ECON 613: Applied Econometrics

Methods for Panel Data

March 26, 2019

Linear Models

Applications

Nonlinear Panel

Introduction (1)

- Data on cross section that is observed over several unit of time.
- In microeconometrics, panel are usually short.

Introduction (2)

- The error is correlated over time..
- Examples
- Open possibilities..

Introduction (3)

Consider the following Model

$$Y_{it} = \alpha_i + \gamma_{j(t)} + \beta X_{it} + \epsilon_{it} \tag{1}$$

- Estimation of fixed effects
- Correlation between the fixed effects
- Estimation issues

Introduction (4)

Consider the following DGP:

- 1,000 individuals over 10 periods.
- $Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$
- Parametrization
  - $\beta = 1$
  - $\alpha_i \sim \mathit{uniform}(0, 1)$
  - $\epsilon_i \sim \mathbb{N}(0, 1)$

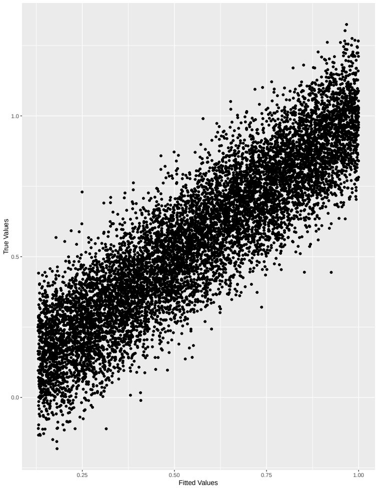
Pooled Estimation

	Model 1
(Intercept)	0.49***
	(0.02)
c(xMat)	0.93***
	(0.00)
R^2	0.87
Adj. R^2	0.87
Num. obs.	10000
RMSE	1.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Fitted Values (1)



Introduction (5)

Consider the following DGP:

- 1,000 individuals over 10 periods.
- $Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it}$
- Parametrization
  - $\beta = 1$
  - $\alpha_i \sim \mathit{uniform}(-10, 10)$
  - $\epsilon_i \sim \mathbb{N}(0, 1)$

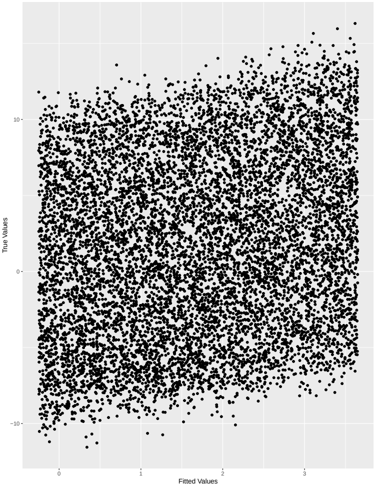
Pooled Estimation

	Model 1
(Intercept)	-0.26^{*}
	(0.11)
c(xMat)	0.40^{***}
	(0.02)
R^2	0.04
Adj. R^2	0.04
Num. obs.	10000
RMSE	5.73

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Fitted Values (2)



Effects

- Pooled Estimation is a good starting point.
- Individual VS Time Effect.

Individual Effects

- Fixed Effects
- Random Effects
- Examples: Return to Education

Time Effects

- Long Panel Case
- Example: Seasonality?

Some Models (1)

- Pooled Estimator

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it} \tag{2}$$

- Problems

Some Models (2)

- Between Estimator

$$\bar{y}_i = \alpha_i + \beta \bar{x}_i + \bar{\epsilon}_i \tag{3}$$

- Problems

Some Models (3)

- Within Estimator

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\epsilon_{it} - \bar{\epsilon}_i) \tag{4}$$

- Problems

Some Models (4)

- First Difference Estimator

$$y_{it} - y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + (\epsilon_{it} - \epsilon_{i,t-1}) \tag{5}$$

- Problems

More Guns, Less Crime

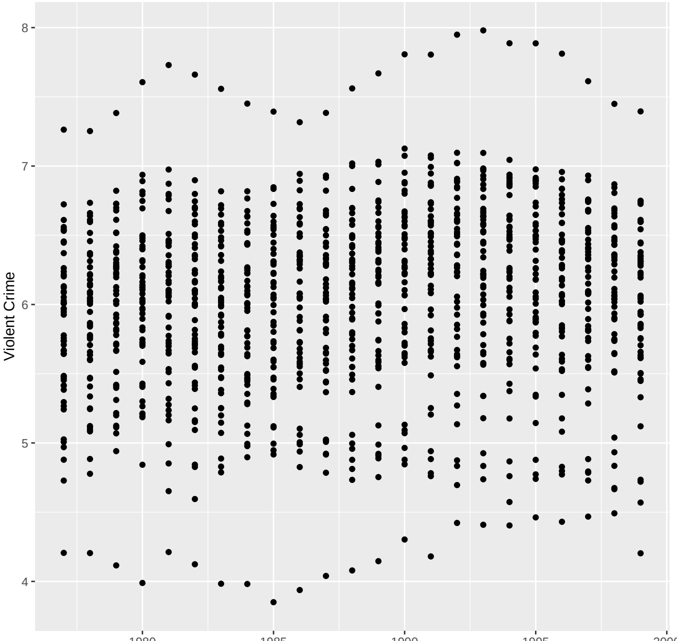
In a remarkable paper published in 1997, John Lott and David Mustard managed to set the agenda for much subsequent work on the impact of guns on crime in America by creating a massive data set of crime across all U.S. counties from 1977 through 1992 and amassing a powerful statistical argument that state laws enabling citizens to carry concealed handguns had reduced crime.1 The initial paper was followed a year later by an even more comprehensive and sustained argument to the same effect in a book solely authored by John Lott entitled More Guns, Less Crime (now in its second edition).

Data: Guns

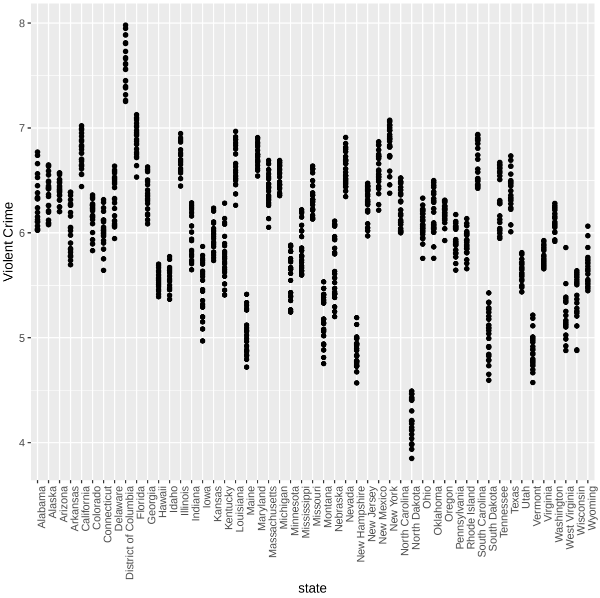
A data frame containing 1,173 observations on 13 variables.

- state: factor indicating state.
- year: factor indicating year.
- violent: violent crime rate (incidents per 100,000 members of the population).
- murder: murder rate (incidents per 100,000).
- robbery: robbery rate (incidents per 100,000).
- prisoners: incarceration rate in the state in the previous year
- afam: percent of state population that is African-American
- cauc: percent of state population that is Caucasian,
- male: percent of state population that is male
- population: state population, in millions of people.
- income: real per capita personal income in the state (US $).
- density population per square mile of land area, divided by 1,000.
- law factor. Does the state have a shall carry law in effect in that year?

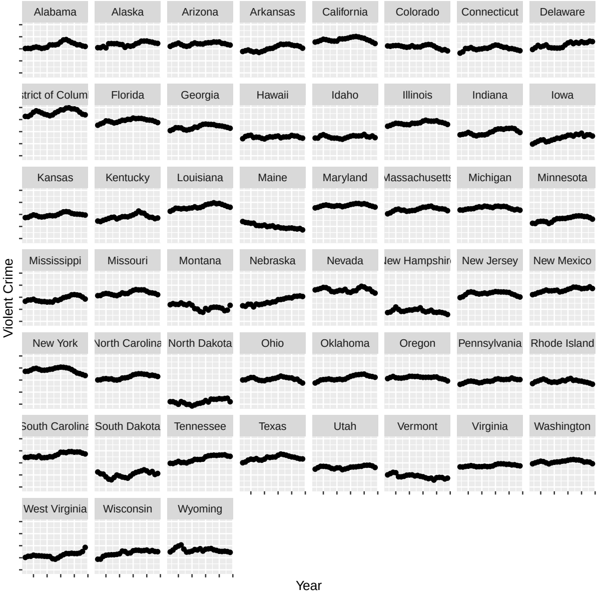
Overtime Variation



Cross-sectional Variation (1)



Cross-sectional Variation (2)



First regressions

	Violent Crime		Robbery	
	Model 1	Model 2	Model 1	Model 2
(Intercept)	6.13^{***}	2.98^{***}	4.87^{***}	0.90
	(0.02)	(0.54)	(0.03)	(0.77)
lawyes	-0.44^{***}	-0.37^{***}	-0.77^{***}	-0.53^{***}
	(0.04)	(0.03)	(0.06)	(0.05)
prisoners		0.00^{***}		0.00^{***}
		(0.00)		(0.00)
density		0.03^{*}		0.09^{***}
		(0.01)		(0.02)
income		0.00		0.00^{***}
		(0.00)		(0.00)
population		0.04^{***}		0.08^{***}
		(0.00)		(0.00)
afam		0.08^{***}		0.10^{***}
		(0.02)		(0.02)
cauc		0.03^{***}		0.03^{*}
		(0.01)		(0.01)
male		0.01		0.03
		(0.01)		(0.02)
R^2	0.09	0.56	0.12	0.60
Adj. R^2	0.09	0.56	0.12	0.59
Num. obs.	1173	1173	1173	1173
RMSE	0.62	0.43	0.90	0.61

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Exploiting the Panel Structure

	Model 1	Model 2	Model 3
(Intercept)	4.04^{***}	3.09^{***}	3.97^{***}
	(0.39)	(0.58)	(0.47)
lawyes	-0.05^{*}	-0.29^{***}	-0.03
	(0.02)	(0.03)	(0.02)
prisoners	-0.00	0.00^{***}	0.00
	(0.00)	(0.00)	(0.00)
density	-0.17^{*}	-0.01	-0.09
	(0.09)	(0.01)	(0.08)
income	-0.00	0.00	0.00
	(0.00)	(0.00)	(0.00)
population	0.01	0.04^{***}	-0.00
	(0.01)	(0.00)	(0.01)
afam	0.10^{***}	0.10^{***}	0.03
	(0.02)	(0.02)	(0.02)
cauc	0.04^{***}	0.04^{***}	0.01
	(0.01)	(0.01)	(0.01)
male	-0.05^{***}	-0.04^{*}	0.07^{***}
	(0.01)	(0.02)	(0.02)
State FE	YES	NO	YES
TIME FE	NO	YES	YES
R^2	0.94	0.59	0.96
Adj. R^2	0.94	0.58	0.95
Num. obs.	1173	1173	1173
RMSE	0.16	0.42	0.14

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Data: EmplUK

Employment and Wages in the United Kingdom

- An unbalanced panel of 140 observations from 1976 to 1984
- firm: firm index
- year: year
- sector: the sector of activity
- emp: employment
- wage: wages
- capital: capital
- output: output

Definitions

- Unbalanced panel: Definition
- What to do: Missing at random?

What to do?

- Testing for missingness at random.
- Missing at random
  - Imputation
  - Full sample
  - Non missing sample
- Not missing at random
  - Understand why?
  - Find an instrument

Description

Table:

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
firm	1,031	73.204	41.233	1	37	110	140
year	1,031	1,979.651	2.216	1,976	1,978	1,981	1,984
sector	1,031	5.123	2.678	1	3	8	9
emp	1,031	7.892	15.935	0.104	1.180	7.020	108.562
wage	1,031	23.919	5.648	8.017	20.636	27.494	45.232
capital	1,031	2.507	6.249	0.012	0.221	1.501	47.108
output	1,031	103.801	9.938	86.900	97.098	110.603	128.365

Linear VS Log Specifications

	Model 1	Model 2
(Intercept)	0.34	8.25^{**}
	(0.86)	(3.11)
log(wage)	-0.37^{***}	
	(0.06)	
log(capital)	0.81^{***}	
	(0.01)	
log(output)	0.48^{**}	
	(0.18)	
wage		-0.32^{***}
		(0.05)
capital		2.11^{***}
		(0.04)
output		0.02
		(0.03)
R^2	0.84	0.69
Adj. R^2	0.84	0.69
Num. obs.	1031	1031

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Fixed VS Random Effects

	Fixed Effects	Random Effect
(Intercept)	2.20^{**}	
	(0.15)	
log(wage)	-0.24^{***}	-0.61^{***}
	(0.05)	(0.03)
log(capital)	0.61^{***}	0.56^{***}
	(0.07)	(0.02)
R^2	0.78	0.99
Num. obs.	1031	1031

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models

Specification Problem

- Choosing between random and fixed effects;
- Durbin - Wu - Hausman Test

$$H = (\beta_{FE} - \beta_{RE})'(Var(\beta_{FE}) - Var(\beta_{RE}))'(\beta_{FE} - \beta_{RE}) \quad (6)$$

- $H \sim \chi_2(rank(Var(\beta_{FE}) - Var(\beta_{RE}))$

Data: US STATES PRODUCTION

- state: state
- year: year
- region: the region
- pcap: public capital stock
- hwy: highway and streets
- water: water and sewer facilities
- util: other public buildings and structures
- pc:private capital stock
- gsp: gross state product
- emp: labor input measured by the employment in nonagricultural payrolls
- unemp: state unemployment rate

Specifications

	Within	Between	First Difference
log(pcap)	-0.03	0.18^{*}	-0.01
	(0.03)	(0.07)	(0.05)
log(pc)	0.29^{***}	0.30^{***}	-0.03
	(0.03)	(0.04)	(0.02)
log(emp)	0.77^{***}	0.58^{***}	0.83^{***}
	(0.03)	(0.06)	(0.04)
unemp	-0.01^{***}	-0.00	-0.01^{***}
	(0.00)	(0.01)	(0.00)
(Intercept)		1.59^{***}	0.01^{***}
		(0.23)	(0.00)
R^2	0.94	0.99	0.69
Adj. R^2	0.94	0.99	0.69
Num. obs.	816	48	768

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table: Statistical models