ECON 613: Applied Econometrics

Methods for Panel Data

Introduction (1)

- Data on cross section that is observed over several unit of time.
- Two motivations for panel data in microeconometrics.
  - Exploiting the panel structure for controlling for unobserved time-invariant heterogeneity in cross-sectional models. Effect of ability on wages.
  - Disentangling components of variance and estimating transition probabilities among states and study the dynamics of cross-sectional populations. Variance of earnings over time.

Introduction (2): Ability bias

- Suppose a cross-sectional model of the form

$$y_{i1} = \beta x_{i1} + \eta_i + \nu_{i1} \tag{1}$$

  with $E(\nu_{i1}|x_{i1}, \eta_i) = 0$.
- If η_i is not observed, then we need an instrument, which is uncorrelated with both η_i and ν_{i1}, but correlated with x_{i1}.
- Suppose instead, we observe

$$y_{i2} = \beta x_{i2} + \eta_i + \nu_{i2} \tag{2}$$

  Then, β is identified using the regression in first-differences

$$y_{i2} - y_{i1} = \beta(x_{i2} - x_{i1}) + \nu_{i2} - \nu_{i1} \tag{3}$$

Introduction (3): Known limitation

Suppose a model of the form

$$y_{it} = \beta x_{it} + \eta_i + \nu_{i1} \tag{4}$$

- y_{it} is the income of individual i at period t.
- x_{it} is the education of individual i at period t.

Lack of overtime variation in x prevents to use the same strategy.

Introduction (4): Error Component's Model

Consider the following Model

$$y_{it} = \mu + \eta_i + \beta X_{it} + \epsilon_{it} \tag{5}$$

- $\epsilon_{it} \sim iid(0, \sigma_\epsilon^2)$.

Introduction (5): Fixed Effect Models

- Unobserved time-invariant individual effect.
- No distribution is required, and the fixed effect does not have to independent of observed characteristics.

Introduction (6): Random Effect Models

- The random effect $\eta_i \sim iid(0, \sigma_\eta^2)$.
- The cross-sectional variance of y_{it} is given $\sigma_\epsilon^2 + \sigma_\eta^2$.
- $\dfrac{\sigma_\eta^2}{\sigma_\epsilon^2 + \sigma_\eta^2}$ is the fraction of the total variance that remains constant over time.
- Earning dynamics, and income inequality.
- Heteroskedascity

Estimation (1)

Consider the following Model

$$Y_{it} = \alpha_i + \gamma_{j(t)} + \beta X_{it} + \epsilon_{it} \tag{6}$$

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

Estimation (2)

Consider the following DGP:

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

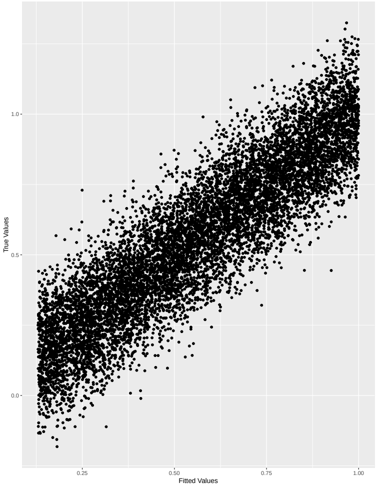
Pooled Estimation

	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
  - $\beta = 1$
  - $\alpha_i \sim \mathit{uniform}(-10, 10)$
  - $\epsilon_i \sim \mathbb{N}(0, 1)$

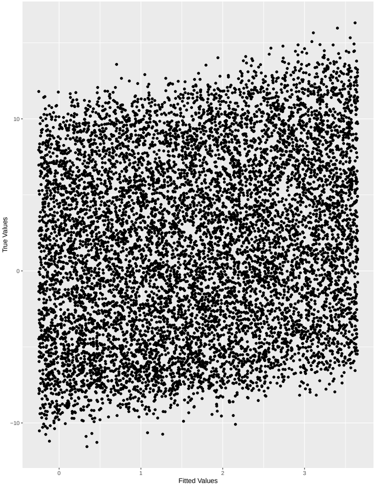
Pooled Estimation

	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

$$Y_{it} = \alpha + \beta X_{it} + \epsilon_{it} \tag{7}$$

- Problems

Some Models (2)

- Between Estimator

$$\bar{y}_i = \alpha_i + \beta \bar{x}_i + \bar{\epsilon}_i \tag{8}$$

- Problems: the cross-sectional information and completely discards the time variation in your data.

Some Models (3)

- Within Estimator

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\epsilon_{it} - \bar{\epsilon}_i) \tag{9}$$

- Problems: the cross-sectional information and completely discards the individual variation in the data.

Some Models (4)

- First Difference Estimator

$$y_{it} - y_{i,t-1} = \beta(x_{it} - x_{i,t-1}) + (\epsilon_{it} - \epsilon_{i,t-1}) \qquad (10)$$

- Problems

Estimation Techniques

- For fixed effect models, OLS is enough.
- For random effect models, heteroskedascity creates issues. Optimal GLS.

Method of Moments

- Orthogonality condition in Linear Models

$$E(x(y' - x)) = 0 \tag{11}$$

- Moment Condition

$$\frac{1}{N}\sum_i x_i(y_i - x'\beta) \tag{12}$$

- Moment Estimator

$$\hat{\beta}_{\text{MM}} = (\sum_i x_i x_i')^{-1}(\sum_i x_i y_i) \tag{13}$$

Nonlinear Model

- Consider
$$Y_i = g(X_i, b_0) + u_i$$
- Orthogonality Condition
$$E[X'(y - g(X, b_0))] = 0$$
- Moments condition
$$E_0 h(Y, X, a_0) = 0$$
- The function h is H-dimensional and the parameter a is of size K.

Formal Idea

Definition

The basic idea of generalized method of moments is to choose a value for a such that the sample mean is closest to zero.

$$\frac{1}{n}\sum_{i=1}^{n} h(Y_i, X_i, a)$$

Formal Definition

Definition

Let $\mathbb{S}_n$ be an $(H \times H)$ symmetric positive definite matrix that may depend on the observations. The generalized method of moments (GMM) estimator associated with $\mathbb{S}_n$ is a solution $\tilde{a}_n(\mathbb{S}_n)$ to the problem

$$min_a \left[\sum_{i=1}^{n} h(Y_i, X_i, a)\right]' \mathbb{S}_n \left[\sum_{i=1}^{n} h(Y_i, X_i, a)\right]$$

Assumptions

- H1 The variables (Y_i, X_i) are independent and identically distributed.
- H2 The expectation $E_0 h(Y, X, a)$ exists and is zero when a is equal to the true value a_0 of the parameter of interest.
- H3 The matrix $\mathbb{S}_n$ converges almost surely to a nonrandom matrix $\mathbb{S}_0$
- H4 The parameter a_0 is identified from the equality constraints, i.e. $E_0 h(Y, X, a)' \mathbb{S}_0 E_0 h(Y, X, a) = 0$
- H5 The parameter value a_0 is known to belong to a compact set $\mathcal{A}$
- H6 The quantity $(1/n) \sum_{i=1}^{n} h(Y_i, X_i, a)$ converges almost surely and uniformly in a to $E_0 h(Y, X, a)$
- H7 The function $h(Y, X, a)$ is continuous in a
- H8 The matrix $\left[E_0 \frac{h(Y, X, a)}{\partial a}\right]' \mathbb{S}_0 \left[E_0 \frac{h(Y, X, a)}{\partial a'}\right]$ is nonsingular, which implies $H \geq K$.

Asymptotic Normality

Under the assumptions, we have

$$\sqrt{n}(\tilde{a}_n(\mathbb{S}_n) - a_0) \sim \mathbb{N}(0, \Sigma(\mathbb{S}_0))$$

where

$$\Sigma(\mathbb{S}_0) = \left(\left[E_0 \frac{h(Y,X,a)}{\partial a} \right]' \mathbb{S}_0 \left[E_0 \frac{h(Y,X,a)}{\partial a'} \right] \right)^{-1}$$
$$\left(\left[E_0 \frac{h(Y,X,a)}{\partial a} \right]' \mathbb{S}_0 V_0(h(Y,X,a_0)) \mathbb{S}_0 \left[E_0 \frac{h(Y,X,a)}{\partial a'} \right] \right)^{-1}$$
$$\left(\left[E_0 \frac{h(Y,X,a)}{\partial a} \right]' \mathbb{S}_0 \left[E_0 \frac{h(Y,X,a)}{\partial a'} \right] \right)^{-1}$$

Optimal GMM

- $\mathbb{S}_0$ is not known.
- Two-step procedure
  - Estimate

$$min_a \left[\sum_{i=1}^{n} h(Y_i, X_i, a) \right]' I \left[\sum_{i=1}^{n} h(Y_i, X_i, a) \right]$$

    where I is the identity matrix, and recover $\hat{a}$.
  - Matrix of variance/covariance

$$\hat{\mathbb{S}} = \frac{1}{N} \sum_{i=1}^{n} h(Y_i, X_i, \hat{a}) h(Y_i, X_i, \hat{a})'$$

Likelihood Approaches

- Under standard conditions, we have

$$y_i|x_i, \eta_i \sim (X_i\beta + \eta_i, \sigma_\epsilon^2) \tag{14}$$

- The log conditional density $f(y_i|x_i, \eta_i)$ is easy to derive

$$-\frac{T}{2}\log\sigma^2 - \frac{1}{2^2}(y_i - X_i\beta - n_i)'(y_i - X_i\beta - n_i) \tag{15}$$

- The log likelihood is given

$$L(\beta, \sigma^2, \eta, y, x) = \sum_i^N log(f(y_i|x_i, \eta_i) \tag{16}$$

More Guns, Less Crime

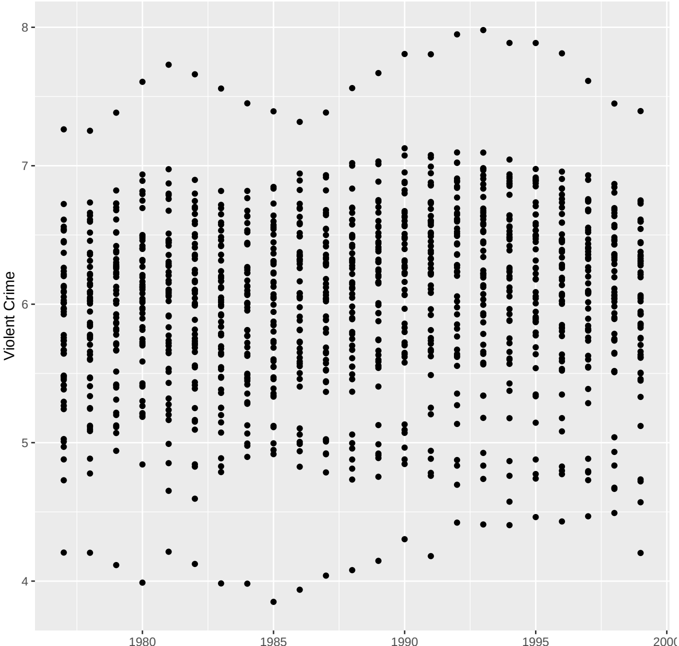
In a remarkable paper published in 1997, John Lott and David Mustard managed to set the agenda for much subsequent work on the impact of guns on crime in America by creating a massive data set of crime across all U.S. counties from 1977 through 1992 and amassing a powerful statistical argument that state 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

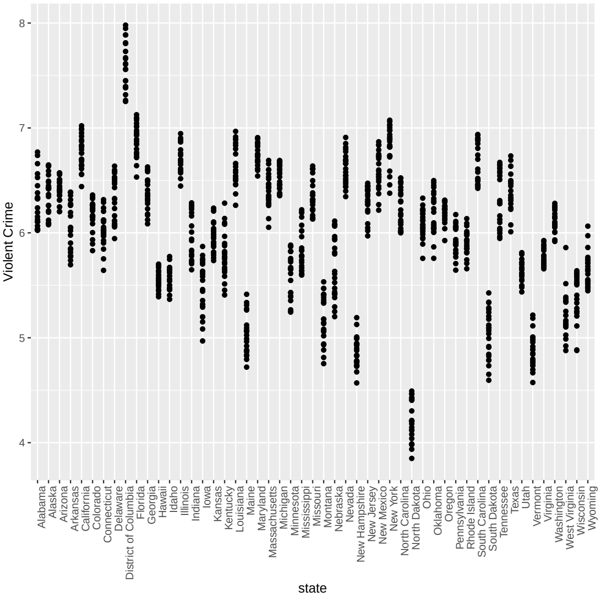
A data frame containing 1,173 observations on 13 variables.

- 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?

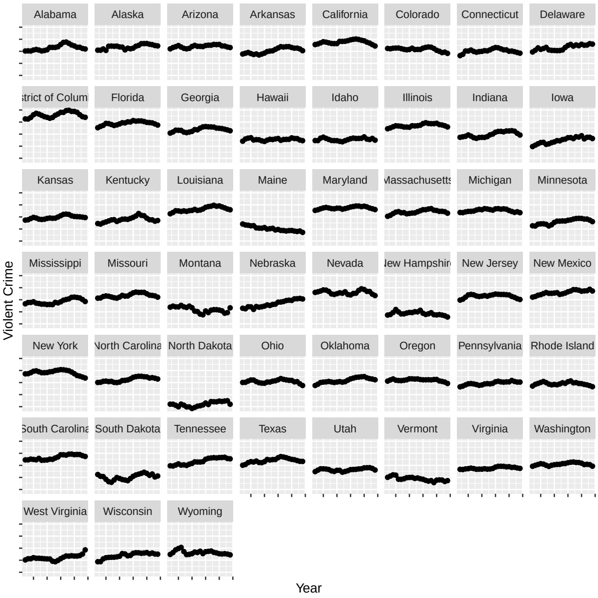
Overtime Variation



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

Unbalanced Panel: Definitions

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

Unbalanced Panel: Solutions

- 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

	Log	Linear
(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

	Random Effect	Fixed Effects
(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

$$H = (\beta_{FE} - \beta_{RE})'(Var(\beta_{FE}) - Var(\beta_{RE}))'(\beta_{FE} - \beta_{RE}) \quad (17)$$

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

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

Statement

- Identify the causal effect of a policy
  - Minimum Wages on Employment
  - Training on Wages
  - Class size on Student Outcomes
  - Welfare on Labor Supply
- Essentially a self selection problem

Evaluation Problem

- Let y_1 denote the outcome with treatment
- Let y_0 denote the outcome without treatment
- We are interested in the average treatment effect

$$ATE = E(y_1 - y_0) \tag{18}$$

- Evaluation problem: An individual can not be in both states, we can not observe both y_0 and y_1.
- Another quantity is the average treatment effect of the treated (Let w be an indicator of treatment).

$$ATT = E(y_1 - y_0 \mid w = 1) \tag{19}$$

Assumptions

Under which assumptions can you do a DiD?

- Stable Unit Treatment Value Assumption (SUTVA): potential outcomes for each person i are unrelated to the treatment status of other individuals
- Random Assignment. The treatment assignment is random i.e we have an independent, identically distributed sample from the population

Threat to Validity

- Pre-trend
- Placebo Effect
- Rubin Effect

DiD as a Linear Regression

Consider

$$Y_{it} = \alpha + \delta Post_t + \gamma D_i + \beta Post_t D_i + \epsilon_{it} \tag{20}$$

Where

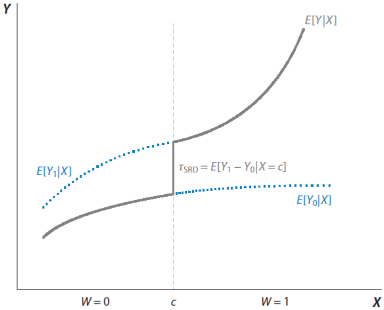
- $D_i = 1$ if treated, 0 otherwise
- $Post_t = 1$ after the implementation of the policy

Alternative Method: Regression discontinuity

The Key intuition for RDD is that we have an understanding of the mechanism which underlies the assignment of treatment. Specifically, assignment to treatment depends on a single variable. In the sharp, regression discontinuity design, the running variable fully determines the treatment

$$D_i = \begin{cases} 1, & \text{If } X_i > X_0 \\ 0, & \text{If } X_i < X_0 \end{cases} \tag{21}$$

Idea



Some examples (1)

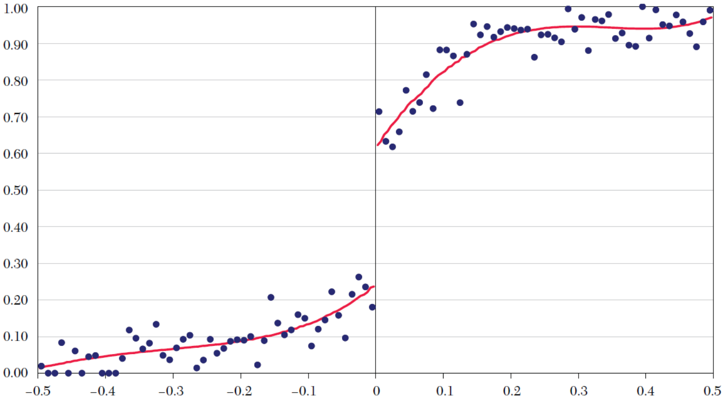
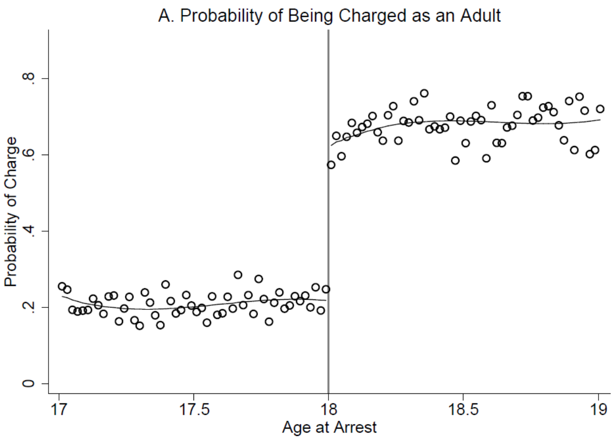
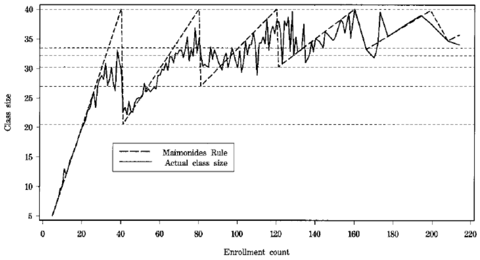


Figure 10. Winning the Next Election, Bandwidth of 0.01 (100 bins)

Some examples (2)



Some examples (3)



Explanation (1)

Twenty five children may be put in charge of one teacher. If the number in the class exceeds twenty five but is not more than forty, he should have an assistant to help with the instruction. If there are more than forty, two teachers must be appointed.

Explanation (2)

- A law prevents class size to exceed say 30
- If cohorts are of average size 90 but fluctuates
- If cohort size is 91-96, we end up four classrooms of size 22 to 24, while if cohort size is 85-90, we end up with three classrooms of size 28 to 30.

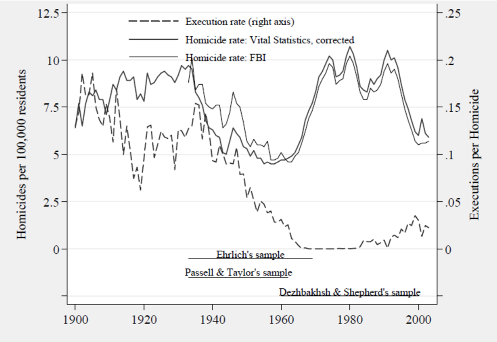
Angrist and Lavy (1999): Comparing test outcomes between students who are randomly assigned to the small vs large classes gives you a credible estimate of the effect of class size on academic performance. 10% decrease in class size increases test score by about 0.2 to 0.3 standard deviations.

Uses and Abuses of Empirical Evidence in the Death Penalty Debate

- Deterrance Effect of the death penalty
- What can we say?

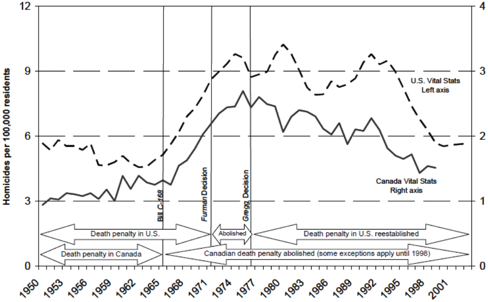
Evidence and Abuse (1)

Figure 1. Homicides and Execution in the United States



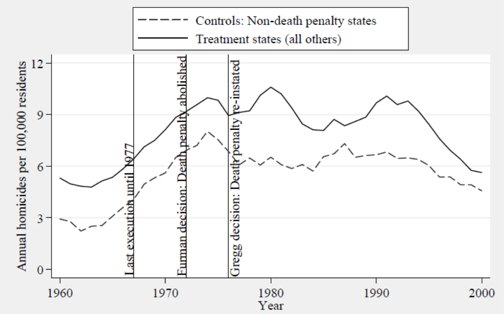
Evidence and Abuse (2)

Figure 2. Homicide Rates and the Death Penalty in the United States and Canada



Evidence and Abuse (3)

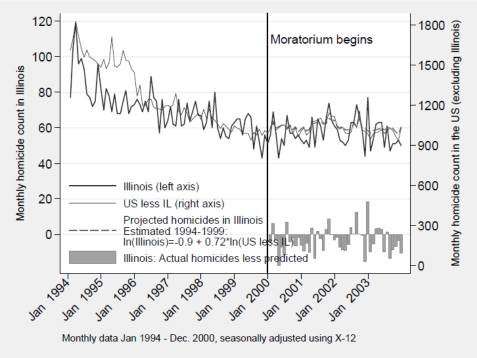
Figure 3. Homicide Rates in the United States



Non-death penalty states are those without a death penalty throughout 1960-2000: AK HI ME MI MN WI

Evidence and Abuse (4)

Figure 5. Homicides Before and After the Illinois Moratorium



Evidence and Abuse (5)

Table 6: Estimating the Impact of Executions on Murder Rates: Reanalyzing Mocan and Gittings: 1977-1997

	Dependent Variable: *Annual Homicides per 100,000 Residents$_{s,t}$*					*Log Homicide Rate$_{s,t}$*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Mocan and Gittings Results: Replication							
Executions$_{t-1}$ per Death Sentence$_{t-7}$	-0.60* (.35)			-0.63* (0.34)	-0.63** (.29)	-0.05* (.03)	-0.05* (.03)
Pardons$_{t-1}$ per Death Sentence$_{t-7}$		0.69** (.32)		0.73** (.30)		0.11*** (.03)	
Death Row Removals$_{t-1}$ per Death Sentence$_{t-6}$			0.17** (.07)		0.18** (.07)		0.02** (0.01)
Sample (1984-1997)	680	693	695	679	690	679	690
Panel B: Correcting Programming Errors							
Executions$_{t-1}$ per Death Sentence$_{t-7}$	-0.50 (.34)			-0.52 (.33)	-0.59 (0.39)	-0.01 (0.03)	-0.02 (0.02)
Pardons$_{t-1}$ per Death Sentence$_{t-7}$		0.63* (.34)		0.71** (.30)		0.09*** (0.03)	
Death Row Removals$_{t-1}$ per Death Sentence$_{t-6}$			0.24*** (.08)		0.17* (0.09)		0.01 (0.01)
Sample (1984-1997)	679	692	691	677	636	677	636
Panel C: Measuring Deterrence Variables with a One-Year Lag on Full Sample							
Executions$_{t-1}$ per Death Sentence$_{t-1}$	0.03 (0.14)			0.01 (0.13)	0.01 (0.14)	0.01 (0.01)	0.01 (0.01)
Pardons$_{t-1}$ per Death Sentence$_{t-1}$		0.41*** (.13)		0.41*** (0.13)		0.05*** (0.01)	
Death Row Removals$_{t-1}$ per Death Sentence$_{t-1}$			0.02 (0.03)		0.02 (0.03)		0.002 (.002)
Sample (1978-1997)	986	984	921	977	918	977	918
Implied Life-Life Tradeoff for Executions[(a)] [95% Confidence Interval]							
Panel A: Replication	4.4 [-1.8, 10.5]			4.6 [-1.4, 10.6]	4.6 [-0.5, 9.7]	2.2 [-1.2, 5.7]	2.3 [-1.3, 6.0]
Panel B: Corrected	3.4 [-2.6, 9.4]			3.6 [-2.2, 9.5]	4.2 [-2.6,11.1]	-0.2 [-3.7, 3.4]	0.5 [-2.7, 3.7]
Panel C: Full Sample	-1.2 [-3.1, 0.7]			-1.1 [-2.8, 0.7]	-1.1 [-3.0, 0.8]	-1.6 [-2.7, -0.5]	-1.6 [-2.8, -0.4]

Summary

- Immigration
- Intergenerational Mobility
- Minimum Wages

Issues

- Causal Inference is hard!
- Problems
  - Data selection
  - Specification

Motivation

- Determinant of an outcome variable which varies over time.
- Example
  - Fertility
  - Retirement
  - Binary decisions with life-cycle component
  - ...

Probit and Logit

- Consider an individual specific effect, such that

$$Pr(y_{it} = 1) = F(x_{it}\beta + \alpha_i) \tag{22}$$

- Likelihood can be written naturally

Fixed effect estimation

- Probit case - kind of complicated - incidental parameter problem.
- Logit case - conditional MLE
- No quasi-differencing estimator

Random effect

- Probit - complicated numerically - requires numerical integration
- Logit - relatively simple

Transition Data

- Panel Data
- Outcome variable is a duration: length of time until an event occurs (or a spell ends)
- Examples
  - Unemployment
  - Strike
  - Time to (buy a house, marry, divorce....)

Basic concepts

- Density: $f(t)$
- Distribution: $F(t)$
- Survival Function: $S(t) = 1 - F(t)$
- Hazard rate: $\lambda(t) = \lim_{h \to 0} \frac{Pr[t < T < t + h \mid T \geq t]}{h} = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)}$

Estimation methods

- Nonparametric: Kaplan Meier
- Parametric: Exponential, Weibull,
- Semi-parametric: Cox Proportional Hazard

Likelihood based Estimation

- Censoring

$$d_i = \begin{cases} 1, & \text{no censoring} \\ 0, & \text{right censoring} \end{cases} \tag{23}$$

- Likelihood

$$\log \mathcal{L}(\theta) = d_i \log f(t_i \mid X, \theta) + (1 - d_i) \log S(t_i \mid X, \theta) \tag{24}$$