Difference in Difference – Part 2

Bill Evans

Making the model more complicated

- So far, a very simple model
  - Two groups
  - Two periods
- However, the "treatment" may cover more than 1 group
- The treatment may happen at very different time periods across groups
- How to generalize this type of model for
  - Many treatments
  - Multiple groups being treated

Example: States as laboratories

- Tremendous variation across states in their laws
  - Variation across states in any given year
  - Variation over time within a state
- Examples
  - Minimum wages, welfare policy, Medicaid coverage, traffic safety laws, use of death penalty, drinking age, cigarette taxes,

Empirical example: Motorcycle Helmet laws

- 1967, Feds require states to have helmet law to get all federal highway money
- By 1975, all states have qualifying law
- 1976, Congress responds to state pressure and eliminate penalties
  - 20 states weaken their law and only require coverage for teens
  - 8 states repeal law completely

- 1991 Federal law again provides incentives for laws covering everyone
  - A bunch of states pass universal laws
- Congress changes its mind and in 1995 eliminate penalties
  - Again many states drop the law
- Currently
  - 20 states have universal law
  - 27 have teen coverage only

5

- Helmets are estimated to reduce the likelihood of death in a motorcycle crash by 37%. (Center for Disease Control)
- http://www.cdc.gov/motorvehiclesafety/pdf/mc2012/MotorcycleSafetyBook.pdf

- Where does this number come from?

6

FARS

- Fatal Accident Reporting System
- Census of motor vehicle accidents that produce a fatality
- Produced since 1975
- Detail information about
  - Accident
  - Vehicles
  - Drivers

7

- Select sample from FARS, 1988-2005
- Unique sample
  - Two riders on motor cycle
  - At least one died (accident was severe enough to produce a death)
  - Where one of the riders used a helmet, the other did not

8

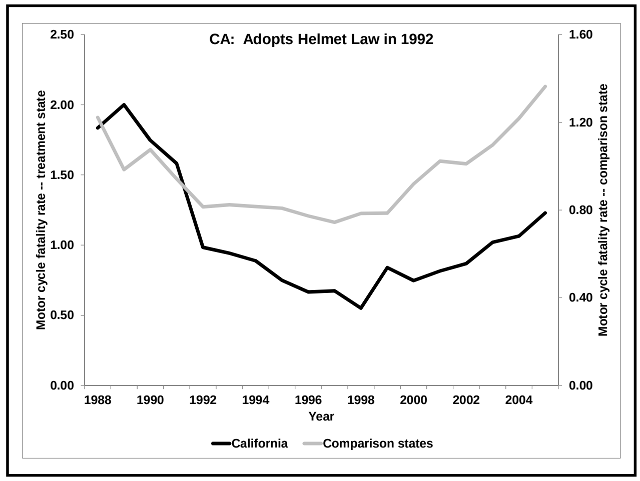
```
+-------------------+
| Key               |
|-------------------|
|     frequency     |
|  row percentage   |
| column percentage |
+-------------------+

           |         fatal
    helmet |         0          1 |     Total
-----------+----------------------+----------
         0 |       240        511 |       751
           |     31.96      68.04 |    100.00
           |     36.31      61.64 |     50.40
-----------+----------------------+----------
         1 |       421        318 |       739
           |     56.97      43.03 |    100.00
           |     63.69      38.36 |     49.60
-----------+----------------------+----------
     Total |       661        829 |     1,490
           |     44.36      55.64 |    100.00
           |    100.00     100.00 |    100.00
```

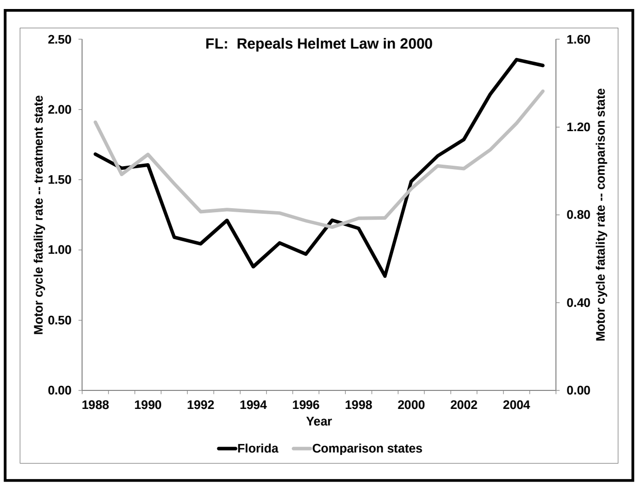
Pr(Die | no helmet) = 0.68

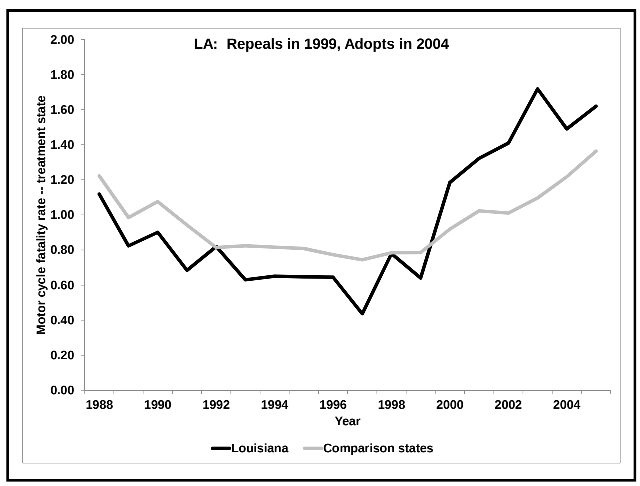
Pr(Die | helmet) = 0.43

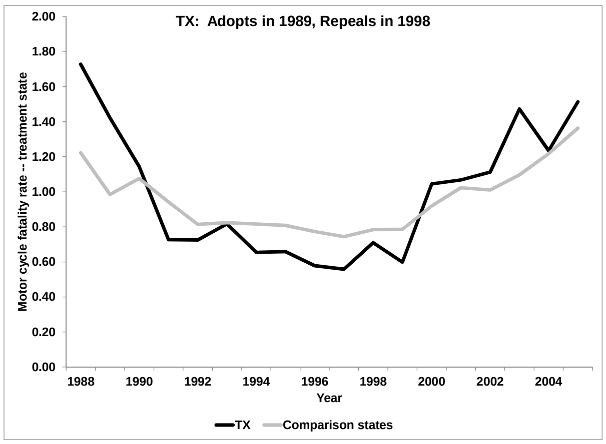
Benefits of a helmet, reduce prob. of death by 37%

(0.43-0.68)/0.68 = -0.37



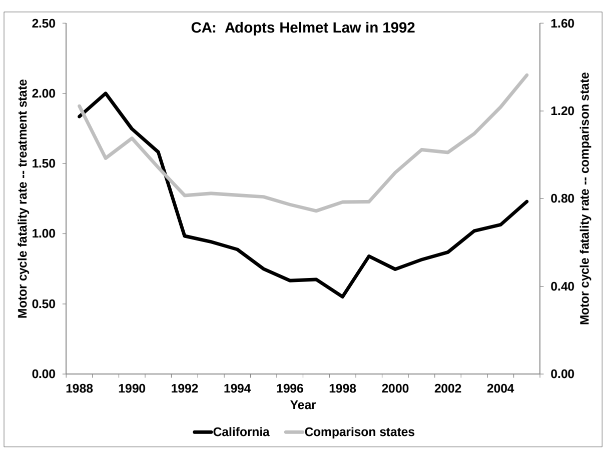






State group	Motor cycles registered per 100,000	Motor cycle death rate per 100,000
Never had helmet law	2366	1.18
Changed helmet law	1525	1.04
Always had helmet law	1224	0.77

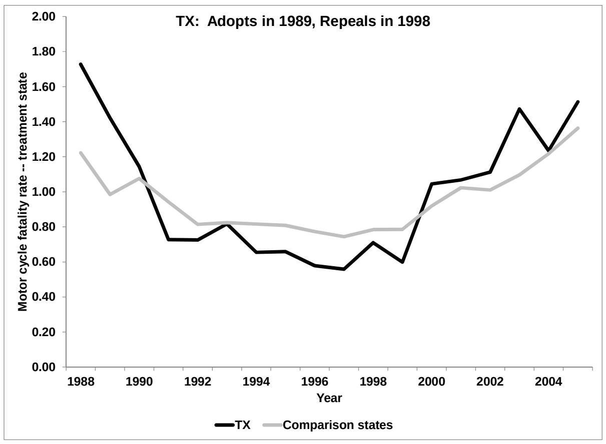
14



Results for CA alone

```
. reg mcdrl speed65 unemp bac_10 trend helmet_law if state=="CA"

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  5,    12) =   22.84
       Model |  2.14855814     5  .429711628           Prob > F      =  0.0000
    Residual |  .225732886    12  .018811074           R-squared     =  0.9049
-------------+------------------------------           Adj R-squared =  0.8653
       Total |  2.37429102    17  .139664178           Root MSE      =  .13715

------------------------------------------------------------------------------
       mcdrl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     speed65 |   .4358863   .1725389     2.53   0.027     .0599564    .8118163
       unemp |   .0340391   .0485866     0.70   0.497    -.0718221    .1399003
      bac_10 |   .3175587    .157151     2.02   0.066    -.0248439    .6599612
       trend |    .065022   .0149018     4.36   0.001     .0325538    .0974902
  helmet_law |  -.8972242    .168596    -5.32   0.000    -1.264563   -.5298851
       _cons |  -.3775448   .2939152    -1.28   0.223    -1.017931    .2628414
------------------------------------------------------------------------------
```

16



Results for TX alone

```
* run a model for Texas
. reg mcdrl speed65 unemp bac_08 trend helmet_law if state=="TX"

      Source |       SS       df       MS              Number of obs =      18
-------------+------------------------------           F(  5,    12) =    9.43
       Model |  1.86115335     5   .37223067           Prob > F      =  0.0008
    Residual |  .473677129    12  .039473094           R-squared     =  0.7971
-------------+------------------------------           Adj R-squared =  0.7126
       Total |  2.33483048    17  .137342969           Root MSE      =  .19868

------------------------------------------------------------------------------
       mcdrl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     speed65 |   .1689669   .2624034     0.64   0.532    -.4027609    .7406947
       unemp |   .1301635   .0747096     1.74   0.107    -.0326147    .2929418
      bac_08 |   .6692796   .2457196     2.72   0.018     .1339026    1.204657
       trend |  -.0405022   .0284775    -1.42   0.180    -.1025494     .021545
  helmet_law |  -.4592142   .1645968    -2.79   0.016    -.8178398   -.1005887
       _cons |  -.5765997    .464616    -1.24   0.238    -1.588911    .4357116
------------------------------------------------------------------------
```

18

Purely Cross Sectional Model (1990)

```
. * run basic OLS model for 1990
. reg mcdrl speed65 unemp bac_08 bac_10 helmet_law if year==1990

      Source |       SS       df       MS              Number of obs =      48
-------------+------------------------------           F(  5,    42) =    4.48
       Model |  2.59400681     5  .518801363           Prob > F      =  0.0023
    Residual |  4.86098775    42  .115737804           R-squared     =  0.3480
-------------+------------------------------           Adj R-squared =  0.2703
       Total |  7.45499457    47  .158616906           Root MSE      =   .3402

------------------------------------------------------------------------------
       mcdrl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     speed65 |   .1692481   .1357937     1.25   0.220    -.1047947    .4432909
       unemp |   .0524134   .0490639     1.07   0.292    -.0466015    .1514283
      bac_08 |  -.0944842   .2372792    -0.40   0.693     -.573333    .3843645
      bac_10 |    .01337 94.1658892      0.08   0.936    -.3213985    .3481573
  helmet_law |  -.4684841   .1038582    -4.51   0.000    -.6780784   -.2588898
       _cons |  -.0643492   .3042114    -0.21   0.833    -.6782726    .5495743
------------------------------------------------------------------------------
```

19

define:

$$i = 1, 2, \ldots n; \quad t = 1, 2, \ldots T$$

$$S_i = 1\,if \;\; state\; i, = 0\; otherwise$$

$$W_t = 1\,if \;\; year\; t, = 0\; otherwise$$

$$Law_{it} = 1\,if \;\; state\; i\; has\; helmet\; law$$

$$in\; year\; t, = 0\; otherwise$$

$$y_{it} = \beta_0 + REFORM_{it}\beta_1 + x_i\beta_2 + \sum_{j=2}^{n} S_j\alpha_j + \sum_{k=2}^{T} W_k\lambda_k + \varepsilon_{it}$$

20

- Why k=2 to N and j=2 to T?
- What does α measure?
- What does λ measure?

21

- Question: impact of MC helmet laws on motorcycle fatalities
- Data: 48 states, 18 years (1988-2005), 864 observations
- Outcome ln(motor cycle death rate)
  - Death rates = deaths/100,000 population
- Treatment variable: =1 if state i has a motor cycle law in year t, =0 otherwise

22

```
Contains data from motorcycles.dta
  obs:           864
 vars:            12                          10 Nov 2012 09:27
 size:        49,248 (99.6% of memory free)
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
year            int    %9.0g                  year
mcfatals        double %9.0g                  total motor cycle fatalities
state           str2   %2s                    2 digit postal code, AL, CA,
                                                etc.
fips            byte   %8.0g                  2 digit numeric fips code
helmet_law      float  %9.0g                  =1 if motorcycle helmet law, =0
                                                otherwise
speed65         float  %9.0g                  =1 if speed limit is 65, 0
                                                otherwise
speed70p        float  %9.0g                  =1 if speed limit is 70 plus, 0
                                                otherwise
bac_10          float  %9.0g                  drunk driving defined as
                                                bac>=0.1, =0 otherwise
bac_08          float  %9.0g                  drunk driving defined as
                                                bac>=0.08, =0 otherwise
unemp           float  %9.0g                  state unemployment rate, 5 is 5%
population      float  %9.0g                  state population
mregs           double %10.0g                 motor cycle registrations
-------------------------------------------------------------------------------
Sorted by:
```

23

```
. * construct dummy variables for state and year
. xi i.state i.year
i.state           _Istate_1-48        (_Istate_1 for state==AL omitted)
i.year            _Iyear_1988-2005    (naturally coded; _Iyear_1988 omitted)
```

24

```
. * run the difference  in difference model
. reg mcdrl speed65 speed70p unemp bac_08 bac_10 _I* helmet_law

      Source |       SS       df       MS              Number of obs =     864
-------------+------------------------------           F( 70,   793) =   30.54
       Model |  139.812929    70  1.99732756           Prob > F      =  0.0000
    Residual |  51.8558902   793  .065392043           R-squared     =  0.7295
-------------+------------------------------           Adj R-squared =  0.7050
       Total |   191.66882   863  .222095967           Root MSE      =  .25572

------------------------------------------------------------------------------
       mcdrl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     speed65 |  -.0577686   .0552537    -1.05   0.296    -.1662293    .0506922
    speed70p |  -.0855586   .0815308    -1.05   0.294    -.2456004    .0744831
       unemp |  -.0117339   .0118625    -0.99   0.323    -.0350195    .0115517
      bac_08 |   .1423512   .0725064     1.96   0.050     .0000241    .2846783
      bac_10 |   .1163134   .0628129     1.85   0.064    -.0069859    .2396127
   _Istate_2 |   -.038139   .0889074    -0.43   0.668    -.2126606    .1363826
                delete some results
  _Istate_47 |   .2392712   .0896769     2.67   0.008     .0632391    .4153033
  _Istate_48 |   .3987819     .09788     4.07   0.000     .2066474    .5909164
 _Iyear_1989 |  -.2367341    .052373    -4.52   0.000    -.3395401   -.1339281
                delete some results
 _Iyear_2005 |   .1032509   .0703676     1.47   0.143    -.0348778    .2413796
  helmet_law |  -.3728078   .0458932    -8.12   0.000    -.4628943   -.2827213
       _cons |   .5393718   .1275965     4.23   0.000     .2889049    .7898387
------------------------------------------------------------------------------
```

25

Lojack

26

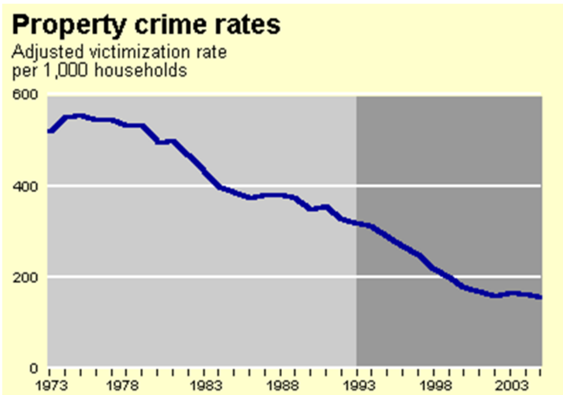
Background

- Transponder installed in cars that is turned on when car is stolen
- Recover 95% of stolen cars, compared to 60% for cars without Lojack
- One-time cost at installation
- Requires working in unison with local police authorities, so market entrance is city-by-city

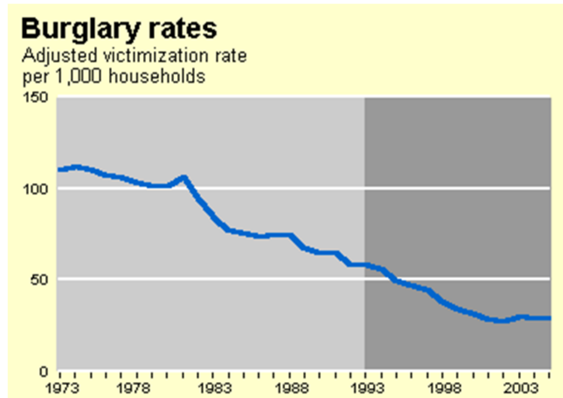
27

- Starts in MA in 1986 and spreads to 12 cities by 1994
- Model: examine changes in crime before/after Lojack is introduced to cities without Lojack
- Time trends are key in this analysis

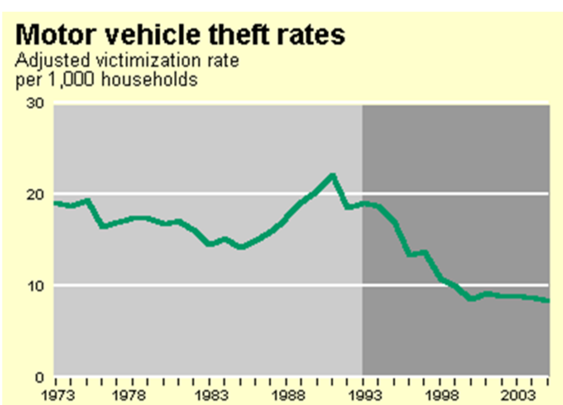
28



29



30



31

TABLE I
MARKETS SERVED BY LOJACK AS OF DECEMBER 1994

Market	Cities > 250,000 covered	Date of entry
Massachusetts	Boston	July 1986
South Florida	Miami	December 1988
New Jersey	Newark	March 1990
Los Angeles County	Los Angeles Long Beach	July 1990
Illinois	Chicago	November 1990
Georgia	Atlanta	August 1992
Virginia	Norfolk Virginia Beach	August 1993
Michigan[a]	Detroit	February 1994
New York	New York City	June 1994
Rhode Island	None	June 1994
Tampa/St. Petersburg	Tampa	July 1994
District of Columbia	Washington, DC	September 1994

32

Dynamics

- Lojack installed in new cars, so market penetration is a function of
  - New car sales
  - Fraction of new cars w/ Lojack
- After 5 yrs, only 2% of all cars have Lojack once it enters an area

33

Potential benefits

- Does not reduce your chance of having your car stolen, but
- Reduces your costs, given that your car is stolen
- Given previous point, will reduce your insurance costs

34

- Chance any car will have Lojack is low.
- If high volume chop shop, will encounter Lojack
- 50 cars annually, 3% market penetration, 78% chance get at least one car with Lojack
- With 100 cars, this rises to 95%

35

- Prob(at least one Lojack car) = 1 – Prob(no Lojack cars)
- Prob car does not have Lojack = 0.97
- All probs are independent
- Prob (non have Lojack) = 0.97^{50} = 0.22

36

Externality

- What is externality?
- How does Lojack generate externalities?
- What does this imply about whether Lojack penetration is too high or low?

37

Data

- 57 cities with pop > 250,000
  - Why only larger cities?
- 1981-1994
- Collect data on local economic conditions, police, age distribution

38

Crime Rates in the US, 2005

	All areas	Metro Areas	Non-Metro
Violent	469	510	374
Property	3,430	3,599	3,998
Auto theft	417	467	195

39

TABLE II
SUMMARY STATISTICS

Variable	Mean	Standard deviation	Minimum	Maximum
All cities in sample:				
Lojack share (% of all vehicles)	.05	.33	0	4.95
Years of Lojack	.17	.85	0	9
City population	764,268	1,045,791	250,720	7,375,097
Auto theft per capita	.012	.008	.002	.054
Robbery, burglary, larceny per capita	.078	.021	.033	.156
Assault, rape, murder per capita	.008	.004	.001	.025
SMSA unemp.	6.3	2.1	2.2	15.9
State per capita real income ($1994)	19,911	2,821	13,720	31,228
% Black	26.0	18.7	1.2	80.7
% Aged 0–17	26.3	2.0	19.7	31.7
% Aged 18–24	11.5	1.3	8.4	15.1
% Aged 25–44	31.4	2.1	26.1	36.4
Sworn officers per capita (×1000)	2.47	.96	1.32	7.81

40

TABLE II
SUMMARY STATISTICS

Variable	Mean	Standard deviation	Minimum	Maximum
Cities with Lojack coverage by 12/94				
Lojack share (% of all vehicles)	.21	.67	0	4.95
Years of Lojack	.83	1.71	0	9
City population	1,402,239	1,959,315	257,617	7,375,097
Auto theft per capita	.018	.011	.002	.05
Robbery, burglary, larceny per capita	.0881	.025	.044	.156
Asault, rape, murder per capita	.011	.006	.001t	
SMSA unemp.	6.5	2.1	2.7	15.9
State per capita real income ($1994)	20,843	3,370	13,932	31,228
% Black	37.5	21.0	10.4	80.7
% Aged 0–17	24.9	2.2	19.7	31.7
% Aged 18–24	11.5	1.5	8.4	15.1
% Aged 25–44	32.0	2.3	26.1	36.4
Sworn officers per capita (×1000)	3.20	1.33	1.40	7.81

41

Mean Values

	All cities	W/ Lojack
Population	764,268	1,402,239
Car theft/pop	0.012	0.018
Unemp rate	6.3	6.5
Per capita inc	$19,911	$20,843
% black	26.0%	37.5%
%18-24	11.5	11.5

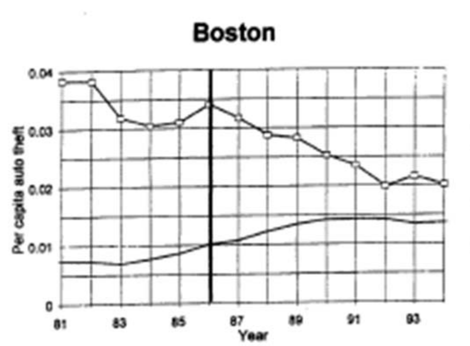
42

The form of the equations estimated in the basic specifications is as follows:

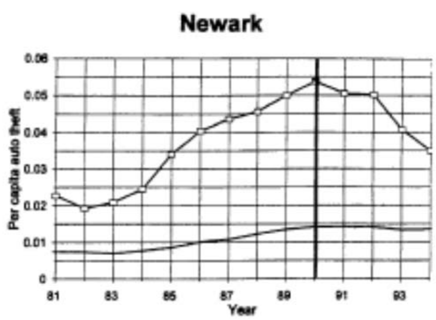
(1) $\ln(AUTO\_THEFT)_{it} = \beta LOJACK_{it} + X'_{it}\Gamma + \lambda_t + \theta_i + \epsilon_{it},$

where *I* indexes cities and *t* corresponds to years. *AUTO_THEFT* is the auto theft rate per capita, *LOJACK* is one of the two Lojack proxies described earlier, and *X* is a vector of controls for SMSA

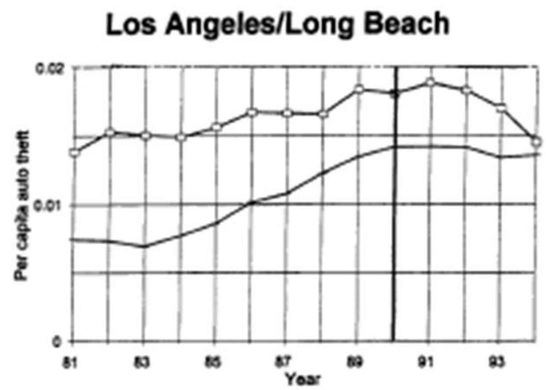
43



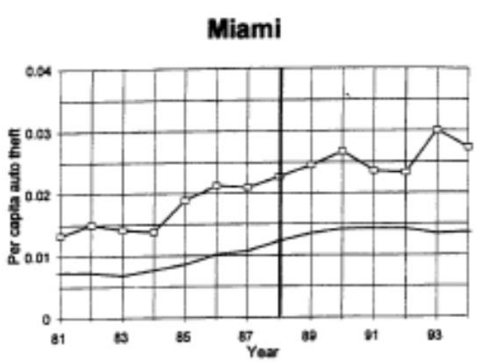
44


Newark
Per capita auto theft
0.06
0.05
0.04
0.03
0.02
0.01
0
81
83
85
87
89
91
93
Year

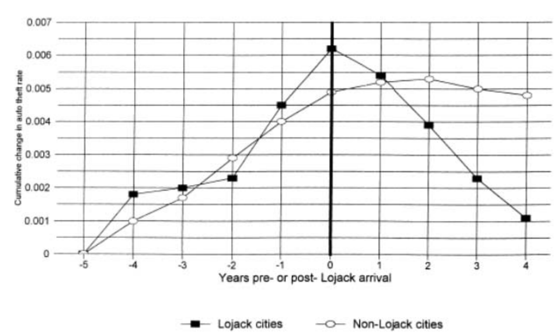
45


Los Angeles/Long Beach
Per capita auto theft
0.02
0.01
0
81
83
85
87
89
91
93
Year

46


Miami
Per capita auto theft
0.04
0.03
0.02
0.01
0
81
83
85
87
89
91
93
Year

47


Cumulative change in auto theft rate
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
-5
-4
-3
-2
-1
0
1
2
3
4
Years pre- or post- Lojack arrival
Lojack cities
Non-Lojack cities

48

TABLE III
IMPACT OF LOJACK ON CITY AUTO THEFT RATES

Variable	(1)	(2)	(3)	(4)
Years of Lojack availability	−.109	−.157	—	—
	(.013)	(.021)		
Lojack share	—	—	−.242	−.463
			(.031)	(.065)
Unemployment rate	.019	.026	.017	.028
	(.009)	(.010)	(.009)	(.010)
State real per capita income (×1000)	.022	.028	.016	.022
	(.014)	(.015)	(.014)	(.016)

Linden and Rockoff

- Megan Kanka
  - 7 year old girl
  - Raped and murdered by neighbor who was convicted sex offender
- No one in the neighborhood knew about neighbor's history
- Lead to passage of "Megan's Law"

Megan's Law

- Sexual Offender (Jacob Wetterling) Act of 1994
  - Sexual offenders required to notify state of change of address
  - Time limits vary across states (10 years after conviction or life)
  - Required of all child sex offenders, some states require of all offenders

Megan's Law

- 1996 Amendment to original law required states to publicly announce location and type of offense of sex offenders

- Indiana site
- http://www.icrimewatch.net/indiana.php

Economic question

- Crime negatively impacts property values
- Problem: crime is not random and neither are home purchases
- Therefore, getting an estimate of the impact of crime on housing prices is tough
- Megan's law
  - Sex offenders will most likely live in poorer areas
  - How to separate thus fact from their impact on house prices?

53

Methodology

- Compare house sales in neighborhoods before and after arrival of sex offender
- Impact should be "local" so comparison sample included homes in the same neighborhood but not near the offender

54

Data: NC Megan's Law Registry

- Between 1/1/1996 – 3/9/2003
- A total of 8287 released offenders required to register
- 1007 left the state
- Of the remaining, 103 (1.4 percent) failed to register

55

Data

- Location of sex offender's address
- Timing of when they moved in
- Matched to home sales data – Charlotte/Mecklenburg county
  - 1994-2004
- Detailed characteristics of home sales
  - 170,000 homes
  - 9,000 within 1/3 miles of a sex offender

56

Consider a simple OLS model

- Cross section of homes (i=1,2,..n)
- p_i =sales price of home i
- $x_{1i}, x_{21}, \ldots x_{ki}$ characteristics of home
  - (rooms, sq feet, brick exterior, Jacuzzi)
- s_i =1 if sex offender lives nearby, =0 otherwise

$$\ln(p_i) = \beta_0 + x_{1i}\beta_1 + \ldots.. x_{ki}\beta_k + s_i\alpha + \varepsilon_i$$

TABLE 1—CHARACTERISTICS OF HOMES SOLD IN MECKLENBURG COUNTY, 1994–2004

	All parcels	Within 1/3 mile of offender	
	Mean (*Standard deviation*)	Mean (*Standard deviation*)	Marginal effect in price regression[1]
Sale price ($100,000)	2.048 (1.324)	1.438 (0.848)	
Square footage (1,000 square feet)	2.075 (0.880)	1.620 (0.595)	0.266 (0.012)*
Quality rating (1 to 6)	3.251 (1.208)	3.066 (0.979)	0.047 (0.006)*
Age (in years)	10.347 (12.090)	16.322 (12.815)	−0.008 (0.001)*
Bedrooms	3.327 (0.648)	3.061 (0.566)	0.028 (0.010)*
Bathrooms	2.018 (0.592)	1.737 (0.539)	0.028 (0.008)*

58

Data set design

- Identify home that eventually get a sex offender resident
- Two types of homes
  - Treated: homes within 1/10th of a mile
  - Control: homes within 0.1 – 0.3 miles
- Two periods – before and after offender moves in
- Why just 0.1 miles?
- Why not all homes – why just 0.1 – 0.3 miles?

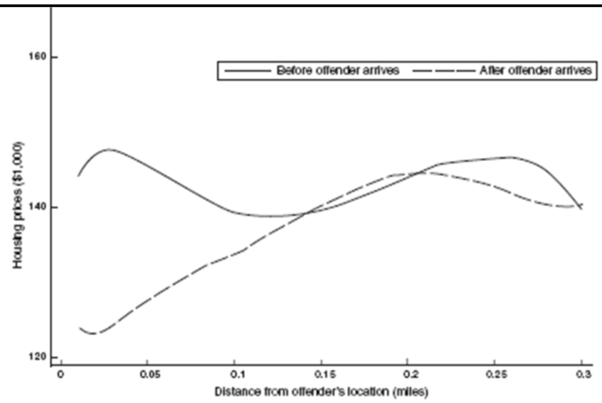


FIGURE 2B. PRICE GRADIENT OF DISTANCE FROM OFFENDER
(Sales during year before and after arrival)

Note: Results from local polynomial regressions (bandwidth = 0.075 miles) of sale price on distance from offender's future/current location.

The Difference-in-Difference model

- Data varies cross homes (i) neighborhoods (j) and time (t)
- $D^{0.1}$ = 1 if home within 0.1 of sex offender
- $D^{0.3}$ = 1 if home within 0.1-0.3 of sex offender
- Post = 1 if after SO arrives in a neighborhood, =0 otherwise
- α_{ji} = 1 if home is from neighborhood j in year t, =0 otherwise

$$\ln(p_{ijy}) = \alpha_{jt} + x_{1ijt}\beta_1 + ...x_{kijt}\beta_k + D_{ijt}^{0.1}\gamma + D_{ijt}^{0.3} * POST_{ijt}\theta + D_{ijt}^{0.1} * POST_{ijt}\pi + \varepsilon_i$$

What does α_{jt} *capture?*

What does $D_{ijt}^{0.1}\gamma$ *capture?*

What does $D_{ijt}^{0.1} * POST_{ijt}\pi$ *capture?*

What does $D_{ijt}^{0.3} * POST_{ijt}\theta$ *capture?*

TABLE 2—PRE- AND POST-ARRIVAL DIFFERENCES IN AVERAGE CHARACTERISTICS OF HOMES SOLD CLOSE TO OFFENDERS' LOCATIONS

Panel A: Pre-arrival differences in sales	Log price	Built in year sold	Age in years	Square feet (1,000s)	Number of bedrooms	Number of bathrooms
Within 0.1 miles of offender	0.007	0.062	−1.081	0.059	0.022	< 0.001
	(0.034)	(0.035)+	(1.117)	(0.047)	(0.034)	(0.036)
Constant	11.745	0.186	16.616	1.589	3.050	1.717
	(0.036)*	(0.030)*	(1.153)*	(0.039)*	(0.028)*	(0.034)*
Sample size	4,497	4,497	4,497	4,497	4,497	4,497
R^2	0.05	0.03	0.04	0.03	0.03	0.03

Sample: homes within of 0.3 of where a sex offender will eventually move

$$x_{1ijt} = \delta_0 + D_{ijt}^{0.1}\delta_1 + \xi_i$$

64

TABLE 3—IMPACT OF SEX OFFENDERS' LOCATIONS ON PROPERTY VALUE AND SALE PROBABILITY

	re- and post-arrival		Probability of sale†
	(5)	(6)	(7)
Within 0.1 miles of offender	−0.006 (0.012)	−0.006 (0.012)	−0.029 (0.035)
Within 0.1 miles × post-arrival	−0.036 (0.021)+	−0.116 (0.059)+	0.126 (0.059)*
Dist* ≤ 0.1 miles × post-arrival (0.1 Miles = 1)		0.107 (0.064)+	
Within 1/3 miles of offender			
Within 1/3 miles × post-arrival	0.003 (0.016)	0.004 (0.016)	−0.055 (0.040)
H_0: within 0.1 miles × post-arrival = 0	*p-value* = 0.0828	*p-value* = 0.0502	*p-value* = 0.0361
	✓	✓	✓
Sample size	9,086	9,086	1,519,364
R^2	0.75	0.75	0.01

65

Questions:

- How is identification achieved?
- What is the key assumption necessary for identification?
- Why might the estimates be an under/over estimate?

66