Session III
Differences in Differences (Dif-in-dif) and Panel Data

Christel Vermeersch

March 2007

Structure of this session

- When do we use Differences-in-Differences? (Diff-in-diff or DD)
- Estimation strategy: a bit of theory
- Examples:
  - Extension of education services in Indonesia
  - Water for life (Argentina)
  - School ranking
  - Progresa (Mexico)

When do we use diff-in-diff?

- We can't always randomize the beneficiaries…
- Estimating the impact of a "past" program
- We can try to find a "natural experiment" that allows us to identify the impact of a policy
  - For example. An unexpected change in policy could be seen as "natural experiment"
  - For example. A policy that only affects 16 year olds but not 15 year olds
- Even in natural experiments, we need to identify which is the group affected by the policy change ("treatment") and which is the group that is not affected ("control").
- The quality of the control group determines the quality of the evaluation.

A simple strategy for natural experiments: Before versus after

With 2 years of data: before (t=0) and after (t=1)

$$Y_{it} = \alpha + \beta.1(t=1) + \varepsilon_{it}$$

$$\beta_{OLS} = \overline{Y_1} - \overline{Y_o}$$

With more than 2 years of data and a policy change at time t=t*

$$Y_{it} = \alpha + \sum_{\tau=1}^{T} \beta_{\tau}.1(t=\tau) + \varepsilon_{it}$$

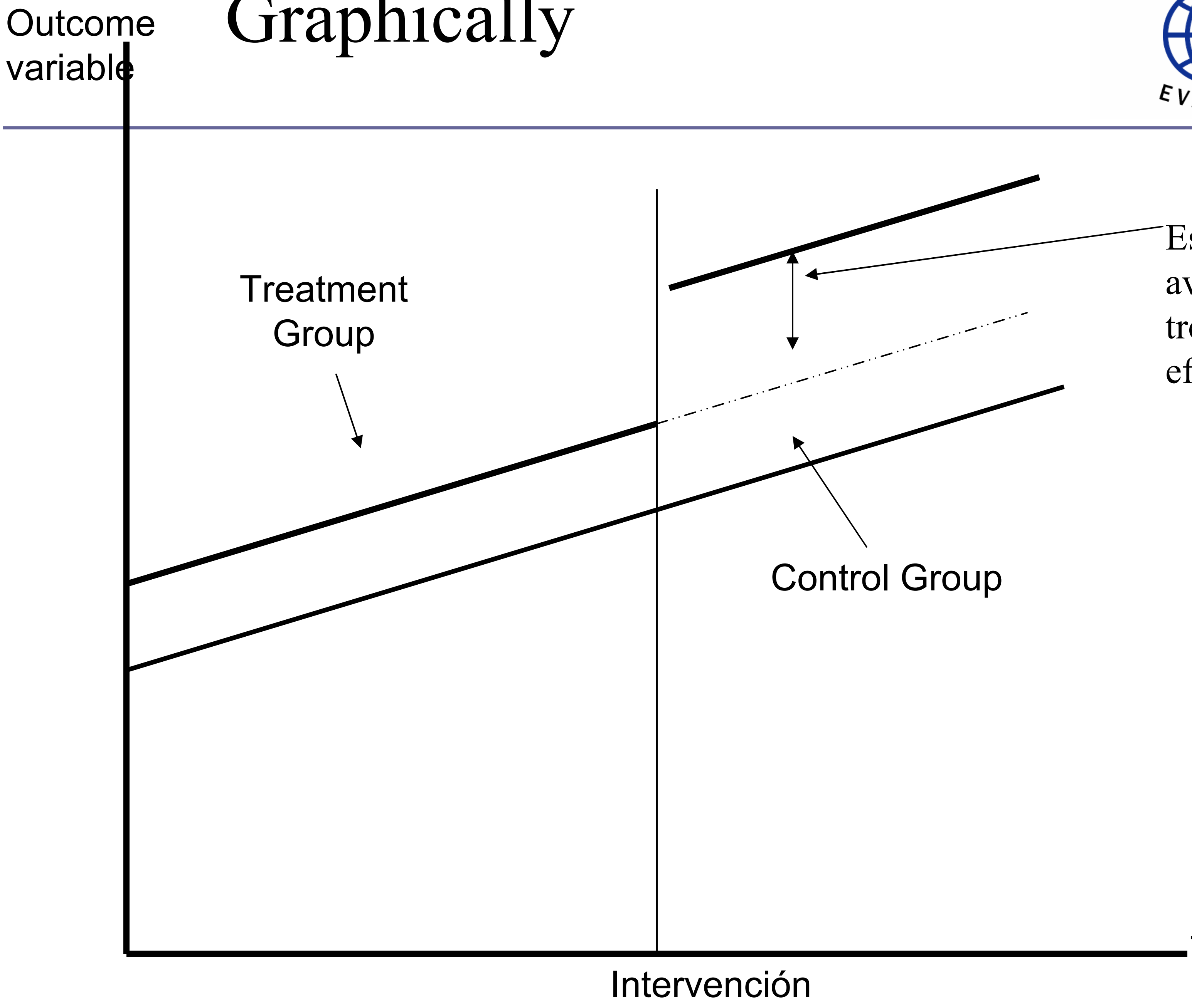
$$\beta_{\tau}^{OLS} = \overline{Y_{\tau}} - \overline{Y_o}$$

? Is there a structural break in the β_{τ}^{OLS} at t*?

Differences in differences

	Group affected by the policy change (treatment)	Group that is not affected by the policy change (comparison)	
After	$Y_{i1} \mid T$	$Y_{i1} \mid C$	
Before	$Y_{i0} \mid T$	$Y_{i0} \mid C$	
	We compare the mean of the outcome variable before and after, for the treatment group $\overline{Y_1} \mid T - \overline{Y_0} \mid T$	We compare the mean of the outcome variable before and after, for the comparison group. $\overline{Y_1} \mid C - \overline{Y_0} \mid C$	We substract those two means $(\overline{Y_1} \mid T - \overline{Y_0} \mid T) - (\overline{Y_1} \mid C - \overline{Y_0} \mid C)$

Graphically



Regression

$$Y_{it} = \alpha + \beta.1(t=1) + \gamma.1(i \in T) + \delta.(t=1).1(i \in T) + \varepsilon_{it}$$

$$\Downarrow$$

$$E(Y_{i1} | T) = ???$$

$$E(Y_{i0} | T) = ???$$

$$E(Y_{i1} | C) = ???$$

$$E(Y_{i0} | C) = ???$$

$$\Downarrow$$

$$DD = (E(Y_{i1} | T) - E(Y_{i0} | T)) - (E(Y_{i1} | C) - E(Y_{i0} | C))$$

$$= ???$$

Regression

$$Y_{it} = \alpha + \beta.1(t=1) + \gamma.1(i \in T) + \delta.1(t=1).1(i \in T) + \varepsilon_{it}$$

$$\Downarrow$$

$$E(Y_{i1} \mid T) = \alpha + \beta.1 + \gamma.1 + \delta.1.1 + E(\varepsilon_{i1} \mid i \in T) = \alpha + \beta + \gamma + \eta$$

$$E(Y_{i0} \mid T) = \alpha + \beta.0 + \gamma.1 + \delta.0.1 + E(\varepsilon_{i0} \mid i \in T) = \alpha + \gamma$$

$$E(Y_{i1} \mid C) = \alpha + \beta.1 + \gamma.0 + \delta.1.0 + E(\varepsilon_{i1} \mid i \in C) = \alpha + \beta$$

$$E(Y_{i0} \mid C) = \alpha + \beta.0 + \gamma.1 + \delta.0.0 + E(\varepsilon_{i0} \mid i \in C) = \alpha$$

$$\Downarrow$$

$$\begin{aligned} DD &= (E(Y_{i1} \mid T) - E(Y_{i0} \mid T)) - (E(Y_{i1} \mid C) - E(Y_{i0} \mid C)) \\ &= (\beta + \delta) - \beta \\ &= \delta \end{aligned}$$

If we have more than 2 time periods /groups

We use a regression with fixed effects for time and group:

$$Y_{it} = \alpha + \sum_{\tau=1}^{T} \beta_{\tau}.1(t=\tau) + \sum_{\iota=1}^{I} \gamma_{\iota}.1(i=\iota) + \delta.T_{it} + \varepsilon_{it}$$

where T_{it} is the intensity of the T treatment in group i in period t.

The identification of the treatment effect is based on the inter-temporal variation between the groups.

I.e.: changes in the outcome variable Y over time, that are specific to the treatment groups.

I.e.: jumps in trends in the outcome variable, that happen only for the treatment groups, not for the comparison groups, exactly at the time that the treatment kicks in.

If we have more than 2 time periods /groups

We use a regression with fized effects for time and group:

$$Y_{it} = \alpha + \sum_{\tau=1}^{T} \beta_{\tau}.1(t=\tau) + \sum_{\iota=1}^{I} \gamma_{\iota}.1(i=\iota) + \delta.T_{it} + \varepsilon_{it}$$

where T_{it} is the intensity of the T treatment in group i in period t.

The identification of the treatment effect is based on the intertemporal variation between the groups.
<u>I.e.</u>: changes in the outcome variable Y over time, that are specific to the treatment groups.
<u>I.e.</u>: jumps in trends in the outcome variable, that happen only for the treatment groups, not for the comparison groups, exactly at the time that the treatment kicks in.

Warnings…

- The fixed effects model is valid only when the policy change has an immediate impact on the oucome variable. If there is a delay in the impact of the policy change, we do need to use laged treatment variables.
- Diff-in-diff/ fixed effects control for: $Y_{i,t-1}$
  - Fixed group effects (eg. Farmers who own their land, farmers who don’t own their land)
  - Effects that are common to all groups at one particular point in time, or “common trends” (eg. The 2006 drought affected all farmers, regardless of who owns the land)
- Diff-in-diff/ fixed effects attributes differences in trends between the treatment and control groups, that occur at the same time as the intervention, to that intervention.
  - If there are other factors that affect the difference in trends between the two groups, then the estimation will be biased!

Quality control for diff-in-diff …

- Perform a « placebo» DD, ie use a « fake » treatment group
  - Ex. for previous years (eg. Years -2, -1).
  - Or using as a treatment group a population you know was NOT affected
  - If the DD estimate is different from 0, the trends are not parallel, and our original DD is likely to be biased.
- Use a different control group.
  - The two DDs should give the same estimates
- Use an outcome variable Y~ which you know is NOT affected by the intervention,
  - using the same control group and treatment year.
  - If the DD estimate is different from zero, we have a problem.

Frequently occuring issues when using DD

- Participation is based in difference in outcomes prior to the intervention
  - "Ashenfelter dip"
- Functional form dependency
- When the size of the response depends in a non-linear way on the size of the intervention, and we compare a group with high treatment intensity, with a group with low treatment intensity.
- When the observation within the unit of time/ group are correlated.

THE WORLD BANK

IMPACT EVALUATION

Examples of Diff-in-Diff and fixed effects estimation

Human Development Network

Middle East and North Africa Region

Spanish Impact Evaluation Fund

Schooling and labor market consequences of school construction in Indonesia: evidence from an unusual policy experiment

Esther Duflo (MIT)

American Economic Review Sept 2001

Research questions

- School infrastructure -> educational achievement ?
- Educational achievement -> salary level ?
- What is the economic return on schooling?

Program description

- 1973-1978: The Indonesian government built 61,000 school (equivalent to one school per 500 children between 5 and 14 years old)
- The enrollment rate increased from 69% to 85% between 1973 and 1978
- The number of schools built in each region depended on the number of children out of school in those regions in 1972, before the start of the program.

Identification of the treatment effect

There are 2 sources of variations in the intensity of the program for a given individual:

- By region: there is variation in the number of schools received in each region

- By age:
  - Children who were older than 12 years in 1972 did not benefit from the program.
  - The younger a child was 1972, the more it benefited from the program – because she spent more time in the new schools

Sources of data

- 1995 population census: individual-level data on:
  - birth date
  - current salary level
- The intensity of the building program in the birth region of each person in the sample
- Sample: men born between 1950 and 1972

A first estimation of the impact…

First step: let's simplify the problem and estimate the impact of the program.

- We simpligy the intensity of the program: high or low
- We simplify the groups of children affected by the program:
  - "Young" cohort of children who benefitted
  - "Older" cohort of children who did not benefit

Let's look at the average of the outcome variables

		Intensity of the Building program	
		Alta	Baja
Age in 1974	2-6 (young cohort)	8.49	9.76
	12-17 (older cohort)	8.02	9.4
	Difference	0.47	0.36

0.12 DD
(0.089)

Let's look at the average of the outcome variables

		Intensity of the Building program		
		High	Low	Difference
Age in 1974	2-6 (young cohort)	8.49	9.76	-1.27
	12-17 (older cohort)	8.02	9.4	-1.39
				0.12 DD (0.089)

Placebo Diff in diff (Cf. p.798, Table 3, panel B)

<u>Idea</u>: Look for 2 groups whom you know did not benefit, compute a DD, and check whether the estimated effect is 0. If it is NOT 0, we're in trouble….

		Intensity of the Building program		
		High	Low	
Age in 1974	12-17	8.02	9.40	
	18-24	7.70	9.12	
	Difference	0.32	0.28	0.034 DD (0.098)

Second step: let's estimate this with a regression

$$S_{ijk} = c + \alpha_j + \beta_k + \gamma.(P_j.T_i) + \delta.(C_j.T_i) + \varepsilon_{ijk}$$

with

S_{ijk} = education level of person i in region j in cohort k

P_j = 1 if the person was born in a region with a high program intensity

T_i = 1 if the person belongs to the "young" cohort

C_j = dummy variable for region j

ε_{ijk} = error term for person i,j,k

Third step: let's use additional information

We will use the intensity of the program in each region:

$$S_{ijk} = c + \alpha_j + \beta_k + \gamma.(P_j.T_i) + \delta.(C_j.T_i) + \varepsilon_{ijk}$$

where P_j = the intensity of building activity in region j

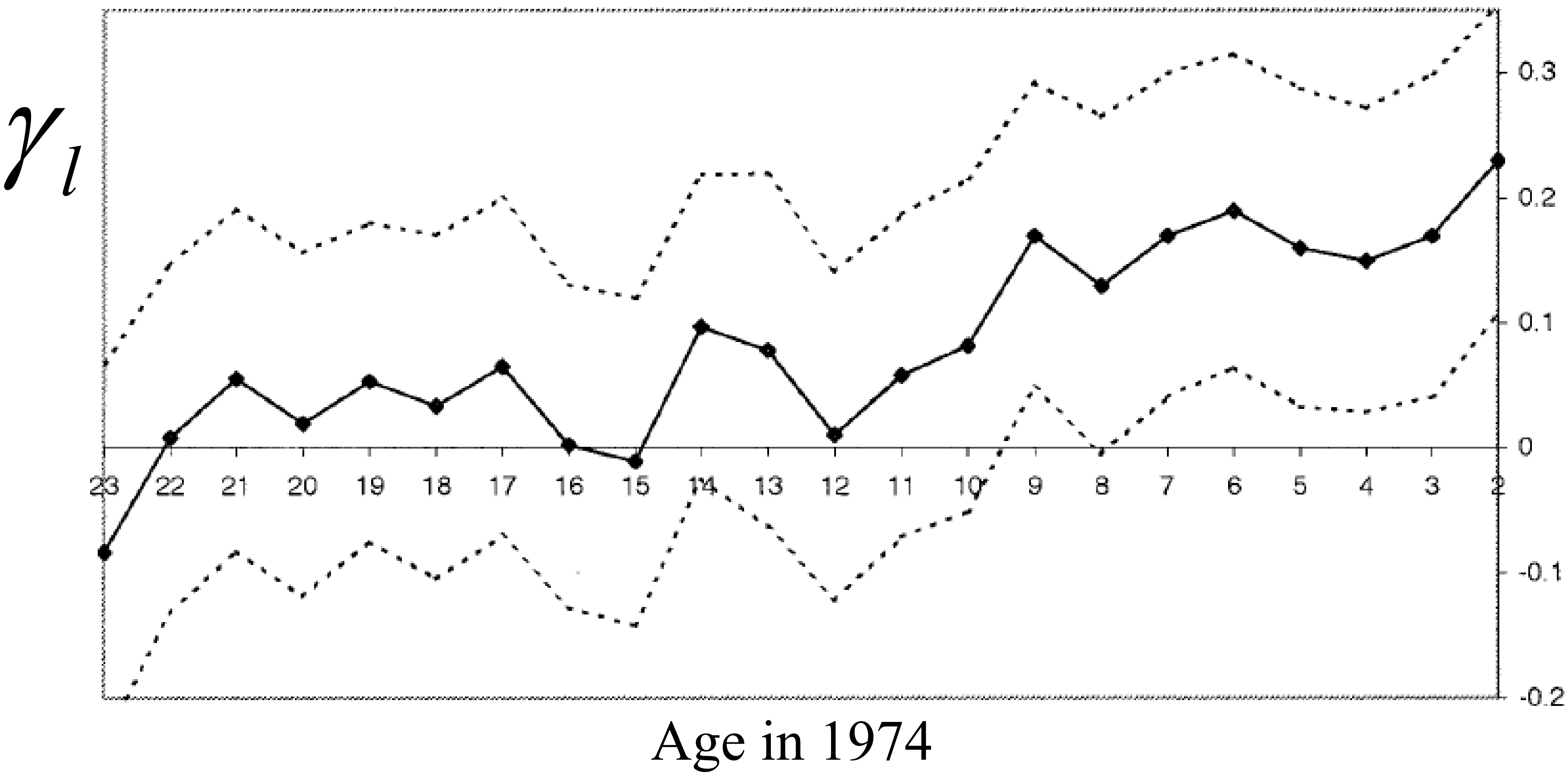
C_j = a vector of regional characteristics

We estimate the effect of the program for each cohort separately:

$$S_{ijk} = c + \alpha_j + \beta_k + \sum_{l=2}^{23} \gamma_l.(P_j.d_i) + \sum_{l=2}^{23} \delta_l C_j T_i + \varepsilon_{ijk}$$

where d_i = a dummy variable for belonging to cohort i

Program effect per cohort



For y=dependent variable =salary

	Log(wages)		
	Level of program in region of birth		
	High (4)	Low (5)	Difference (6)
Panel A: Experiment of Interest			
Aged 2 to 6 in 1974	6.61 (0.0078)	6.73 (0.0064)	−0.12 (0.010)
Aged 12 to 17 in 1974	6.87 (0.0085)	7.02 (0.0069)	−0.15 (0.011)
Difference	−0.26 (0.011)	−0.29 (0.0096)	0.026 (0.015)
Panel B: Control Experiment			
Aged 12 to 17 in 1974	6.87 (0.0085)	7.02 (0.0069)	−0.15 (0.011)
Aged 18 to 24 in 1974	6.92 (0.0097)	7.08 (0.0076)	−0.16 (0.012)
Difference	0.056 (0.013)	0.063 (0.010)	0.0070 (0.016)

Conclusion

- Results: For each school built per 1000 students
  - The average educational achievement increase by 0.12- 0.19 years
  - The average salaries increased by 2.6 – 5.4 %
- Making sure the DD estimation is accurate:
  - A placebo DD gave “0” estimated effect
  - Use various alternative specifications
  - Check that the impact estimates for each age cohort make sense

THE WORLD BANK

IMPACT EVALUATION

Water for Life: The Impact of the Privatization of Water Services on Child Mortality

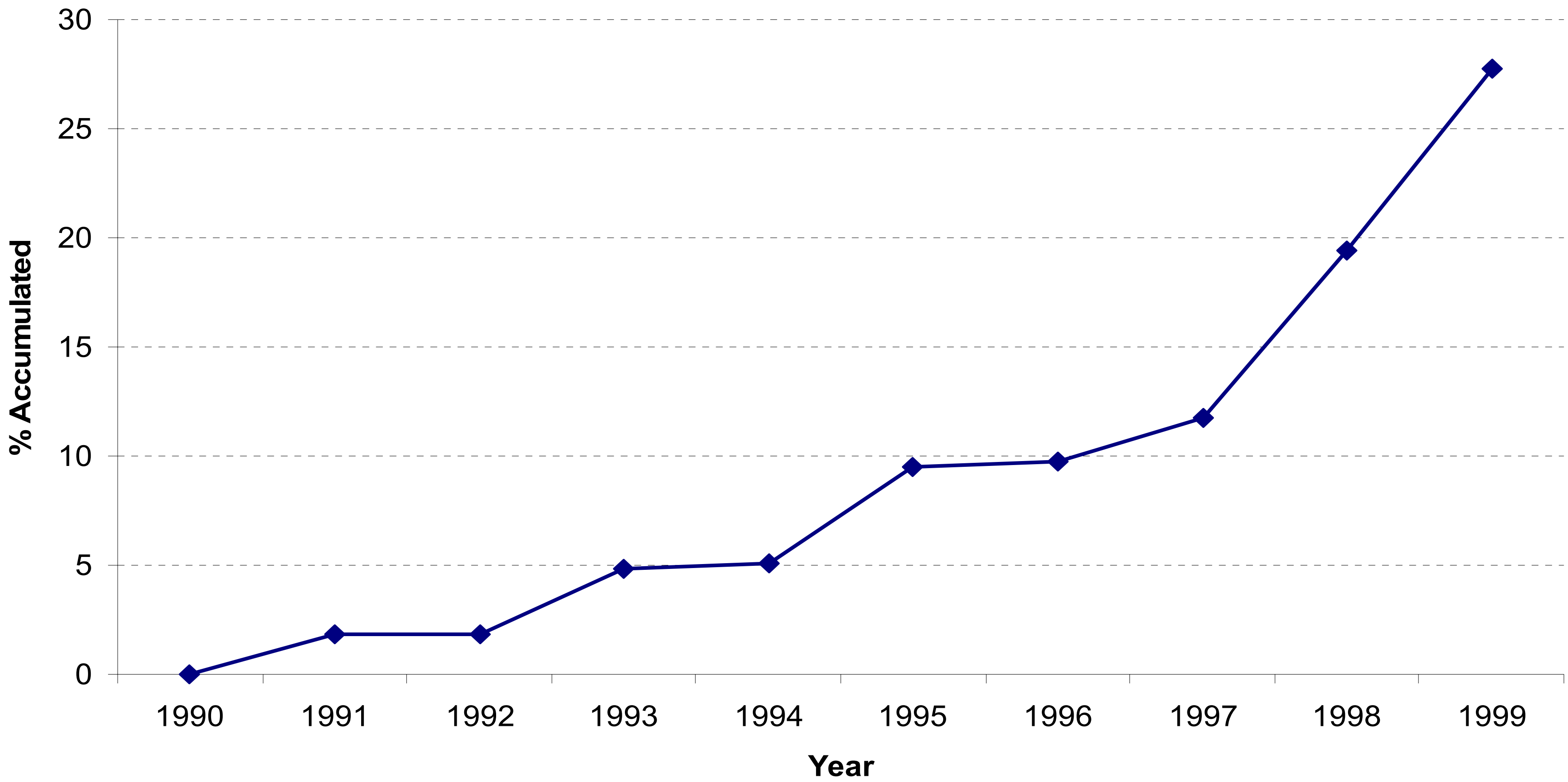
Sebastián Galiani, Universidad de San Andrés
Paul Gertler, UC Berkeley
Ernesto Schargrodsky, Universidad Torcuato Di Tella

JPE (2005)

Changes in water services delivery 1990-1999

Type of provision methods	**Number of municipalities**	**%**
Always public	196	39.7%
Always a not-for-profit cooperative	143	28.9%
Converted from public to private	138	27.9%
Always private	1	0.2%
No information	16	3.2%
TOTAL	**494**	**100.0%**

Figure1: Percentage of Municipalities with Privatized Water Systems



Use “outside” factors to determine who privatizes…

- The political party that governs the municipality
  - Federal, Peronist y Provincial parties: allowed privatization
  - Radical party: did not allow privatization
- Which party is in power / whether the water gets privatized does not depend on:
  - Income, unemployment, inequality at the municipal level
  - Recent changes in infant mortality rates

Regression

$$y_{it} = \alpha \, dI_{it} + \beta \, \boldsymbol{x}_{it} + \lambda_t + \mu_i + \varepsilon_{it}$$

where

y_{it} = infant mortality rate in munic. i in year t

dI_{it} = dummy variable that takes value 1 if municipality i has private water provider in year t

x_{it} = vector of covariates

λ_t = fixed time effect

μ_i = fixed municipality effect

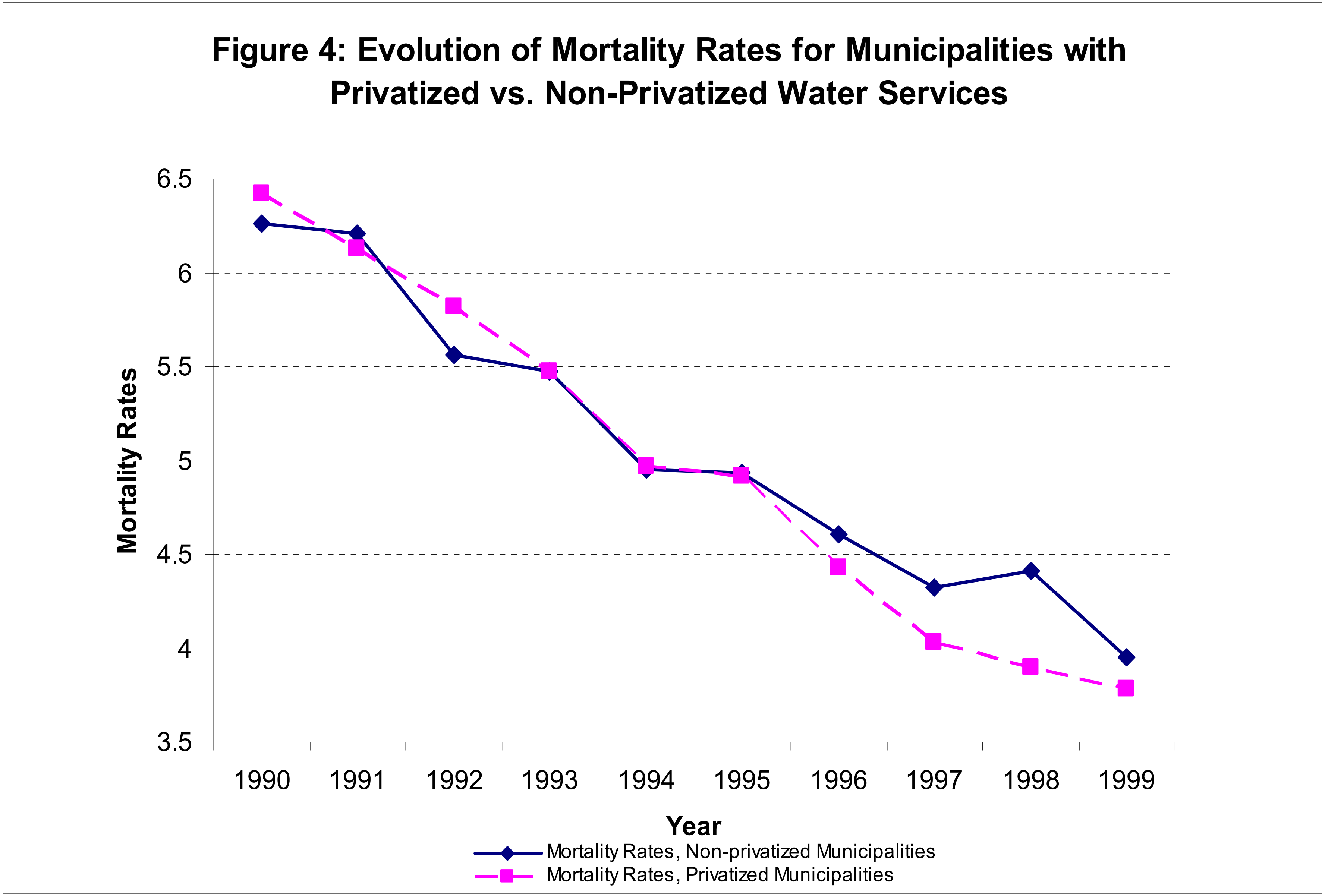


Figure 4: Evolution of Mortality Rates for Municipalities with Privatized vs. Non-Privatized Water Services
Mortality Rates
6.5
6
5.5
5
4.5
4
3.5
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
Year
Mortality Rates, Non-privatized Municipalities
Mortality Rates, Privatized Municipalities

DD results: Privatization reduced infant mortality

	Full Sample			Common Support			Matched
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Private Water (=1)	- 0.33 **	- 0.32 *	- 0.29 *	- 0.54 ***	- 0.54 ***	- 0.53 ***	- 0.60 ***
% Δ in Mortality	*- 5.3 %*	*- 5.1 %*	*- 4.5 %*	*- 8.6 %*	*- 8.6 %*	*- 8.4 %*	*- 10.0 %*
Real GDP/Capita		0.01	0.01		0.01	0.01	
Unemployment Rate		- 0.56	-0.64		-0.78	-0.84	
Inequality (Gini)		5.17 *	5.09 *		3.05	3.05	
Public Spending/Cap		- 0.03	- 0.04		-0.07 *	- 0.07 *	
Radical Party (=1)			0.48 *			0.17	
Peronist Party (=1)			- 0.20			- 0.17	
F-Stat Municipal FE	13.84***	11.92***	11.51***	10.39***	8.65***	8.32***	
F-Stat for year FE	55.03***	19.88***	18.25***	52.25***	15.59***	12.98***	

Quality checks on the DD

1. Check that the trends in infant mortality were identical in the two types of municiplaities BEFORE privatization
   - You can do this by running the same equation, using only the years before the intervention – the treatment effect should be zero for those years
   - Found that we cannot reject the null hypothesis of equal trends between treatment and controls, in the years before privatization

2. Check that privatization only affects mortality through reasons that are logically related to water and sanitation issues.
   - For example, there is no effect of privatization on death rate from cardiovascular disease or accidents.

Impact of privatization on death from various causes D-in-D on common support

	1990 Mean Mortality Rate	Estimated Impact Coefficients	%Δ in Mortality Rate
Infectious and parasitic diseases	.565	−.103 (.048)** [.055]* {.068}	−18.2
Perinatal deaths	2.316	−.266 (.105)** [.107]** {.123}**	−11.5
All other causes in aggregate	2.565	−.082 (.114) [.101] {.109}	−3.2
All other causes disaggregated:			
Accidents	.399	−.004 (.057)	...
Congenital anomalies	.711	−.022 (.056)	...
Skin and soft-tissue diseases	.000	.000 (.001)	...
Blood and hematologic diseases	.024	−.002 (.008)	...
Nervous system disorders	.163	.025 (.026)	...
Cardiovascular diseases	.236	.006 (.030)	...

Privatization has a larger effect in poor and very poor municipalities than in non-poor municipalities

Municipalities	*Average mortality per 100, 1990*	*Estimated impact*	*% change in mortality*
Non-poor	5.15	0.105	…
Poor	7.17	-0.767***	-10.7%
Very poor	9.46	-2.214***	-23.4%

Conclusion: Using a combination of methods, we found that….

- Privatization of water services is associated with a reduction in infant mortality of 5-7 percent.
- The reduction of mortality is:
  - Due to fewer deaths from infectious and parasitic diseases.
  - Not due to changes in death rates from reasons not related to water and sanitation
- The largest decrease in infant mortality occurred in low income municipalities

THE WORLD BANK

IMPACT EVALUATION

The central role of noise in evaluating interventions that use test scores to rank schools

Chay, McEwan and Urquiola
American Economic Review, 2005

P900 education program in Chile

- Program details
  - 1990/1991: provision of infrastructure and learning materials
  - 1992 and later: teacher training and remedial for students
- Selection of schools:
  - 900 schools with the lowest average scores in the 1988 4th grade standardized test
- Why is this paper interesting?
  - It illustrates very nicely the bias that arises when using DD when there is an "Ashenfelter dip"+ it suggests an alternative methodology that solves the problem

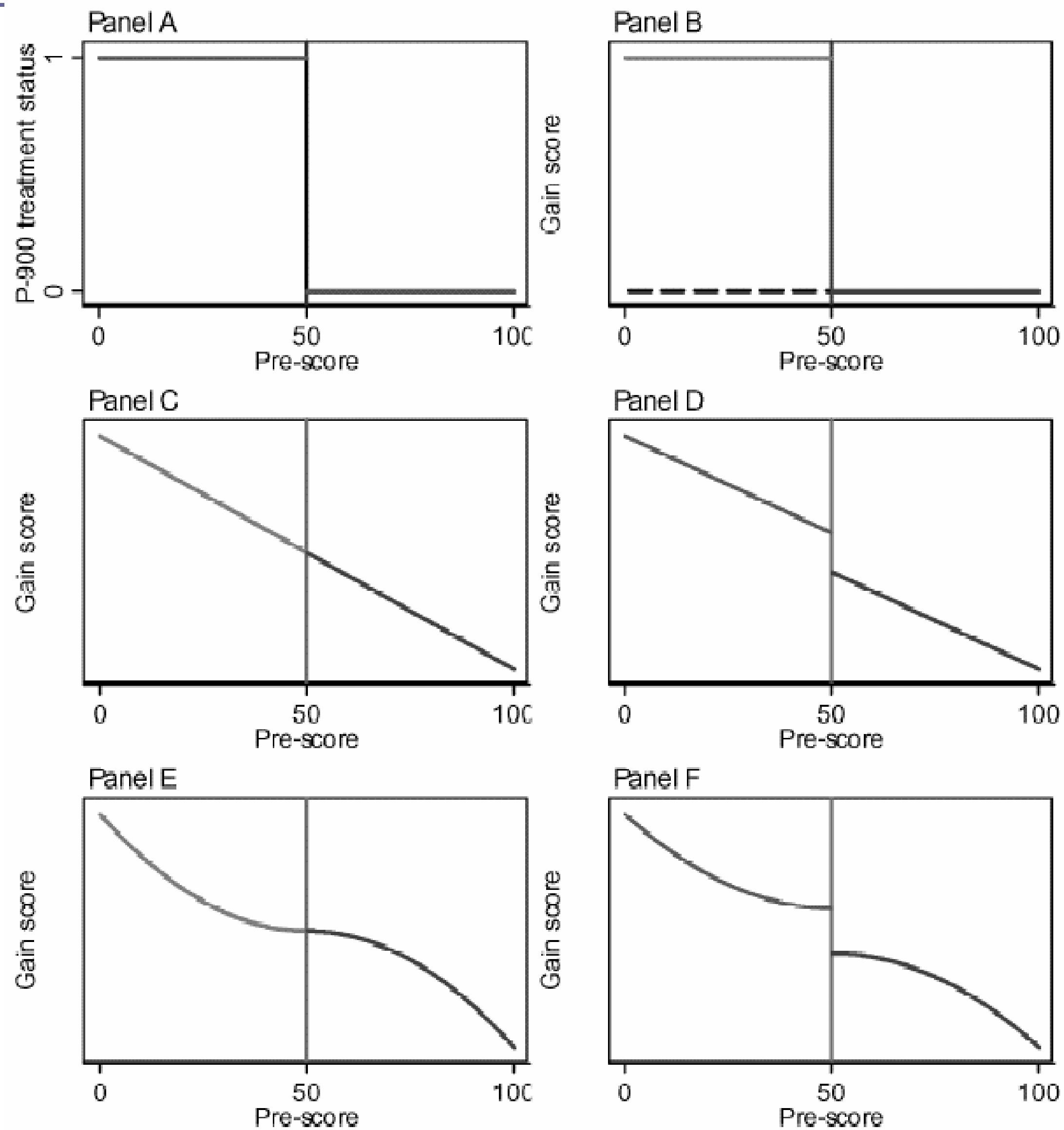
Reversion to the mean and Diff in Diff (the "Ashenfelter dip")

- Schools that performed poorly in the 1988 test, did so for two basic reasons:
  - Bad luck (low lambda)
  - Poor intrinsic quality (low u or alpha)

$$y_{ij}^{88} = \lambda_j + u_j^{88} + \alpha_i^{88}$$

- If a school had "bad luck"in 1988, this does not predict whether the school is going to have bad luck again in 1990. (Otherwise it would not be bad "luck")
- So, on average, schools that had a low outcome in 1988, will have a higher outcome in 1990.
- This is called "reversion to the mean".

Hypothetical assignment of schools, and its effect on the schools'outcomes in the test



P-900 Effects on [redacted] Gain Scores

	1988-1990 Gain score				
	(1)	(2)	(3)	(4)	(5)
Panel A: Mathematics					
P-900	2.28*** (0.40)	-0.02 (0.47)	-0.11 (0.46)	-0.16 (0.51)	0.25 (0.53)
Score relative to cutoff		-0.16*** (0.02)			
σ^2_λ			142.32*** (18.36)		
SES index, 1990					0.15*** (0.01)
Cubic in 1988 score	N	N	N	Y	Y
Region dummies	N	N	N	N	Y
Adjusted R^2	0.013	0.041	0.046	0.041	0.130
Sample size	2,644	2,644	2,644	2,644	2,644

Huber-White standard errors are in parentheses.
*** significant at 1% ** significant at 5% * significant at 10%

Simple DD

Controls Reversion To the Mean

P-900 Effects on [redacted]ain Scores

	1988-1992 Gain score				
	(6)	(7)	(8)	(9)	(10)
Panel A: Mathematics					
P-900	3.74*** (0.44)	1.61*** (0.50)	1.48*** (0.48)	1.79*** (0.56)	2.09*** (0.60)
Score relative to cutoff		-0.15*** (0.02)			
σ^2_λ			141.65*** (34.01)		
SES index, 1990					0.18*** (0.01)
Change in SES, 1990-1992					0.07*** (0.01)
Cubic in 1988 score	N	N	Y	Y	Y
Region dummies	N	N	N	N	Y
Adjusted R^2	0.031	0.053	0.060	0.053	0.140
Sample size	2,591	2,591	2,591	2,591	2,591

Huber-White standard errors are in parentheses.
*** significant at 1% ** significant at 5% * significant at 10%

Simple DD

Controls Reversion To the Mean

P-900 Effects on 1[redacted]ain Scores

	1988-1990 Gain score				
	(1)	(2)	(3)	(4)	(5)
Panel B: Language					
P-900	4.25*** (0.39)	0.25 (0.44)	0.18 (0.41)	-0.02 (0.48)	0.54 (0.49)
Score relative to cutoff		-0.28*** (0.02)			
σ^2_λ			68.79*** (5.55)		
SES index, 1990					0.13*** (0.01)
Cubic in 1988 score	N	N	N	Y	Y
Region dummies	N	N	N	N	Y
Adjusted R^2	0.050	0.147	0.151	0.155	0.230
Sample size	2,644	2,644	2,644	2,644	2,644

Simple DD

Controls Reversion To the Mean

Huber-White standard errors are in parentheses.
*** significant at 1% ** significant at 5% * significant at 10%

P-900 Effects on [redacted] Gain Scores

	1988-1992 Gain score				
	(6)	(7)	(8)	(9)	(10)
Panel B: Language					
P-900	5.94*** (0.39)	2.24*** (0.44)	2.09*** (0.43)	1.67*** (0.48)	2.10*** (0.52)
Score relative to cutoff		-0.26*** (0.02)			
σ^2_λ			62.32*** (11.21)		
SES index, 1990					0.16*** (0.01)
Change in SES, 1990-1992					0.07*** (0.01)
Cubic in 1988 score	N	N	Y	Y	Y
Region dummies	N	N	N	N	Y
Adjusted R^2	0.089	0.163	0.175	0.173	0.250
Sample size	2,591	2,591	2,591	2,591	2,591

Huber-White standard errors are in parentheses.
*** significant at 1% ** significant at 5% * significant at 10%

Simple DD

Controls Reversion To the Mean

P-900 Effects on [redacted] Gain Scores, within Narrow Bands of the Selection Threshold

	± 5 Points		± 3 Points		± 2 Points	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: [redacted]						
P-900	1.50**	1.82***	1.79***	2.00***	2.37***	2.39***
	(0.60)	(0.66)	(0.73)	(0.77)	(0.84)	(0.85)
SES index, 1990		0.14***		0.13***		0.12***
		(0.02)		(0.03)		(0.03)
Change in SES, 1990-1992		0.08***		0.09***		0.06***
		(0.02)		(0.02)		(0.02)
Cubic in 1988 score	N	Y	N	Y	N	Y
R^2	0.007	0.067	0.011	0.074	0.021	0.080
Sample size	883	883	553	553	363	363

Huber-White standard errors are in parentheses.
*** significant at 1% ** significant at 5% * significant at 10%

P-900 Effects on [redacted] Gain Scores, within Narrow Bands of the Selection Threshold

	± 5 Points		± 3 Points		± 2 Points	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B: [redacted]						
P-900	2.78***	2.23***	2.10***	1.96***	2.62***	2.48***
	(0.54)	(0.57)	(0.69)	(0.70)	(0.80)	(0.75)
SES index, 1990		0.13***		0.12***		0.12***
		(0.02)		(0.03)		(0.03)
Change in SES, 1990-1992		0.07***		0.09***		0.06***
		(0.02)		(0.02)		(0.02)
Cubic in 1988 score	N	Y	N	Y	N	Y
R^2	0.030	0.111	0.017	0.101	0.029	0.111
Sample size	883	883	553	553	363	363

Huber-White standard errors are in parentheses.
*** significant at 1% ** significant at 5% * significant at 10%

References

- Duflo, E. (2001). “Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence From an Unusual Policy Experiment,” *American Economic Review*, Sept 2001
- Sebastian Galiani, Paul Gertler and Ernesto Schargrodsky (2005): “Water for Life: The Impact of the Privatization of Water Services on Child Mortality,” *Journal of Political Economy,* Volume 113, pp. 83-120.
- Chay, Ken, McEwan, Patrick and Miguel Urquiola (2005): “The central role of noise in evaluating interventions that use test scores to rank schools,” *American Economic Review*, 95, pp. 1237-58.
- Secretaría de Desarrollo Social Mexico (1999): Progresa: Evaluación de Resultados del Programa de Educación, Salud y Alimentación.
- Gertler, Paul (2004): “Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA’s Control Randomized Experiment, *American Economic Review*, 94, pp. 336-41.