11th Economics Summer Seminars

Pamukkale University

Denizli

Turkey

Applied Spatial Econometrics

Bernard Fingleton
University of Cambridge
UK

Spatial panels

What does panel (or longitudinal) data look like?

- Each of N individual's data is measured on T occasions
- Individuals may be people, firms, countries etc
- Some variables change over time for $t = 1,\ldots,T$
- Some variables may be fixed over the time period, such as gender, the geographic location of a firm or a person's ethnic group
- When there are no missing data, so that there are NT observations, then we have a balanced panel (less than NT is called an unbalanced panel)
- Typically N is large relative to T, but not always

Why are panel data useful?

- With observations that span **both** time and individuals in a cross-section, more information is available, giving <u>more efficient</u> estimates.
- The use of panel data allows empirical tests of a <u>wide range of hypotheses</u>.
- With panel data we can control for :
  - Unobserved or unmeasurable sources of individual heterogeneity that vary across individuals but do not vary over time
  - omitted variable bias

Key Reading

- **Stock and Watson (2007), Chapter 10: Regression with panel data**
- **Baltagi(2002) Econometrics 3rd Edition**
- **Baltagi(2005) Econometric Analysis of Panel Data**

Assumptions of fixed effects

1. The slopes of the regression lines are the same across individuals
2. The fixed effects capture entirely the time-constant omitted variables
    - This means we can soak up unmodelled heterogeneity across individuals and thus avoid misspecification error
    - But if there are <u>time-varying</u> omitted variables, their effects would not be captured by the fixed effects
    - Fixed <u>time</u> effects are also possible
3. Inference is with respect to the particular set of individuals represented by the fixed effects, and unrelated to any larger population

Disadvantage of fixed effects

- Fixed effects wipe out explanatory variables that do not vary within an individual (ie are time-invariant, such as gender, race)
- We are often interested in in the effects of these separate sources of individual heterogeneity

The error components model : random effects

- The alternative to the fixed effects model is the <u>random effects model</u>
  - Individual effects unobserved but their effect felt in the residuals/errors
  - Error variance altered by presence of individual effects
  - Capture effects by components of error variance attributable to unobserved individual effects

Error components : random effects

We can write our model as an error components model,

so that

$$y_{it} = \beta_1 x_{1it} + ... \beta_K x_{Kit} + \varepsilon_{it}$$

$$\varepsilon_{it} = \alpha_i + u_{it}$$

$\alpha_i =$ individual specific permanent components

u_{it} = remainder transient components, a 'traditional' error term

ε_{it} = disturbance term

the disturbance term is a composite of the

two error components

The random effects model

- In the fixed effects approach, we do not make any hypotheses about the individual specific effects
- beyond the fact that they exist — and that can be tested
- Once these effects are swept out by taking deviations from the group means, or by dummy variables, the remaining parameters can be estimated
- So fixed effects model gives results conditional on a particular set of fixed effects (the population)

The random effects model

- the random effects approach attempts to model the unobservable individual effects as drawings from a probability distribution instead of removing them
- So inference is with respect to a larger population of possible outcomes of which the data are one sample
- With random effects, the <u>individual effects</u> are part of the <u>disturbance term</u>, that is, zero-mean random variables, uncorrelated with the regressors.

The random effects model

- The composite disturbance term means that OLS is not appropriate
- We therefore use GLS (generalised least squares)
- There are various GLS estimators, but all are asymptotically efficient as T and N become large

The random effects model

- the fixed-effects estimator “always works”, but at the cost of not being able to estimate the effect of time-invariant regressors.
  - This is because time-invariant regressors are perfectly correlated with the fixed effect dummies
- the random-effects estimator : time-invariant regressors can be estimated,
- but if individual effects (captured by the disturbance) are correlated with explanatory variables, then the random-effects estimator would be inconsistent, while fixed-effects estimates would still be valid.
- In contrast, the fixed effects are explicit (dummy) variables and can be correlated with the other X variables

The random effects model

- The random effects specification is appropriate if we assume the data are a representative sample of individuals *N* drawn at random from a large population
- Each individual effect is modelled as a random drawing from a probability distribution with mean 0 and with constant variance
- We are assuming that the composite disturbance term ε has a value for a particular individual at a specific time which is made up of two components

The random effects model

- Two components
- A <u>permanent component</u>, which allows for individual heterogeneity
  - This varies across individuals but is constant over time, reflecting the individual specific effect which is time-constant
- A <u>transient component</u>
  - this varies across individuals and across time and represents other unmodeled effects occurring at random

The random effects model

$$y_{it} = \beta_1 x_{1it} + \ldots \beta_K x_{Kit} + \varepsilon_{it}$$

$$\alpha_i \sim iid(0, \sigma_\alpha^2)$$

$$u_{it} \sim iid(0, \sigma_u^2)$$

$$\varepsilon_{it} = \alpha_i + u_{it}$$

$$\operatorname{cov}(\alpha_i ; u_{it}) = 0$$

$$\operatorname{cov}(x_1, \ldots, x_K ; \varepsilon_{it}) = 0$$

The random effects model

for OLS to be BLUE (the best linear unbiased estimator)

we require that

$E(\varepsilon_{it}^2) =$ a constant σ_ε^2 for all i and t

$E(\varepsilon_{it}, \varepsilon_{is}) = 0$ for $s \neq t$

$E(\varepsilon_{it}, \varepsilon_{jt}) = 0$ for $i \neq j$

If these assumptions are not met, and they are unlikely to be met

in the context of panel data, OLS is not the most efficient estimator.

Greater efficiency may be gained using generalized least

squares (GLS), taking into account the covariance structure of the error term.

The random effects model

$$\alpha_i \sim iid(0, \sigma_\alpha^2)$$

$$u_{it} \sim iid(0, \sigma_u^2)$$

$$\varepsilon_{it} = \alpha_i + u_{it}$$

$$\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = \text{var}(\varepsilon_{it}) = \sigma_\alpha^2 + \sigma_u^2 \text{ for } i = j \text{ and } t = s$$

$$\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = \sigma_\alpha^2 \text{ for } i = j \text{ and } t \neq s$$

$$\text{cov}(\varepsilon_{it}, \varepsilon_{js}) = 0 \text{ for } i \neq j$$

thus there is serial correlation over time between disturbances

of the same individual

these variances and covariances form the elements

of an NT by NT variance-covariance matrix Ω

which is the basis of GLS estimation (ie weighted least squares)

Error Covariance Structure Ω

OLS

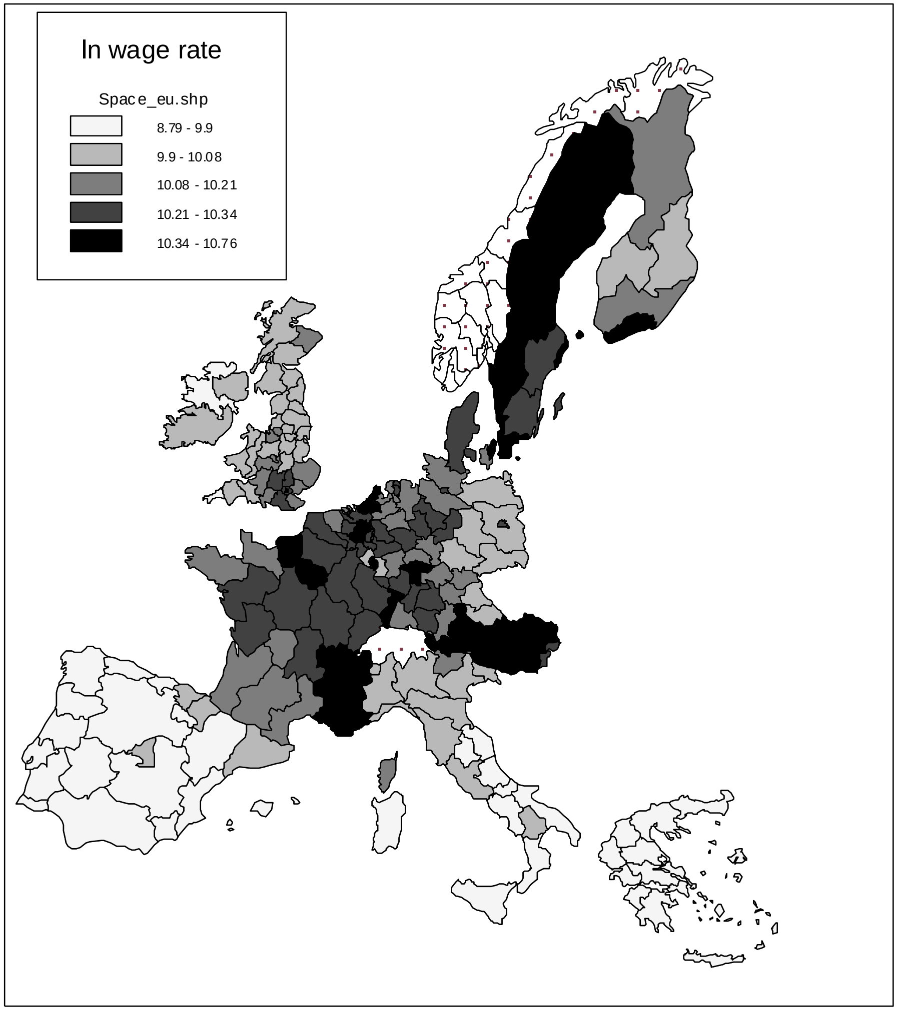
		i =1	i =1	i =2	i =2
		t =1	t =2	t =1	t =2
i =1	t =1	σ_ε^2	0	0	0
i =1	t =2	0	σ_ε^2	0	0
i =2	t =1	0	0	σ_ε^2	0
i =2	t =2	0	0	0	σ_ε^2

GLS

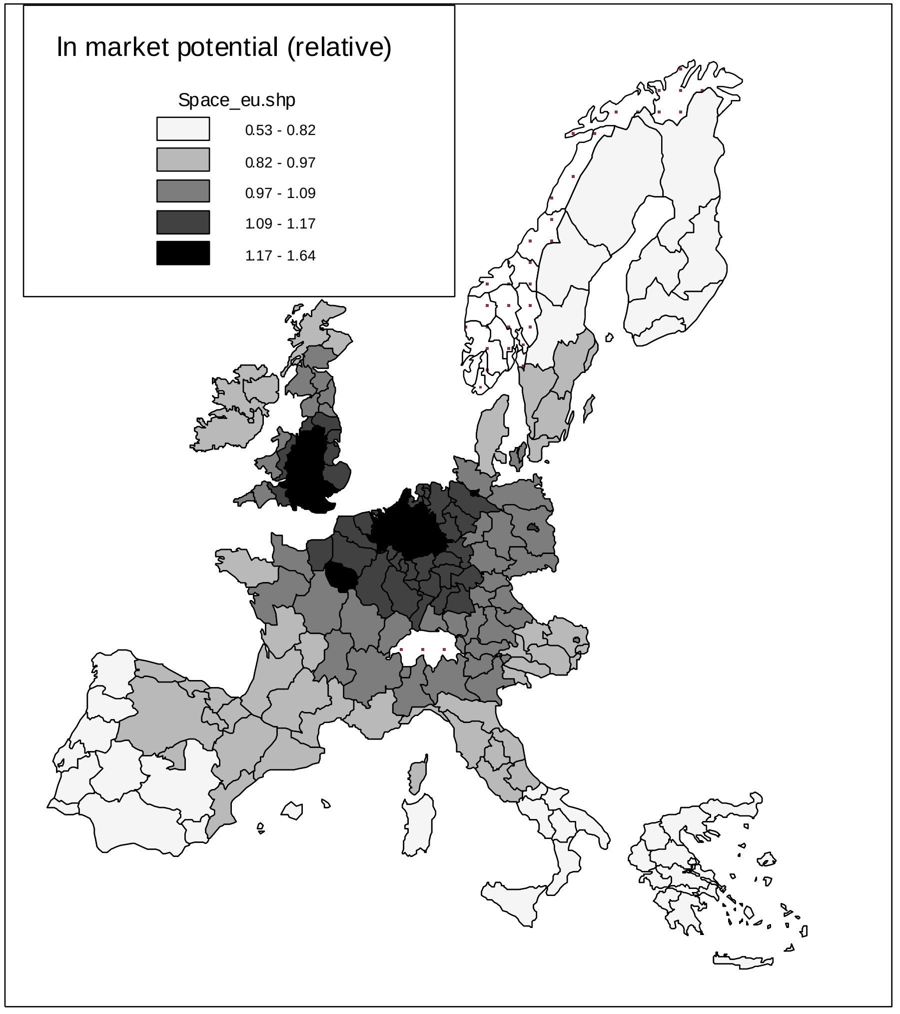
		i =1	i =1	i =2	i =2
		t =1	t =2	t =1	t =2
i =1	t =1	$\sigma_\alpha^2 + \sigma_u^2$	σ_α^2	0	0
i =1	t =2	σ_α^2	$\sigma_\alpha^2 + \sigma_u^2$	0	0
i =2	t =1	0	0	$\sigma_\alpha^2 + \sigma_u^2$	σ_α^2
i =2	t =2	0	0	σ_α^2	$\sigma_\alpha^2 + \sigma_u^2$

The random effects model

- We gain degrees of freedom
- We can introduce time invariant regressors (gender, race, religion etc) which are not wiped out by the presence of the fixed effect dummies
- Greater efficiency may be gained using generalized least squares (GLS), taking into account the covariance structure of the error term.



ln wage rate
Space_eu.shp
8.79 - 9.9
9.9 - 10.08
10.08 - 10.21
10.21 - 10.34
10.34 - 10.76



Fingleton B, Fischer M (2010) Neoclassical Theory versus New Economic Geography. Competing explanations of cross-regional variation in economic development, *Annals of Regional Science*, 44 467-491

Data layout 255 EU regions

CODE	NAME	lnGVApw	ln_adj_p_g	lns	lnMPa	lnHed	year_1995(1)-2003(9)	CZ	Eesti
AT11	Burgenland	10.57923	-2.89944	-1.34807	9.297275	1.811864	1	0	0
AT12	Niederösterreicl	10.70796	-2.8929	-1.43019	9.25791	2.035079	1	0	0
AT13	Wien	10.93456	-3.11427	-1.71378	10.17165	2.625214	1	0	0
AT21	Kärnten	10.67353	-2.94139	-1.46771	9.291567	1.949171	1	0	0
AT22	Steiermark	10.62469	-3.00101	-1.43836	9.236481	2.175438	1	0	0
AT31	Oberösterreich	10.70373	-2.91851	-1.46739	9.284781	1.891296	1	0	0
AT32	Salzburg	10.76274	-2.86082	-1.47764	9.330464	2.21531	1	0	0
AT33	Tirol	10.70657	-2.83274	-1.32531	9.334951	1.827362	1	0	0
AT34	Vorarlberg	10.7661	-2.85989	-1.42534	9.581739	1.872076	1	0	0
BE10	Région de Brux	11.00334	-2.9784	-1.7951	10.7211	2.893847	1	0	0
BE21	Prov. Antwerpe	10.96942	-2.9369	-1.61928	9.706519	2.588043	1	0	0
BE22	Prov. Limburg (	10.80939	-2.83605	-1.50151	9.653046	2.285049	1	0	0
BE23	Prov. Oost-Vlaa	10.81587	-2.94067	-1.53618	9.660629	2.595616	1	0	0
BE24	Prov. Vlaams B	11.00496	-2.84727	-1.63594	9.754418	2.875688	1	0	0
BE25	Prov. West-Vla	10.76332	-2.94501	-1.51328	9.632094	2.397464	1	0	0
BE31	Prov. Brabant V	10.95707	-2.74243	-1.60496	9.803192	2.989779	1	0	0
BE32	Prov. Hainaut	10.74821	-3.01044	-1.80255	9.576635	2.400279	1	0	0
BE33	Prov. Liège	10.76181	-3.00145	-1.72772	9.612037	2.532512	1	0	0
BE34	Prov. Luxembo	10.64172	-2.79525	-1.51756	9.503033	2.491419	1	0	0
BE35	Prov. Namur	10.67695	-2.86625	-1.81862	9.519095	2.682265	1	0	0
CH01	Région lémanic	11.06256	-2.66348	-1.44457	9.482571	2.615551	1	0	0
CH02	Espace Mittella	10.94564	-2.99998	-1.37025	9.483919	2.432667	1	0	0
CH03	Nordwestschwe	11.10428	-2.62383	-1.49313	9.764912	2.457364	1	0	0
CH04	Zürich	11.12237	-2.66996	-1.50482	9.8488	2.652089	1	0	0
CH05	Ostschweiz	10.98804	-2.74371	-1.4115	9.464399	2.257391	1	0	0
CH06	Zentralschweiz	11.09763	-2.76917	-1.54572	9.566808	2.396922	1	0	0
CH07	Ticino	10.83974	-2.88874	-1.21977	9.524018	2.322914	1	0	0
CZ01	Praha	9.404299	-3.06956	-1.1449	9.492706	2.5983	1	1	0
CZ02	Strední Cechy	8.805276	-3.02197	-1.19799	9.240543	1.383584	1	1	0
CZ03	Jihozápad	8.896247	-2.99863	-0.87192	9.244868	1.713004	1	1	0
CZ04	Severozápad	8.919282	-2.98325	-1.27958	9.274141	1.291539	1	1	0

Created by demo_4.m

```
***** Fixed Effects Model *****
Dependent Variable =        lnGVApw
R-squared       =    0.6282
Rbar-squared    =    0.6282
sige            =    0.0079
Time           =    0.0680
Nobs, Nvars     =   2295,      1
******************************************************************
Variable        Coefficient       t-statistic    t-probability
lnMPa              0.387843         62.252270         0.000000


 ***** Random Effects Model *****
Dependent Variable =        lnGVApw
R-squared       =    0.5985
Rbar-squared    =    0.5984
sige            =    0.0091
sigu            =    0.2546
Time           =    0.0700
Nobs, Nvars     =   2295,      2
******************************************************************
Variable        Coefficient       t-statistic    t-probability
lnMPa              0.390363         58.468376         0.000000
constant           6.597649         90.238169         0.000000


 ***** Random Effects Model *****
Dependent Variable =        lnGVApw
R-squared       =    0.6416
Rbar-squared    =    0.6413
sige            =    0.0082
sigu            =    0.2546
Time           =    0.0730
Nobs, Nvars     =   2295,      3
******************************************************************
Variable        Coefficient       t-statistic    t-probability
lnMPa              0.387763         61.437598         0.000000
ne                -0.086977        -16.601518         0.000000
constant           6.836775         96.868502         0.000000
```

Time Fixed Effects

- An omitted variable might vary over time but not across regions/countries/individuals:
  - E.G. legislation at EU level (employment, environment etc.)
- These produce intercepts that change over time

Time Fixed Effects

- The time fixed effects are introduced in exactly the same way as the individual fixed effects, with N-1 dummies (plus constant) or N (without constant) or demeaning
- In this case, the dummies are set to 1 for a specific time period, and zero otherwise
    - For example, the dummy variable for 1970 would have 1s for all the EU regions for 1970, and zeros for all other times
    - In contrast a region specific fixed effect has 1s for the region for all times, and zeros for all the other regions.
- Demeaning is with reference to time means not region means.

Time Fixed Effects

with both individual effects α_i

and time effects ϕ_t,

the model is

$$y_{it} = x_{it}\beta + \varepsilon_{it}$$

$$\varepsilon_{it} = \alpha_i + \phi_t + u_{it}$$

Created by demo_4.m

```
***** Fixed Effects Model *****
Dependent Variable =        lnGVApw
R-squared      =    0.6560
Rbar-squared   =    0.6548
sige           =    0.0074
Time         =    0.0310
Nobs, Nvars    =   2295,      9
*******************************************************************
Variable          Coefficient      t-statistic    t-probability
lnMPa                0.434074        35.806574         0.000000
year 2               0.052266         6.945393         0.000000
year 3               0.036084         5.760815         0.000000
year 4              -0.035705        -6.628769         0.000000
year 5              -0.023523        -4.371483         0.000013
year 6              -0.022299        -3.836978         0.000128
year 7              -0.021857        -3.396842         0.000693
year 8               0.000661         0.097962         0.921971
year 9               0.014775         2.195126         0.028255
```

Created by demo_4.m

```
***** Random Effects Model *****
Dependent Variable =         lnGVApw
R-squared      =    0.6670
Rbar-squared   =    0.6655
sige           =    0.0076
sigu           =    0.2547
Time         =    0.0940
Nobs, Nvars    =   2295,    11
******************************************************************
Variable       Coefficient      t-statistic    t-probability
lnMPa             0.379598         5.221714         0.000000
ne               -0.083983       -16.000748         0.000000
year 2            0.052886         6.858080         0.000000
year 3            0.045914         3.111910         0.001882
year 4           -0.013095        -0.427637         0.668955
year 5            0.002068         0.059953         0.952198
year 6            0.011946         0.260738         0.794318
year 7            0.018345         0.341823         0.732516
year 8            0.043305         0.761145         0.446649
year 9            0.057286         1.010002         0.312601
constant          6.512891        31.107920         0.000000
```