The random filter identification strategy

Pierce Donovan

Department of Economics, University of Nevada, Reno

University of Nevada, Reno
February 2024

Toy example: Does financial inclusion lead to increased savings?

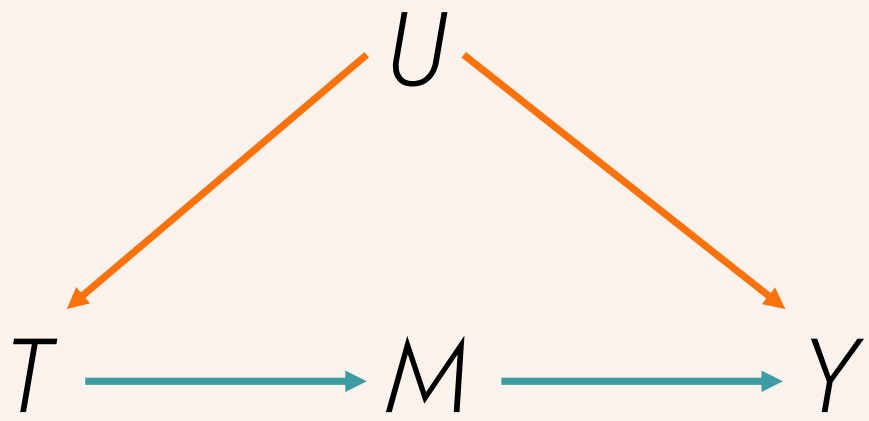
Setting: we want to measure the effectiveness of a new program that provides low fee savings accounts for El Salvadorans.

Identification problem: (selection) those who sign up are likely to be better off financially even without this particular account.

Mechanism: this treatment will only have an impact if the customer is able to interact with the bank.*

*For the sake of this example, imagine that customers are occasionally—and randomly—unable to deposit, withdraw, or transfer their funds due to an imperfect rollout of this banking intervention.

We can simulate a dataset with a few lines of code. The treatment effect is a $200 increase in yearly savings.



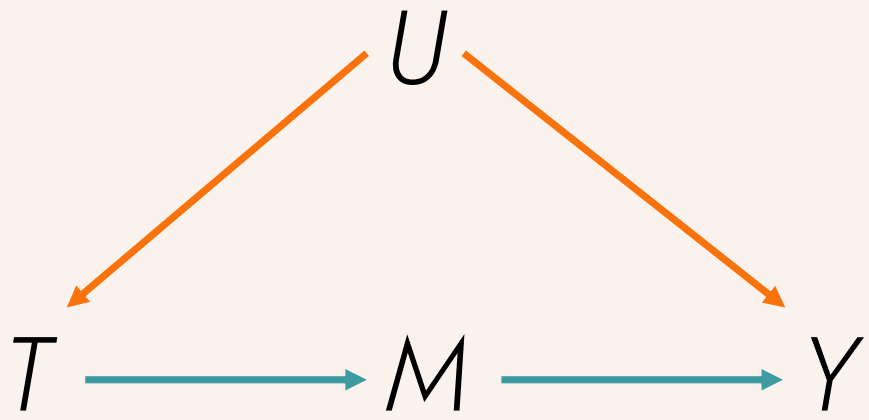
$$U \sim Bern(0.5) \quad \text{financial savvy}$$

$$T \sim Bern(0.25 + 0.5 \cdot U) \quad \text{savings account}$$

$$M \sim Bern(0.67 \cdot T) \quad \text{usable account}$$

$$Y \sim N(500 + 500 \cdot U + 300 \cdot M, 25^2) \quad \text{yearly savings}$$

The data generating process for this story suggests that the treatment effect is facilitated by an exogenous mediator M.



There are three properties of M that stand out to me:

1. M facilitates the full effect of T on Y
2. M is independent of U, thus $T \to M$ seems identifiable
3. $M \to Y$ also seems identifiable after controlling for T

The exogenous mediator allows us to separately estimate the impacts of treatment on the mediator and mediator on the outcome, scale the former by the latter, and avoid confounding factors.

Simplest implementation (seemingly unrelated regression):

$$SUR: \begin{cases} M_i = \pi + \lambda \cdot T_i + \vartheta_i \\ Y_i = \alpha + \gamma \cdot M_i + \delta \cdot T_i + \varepsilon_i \end{cases}$$

Then $\hat{\beta}_{RF} = \hat{\lambda} \cdot \hat{\gamma}$. (You can bootstrap or delta-method the error.)

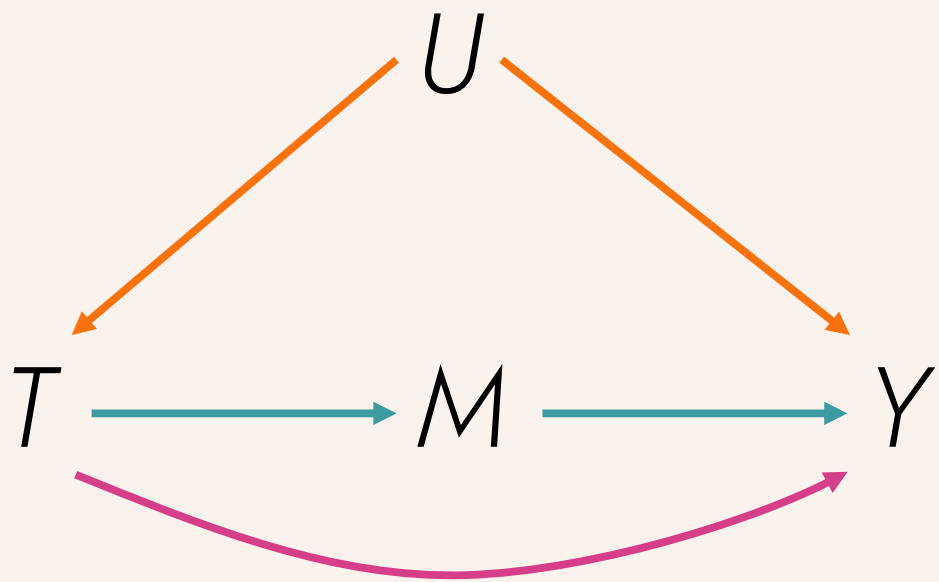
This "filtering" is useful in settings with selection into treatment. It opens up an additional avenue for effect identification when the assumptions underlying other popular strategies are not met.

The "random filter" identification strategy is able to remove selection bias by using the exogenous mediator.

	OLS		Random Filter	
	naïve	benchmark	first	second
	Y	Y	M	Y
T	453.77	**199.74**	0.67	245.67
	(4.81)	**(2.37)**	(0.01)	(6.19)
M	.	.	.	310.54
				(6.57)
U	.	503.83	.	.
		(2.37)		
cons	623.86	498.95	0.00	623.86
	(3.38)	(1.55)	(0.00)	(3.06)

$\beta = \$200$. Random Filter estimate: $0.67 \cdot \$310.54 = \$208.06$.

Additional direct or indirect causal mechanisms—like a sign-up bonus—could impact savings, but aren't intercepted by the mediator. What will the random filter identify here?



$$U \sim Bern(0.5) \quad \text{financial savvy}$$

$$T \sim Bern(0.25 + 0.5 \cdot U) \quad \text{savings account}$$

$$M \sim Bern(0.67 \cdot T) \quad \text{usable account}$$

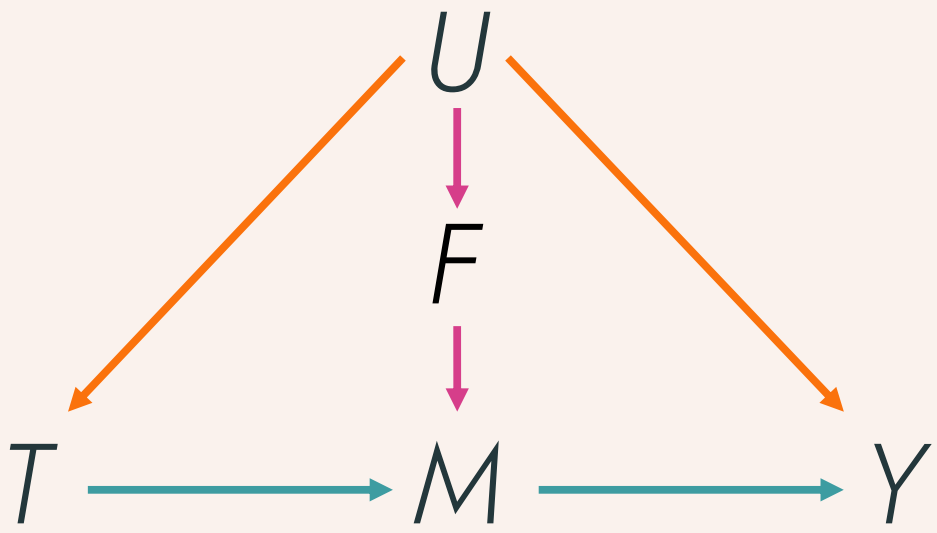
$$Y \sim N(500 + 500 \cdot U + 300 \cdot M + 30 \cdot T, 25^2) \quad \text{yearly savings}$$

The random filter can only measure the effect that M intercepts.

	OLS		Random Filter	
	naïve	benchmark	first	second
	Y	Y	M	Y
T	482.77	**229.74**	0.67	275.67
	(4.81)	**(2.37)**	(0.01)	(6.19)
M	.	.	.	310.54
				(6.57)
U	.	503.83	.	.
		(2.37)		
cons	623.86	498.95	0.00	623.86
	(3.38)	(1.55)	(0.00)	(3.06)

$\beta_{total} = \$230$, $\beta_M = \$200$, $\hat{\beta}_{RF} = \$208.06$ (same as before).

What if some errors aren't exogenously-driven? Then the mediator is only conditionally-exogenous.



$$U \sim Bern(0.5) \quad \text{financial savvy}$$

$$T \sim Bern(0.25 + 0.5 \cdot U) \quad \text{savings account}$$

$$F \sim Bern(0.25 + 0.5 \cdot U) \quad \text{banking familiarity}$$

$$M \sim Bern((0.34 + 0.67 \cdot F) \cdot T) \quad \text{usable account}$$

$$Y \sim N(500 + 500 \cdot U + 300 \cdot M, 25^2) \quad \text{yearly savings}$$

Running both stages of the random filter while controlling for familiarity restores the ability to identify the treatment effect.

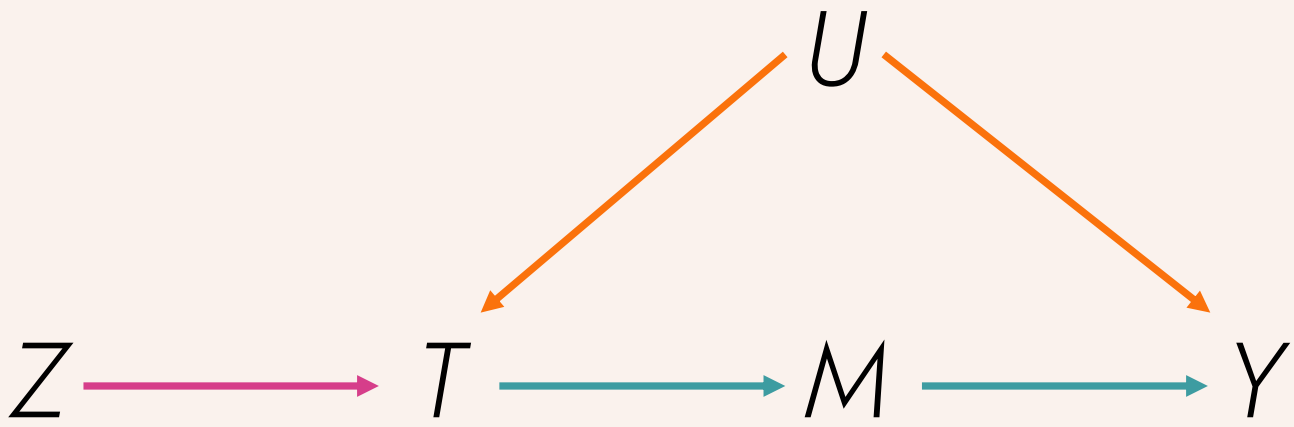
	OLS		Random Filter	
	naïve	benchmark	first	second
	Y	Y	M	Y
T	478.82	200.62	0.67	194.24
	(5.10)	(2.11)	(0.01)	(6.49)
M	.	.	.	306.84
				(7.61)
U	.	551.76	.	.
		(2.11)		
F	.	.	0.33	197.70
			(0.01)	(4.77)
cons	623.78	486.99	-0.12	550.26
	(3.59)	(1.39)	(0.00)	(3.27)

$$\beta = \$200,\ \hat{\beta}_{RF} = \$205.58.$$

Takeaway #1: the random filter can identify treatment effects amid selection into treatment.

So, which treatment effect are we measuring?

This is not instrumental variables. An instrument plays a fundamentally different role in the data generating process.

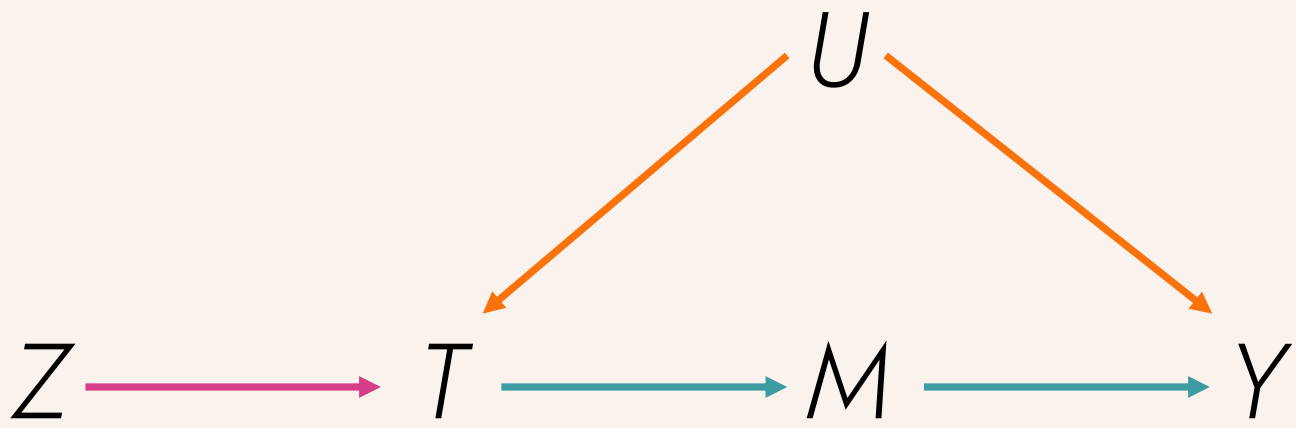


A good instrument (Z) exhibits the following traits:

1. Z affects T (M clearly does not)
2. Z affects Y only through its effect on T (M fails here too)
3. Z is independent of U or any other confounder, thus $Z \to T$ and $Z \to Y$ seem identifiable

Takeaway #2: the random filter may be applicable when an instrumental variables strategy is unavailable.

We can add some random variation in enrollment to determine the difference in instrumental variable and random filter behavior.



$U \sim Bern(0.5)$	financial savvy
$Z \sim Bern(0.5)$	assignment
$C \sim Bern(0.5)$	complier (latent)
$T \sim (C = 0) \cdot Bern(0.25 + 0.5 \cdot U) + (C = 1) \cdot Z$	savings account
$M \sim Bern(0.67 \cdot T)$	usable account
$Y \sim N(500 + 500 \cdot U + (450 - 300 \cdot C) \cdot M, 25^2)$	yearly savings

IV only picks up the average treatment effect on compliers (ATC); RF can measure the average treatment effect on the treated (ATT)!

	naïve	benchmark	complier	rf1	rf2	iv1	iv2
	Y	Y	Y	M	Y	T	Y
T	335.04	300.29	99.42	0.66	130.51	.	.
	(5.74)	(3.44)	(1.57)	(0.01)	(7.62)		
M	.	.	.	.	307.65	.	.
					(8.14)		
Z	.	.	.	.	.	51.21	0.50
						(6.63)	(0.01)
U	.	499.20	500.89	.	.	.	.
		(2.45)	(1.56)				
$T \cdot C$	.	-200.84	.	.	.	.	.
		(4.71)					
cons	683.36	494.45	500.14	0.00	683.36	822.04	0.25
	(4.01)	(2.32)	(1.33)	(0.00)	(3.75)	(4.63)	(0.01)

$$\beta_{ATT} = \$200,\ \beta_{ATC} = \$100,\ \hat{\beta}_{RF} = \$203.05,\ \hat{\beta}_{IV} = \$102.42.$$

Takeaway #3: the random filter can identify an ATE (or ATT)—even in cases where IV would have identified a LATE.

Random filter background and theory

Judea Pearl, a computer scientist, is credited with the discovery of this estimation method, which he calls the *Front Door Criterion*.

The FDC isn't used in economics because it relies on prior knowledge of Directed Acyclic Graphs (DAGs) or do-calculus.

It can be explained more thoroughly using econometric theory.

The first empirical example of the random filter asks if [authorizing] ridesharing on ride-hailing apps changes tipping behavior.

Selection: frugal riders choose the cheaper option and are already less likely to tip.

Random filter: authorization does not guarantee a shared ride. Actually sharing is conditionally-exogenous.

Outcome: tipping didn't decrease in share-authorized rides.

[†]Bellemare, Bloem, and Wexler (2024). *The Paper of How: Estimating Treatment Effects Using the Front-Door Criterion*.

Relative to *The Paper of How*, my paper derives the assumptions needed for the random filter to generate an ATE or ATT. Pearl's approach to causal modeling cannot address this.

TPOH uses simulation to refine a proposed assumption by Pearl, but they do not prove its necessity.

A complete explanation requires counterfactual language. I took a Rubinesque approach.

A few lessons from the "Random Filter Theorem"

The random filter picks up "the treatment effect for those impacted by the mediator." Does this create another LATE?

Not if these assumptions are met:

1. $Y^T_{Mi} = Y_{Mi}$ (M intercepts $T \to Y$)
2. $M_{1i}, M_{0i} \perp T_i$ (M unconfounded with T)
3. $Y_{1i}, Y_{0i} \perp M_i \mid T_i$ (M unconfounded with Y, given T)
4. $0 < P(M_{Ti} = 1) < 1$ (common support, ATE)

Two definitions round out the primitives:

- $M_i = M_{0i} + (M_{1i} - M_{0i}) \cdot T_i$
- $Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) \cdot M_i$ (given (1) above)

Assumption 1: "only through."

If we know the value of M, then T provides no additional information about which potential outcome of Y is observed:

$$Y_{Mi}^{T} = Y_{Mi}$$

Implication: if true, we measure the full effect of T on Y.

T changes the "lottery" that determines M, and only the *result* of this lottery drives the value of Y:

$$M_i = M_{0i} + (M_{1i} - M_{0i}) \cdot T_i$$
$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i}) \cdot M_i$$

Assumptions 2/3: "exogeneity."

M is as good as randomly assigned w.r.t. T, Y:

$$M_{1i}, M_{0i} \perp T_i$$
$$Y_{1i}, Y_{0i} \perp M_i \mid T_i$$

This allows for statements like these:

a) $E[M_i \mid T_i = 1] - E[M_i \mid T_i = 0] = E[M_{1i} - M_{0i}]$

b) $E[Y_i \mid M_i = 1, T_i = 1] - E[Y_i \mid M_i = 0, T_i = 1] = E[Y_{1i} - Y_{0i} \mid T_i = 1]$

Implication: unbiased identification of $T \to M$ and $M \to Y$.

Assumption 4: "common support."

M must vary for treated and untreated to identify an ATE:

$$0 < P(M_{Ti} = 1) < 1$$

Implication: (e.g.) if $M_{0i} = 0$ always, then $M \to Y$ is only estimable for the treated (ATT).

Pearl thought this simpler data condition would suffice:

$$P(M_i = m \mid T_i = t) > 0, \ \forall t, m.$$

(But we still have to make an assumption about the population to know if what we measured is what we wanted.)

The Random Filter Theorem is the core contribution of my paper.

Random Filter Theorem *Provided (1)-(4) hold, then the random filter estimand,*

$$\beta_{RF} = \{E[M_i \mid T_i = 1] - E[M_i \mid T_i = 0]\} \cdot E\left[E[Y_i \mid M_i = 1, T_i = T] - E[Y_i \mid M_i = 0, T_i = T]\right] ,$$

is equivalent to the ATE, $E\left[Y_i^1 - Y_i^0\right]$.

(There's a corollary for the ATT, $E\left[Y_i^1 - Y_i^0 \mid T_i = 1\right]$.)

Why doesn't the random filter reduce the external validity of the estimate in the same way that IV does?

The first stage of the RF uses information from the full sample:

$$E[M_i \mid T_i = 1] - E[M_i \mid T_i = 0] = P(M_{1i} > M_{0i}) - P(M_{1i} < M_{0i}).$$

T must impact M in order to affect Y. Those whose M isn't driven by T pull this first stage towards zero. In IV, those unaffected by Z receive *zero weighting*, which forces the LATE.

The second stage is just a weighted average of the mediator effects for treated and untreated, so no one is left out here either:

$$E\left[E[Y_i \mid M_i = 1, T_i = T] - E[Y_i \mid M_i = 0, T_i = T]\right].$$

"Intuitive" non-LATE: While instrumental variables selects *some* exogenous variation in treatment, the random filter removes only the endogenous variation.

Application: Chivo Wallet

Chivo, the great banking innovation.

[†] Armin Kübelbeck, *Goat, located in Fiesch, Valais* (2013)

The beginnings of a banking app

Nayib Bukele, the *millennial dictator*, is known for suspending constitutional rights, forcibly removing uncooperative politicians and judges, and cryptocurrency evangelism.

He needed to find a way to pay for large foreign debts, but with zero credit, he could only pander to wealthy crypto investors.

[†]Marvin Recinos, *Nayib Bukele at a surfing competition* (2021)

Bukele's crypto scheme had nothing to do with financial inclusion, but it had the potential to benefit El Salvadorans.

Chivo Wallet is a state-run bank that touts zero fees on deposits and withdrawals, remittances, purchases, and peer-to-peer transfers. It was created to promote cryptocurrency adoption.

This was deployed in a country where 70% of the population was unbanked, but 65% had a cell phone ready to link to an account.

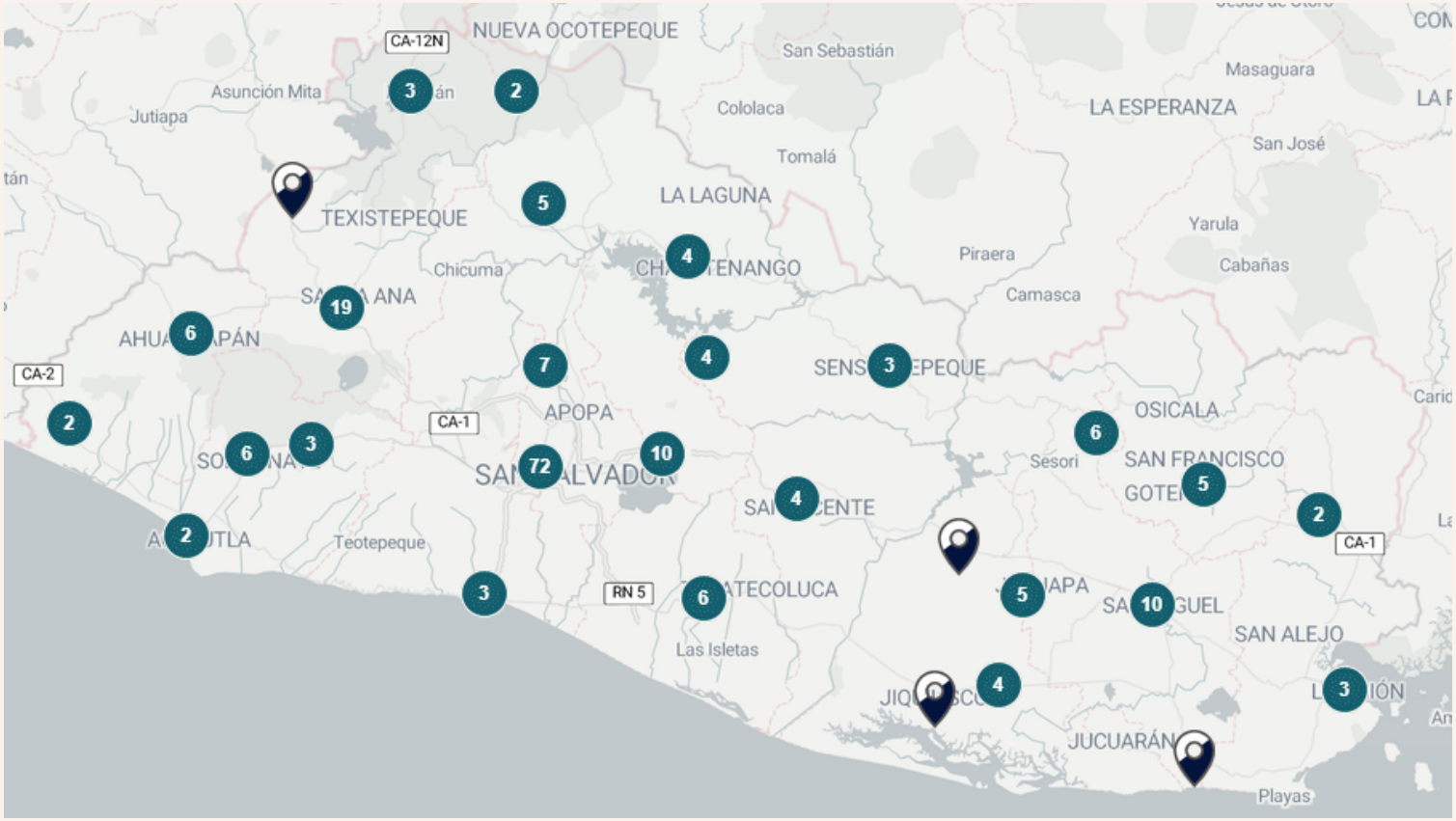
President Bukele incentivized take-up of Chivo with $30 of BTC, approximately three days of wages for an El Salvadoran. 53% of the adult population downloaded the app in three months.

[†] Press Secretariat of the Republic of El Salvador, *Nayib Bukele announcing a $30 airdrop of BTC to those who sign up for Chivo* (2021)

Access to the Chivo network is extended to users without a cell phone, too. Hundreds of ATMs now exist all over the country and were initially attended by agents to help users with transactions.

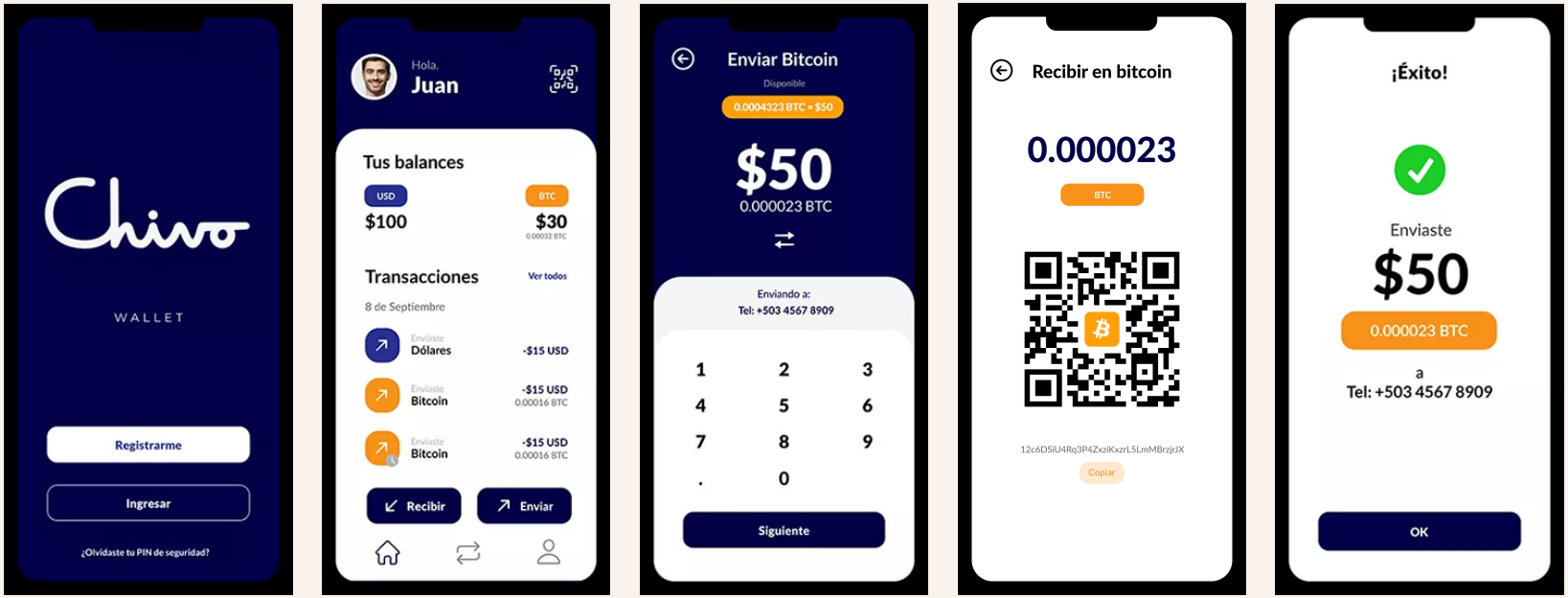
[†]Marvin Recinos, *A man wearing a Nayib Bukele facemask next to a bitcoin ATM* (2021)

There is at least one Chivo ATM within 20 miles of everyone in the country (although distance $\neq$ travel time in El Salvador).



[†]Coin ATM Radar, *Chivo ATM locations* (2023)

The Chivo app had a great design team, but an impossible release date. The interface and backend glitched constantly; many users could not sign up, make purchases, or transfer money to others.



[†]Chivo, *Chivo App Screenshots* (2021)

This dumpster fire of a rollout led to 80% of users and businesses abandoning the app (and related infrastructure). Almost no new downloads of the app have occurred since the initial release.

[†]Ivan Manzano, *Bitcoin ATM kiosk on fire* (2021)

This setting created an interesting opportunity to use the random filter to measure the effect of Chivo on financial well-being.

Max Edelstein—my student at Colgate University—and I set up a survey with the help of CIDGallup, a prominent enumeration company in Latin America:

- 700 individuals surveyed
- demographic information
- degree of financial inclusion
- changes in financial well-being
- involvement with Chivo

Max also collected some qualitative data through interviews with the directors of several organizations providing financial services to disadvantaged communities in El Salvador.

The data collected supported the hypothesis that any naïve comparison would be due to selection bias.

Selection bias would distort any comparison of well-being. (1/3)

	Downloaded Chivo		
Variable	$T = 0$	$T = 1$	Difference
no savings	0.347	0.321	-0.026
	(0.477)	(0.467)	(0.041)
money in bank	0.318	0.392	0.075*
	(0.467)	(0.489)	(0.043)
money in house	0.324	0.245	-0.078**
	(0.469)	(0.431)	(0.039)
unbanked	0.512	0.389	-0.123***
	(0.501)	(0.488)	(0.043)
bank account	0.424	0.511	0.088**
	(0.496)	(0.500)	(0.044)
bank trips/month	1.244	1.949	0.705***
	(1.728)	(2.144)	(0.181)
Observations	170	530	700

Selection bias would distort any comparison of well-being. (2/3)

	Downloaded Chivo		
Variable	$T = 0$	$T = 1$	Difference
bank travel time	43.556	36.954	-6.602
	(39.343)	(42.611)	(4.223)
credit card	0.112	0.145	0.034
	(0.316)	(0.353)	(0.030)
remittance change	2.435	2.527	0.092
	(0.866)	(1.107)	(0.142)
cash vs digital	1.854	2.517	0.663***
	(1.235)	(1.476)	(0.132)
only uses cash	0.609	0.404	-0.205***
	(0.490)	(0.491)	(0.046)
dollars vs bitcoin	1.066	1.391	0.325***
	(0.275)	(0.769)	(0.064)
Observations	170	530	700

Selection bias would distort any comparison of well-being. (3/3)

	Downloaded Chivo		
Variable	$T = 0$	$T = 1$	Difference
female	0.524	0.368	-0.156***
	(0.501)	(0.483)	(0.043)
$18 \leq$ age ≤ 39	0.382	0.564	0.182***
	(0.487)	(0.496)	(0.044)
$40 \leq$ age ≤ 62	0.482	0.355	-0.128***
	(0.501)	(0.479)	(0.043)
age ≥ 63	0.135	0.081	-0.054**
	(0.343)	(0.273)	(0.026)
primary school	0.447	0.325	-0.123***
	(0.499)	(0.469)	(0.042)
high school	0.241	0.360	0.119***
	(0.429)	(0.481)	(0.041)
college	0.312	0.315	0.003
	(0.465)	(0.465)	(0.041)
Observations	170	530	700

Those divided by Chivo app usability looked more similar. (1/3)

	Experienced no Chivo issue		
Variable	$M = 0$	$M = 1$	Difference
no savings	0.301	0.330	0.029
	(0.460)	(0.471)	(0.044)
money in bank	0.380	0.398	0.017
	(0.487)	(0.490)	(0.046)
money in house	0.276	0.232	-0.044
	(0.448)	(0.422)	(0.041)
unbanked	0.429	0.371	-0.059
	(0.497)	(0.484)	(0.046)
bank account	0.442	0.542	0.101**
	(0.498)	(0.499)	(0.047)
bank trips/month	1.908	1.967	0.059
	(2.390)	(2.029)	(0.202)
Observations	163	367	530

Those divided by Chivo app usability looked more similar. (2/3)

	Experienced no Chivo issue		
Variable	$M = 0$	$M = 1$	Difference
bank travel time	39.443	35.941	-3.502
	(48.681)	(39.917)	(4.418)
credit card	0.172	0.134	-0.038
	(0.378)	(0.341)	(0.033)
remittance change	2.506	2.535	0.029
	(1.193)	(1.075)	(0.143)
cash vs digital	2.426	2.557	0.131
	(1.419)	(1.501)	(0.142)
only uses cash	0.419	0.397	-0.022
	(0.495)	(0.490)	(0.047)
dollars vs bitcoin	1.283	1.438	0.155**
	(0.665)	(0.806)	(0.074)
Observations	163	367	530

Those divided by Chivo app usability looked more similar. (3/3)

	Experienced no Chivo issue		
Variable	$M = 0$	$M = 1$	Difference
female	0.423	0.343	-0.080*
	(0.496)	(0.475)	(0.045)
$18 \leq$ age ≤ 39	0.528	0.580	0.053
	(0.501)	(0.494)	(0.047)
$40 \leq$ age ≤ 62	0.337	0.362	0.025
	(0.474)	(0.481)	(0.045)
age ≥ 63	0.135	0.057	-0.078***
	(0.343)	(0.233)	(0.026)
primary school	0.288	0.341	0.052
	(0.454)	(0.475)	(0.044)
high school	0.387	0.349	-0.038
	(0.488)	(0.477)	(0.045)
college	0.325	0.311	-0.015
	(0.470)	(0.463)	(0.044)
Observations	163	367	530

So, what did the random filter have to say about the effect of Chivo Wallet?

Downloading a broken app had no effect on financial well-being.

	OLS	SUR		Random Filter
	finance score	functional Chivo	finance score	finance score
intercept	3.073***	0.000	3.073***	.
	(0.089)	(0.031)	(0.088)	
download Chivo	0.301***	0.691***	0.316***	-0.015
	(0.102)	(0.036)	(0.126)	(0.075)
functional Chivo	.	.	-0.022	.
			(0.107)	
Observations	690	690	530	

A naïve OLS regression of a Likert-scale welfare measure on downloading Chivo reveals a positive impact. The random filter estimation suggests this initial correlation is driven by selection, since the *functional* app has no effect.

[Mini] discussion

Where does the random filter fit in our empirical toolkit?

1. The RF identifies treatment effects amid selection into treatment without explicit data on the drivers of selection.
2. The RF may be useful when IV (or RD, DiD, etc.) is unavailable. *We will learn to spot good mediators.*
3. The RF can identify an ATE or ATT—even in cases where IV would have identified a LATE. (**Random Filter Theorem**)

When we hope for $M = T$—but things don't go as planned—the random filter may facilitate effect identification.

The random filter highlights not only the resulting effect $M \to Y$, but the effectiveness of the *intended* intervention, $T \to Y$.

In the Chivo example, we verified that an imperfect intervention had no effect because the program itself was poorly-conceived.

Thank you!

pierce.donovan@unr.edu

[†]Marvin Recinos, *Nayib Bukele at a surfing competition* (2021)