effects estimators** (RE estimators). Other estimators of σ_a^2 and σ_u^2 (and therefore λ) are available. The particular version we discussed is the Swamy-Arora estimator.

Under the random effects assumptions* the estimators are consistent, but not unbiased.

They are also asymptotically normal as $N \to \infty$ for fixed T.

However, for small N and large T the properties of the RE estimator are largely unknown.

*The ideal random effects assumptions include FE.1, FE.2, FE.4–FE.6.

FE.3 is replaced with
RE.3: There are no perfect linear relationships among the explanatory variables.
RE.4: In addition of FE.4, $\mathbb{E}[a_i|X_i] = 0$.

Note that $\lambda = 0$ in (29) corresponds to pooled regression and $\lambda = 1$ to FE, such that for $\sigma_u^2 \ll \sigma_a^2$ ($\lambda \approx 1$) RE estimates will be similiar to FE estimates, whereas for $\sigma_u^2 \gg \sigma_a^2$ ($\lambda \approx 0$) RE estimates will resemble pooled OLS estimates.

Example 7 (continued.)

Note that the constant dummies `black` and `hisp` and the variable with constant change `exper`, which dropped out with the FE method, can be estimated with RE.

$$\hat{\lambda} = 1 - \left(\frac{0.351^2}{0.351^2 + 8 \cdot 0.3246^2} \right)^{1/2} = 0.643,$$

such that the RE estimates lie closer to the FE estimates than to the pooled OLS estimates.

Applying RE is probably not appropriate in this case, because, as discussed earlier, the unobservable a_i is probably correlated with some of the explanatory variables.

EViews output for RE estimation:

Dependent Variable: LWAGE
Method: Panel EGLS (Cross-section random effects)
Date: 11/26/12 Time: 12:26
Sample: 1980 1987
Periods included: 8
Cross-sections included: 545
Total panel (balanced) observations: 4360
Swamy and Arora estimator of component variances

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.023586	0.150265	0.156965	0.8753
EDUC	0.091876	0.010631	8.642166	0.0000
BLACK	-0.139377	0.047595	-2.928388	0.0034
HISP	0.021732	0.042492	0.511429	0.6091
EXPER	0.105755	0.015326	6.900482	0.0000
EXPERSQ	-0.004724	0.000688	-6.869682	0.0000
MARRIED	0.063986	0.016729	3.824781	0.0001
UNION	0.106134	0.017806	5.960582	0.0000
D81	0.040462	0.024628	1.642894	0.1005
D82	0.030921	0.032255	0.958646	0.3378
D83	0.020281	0.041471	0.489036	0.6248
D84	0.043119	0.051179	0.842509	0.3995
D85	0.057815	0.061068	0.946733	0.3438
D86	0.091948	0.071039	1.294334	0.1956
D87	0.134929	0.081096	1.663821	0.0962

Effects Specification

	S.D.	Rho
Cross-section random	0.324603	0.4610
Idiosyncratic random	0.350990	0.5390

Weighted Statistics

R-squared	0.180618	Mean dependent var	0.588893
Adjusted R-squared	0.177977	S.D. dependent var	0.388166
S.E. of regression	0.351932	Sum squared resid	538.1558
F-statistic	68.41243	Durbin-Watson stat	1.589754
Prob(F-statistic)	0.000000		

Unweighted Statistics

R-squared	0.182847	Mean dependent var	1.649147
Sum squared resid	1010.433	Durbin-Watson stat	0.846702

Random effects or fixed effects?

FE is widely considered preferable because it allows correlation between a_i and x variables.

Given that the common effects, aggregated to a_i is not correlated with x variables, an obvious advantage of the RE is that it allows also estimation of the effects of factors that do not change in time (like education in the above example).

Typically the condition that common effects a_i is not correlated with the regressors (x-variables) should be considered more like an exception than a rule, which favors FE.

Whether this condition is met, can be tested with the Hausman test to be discussed in the following.

Hausman specification test

Hausman (1978) devised a test for the orthogonality of the common effects (a_i) and the regressors.

The basic idea of the test relies on the fact that under the null hypothesis of orthogonality both OLS and GLS are consistent, while under the alternative hypothesis GLS is not consistent. Thus, under the null hypothesis OLS and GLS estimates should not differ much from each other.

The Hausman test statistic is a transformation of the differences between the parameter estimates obtained from RE and FE estimation, which becomes asymptotically χ^2-distributed under the null hypothesis (26)

$$H_0: \ \mathbb{C}\mathrm{ov}[x_{it}, a_i] = 0, \ t = 1, \ldots, T.$$

The degrees of freedom are the number of regressors, where only those regressors may be included which are estimable with FE, that is, time-constant variables must be dropped (also constant time changes, if year dummies are included).

In order to perform the Hausman test in EViews, first estimate the model with RE including only those regressors, which are estimable with FE as well. Then select

View/ Fixed/Random Effects Testing/ Correlated Random Effects - Hausman Test.

The first part of the output is the Hausman test statistic with degrees of freedom and p-value. The second part lists the parameter estimates for both FE and RE estimation. The final third part is more detailed FE estimation output.

Example 7 (continued.)

As expected, the Hausman test strongly rejects the null hypothesis, that a_i would be uncorrelated with all explanatory variables. Therefore, RE is inappropriate and we must use FE parameter estimates instead.

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

Test Summary	Chi-Sq. Statistic	Chi-Sq. d.f.	Prob.
Cross-section random	35.992523	10	0.0001

Cross-section random effects test comparisons:

Variable	Fixed	Random	Var(Diff.)	Prob.
EXPERSQ	-0.005185	-0.003139	0.000000	0.0001
MARRIED	0.046680	0.078034	0.000054	0.0000
UNION	0.080002	0.103974	0.000051	0.0008
D81	0.151191	0.133626	0.000016	0.0000
D82	0.252971	0.214562	0.000081	0.0000
D83	0.354444	0.290904	0.000228	0.0000
D84	0.490115	0.398106	0.000491	0.0000
D85	0.617482	0.494105	0.000915	0.0000
D86	0.765497	0.606478	0.001553	0.0001
D87	0.925025	0.724616	0.002467	0.0001

Cross-section random effects test equation:
Dependent Variable: LWAGE
Method: Panel Least Squares
Date: 12/12/12 Time: 11:41
Sample: 1980 1987
Periods included: 8
Cross-sections included: 545
Total panel (balanced) observations: 4360

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.426019	0.018341	77.74835	0.0000
EXPERSQ	-0.005185	0.000704	-7.361196	0.0000
MARRIED	0.046680	0.018310	2.549385	0.0108
UNION	0.080002	0.019310	4.142962	0.0000
D81	0.151191	0.021949	6.888319	0.0000
D82	0.252971	0.024418	10.35982	0.0000
D83	0.354444	0.029242	12.12111	0.0000
D84	0.490115	0.036227	13.52914	0.0000
D85	0.617482	0.045244	13.64797	0.0000
D86	0.765497	0.056128	13.63847	0.0000
D87	0.925025	0.068773	13.45039	0.0000

Effects Specification

Cross-section fixed (dummy variables)

R-squared	0.620912	Mean dependent var	1.649147
Adjusted R-squared	0.565718	S.D. dependent var	0.532609
S.E. of regression	0.350990	Akaike info criterion	0.862313
Sum squared resid	468.7531	Schwarz criterion	1.674475
Log likelihood	-1324.843	Hannan-Quinn criter.	1.148946
F-statistic	11.24956	Durbin-Watson stat	1.821184
Prob(F-statistic)	0.000000		