Applying Panel Data Methods to Other Data Structures

Professor Jan Minx

Hertie School of Governance

Stats II, Lecture ?, ? 2015

Applying Panel Data Methods to Other Data Structures

Motivation

- Panel data methods can be used with data structures that do not involve time
- Hierarchical data structures contain clusters of observation which share common characteristics
- When these characteristics are unobservable and correlated with other explanatory variables, pooled OLS will give us estimates that are biased and inefficient

Applying Panel Data Methods to Other Data Structures

Motivation

- Consider a geographical dataset that observes variables for small areas (in this case MSOAs, or Middle Layer Super Output Areas)
- Each small area belongs to a local authority
- If local authority attributes that we cannot observe affect our other variables, we will get biased and inefficient estimates using OLS

Applying Panel Data Methods to Other Data Structures

Motivation

Remember that OLS regression is estimated using the equation

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

When we use panel methods across time, our equation becomes

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

Here the variable a_i captures all unobserved, time-constant factors that affect y_{it}

Applying Panel Data Methods to Other Data Structures

Motivation

By constructing our dataset and a fixed effects model carefully, we can also account for fixed effects given by local authorities with the equation

$$y_{pc} = \beta_0 + \beta_1 x_{pc} + a_p + u_{pc}$$

where, given our hierarchical data structure, p indexes the parent (local authority) and c indexes the child (MSOA)

Here the local authority fixed effect is given by a_p, and the coefficient $\beta_1 x_{pc}$ describes the effect of our explanatory variable on our independent variable x *within* local authorities.

Applying Panel Data Methods to Other Data Structures

Pooled OLS

When we use a pooled OLS regression on our dataset to estimate the effect of income on household energy consumption, we get the following results

```
reg energy_consumption income_est
outtex, file(ols.tex) labels level detail legend key(stab) replace est store ols
```

Table : Estimation results : regress

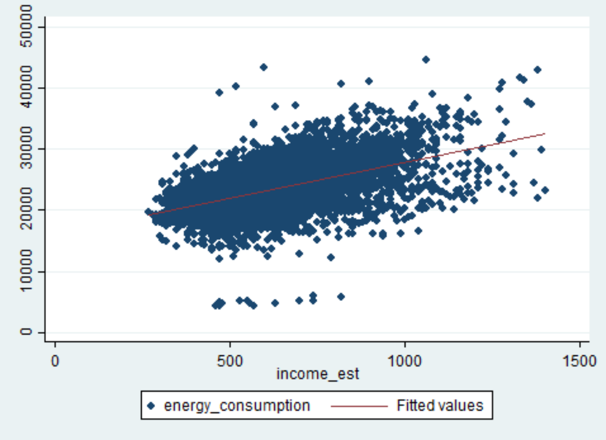
Variable	Coefficient	(Std. Err.)
income_est	11.681**	(0.222)
Intercept	16142.067**	(139.382)
N	7133	
R^2	0.28	
$F_{(1,7131)}$	2766.551	

Significance levels : † : 10% ∗ : 5% ∗∗ : 1%

Applying Panel Data Methods to Other Data Structures

Pooled OLS

```
graph twoway (scatter energy_consumption income_est) ///
(lfit energy_consumption income_est)
```



Applying Panel Data Methods to Other Data Structures

Fixed Effects

When we use a fixed effects model to estimate the effect of income on household energy consumption *within* local authorities, the size of the effect changes.

```
xtset LA_CODE MSOA_CODE
xtreg energy_consumption income_est, fe
outtex, file(fe.tex) labels level detail legend key(stab) replace
est store fe
```

Table : Estimation results : xtreg

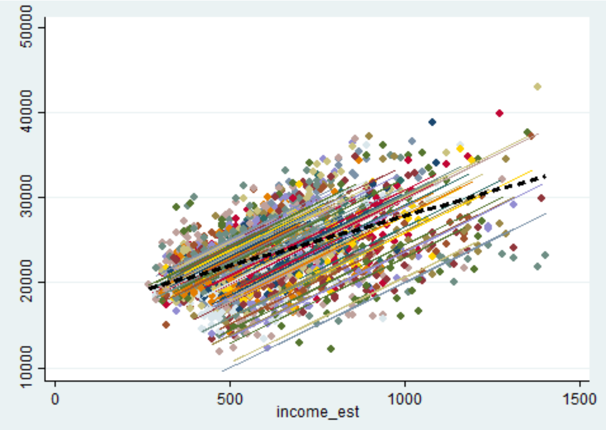
Variable	Coefficient	(Std. Err.)
income_est	20.111**	(0.237)
Intercept	11040.143**	(145.751)
N	7133	
R^2	0.516	
$F_{(376,6756)}$	7201.692	

Significance levels : † : 10% * : 5% ** : 1%

Applying Panel Data Methods to Other Data Structures

Fixed Effects

```
xi: regress energy_consumption income_est i.LA_CODE
predict energy_consumption_fitted
separate energy_consumption, by(LA_CODE)
separate energy_consumption_fitted, by(LA_CODE)
graph twoway (scatter energy_consumption1-energy_consumption80 income_est) ///
(line energy_consumption_fitted1-energy_consumption_fitted80 income_est) ///
(lfit energy_consumption income_est, ///
color(black) lwidth(thick) lpattern(dash)), legend(off) , legend(off)
```



Applying Panel Data Methods to Other Data Structures

Comparing OLS with Fixed Effects Models

```
esttab ols fe using table1.tex, replace
```

	(1) energy_consumption	(2) energy_consumption
income_est	11.68***	20.11***
	(52.60)	(84.86)
_cons	16142.1***	11040.1***
	(115.81)	(75.75)
N	7133	7133

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$