Design of Two-Stage Experiments
with an Application to Spillovers in Tax Compliance

Guillermo Cruces
Nottingham & CEDLAS-UNLP

Dario Tortarolo
Nottingham & IFS

Gonzalo Vazquez-Bare
UC Santa Barbara

78th IIPF Annual Congress, Linz Austria

August 11, 2022

Motivation

- Social interactions among units targeted/non-targeted by policies are common
- This poses challenges for the design and evaluation of RCTs
- Early literature: ex-post analysis of untreated units [e.g., Miguel & Kremer'04]
- Moreover, in public finance, interference/spillovers among tax units is understudied

Contribution of our paper

- We make two contributions:

  1. **Methodological:** develop a framework for Partial Population experiments in samples where units are grouped into mutually exclusive clusters [e.g., Duflo & Saez, 2003]
  2. **Empirical:** large-scale RCT designed to capture spillovers in property tax compliance

- **Key:** experimental design with built-in spillovers—instead of as an afterthought

Design of Partial Population Experiments

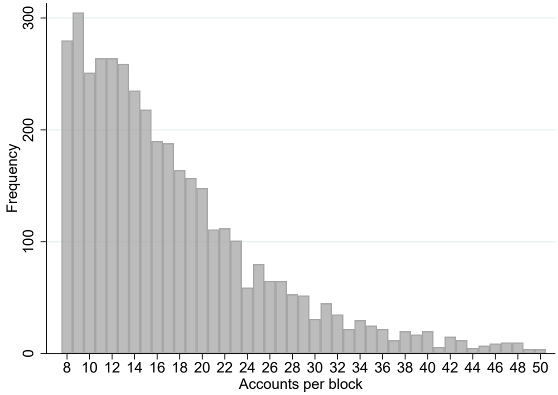
- Goal: estimate within-group spillovers (e.g., employees in firms)
- Partial Population (PP) experiments:
  - Groups randomly divided into different "intensities" (saturations)
  - Units within each group randomly assigned to treatment and control
- Intuition: compare units across groups with different treatment intensities

Challenges for Designing PP Experiments

- Two-stage design
- Multiple treatments
  - Compare units exposed to different treatment intensities
- Within-group correlations (clustering)
- Heterogeneity in group sizes
  - Group sizes tend to vary widely in practice (e.g., electoral precincts, schools)
  - Literature and software (e.g., Stata's **power**) make restrictive assumptions (e.g., equally-sized groups, N_T proportional to N_C...)

Group size heterogeneity is commonplace

Taxable properties per street-block in *Tres de Febrero*



Two practical implications:

1. $\mathbb{V}[\hat{\beta}]$ needs an adjustment term. Otherwise:
   $\Rightarrow$ Power is overestimated
   $\Rightarrow$ MDEs underestimated

2. Can affect the accuracy of the large sample normal approx
   $\Rightarrow$ Power calculations misleading

Methodological Contribution

- We derive an **asymptotic variance approximation** that allows for:
  - Multiple treatment intensities
  - General forms of intracluster correlation and heteroskedasticity
  - Cluster size heterogeneity

Methodological Contribution

- We derive an **asymptotic variance approximation** that allows for:
  - Multiple treatment intensities
  - General forms of intracluster correlation and heteroskedasticity
  - Cluster size heterogeneity
- These factors affect $\mathbb{V}[\hat{\beta}]$... but have been overlooked by the literature
  - Using data from existing studies we show that corrected MDEs can be 20% to 30% larger!

Methodological Contribution

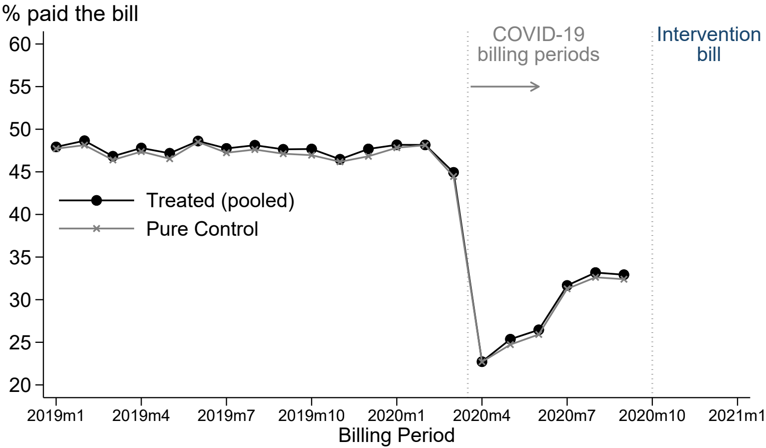
- We derive an **asymptotic variance approximation** that allows for:
  - Multiple treatment intensities
  - General forms of intracluster correlation and heteroskedasticity
  - Cluster size heterogeneity
- These factors affect $\mathbb{V}[\hat{\beta}]$... but have been overlooked by the literature
  - Using data from existing studies we show that corrected MDEs can be 20% to 30% larger!
- Our formula nests other cases [e.g., Duflo et al, 2007; Hirano & Hahn, 2010; Baird et al, 2018] and can be applied in a wide range of designs (e.g., PP, clustered, stratified...)

Spillovers in Property Tax Compliance

- Ample evidence on **direct effects** of tax compliance interventions [Antinyan & Asatryan'19]
- We know little about **interference among tax units**
- We teamed up with a large municipality in Argentina (Tres de Febrero)
  Neighbors must pay a monthly bill on their real estate (∼70k units)
- Context: arrears mainly due to COVID-19 lockdown
  ⇒ we devised an intervention in Oct'2020, when mobility restrictions started to ease

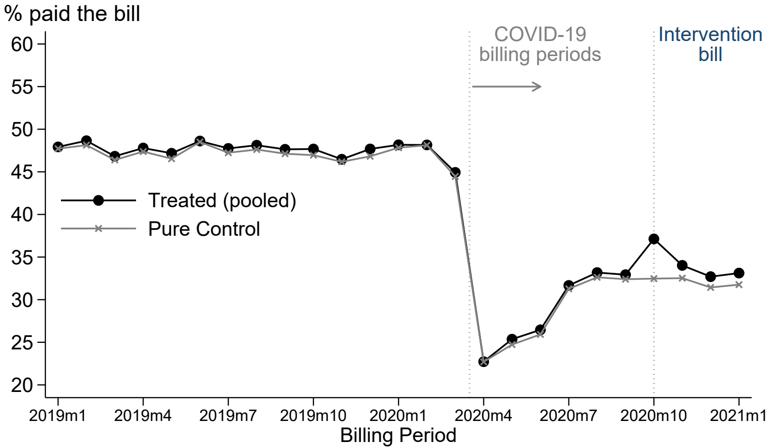
Timely payments of **treated** units increased due to our intervention

What about untreated neighbors in treated blocks?



Timely payments of **treated** units increased due to our intervention

What about untreated neighbors in treated blocks?



What exactly did we do?

- We sent ∼25,000 personalized letters to randomly selected accounts with information about the Oct'20 bill, due dates, past due debt, and payment methods
- Critically, we designed the experiment using our framework to maximize the chance of capturing spillover effects

Design and example of the letter


Tus impuestos municipales ahora vienen en la BOLETA DIGITAL

ID: XXXXX

TITULAR:
DIRECCIÓN: CAP. MADARIAGA N°
LOCALIDAD: 11 de Septiembre
C.P.: 1657
PARTIDA: XXXXXX/7

Te queremos contar que ahora en Tres de Febrero tu boleta municipal de la Tasa por Servicios Generales (TSG) es 100% digital. O sea, ya no se usa más el papel. Podés acceder a ella y pagarla desde el celular o la computadora. De esta manera, nos cuidamos entre todos al reducir la circulación y también cuidamos el medio ambiente. Es una situación difícil y te agradecemos el esfuerzo que estás haciendo para estar al día con tus impuestos, porque eso se transforma directamente en obras y servicios que no paran en tu barrio. Te informamos el estado de tu cuenta y te mostramos lo fácil que es:

PARTIDA: XXXXX/7	
Cuota 10 vencimiento 10 de octubre 2020:	347,29
Deuda año en curso*:	1.702,58
Deuda años anteriores*:	289,54

* Al 15/09/2020

¿CÓMO PAGAR?

Ingresando a **tasas.tresdefebrero.gov.ar** completá los datos:

CLICKEÁ ESTE BOTÓN y recibíla todos los meses en tu mail.

También podés entrar a **miboleta.tresdefebrero.gov.ar**

1) Podés pagar **ONLINE** con
mercado pago → En el momento desde nuestra web.
PagoMisCuentas, LINK → Obteniendo el **código de pago electrónico** para pagar desde **la plataforma de tu banco o cajero automático.**

2) Podés pagar en **EFECTIVO** en
Pago Fácil, rapipago → **DESCARGALA** o llevá tu **NÚMERO DE PARTIDA.**

Por dudas comunicate con nosotros a **reclamos.mistasas@tresdefebrero.gov.ar**
Si esta carta llegó por error a tu domicilio, informanos en ese mismo correo electrónico

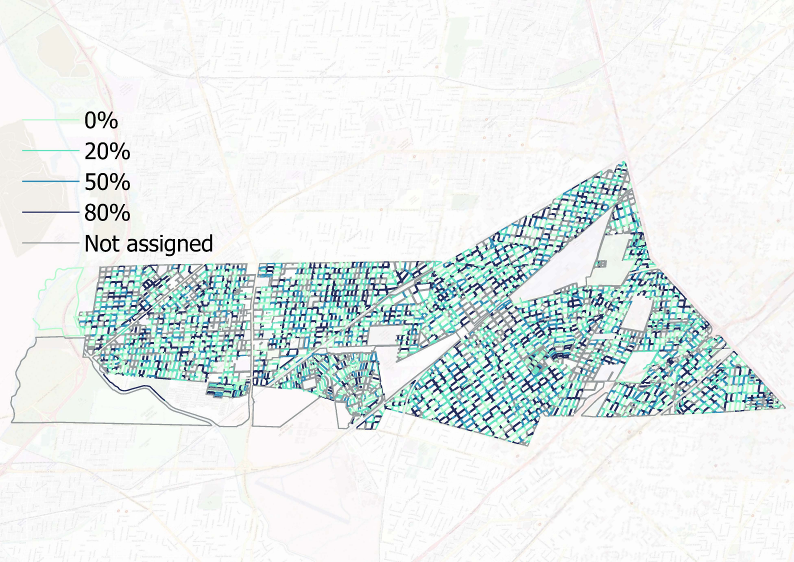
¡Muchas gracias!

Randomization in 2 stages:

1) Randomly divide 3,982 street-blocks into 4 categories with $\neq$ treatment intensity:
   - $T_g = 0$: pure controls
   - $T_g = 1$: blocks with **20%** of properties treated
   - $T_g = 2$: blocks with **50%** of properties treated
   - $T_g = 3$: blocks with **80%** of properties treated
2) Within treated street-blocks, randomly assign accounts to **treated** (letter) or **untreated**

▸ KITCHEN: Treatment Assignment, Power, MDE

Map of the municipality & the experimental design



Empirical strategy

- In multi-treatment experiments, effects on outcome Y_{ig} are commonly estimated through saturated OLS regressions:

$$Y_{ig} = \alpha + \sum_{t=1}^{3} \beta_{0t} \mathbb{1}(T_g = t)(1 - D_{ig}) + \sum_{t=1}^{3} \beta_{1t} \mathbb{1}(T_g = t) D_{ig} + \varepsilon_{ig}$$

where

$$\beta_{0t} = \mathbb{E}[Y_{ig} | D_{ig} = 0, T_g = t] - \mathbb{E}[Y_{ig} | D_{ig} = 0, T_g = 0]$$

Spillover effects on untreated units

and

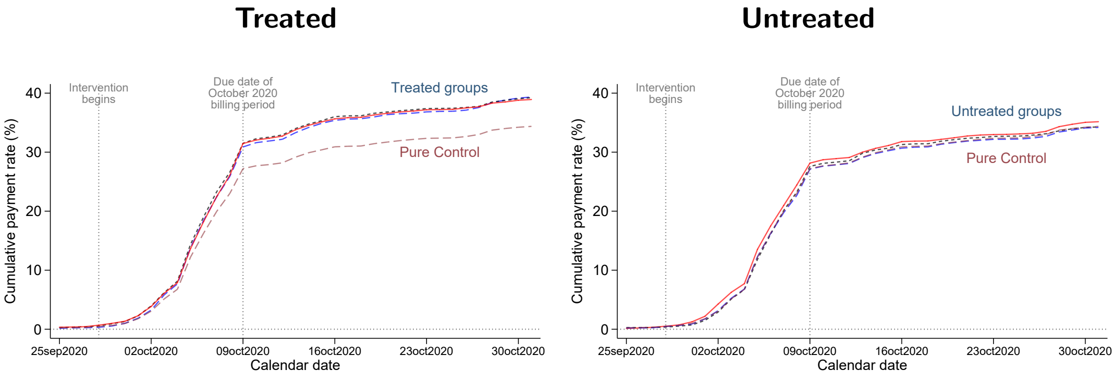
$$\beta_{1t} = \mathbb{E}[Y_{ig} | D_{ig} = 1, T_g = t] - \mathbb{E}[Y_{ig} | D_{ig} = 0, T_g = 0]$$

Total effects on treated units

- We allow ε_{ig} to be correlated within blocks and use a cluster-robust variance estimator

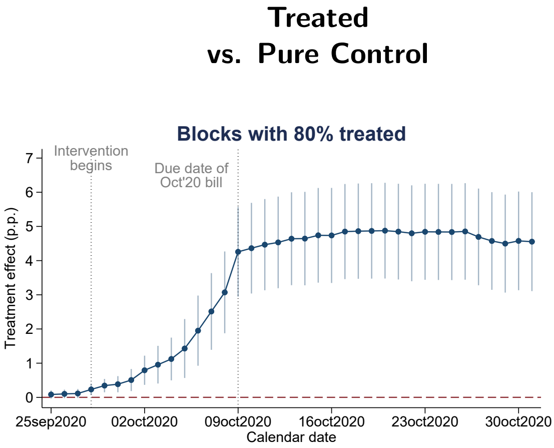
Daily payment rates of the Oct'2020 bill

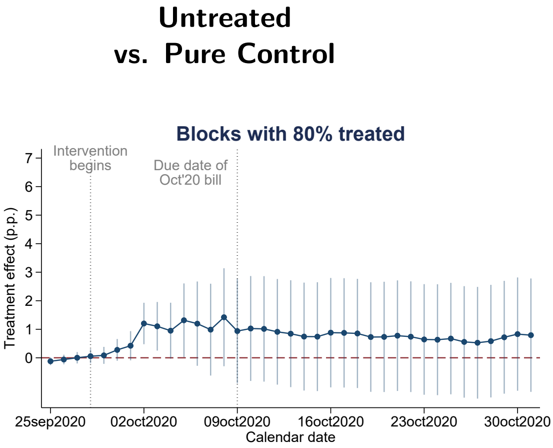
Difference



Daily payment rates of the Oct'2020 bill

Blocks with 80% treated

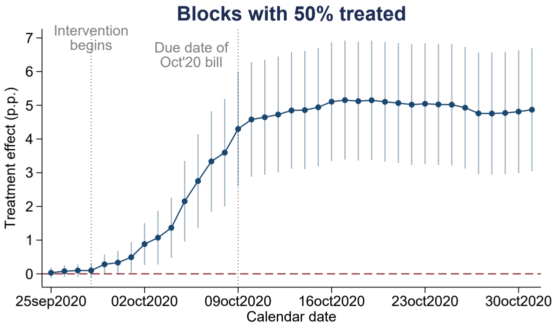




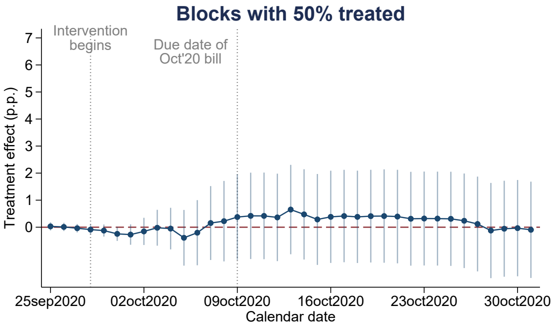
Daily payment rates of the Oct'2020 bill

Blocks with 50% treated

Treated vs. Pure Control



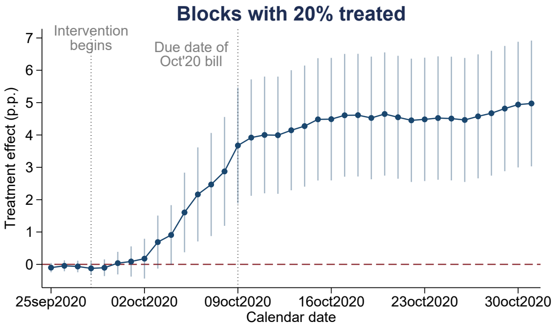
Untreated vs. Pure Control



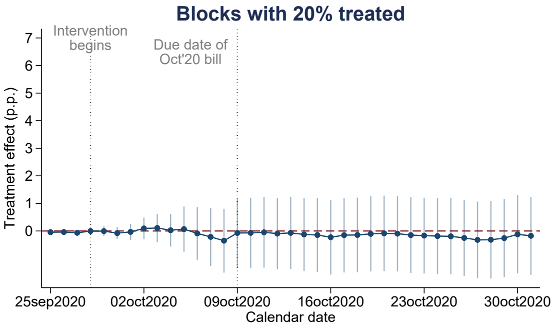
Daily payment rates of the Oct'2020 bill

Blocks with 20% treated

Treated vs. Pure Control



Untreated vs. Pure Control



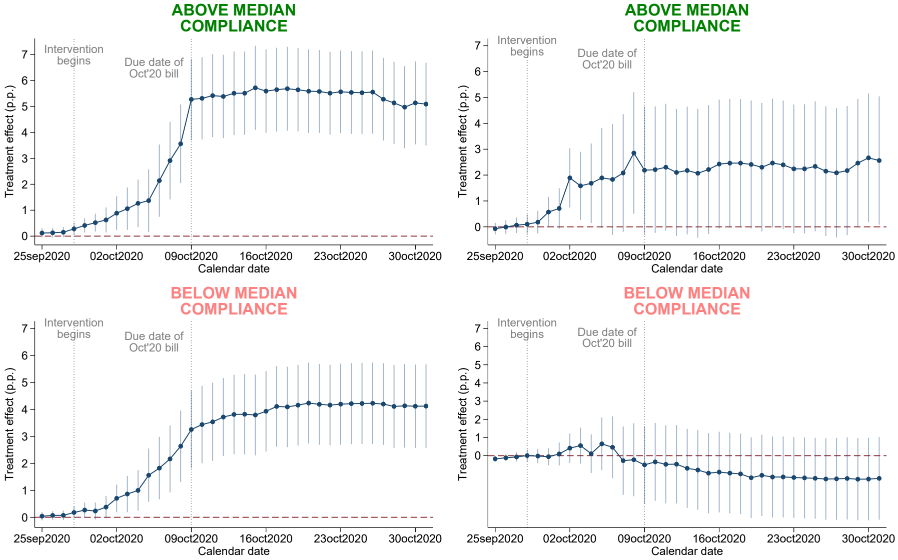
Total and spillover effects for bill payments

▸ Placebo Figs ▸ Table het

Dependent variable:	**Placebo bill:**	**Intervention bill:**	
Pr(pay the bill)	Sep'20	Early	By Oct 31
	(1)	(2)	(3)
A. Blocks with 80% treated			
Treated	0.12	0.96***	4.55***
	(0.69)	(0.28)	(0.74)
Untreated	-0.30	1.10**	0.79
	(0.95)	(0.43)	(1.01)
B. Blocks with 50% treated			
Treated	0.76	1.07***	4.87***
	(0.88)	(0.41)	(0.93)
Untreated	0.26	-0.02	-0.10
	(0.88)	(0.34)	(0.91)
C. Blocks with 20% treated			
Treated	0.85	0.69*	4.97***
	(0.93)	(0.42)	(0.99)
Untreated	0.07	0.11	-0.18
	(0.68)	(0.26)	(0.72)
Payment Rate of Pure Control	29.70	5.15	34.37
Observations	68,806	68,806	68,806
Number of clusters (blocks)	3,981	3,981	3,981

Above/Below 2019 compliance – Blocks 80%

▸ Distribution



Conclusions

- General framework to conduct experiments to estimate spillovers
  - Allows for group size heterogeneity, heteroskedasticity, ICC, ...
  - Derive optimal choice of group-level probabilities
- Application to property tax compliance in Argentina
  - Our letters increased payment rates of both **treated** and **untreated** neighbors
  - **Direct effects:** 4.5 p.p. (16% of the payment rate in pure control blocks)
  - **Spillover effects:** more modest in magnitude, precisely estimated
    Larger in "good payer" blocks with 80% treated

Thank you!

Dario Tortarolo

Setup ▸ Back

- Groups $g = 1, \ldots, G$ with units $i = 1, \ldots, n_g$
- Total sample size $n = \sum_g n_g$
- Multi-valued unit-level treatment $A_{ig} = \{a_0, a_1, a_2, \ldots, a_K\}$
- Assignment probabilities:

$$\pi_g(a_k) = \mathbb{P}_g[A_{ig} = a_k], \quad \pi_g(a_k, a_l) = \mathbb{P}_g[A_{ig} = a_k, A_{jg} = a_l]$$

- Moments:

$$\sigma^2(a_k) = \mathbb{V}[Y_{ig} | A_{ig} = a_k]$$
$$\rho(a_k, a_l) = cor(Y_{ig}, Y_{jg} | A_{ig} = a_k, A_{jg} = a_l)$$

Setup ▸ Back

- Empirical strategy: estimate

$$Y_{ig} = \alpha + \sum_{k=1}^{K} \beta_k \mathbb{1}(A_{ig} = a_k) + \varepsilon_{ig}$$

by OLS, where

$$\beta_k = \mathbb{E}[Y_{ig}|A_{ig} = a_k] - \mathbb{E}[Y_{ig}|A_{ig} = a_0]$$

and

$$\hat{\beta}_k = \bar{Y}_k - \bar{Y}_0$$

- Error terms correlated within groups

Main Result

Back

Asymptotic Approximation

Under regularity conditions, if

$$\max_{g \leq G} \frac{n_g^2}{n} \to 0, \quad \frac{\sum_{g=1}^{G} n_g^4}{n^2} \leq C < \infty,$$

then $\hat{\beta}_k \overset{a}{\sim} \mathcal{N}(\beta_k, V_k)$ where:

$$\begin{aligned} V_k = {} & \frac{\sigma^2(a_k)}{\sum_g n_g \pi_g(a_k)} \left\{ 1 + \rho(a_k, a_k) \frac{\sum_g n_g(n_g - 1)\pi_g(a_k, a_k)}{\sum_g n_g \pi_g(a_k)} \right\} \\ & + \frac{\sigma^2(a_0)}{\sum_g n_g \pi_g(a_0)} \left\{ 1 + \rho(a_0, a_0) \frac{\sum_g n_g(n_g - 1)\pi_g(a_0, a_0)}{\sum_g n_g \pi_g(a_0)} \right\} \\ & - 2\sigma(a_k)\sigma(a_0)\rho(a_k, a_0) \frac{\sum_g n_g(n_g - 1)\pi_g(a_k, a_0)}{\sum_g n_g \pi_g(a_k) \sum_g n_g \pi_g(a_0)} \end{aligned}$$

Main Result: Intuition ▸ Back

- The formula is an explicit version of

$$\mathbb{V}[\bar{Y}_k - \bar{Y}_0] = \mathbb{V}[\bar{Y}_k] + \mathbb{V}[\bar{Y}_0] - 2\mathbb{C}\text{ov}(\bar{Y}_k, \bar{Y}_0)$$

  allowing for:

  - Intracluster correlation
  - Heteroskedasticity
  - Unequal probabilities between groups
  - Group size heterogeneity

Main Result: Intuition

Back

- Condition:

$$\max_{g \leq G} \frac{n_g^2}{n} \to 0$$

restricts the relative size of the largest group

  - Ensures that no group "dominates" the sample

- Condition:

$$\frac{\sum_{g=1}^{G} n_g^4}{n^2} \leq C < \infty$$

bounds the fourth moment of the distribution

  - Rules out fat tails (outliers)

Why is group size heterogeneity important? ▸ Back

- It affects the variance of estimators

$$\mathbb{V}[\hat{\beta}] \approx \sigma^2[1 + \rho(ICC, \bar{n}, Var(n_g))]$$

  - Ignoring $Var(n_g)$ underestimates $\mathbb{V}[\hat{\beta}]$ $\Rightarrow$ overestimates power

- It affects inference and power calculations

  - Normal approx may be inaccurate if groups are "too heterogeneous"
  - Carter et al (2017), Djogbenou et al (2019), Hansen and Lee (2019)

Illustration using data from four published studies ▸ Back

- Ichino & Schundeln (2012), Haushofer & Shapiro (2016), Gine & Mansuri (2018) and Imai, Jiang & Malani (2021)
- In common: clusters randomly assigned to $\neq$ treatment intensities to estimate spillovers
- We calculate standard errors and MDEs accounting for cluster size heterogeneity using the median values of number of groups, $G = 95$, average group size, $\bar{n} = 23.3$, and group size SD, $sd(n_g) = 15.2$.
- We compare "adjusted" standard errors and MDEs with "undajusted" ones—those obtained if (incorrectly) ignoring cluster size heterogeneity.

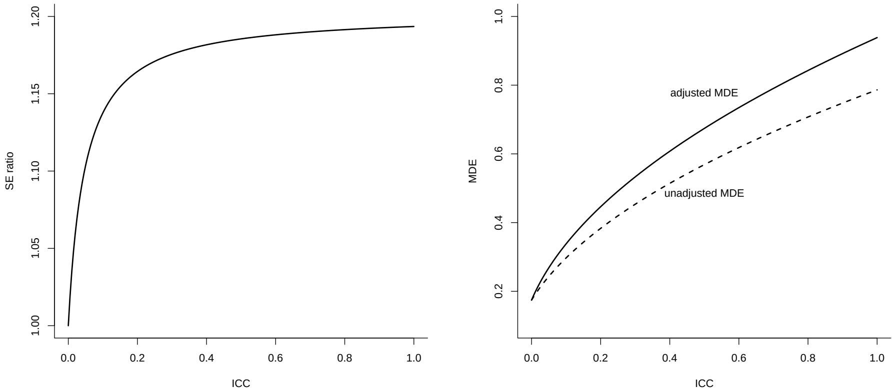
Table A4: Sample sizes in existing literature

	Sample size	No. of groups	Ave. group size	Sd. group size
Giné and Mansuri (2018)	2,736	67	39.4	16.7
Haushofer and Shapiro (2016)	1,440	123	23.4	14.8
Ichino and Schündeln (2012)	868	39	22.3	9.6
Imai, Jiang and Malani (2021)	10,030	434	23.1	15.5
Mean	3,769	165.8	27.05	14.2
Median	2,088	95	23.3	15.2

Table A5: Numerical results

	Standard error			MDE		
	Adj.	Unadj.	Ratio	Adj.	Unadj.	Ratio
$\rho = 0.1$						
GM	0.1262	0.1181	1.0687	0.3536	0.3308	1.0689
HS	0.1053	0.0932	1.1307	0.2951	0.2610	1.1307
IS	0.1768	0.1667	1.0608	0.4954	0.4670	1.0608
IJM	0.0569	0.0497	1.1453	0.1595	0.1393	1.1450
$\rho = 0.5$						
GM	0.2593	0.2393	1.0835	0.7265	0.6705	1.0835
HS	0.2098	0.1783	1.1761	0.5877	0.4997	1.1761
IS	0.3437	0.3171	1.0840	0.9630	0.8884	1.0840
IJM	0.1136	0.0950	1.1961	0.3183	0.2661	1.1962
$\rho = 0.8$						
GM	0.3252	0.2997	1.0851	0.9112	0.8397	1.0851
HS	0.2622	0.2218	1.1818	0.7345	0.6215	1.1818
IS	0.4284	0.3941	1.0869	1.2002	1.1042	1.0869
IJM	0.1420	0.1181	1.2024	0.3979	0.3309	1.2025

Figure: Adjusted and unadjusted standard errors and MDEs ▸ Back



(a) **Ratio of adjusted to unadjusted SEs** (b) **MDE**

Notes: Adjusted magnitudes account for group size variability. Unadjusted magnitudes assume no group size variability, i.e. zero variance of group size.

PP Experiment: Design

Back

- We randomly divide city street-blocks into four categories:
  - $T_g = 0$: pure controls with prob q_0
  - $T_g = 1$: blocks with **20%** of properties treated with prob q_1
  - $T_g = 2$: blocks with **50%** of properties treated with prob q_2
  - $T_g = 3$: blocks with **80%** of properties treated with prob q_3
- Goal: choose $q_t = \mathbb{P}[T_g = t]$, $t = 0, 1, 2, 3$
- We set up a system of eqs to incorporate constraints on $\{q_t\}_t$

Constrained choice of $\{q_t\}_t$ Back

- Choose q_1, q_2, q_3, with $q_0 = 1 - q_1 - q_2 - q_3$
- Total sample size $n = \sum_g n_g$
- The total number of letters sent (L) should equal the expected number of treated:

$$L = n(0.2q_1 + 0.5q_2 + 0.8q_3)$$

- Categories $T_g = 1$ and $T_g = 3$ are symmetric, so $q_1 = q_3$
- This leaves two probabilities to be determined: q_2 and q_3
- Idea: balance variances across assignments

Constrained choice of $\{q_t\}_t$ ▸ Back

- The "hardest" effects (smallest cells) to estimate are θ_3 and τ_1
  - Spillover effect in 80% groups and direct effect in 20% groups
- We choose q_2 and q_3 by setting:

$$V(\hat{\theta}_3) = V(\hat{\theta}_2)$$

  based on our variance approximation
- We assume $\sigma^2(0,2) \approx \sigma^2(0,3) = \sigma^2$ and $\rho \approx 0.1$

Power calculations

▸ Back

Probabilities:

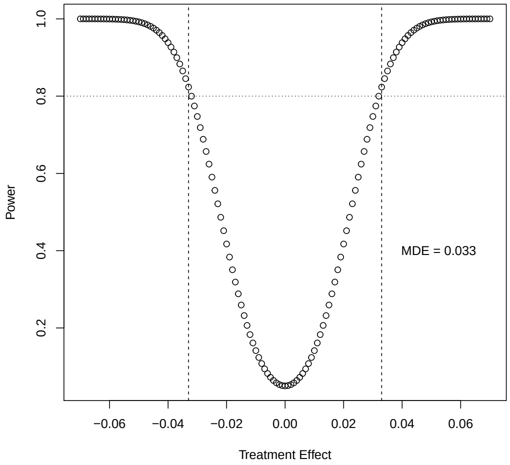
	Prob
q_0	0.273
q_1	0.302
q_2	0.121
q_3	0.302

Expected sample sizes:

	Blocks	Control Obs	Treated Obs
$T_g = 0$	$1,087$	$18,870$	0
$T_g = 1$	$1,205$	$16,530$	$4,236$
$T_g = 2$	483	$4,192$	$4,184$
$T_g = 3$	$1,205$	$4,024$	$16,772$
Total	$3,980$	$43,616$	$25,192$

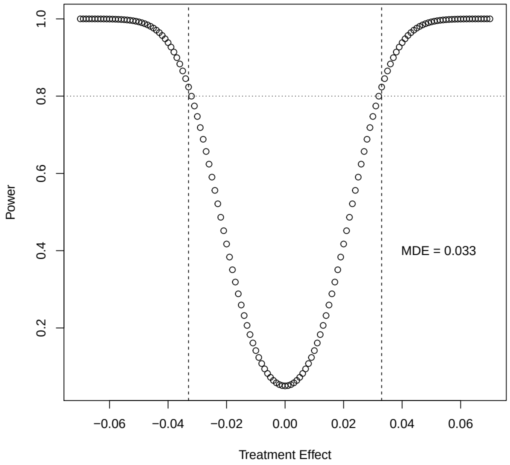
Power functions and MDE

Back



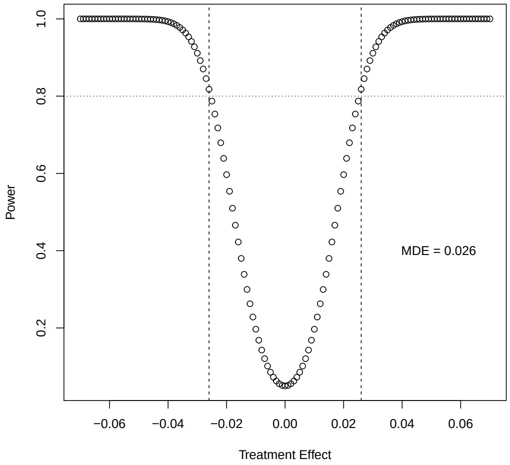
Power function for $(d, t) = (0, 3)$ and $(d, t) = (1, 1)$

Power functions and MDE ▸ Back



Power function for $(d, t) = (0, 2)$ and $(d, t) = (1, 2)$

Power functions and MDE ▸ Back



Power function for $(d, t) = (1, 3)$ and $(d, t) = (0, 1)$

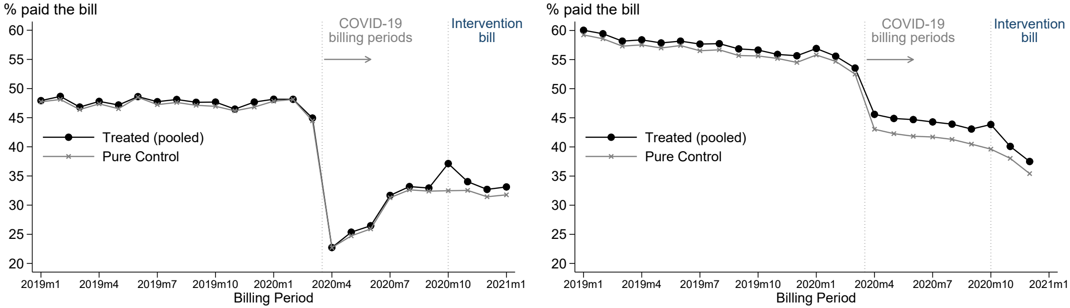
Balance

	Property Value	Front Metres	House type	Tenant Male	Tenant Age	Bill amount	N Bills paid 2019	Digital payment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Blocks with 80% treated:								
Treated	0.01	−8.27	−0.00	−0.00	−0.14	2.81	0.05	−0.00
	(0.02)	(17.77)	(0.00)	(0.01)	(0.40)	(7.81)	(0.09)	(0.01)
Untreated	0.00	−1.76	0.00	0.00	−0.53	6.27	−0.06	−0.00
	(0.02)	(20.70)	(0.01)	(0.01)	(0.53)	(12.95)	(0.12)	(0.01)
B. Blocks with 50% treated:								
Treated	0.01	12.65	−0.00	−0.00	−0.47	1.16	0.03	0.00
	(0.02)	(20.38)	(0.01)	(0.01)	(0.50)	(9.21)	(0.11)	(0.01)
Untreated	0.01	25.30	−0.00	−0.00	−0.42	1.88	0.02	0.01
	(0.02)	(20.66)	(0.01)	(0.01)	(0.48)	(9.66)	(0.11)	(0.01)
C. Blocks with 20% treated:								
Treated	0.02	32.57*	−0.01	0.01	0.10	5.94	0.07	−0.01
	(0.02)	(16.79)	(0.01)	(0.01)	(0.54)	(9.55)	(0.12)	(0.01)
Untreated	0.02	19.14	−0.01	−0.01	0.12	1.32	0.00	0.00
	(0.02)	(14.05)	(0.00)	(0.01)	(0.40)	(7.77)	(0.09)	(0.01)
Mean Pure Control	13.64	841.50	0.91	0.62	19.15	368.66	6.71	0.35
Observations	64,932	68,808	68,808	46,419	52,714	68,808	68,808	38,112
Number of clusters	3,979	3,981	3,981	3,973	3,976	3,981	3,981	3,968

Direct effect on treated neighbors

Timely payments (left) and w/past-due payments (right)

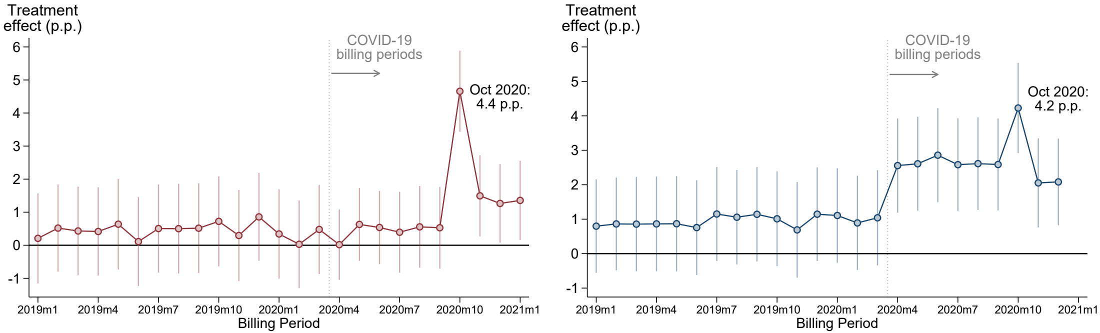
(a) Payment rates in levels



Direct effect on treated neighbors

Timely payments (left) and w/past-due payments (right)

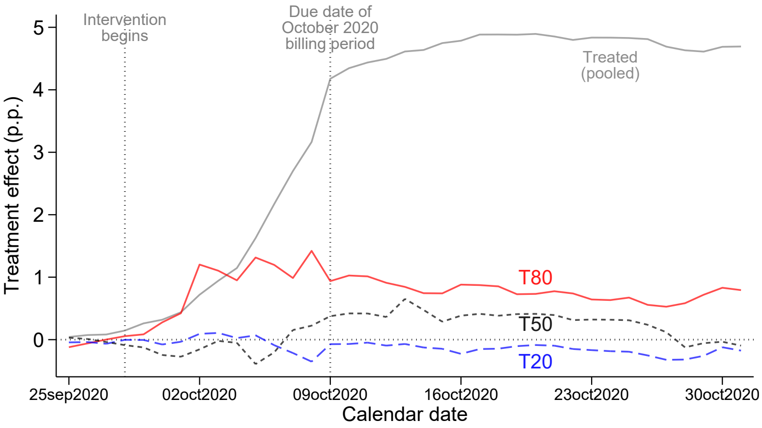
(b) Difference relative to pure control group



Payment rate of the Oct'2020 bill

▸ Back

Figure: Difference relative to pure control group

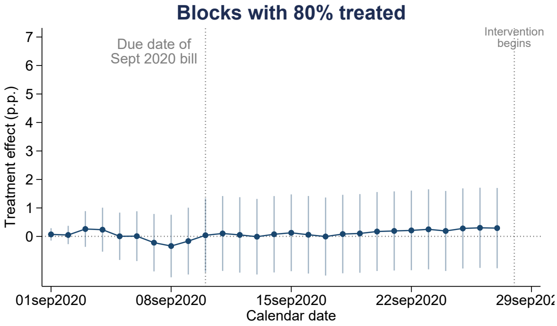


Daily payment rates of the **Sep'2020** bill [PLACEBO]

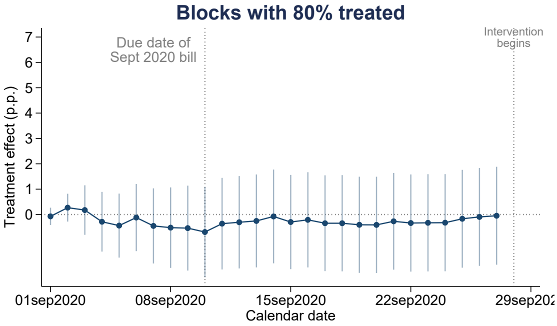
Blocks with 80% treated

▸ Back

Treated vs. Pure Control



Untreated vs. Pure Control

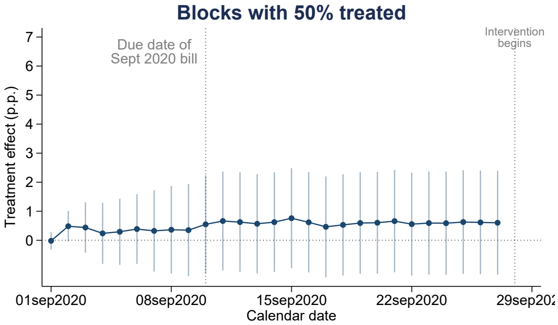


Daily payment rates of the **Sep'2020** bill [PLACEBO]

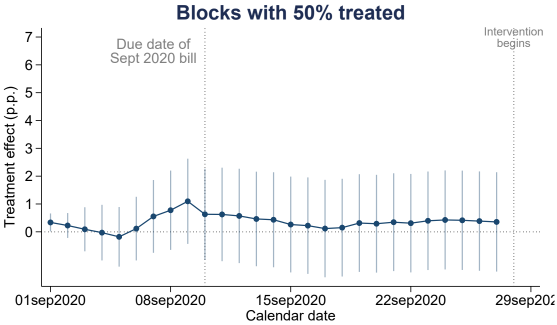
Blocks with 50% treated

Back

Treated vs. Pure Control



Untreated vs. Pure Control

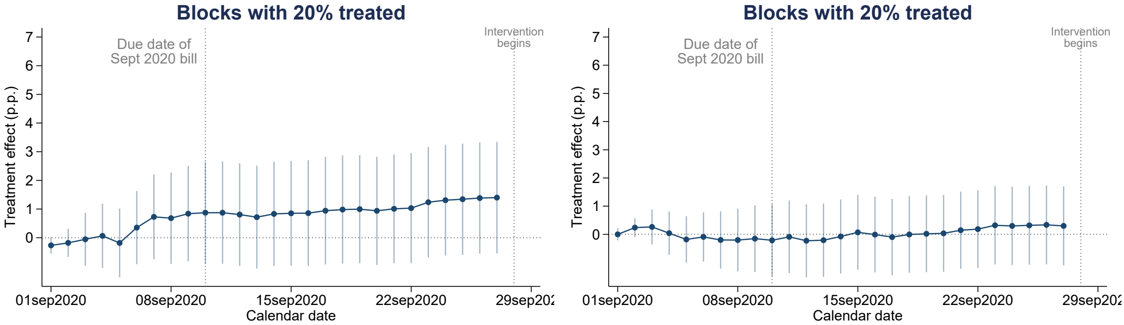


Daily payment rates of the **Sep'2020** bill [PLACEBO]

Blocks with 20% treated

▸ Back

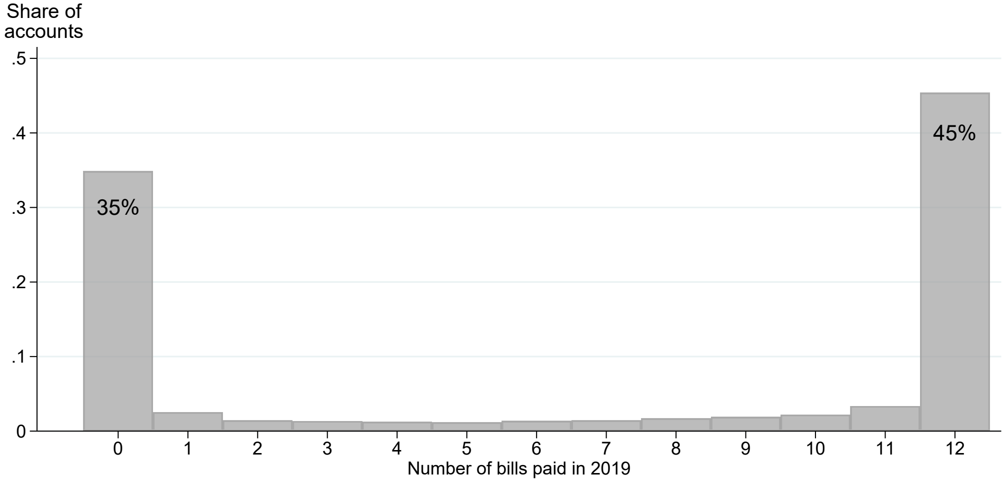
Treated vs. Pure Control	**Untreated vs. Pure Control**



Tax compliance in 2019: always payers and never payers

Stylized fact in property taxation

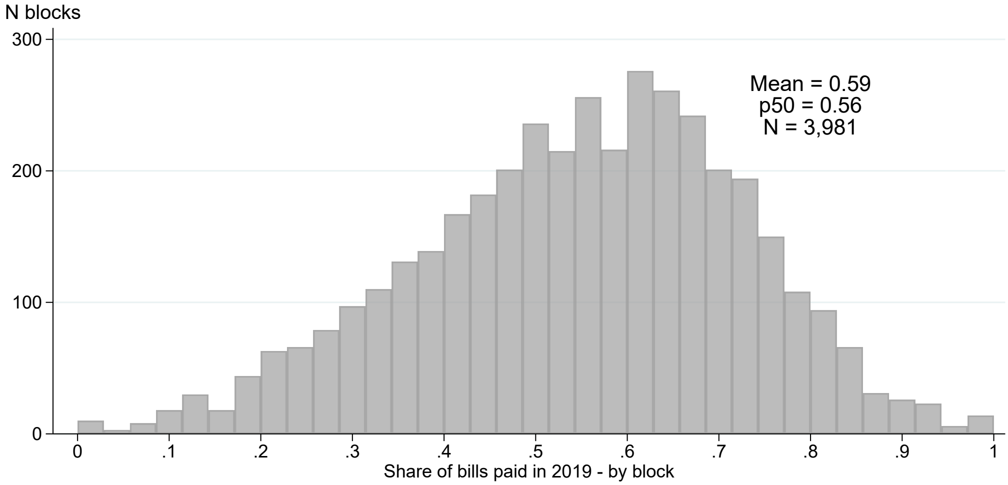
▸ Back



Tax compliance in 2019: always payers and never payers

Stylized fact in property taxation

▸ Back



Heterogeneous Effects (pre-registered!)

▸ Back

	Placebo bill:		**Intervention bill:**			
	Sep'20		Early		By Oct 31	
	Below Median	Above Median	Below Median	Above Median	Below Median	Above Median
	(1)	(2)	(3)	(4)	(5)	(6)
A. Blocks with 80% treated						
Treated	0.10	0.28	0.86**	1.06**	4.12***	5.09***
	(0.73)	(0.81)	(0.34)	(0.42)	(0.79)	(0.81)
Untreated	-1.55	0.78	0.55	1.58**	-1.25	2.56**
	(1.09)	(1.24)	(0.50)	(0.67)	(1.16)	(1.27)
B. Blocks with 50% treated						
Treated	1.54	0.69	1.24**	1.02	4.81***	5.67***
	(0.99)	(1.12)	(0.50)	(0.62)	(1.07)	(1.08)
Untreated	0.81	0.36	0.10	-0.03	1.34	-0.76
	(0.94)	(1.15)	(0.43)	(0.50)	(1.00)	(1.14)
C. Blocks with 20% treated						
Treated	1.32	0.27	0.85*	0.52	5.41***	4.40***
	(1.11)	(1.24)	(0.52)	(0.63)	(1.21)	(1.27)
Untreated	0.27	-0.32	0.68**	-0.42	0.61	-1.09
	(0.72)	(0.80)	(0.33)	(0.38)	(0.77)	(0.82)
Payment Rate of Pure Control	20.05	38.19	3.63	6.49	23.53	43.91
Observations	32,361	36,445	32,361	36,445	32,361	36,445
Number of clusters (blocks)	2,013	1,968	2,013	1,968	2,013	1,968