



The  
World Bank



Human Development  
Network



Spanish Impact  
Evaluation Fund

[www.worldbank.org/hdchiefeconomist](http://www.worldbank.org/hdchiefeconomist)



# **DOUBLE DIFFERENCES**

---

**& DONNEES DE PANEL**

---

## **Session Technique III**

**Phillippe Leite**  
**La Banque mondiale**

# Structure de cette session

- 1 Quand utilise-t-on les Doubles Différences? (DD)
- 2 Stratégie d'estimation: 3 façons d'examiner les doubles différences
- 3 Exemples:
  - Extension des services de l'enseignement (Indonésie)
  - Eau pour la vie (*Argentine*)

1

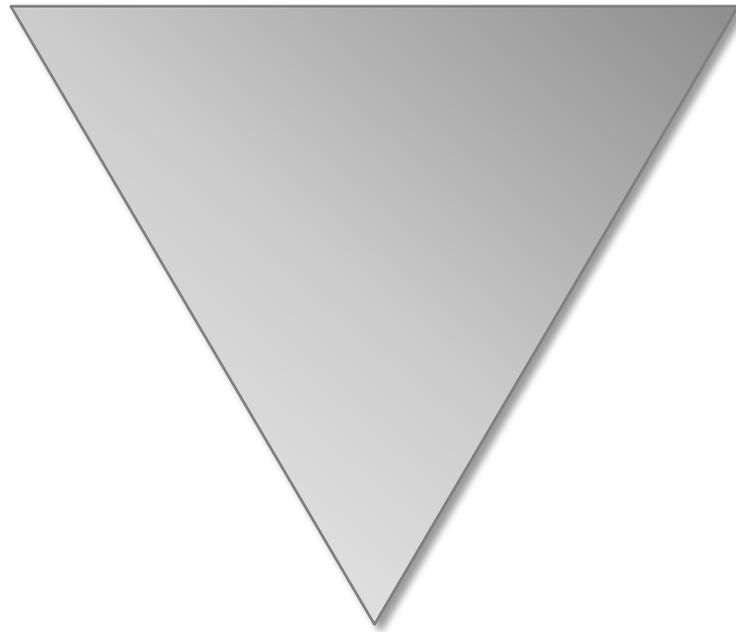
# Quand utilise-t-on doubles différences?

- On ne peut pas toujours sélectionner aléatoirement les bénéficiaires...  
Ex: estimer l'impact d'un programme «passé»
- Avant toute chose, il est nécessaire d'identifier
  - Quel groupe est affecté par le changement de politique ("*traitement*"), et
  - Quel groupe n'est pas affecté («*témoin*»),
- Nous pouvons essayer de trouver une «expérience naturelle» qui nous permette d'identifier l'impact d'une politique
  - Ex: un changement de politique inattendu
  - Ex: une politique affectant seulement les jeunes âgés de 16 ans mais pas ceux âgés de 15 ans
- La qualité du groupe témoin détermine la qualité de l'évaluation.

## 2 3 façons de regarder les Diff-in-Diff

Avec la "Boîte"

Graphiquement



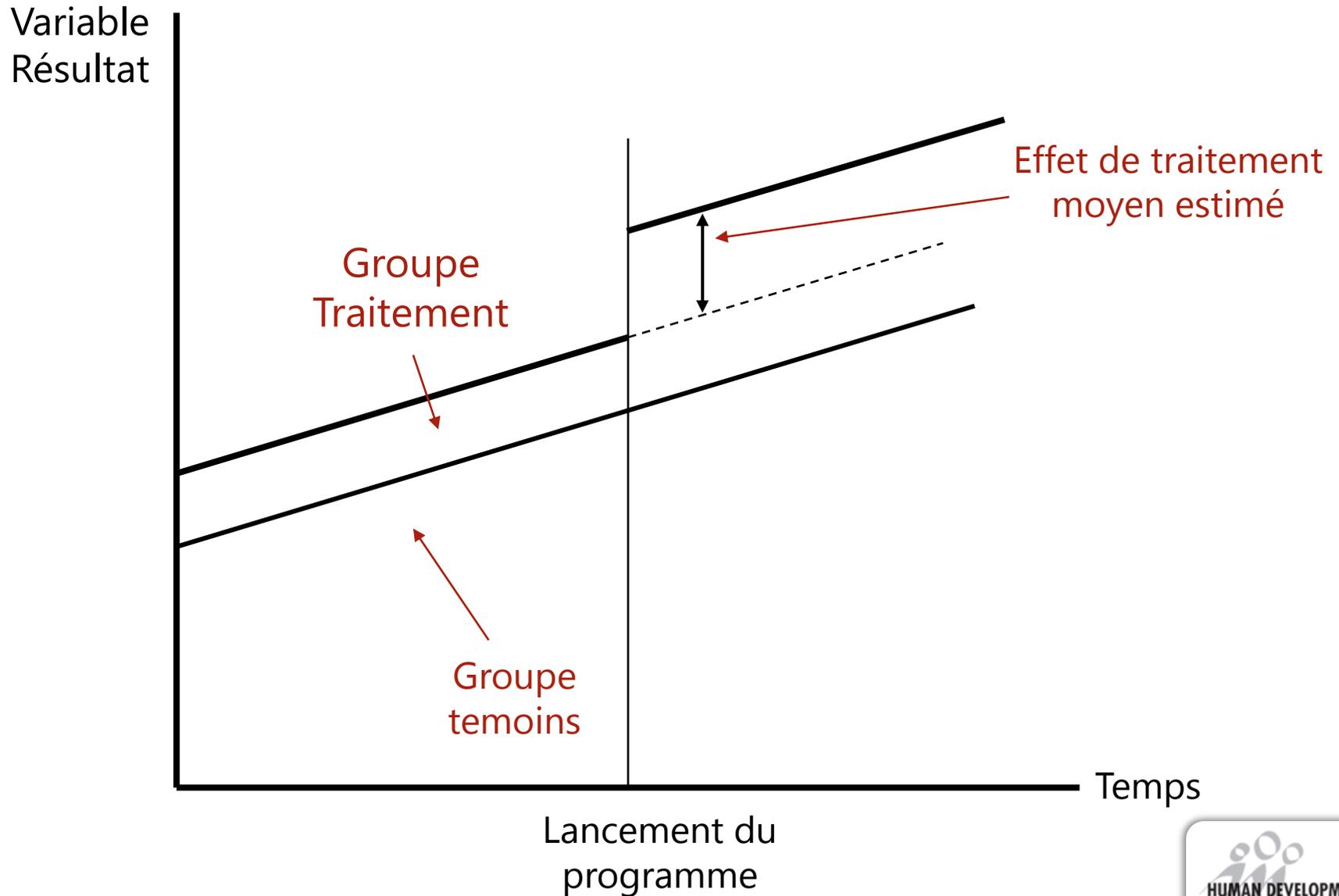
Dans une  
regression

# La boîte

	<b>Groupe affecté par le changement de politique</b> ( <i>traitement</i> )	<b>Groupe qui n'est pas affecté par le changement politique</b> ( <i>comparaison</i> )
Après lancement du programme	$Y_1(u_i) \mid D_i=1$	$Y_1(u_i) \mid D_i=1$
Avant lancement du programme	$Y_0(u_i) \mid D_i=1$	$Y_0(u_i) \mid D_i=1$
	$(\bar{Y}_1 \mid D=1) - (\bar{Y}_0 \mid D=1)$	$(\bar{Y}_1 \mid D=0) - (\bar{Y}_0 \mid D=0)$

$$DD = [(\bar{Y}_1 \mid D=1) - (\bar{Y}_0 \mid D=1)] - [(\bar{Y}_1 \mid D=0) - (\bar{Y}_0 \mid D=0)]$$

# Graphiquement



# Régression (pour 2 périodes temporelles)

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

⇓

$$E(Y_{i1} | D_i = 1) = ???$$

$$E(Y_{i0} | D_i = 1) = ???$$

$$E(Y_{i1} | D_i = 0) = ???$$

$$E(Y_{i0} | D_i = 0) = ???$$

⇓

$$\begin{aligned} DD &= (E(Y_{i1} | D_i = 1) - E(Y_{i0} | D_i = 1)) - (E(Y_{i1} | D_i = 0) - E(Y_{i0} | D_i = 0)) \\ &= ??? \end{aligned}$$

# Régression (pour 2 périodes temporelles)

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

⇓

$$E(Y_{i1} | D_i = 1) = \alpha + \beta.1 + \gamma.1 + \delta.1.1 + E(\varepsilon_{i1} | D_i = 1) = \alpha + \beta + \gamma + \delta$$

$$E(Y_{i0} | D_i = 1) = \alpha + \beta.0 + \gamma.1 + \delta.0.1 + E(\varepsilon_{i0} | D_i = 1) = \alpha + \gamma$$

$$E(Y_{i1} | D_i = 0) = \alpha + \beta.1 + \gamma.0 + \delta.1.0 + E(\varepsilon_{i1} | D_i = 0) = \alpha + \beta$$

$$E(Y_{i0} | D_i = 0) = \alpha + \beta.0 + \gamma.0 + \delta.0.0 + E(\varepsilon_{i0} | D_i = 0) = \alpha$$

⇓

$$DD = (E(Y_{i1} | D_i = 1) - E(Y_{i0} | D_i = 1)) - (E(Y_{i1} | D_i = 0) - E(Y_{i0} | D_i = 0))$$

$$= (\beta + \delta) - \beta$$

$$= \delta$$

# Si nous avons plus de périodes de temps/groupes

Nous utilisons une régression avec effets fixes pour le temps et le groupe...

$$Y_{it} = \alpha + \sum_{\tau=1}^T \beta_{\tau} \cdot 1(t = \tau) + \sum_{l=1}^I \gamma_l \cdot 1(i = l) + \delta \cdot D_{it} + \varepsilon_{it}$$

where  $D_{it}$  is the intensity of the  $D$  treatment in group  $i$  in period  $t$ .

# Identification en Doubles Différences

- L'identification de l'effet de traitement repose sur la variation intertemporelle entre les groupes.
- *i.e.*: les changements dans la variable de résultat Y dans le temps, qui sont spécifiques aux groupes de traitement,
- *i.e.* des changements de tendance de la variable de résultat affectant seulement les groupes de traitement, et non les groupes de comparaison/de contrôle, exactement au moment où le traitement prend effet.

# Avertissements

- Doubles différences/ contrôle d'effets fixes pour
  - Effets de groupes fixes. *Ex, Agriculteurs propriétaires de leurs terres, agriculteurs non-propriétaires de leurs terres*
  - Effets communs à tous les groupes en un point particulier dans le temps, ou "tendances communes"  
*Ex. La sécheresse de 2006 affectant tous les agriculteurs, indépendamment du statut de propriété*

- Valide seulement lorsque le changement de politique a un **impact immédiat sur la variable de résultat.**

S'il existe un délai de l'impact du changement de politique, nous devons utiliser les variables de traitement retardé.

# Avertissements

- La méthode des Doubles Différences attribue toute différence de tendance entre les groupes de traitement et témoins, survenant en même temps que l'intervention, à cette intervention.
- S'il existe d'autres facteurs affectant la différence de tendance entre deux groupes, alors l'estimation sera **biaisée!**

# Contrôle qualité pour la méthode des Doubles différences

- Exécuter une DD «placébo», ex, utiliser un « faux » groupe de traitement
  - Ex. pour les années précédentes (ex, Années -2, -1),
  - Ou prendre comme groupe de traitement une population que vous savez ne PAS avoir été affectée
  - Si l'estimation DD est différente de 0, les tendances ne sont pas parallèles, et notre estimateur DD initial est probablement biaisé.
- Utiliser un groupe témoin différent  
Les deux DD devraient donner les même estimations.
- Utiliser une variable de résultat  $Y_2$  que vous savez ne pas être affectée par l'intervention:
  - Utiliser le même groupe témoin et la même année de traitement
  - Si l'estimation DD est différente de 0, il y a un problème

# Problèmes fréquents en Doubles Différences (DD)

- La participation est basée sur les différences de résultats avant l'intervention : "*Ashenfelter dip*"
- Forme fonctionnelle de dépendance?
- Quand l'ampleur de la réponse dépend de façon non-linéaire à l'ampleur de l'intervention, et que nous comparons un groupe à forte intensité de traitement avec un groupe à faible intensité de traitement
- Quand les observations dans les limites de l'unité de temps/groupe sont corrélées,



## **Exemple 1**

# **Conséquences de la construction d'écoles sur la scolarisation et le marché du travail en Indonésie: constat d'une expérience politique non- ordinaire**

---

Esther Duflo, MIT  
**American Economic Review, Sept 2001**

# Questions de recherche

Infrastructure  
scolaire



réussite scolaire?

Réussite  
scolaire



niveau de salaire?

Quel est les retombées  
économiques de la  
scolarisation?

# Description du programme

- **1973-1978**: Le gouvernement indonésien a construit 61,000 écoles équivalent à **une école pour 500 enfants** d'âge compris entre **5 et 14 ans**
- Le taux de scolarisation est passé de 69% à **85%** entre 1973 et 1978
- Le nombre d'écoles construites dans chaque région était fonction du nombre d'enfants non-scolarisés dans ces régions en 1972 avant le lancement du programme.

# Identification de l'effet de traitement

Il existe 2 sources de variation dans l'intensité du programme pour un individu donné:

- Par région:  
il existe une variation dans le nombre d'écoles reçues dans chaque région.
- Par âge:
  - Les enfants âgés de plus de 12 ans en 1972 n'ont pas bénéficié du programme.
  - Plus un enfant était jeune en 1972, plus il bénéficiait du programme en raison du nombre d'années passées dans les nouvelles écoles.

# Sources des données

- 1995 recensement de la population : données individuelles :
  - date de naissance
  - niveau de salaire en 1995
  - niveau d'éducation en 1995
- L'intensité du programme de construction dans la région de naissance pour chaque personne de l'échantillon.
- **Echantillon:** hommes nés entre 1950 et 1972.

# Une première estimation de l'impact

**Première étape:** simplifier le problème et estimer l'impact du programme.

- Nous simplifions l'intensité du programme: forte ou faible.
- Nous simplifions les groupes d'enfants affectés par le programme:
  - les jeunes cohorte d'enfants bénéficiaires
  - les personnes âgées cohorte d'enfants non-bénéficiaires

# Regardons la moyenne des variables de résultat des années de scolarisation

## Intensité du programme de construction

Age en 1974	Forte	Faible	
2-6 (cohorte jeunes)	8,49	9,76	
12-17 (cohorte âgé)	8,02	9,4	
<b>Différence</b>	<b>0,47</b>	<b>0,36</b>	<b>0,12 DD (0,089)</b>

# Regardons la moyenne des variables de résultat des années de scolarisation

## Intensité du programme de construction

Age en 1974	Forte	Fible	Différence
2-6 (cohorte jeune)	8,49	9,76	<b>-1,27</b>
12-17 (cohorte âgé)	8,02	9,4	<b>-1,39</b>
			<b>0,12 DD</b> <b>(0,089)</b>

# Placébo doubles differences

(Cf. p.798, Table 3, panel B)

## Idée:

- Chercher 2 groupes dont nous savons qu'ils n'ont pas bénéficié du programme, calculer une DD, et vérifier si l'effet estimé est nul.
- S'il n'est PAS nul, nous avons un problème...

## Intensité du programme de construction

Age en 1974	Forte	Faible	
12-17	8,02	9,40	
18-24	7,70	9.12	
<b>Différence</b>	<b>0,32</b>	<b>0,28</b>	<b>0,034 DD</b> <b>(0,098)</b>

## Seconde étape: estimons ceci à l'aide d'une régression

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

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  $j$

$\varepsilon_{ijk}$  = error term for person  $i$  in region  $j$  in cohort  $k$

# Troisième étape: utilisation d'information additionnelle

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

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

where  $P_j$  = the intensity of building activity in region j

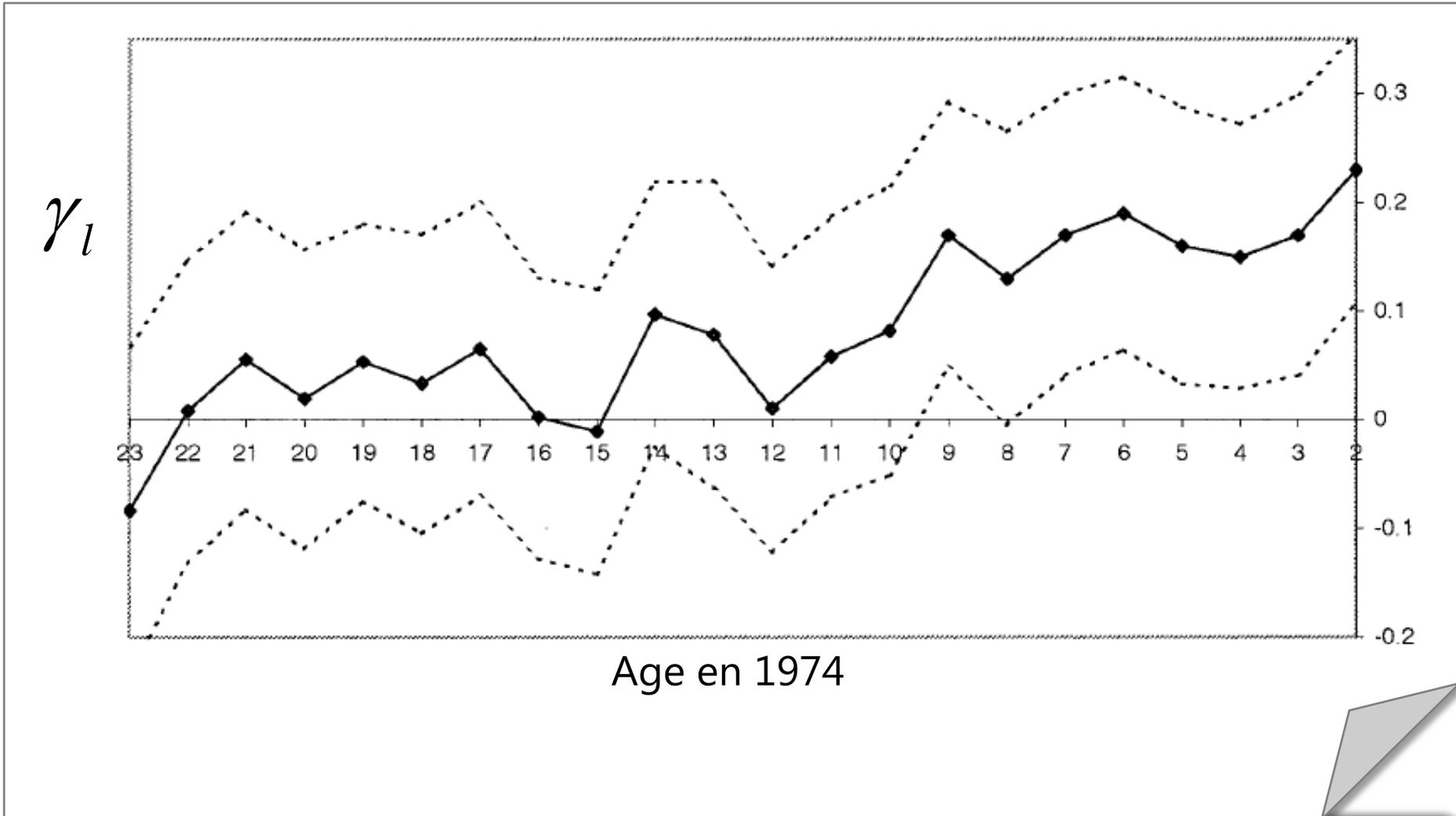
$C_j$  = a vector of regional characteristics

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

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

where  $d_i$  = a dummy variable for belonging to cohort i

# Effet du programme par cohorte



# Pour $y$ =variable dépendante =salaire

	Log(wages)		
	Level of program in region of birth		
	High (4)	Low (5)	Difference (6)
<i>Panel A: Experiment of Interest</i>			
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)
<i>Panel B: Control Experiment</i>			
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

- **Résultats:** Pour chaque école construite pour 1000 élèves;
  - Le niveau moyen d'éducation augmente de **0,12-0,19 années**
  - Les salaires moyens ont augmenté de **2,6 –5,4 %**
- Vérification de l'exactitude de l'estimation DD:
  - un placebo DD a donné un effet estimé **0**
  - Utiliser diverses spécifications alternatives
  - Vérifier que l'estimation d'impact pour chaque cohorte d'âge est plausible



## **Exemple 2**

# **Eau pour la vie: L'Impact de la privatisation du Service des Eaux sur la Mortalité Infantile**

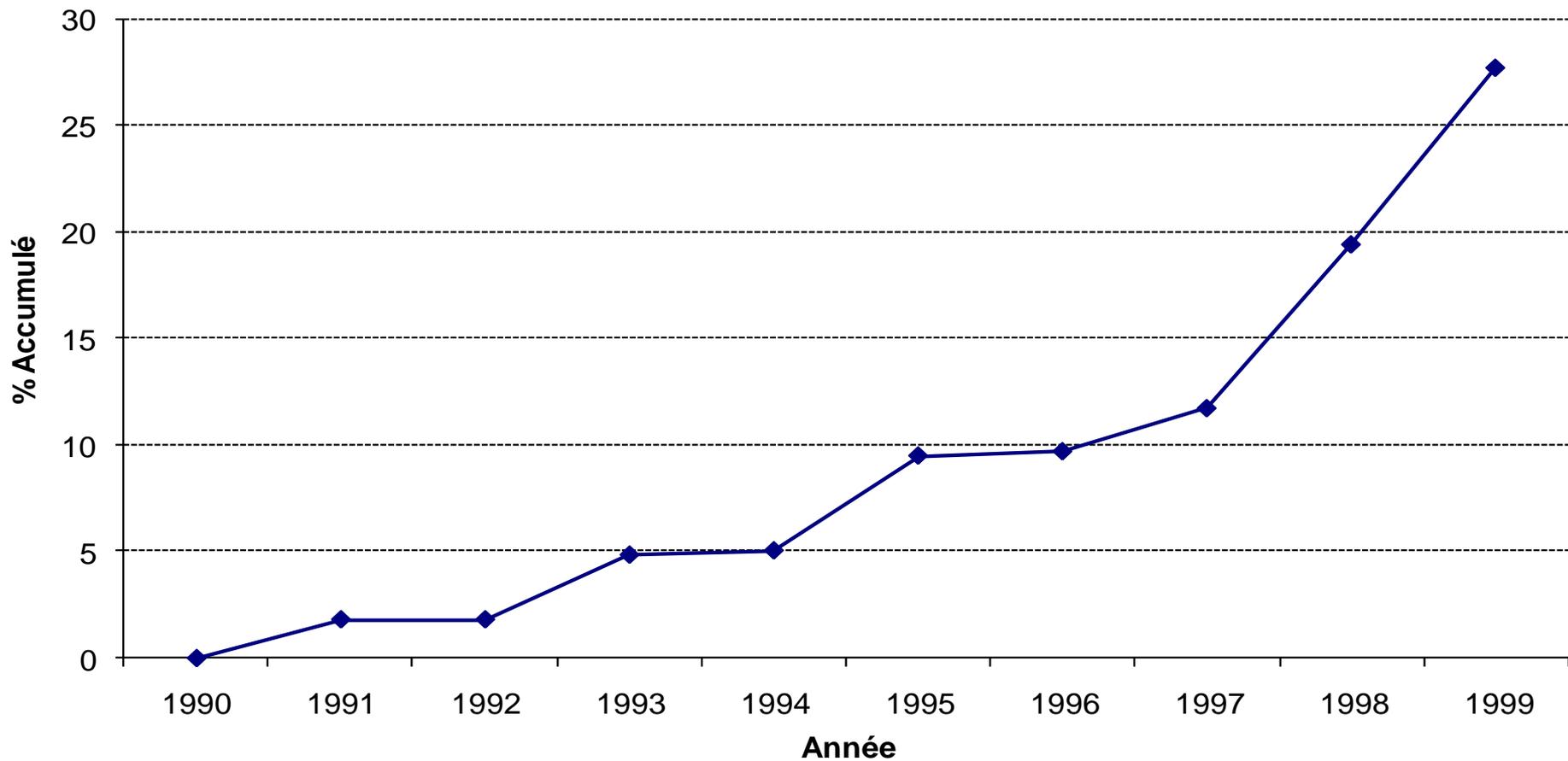
---

Sebastián Galiani, Universidad de San Andrés  
Paul Gertler, UC Berkeley  
Ernesto Schargrodsky, Universidad Torcuato Di Tella  
**JPE (2005)**

# Changements des modalités de l'approvisionnement en eau: 1990-1999

<b>Modalités de distribution</b>	<b>Nombre de municipalités</b>	<b>%</b>
Toujours publique	196	39,7%
Toujours une coopérative à but non-lucratif	143	28,9%
Convertie du publique au privée	138	27,9%
Toujours privée	1	0,2%
Pas d'information	16	3,2%
<b>Total</b>	<b>494</b>	<b>100%</b>

# Figure1: Pourcentage de municipalités avec des systèmes des eaux privés



# Utilisation de facteurs "externes" pour déterminer qui privatise

- Le parti politique qui gouvernait la municipalité
  - Les partis fédéral, péroniste et provincial: **ont permis la privatisation**
  - Le parti radical: **n'a pas permis la privatisation**
- La privatisation ou non-privatisation ne dépendait pas du parti au pouvoir:
  - Revenu, chômage, inégalité au niveau municipal
  - Récents changements dans les taux de mortalité infantile

# Régression

$$y_{it} = \alpha dI_{it} + \beta x_{it} + \lambda_t + \mu_i + \varepsilon_{it}$$

*where*

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

$dI_{it}$  = dummy variable that takes value 1 if

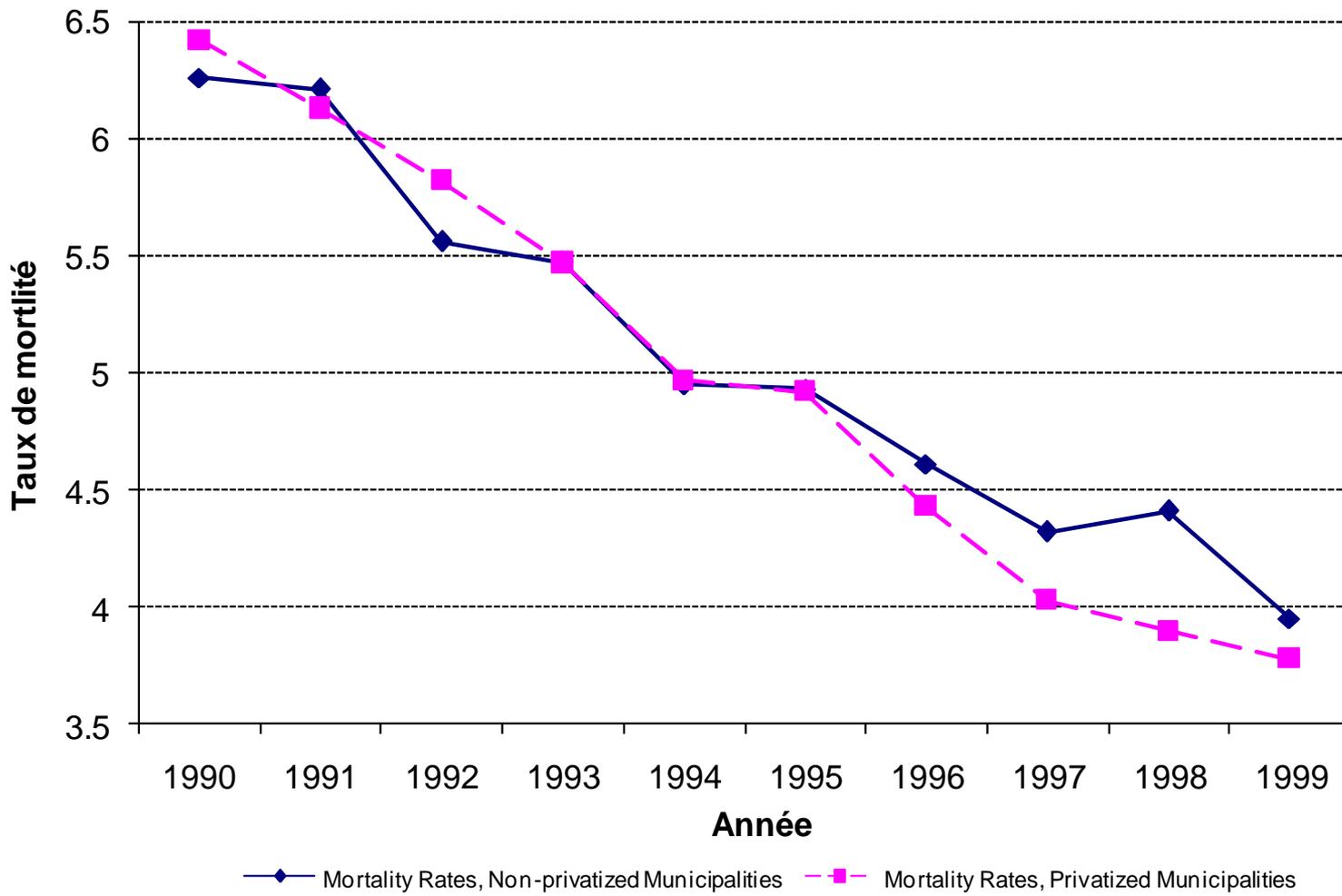
municipality  $i$  has private water provider in year  $t$

$x_{it}$  = vector of covariates

$\lambda_t$  = fixed time effect

$\mu_i$  = fixed municipality effect

**Figure 4: Evolution comparée du taux de mortalité pour les municipalités avec services des eaux privatisés et non-privatisés**



# Résultats DD : La privatisation a entraîné une réduction de la mortalité infantile

	Echantillon Complet			Support Commun			Ajusté
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
<b>Eau privée (=1)</b>	- 0.33 **	- 0.32 *	- 0.29 *	- 0.54 ***	- 0.54 ***	- 0.53 ***	- 0.60 ***
% $\Delta$ <i>en mortalité</i>	- 5.3 %	- 5.1 %	- 4.5 %	- 8.6 %	- 8.6 %	- 8.4 %	- 10.0 %
<b>PIB réel/tête</b>		0.01	0.01		0.01	0.01	
<b>Taux de Chômage</b>		- 0.56	-0.64		-0.78	-0.84	
Inégalité (Gini)		5.17 *	5.09 *		3.05	3.05	
Dépense publique/tête		- 0.03	- 0.04		-0.07 *	- 0.07 *	
Parti radical (=1)			0.48 *			0.17	
Parti péronist (=1)			- 0.20			- 0.17	
F-Stat Municipal FE	13.84***	11.92***	11.51***	10.39***	8.65***	8.32***	
F-Stat pour année FE	55.03***	19.88***	18.25***	52.25***	15.59***	12.98***	

# Contrôle de qualité sur les DD

1

Vérifier que les tendances en mortalité infantile étaient identiques dans les deux types de municipalités **AVANT** la privatisation

- Peut être fait en testant la même équation en utilisant seulement les années avant l'intervention – l'effet de traitement devrait être nul pour ces années
- Avons observé que nous ne pouvons pas rejeter l'hypothèse nulle de tendances égales entre les groupes de traitements et témoin pour les années précédant la privatisation

2

Vérifier que la privatisation affecte seulement la mortalité pour des raisons logiquement liées aux problèmes d'eau et d'assainissement.

Par exemple, la privatisation n'a aucun effet sur la proportion de décès dus à des maladies ou accidents cardio-vasculaires.

# Impact de la privatisation sur les décès de causes diverses

## DD sur support commun

	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)	...

# La privatisation a un effet plus important dans les municipalités pauvres et très pauvres que dans les municipalités non-pauvres

Municipalités	Mortalité moyenne pour 100, 1990	Impact estimé	% de variation de mortalité
Non-pauvres	5,15	0,105	...
Pauvres	7,15	-0,767***	-10,7%
Très pauvre	9,46	-2,214***	-23,4%

# Conclusion

En combinant des méthodes, nous avons conclu que:

- La privatisation est associée avec une réduction de la mortalité infantile de 5-7%.
- La réduction de la mortalité est:
  - Due à une réduction du nombre de morts causées par des maladies parasitaires et infectieuses.
  - N'est pas due à des variations des taux de mortalité pour des causes qui ne sont pas liées à l'eau et l'assainissement
- La plus grande réduction de la mortalité infantile est survenue dans les municipalités à faible revenu.

# Références

- 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 Schargrotsky (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.
- Gertler, Paul (2004): "Do Conditional Cash Transfers Improve Child Health? Evidence from PROGRESA's Control Randomized Experiment," *American Economic Review*, 94, pp. 336-41.



**Je vous remercie**



**Q&R**