The random filter identification strategy

Pierce Donovan[†]

February 2024

Abstract

In 2021, the El Salvadoran government launched Chivo Wallet, a mobile banking application that allowed citizens to save money securely, transact with businesses, and send money to others with very low transaction costs. A majority of the adult population downloaded the application within four months. Despite this interest, I find little evidence that this push toward financial inclusion made a positive impact on financial well-being. While treatment was not randomly-assigned, Chivo Wallet's programming was fraught with errors that precluded its use, and this post-treatment exogenous mediation allows us to identify a treatment effect using a new random filter identification strategy.

Keywords: random filter, front door criterion, identification strategies, financial inclusion, digital banking, mobile money, Chivo Wallet (JEL: C13, D14, G21, G28)

[†]Assistant Professor, Department of Economics, University of Nevada
email: pierce.donovan@unr.edu
web: piercedonovan.github.io

[††]I would like to thank Max Edelstein for his help in facilitating fieldwork in El Salvador. That fieldwork received financial support from the Lampert Institute at Colgate University.

1 Random filters and the front-door criterion

In this paper, I highlight an identification strategy with wide applicability to empirical research in economics. The *random filter* approach uses the presence of an exogenous mediator variable between a treatment and outcome of interest to select variation in treatment that is as good as randomly assigned with respect to potentially confounding factors. This filtering—which is particularly useful in settings with selection into treatment—opens up an additional avenue for causal effect identification when the assumptions underlying other popular strategies are not met.

The identifying variation selected by the random filter may sound a bit like what an instrumental variable picks up, so a brief comparison of methods is in order. In absence of random assignment to treatment, a common approach is to look for situations where the eligibility for treatment was randomly-assigned. In these cases, we can use the variation in treatment that is attributable to eligibility and exclude the non-exogenous variation driven by other factors. For this strategy to work, we require that an instrument has no additional impact on the outcome of interest, either directly or through some other unobservable channel—i.e. the eligibility for treatment may only impact the outcome through its effect on treatment status. If there was another causal channel from instrument to outcome, any resulting estimates mistakenly attribute this additional effect to the treatment.

In contrast, when an exogenous mediator variable facilitates the treatment effect, we can remove the influence of any confounding factors between treatment and outcome by separately estimating the impacts of treatment on the mediator and mediator on the outcome, then scaling the former by the latter. This approach "filters out" the endogenous variation in treatment generated by the confounding factor rather than selecting the exogenous variation directly—as an instrument would. The roles of the instrument and filter are quite different, and a filter would make for a terrible instrument as it affects outcomes instead of treatment status. The critical assumption needed for the random filter approach is that the mediator must intercept the full treatment effect—i.e. another causal channel through which treatment has an effect on the outcome may not exist. An instrumental variable clearly violates this assumption and would not be a viable filter.

This approach was brought to light as the "Front-Door Criterion" (FDC) (Pearl, 1995). The FDC has not seen employment in social science research because the typical presentation relies on prior knowledge of Directed Acyclic Graphs (DAGs) or "do-calculus" (Pearl, 2000), and applications are difficult to imagine without an existing collection of examples (Imbens, 2020; Heckman and Pinto, 2022; Huntington-Klein, 2022b; Donovan, 2024). But the FDC can be explained without relying on jargon and theory from outside economics—

and more completely with econometric theory. Thus the objective of this paper is to provide a more rigorous introduction to the FDC and demonstrate the identification of a treatment effect in a setting where the [refined] FDC assumptions are clearly met.

To ameliorate the FDC's inaccessibility, I am proposing an alternative term—the "random filter"—for use in econometrics. My term distinctly characterizes the identification strategy and provides a convenient mnemonic that aids in its understanding. Beyond nomenclature, I answer a previously unresolved question about the scope of the treatment effect identified by the method. I first discuss the necessary identification assumptions for the random filter, then show that it either estimates the average treatment effect (ATE) or the average treatment effect on the treated (ATT)—even when unobservable confounding factors pollute the relationship between treatment and outcome.

I then employ the random filter to determine whether the most promising facet of El Salvador's gamble on cryptocurrency and quasi-decentralized banking had a positive impact on financial well-being. As part of this agenda, the El Salvadoran government released a low-fee financial product for savings, transactions, peer-to-peer transfers, and remittances. A majority of the adult population voluntarily accessed the "Chivo Wallet" banking network within three months of its launch due to a large financial incentive. However, the customer-facing app, ATMs, and back-end of the network were all laden with coding errors, making the network all but unusable for reasons outside of users' control. I find little evidence that this push for financial inclusion made a positive impact—as might be expected—using the random filter approach, while other methods would have produced positively-biased results.

The single contemporary use of the random filter is Bellemare et al. (2024), which presents the first application to observational data. The core application estimates the effect of authorizing ride-sharing on an Uber or Lyft trip on tipping behavior. In this setting, unobserved rider characteristics like frugality will partially influence both tipping and ride-sharing decisions, thus a filter is needed—and one exists. When someone authorizes ride sharing, they will not necessarily share a ride with another passenger. Plenty of would-be ride-sharers go unmatched. Because the only plausible mechanism through which authorization would impact tipping behavior is through the matching process, and because the matching algorithm is known, they can use this conditionally-exogenous mediator to filter any confounding variation between authorization and tipping.

Section 2 provides a rigorous introduction to the random filter identification strategy. Section 3 details the Chivo Wallet setting and estimates the treatment effect of signing up for an account on near-term financial outcomes. Section 4 concludes with suggestions for finding new empirical research opportunities that can make use of the random filter.

2 The random filter identification strategy

In this section, I introduce the necessary identification assumptions for the random filter approach. This discussion does not require any understanding of DAGs or do-calculus—the traditional approaches that motivate these assumptions—as neither causal inference tool is necessary to demonstrate the random filter. If desired, these alternatives can be found in Donovan (2024) and Bellemare et al. (2024), respectively—although both explanations are incomplete. Instead, this introduction only uses potential outcomes language already familiar to economists.[1] I then prove that the approach can identify the average treatment effect (ATE) or the average treatment effect on the treated (ATT) amid unobserved confounding factors. This makes the random filter approach a valuable addition to any empiricist's toolkit and expands the set of data generating processes that can be leveraged for causal inference. I leave further discussion in this vein for Section 4.

2.1 Selection bias, unobservable confounders, and an estimable ATE

In the typical empirical setting, we are interested in determining the impact of some treatment (T) on an outcome of interest (Y) for a group of treated individuals. The key identification obstacle to overcome is that each treated individual only reveals their outcome under treatment $\left(Y_i^1\right)$, and the untreated counterfactual $\left(Y_i^0\right)$ remains unobserved (Rubin, 1974; Holland, 1986). The individual treatment effect is of course determined by the difference of these two outcomes, and the missing data problem is evident.

A separate untreated group often provides this missing data. If we focus on the difference in group average outcomes, we can relax the need for matching each treated individual and rely on group-level similarity to justify our comparison. But outside of an experimental setting, it is very likely that individuals will self-select into groups, and this can drive a difference in outcomes that is not due to treatment. In these cases, this selection bias cannot be disentangled from the treatment effect and implies that the untreated will not produce a credible counterfactual outcome for the treated (Duflo et al., 2006; Angrist and Pischke, 2009). This can be represented mathematically by Equation 1,

$$E\left[Y_i^0 \mid T_i = 1\right] \neq E\left[Y_i^0 \mid T_i = 0\right]. \tag{1}$$

This bias prevents a comparison of outcomes from having a causal interpretation,

$$E[Y_i \mid T_i = 1] - E[Y_i \mid T_i = 0] \neq ATT. \tag{2}$$

[1] While I take advantage of a Rubinesque treatment-control style throughout this paper, the scope of the random filter is not limited to experimental settings or data with binary treatment or mediator designations.

If the confounding factor driving selection (U) is observed, a matching strategy will generate an otherwise-similar untreated group. This works by making comparisons of the treated and untreated groups conditional on each value of U where treated and untreated individuals are both observed, then averaging over these conditional average treatment effects (Heckman et al., 1997, 1998). Equation 3 is thus estimable on the common support of U, $S_T(U_i) = \text{supp}(U_i \mid T_i = 1) \cap \text{supp}(U_i \mid T_i = 0)$,

$$E\left[E[Y_i \mid T_i = 1, U_i = u] - E[Y_i \mid T_i = 0, U_i = u] \mid u \in S_T(U_i)\right] = ATT.^2 \tag{3}$$

However, in many cases, U is unobserved, and this strategy is unworkable. Apart from randomized trials, selection into (or out of) treatment should be expected, even in cases where some variation in treatment status is purported to be exogenously-driven.

In response to the selection threat, another common approach is to find some instrumental variable (Z) that predicts some of the variation in T and is unconfounded with both T and Y. Z will therefore be unresponsive to the variation in the confounding factor U. In absence of treatment, the two groups separated by the instrument will have similar average outcomes, and their comparison will provide a causally-interpretable estimate of the intent to treat effect (ITT) (Duflo et al., 2006),[3]

$$E[Y_i \mid Z_i = 1] - E[Y_i \mid Z_i = 0] = ITT. \tag{4}$$

The ITT can be shown to equal the impact of Z on the probability of treatment, times the treatment effect for those treated because of Z (Imbens and Angrist, 1994). To isolate the latter effect, the ITT can be reduced by dividing by the impact of Z on T, as long as Z has no additional impact on Y through some other mechanism. This "exclusion restriction" assumption assures that no other effects on Y are misattributed to T. The reduction provides the local average treatment effect (LATE)—an average treatment effect weighted by individual susceptibility to the instrument. In the case of binary Z and T, this is the average treatment effect on compliers (ATC), as those not induced to take up treatment by the instrument will receive zero weighting, and those who are will receive full weighting (Huntington-Klein, 2022a). The LATE theorem result makes use of an additional as-

[2]The outer expectation is taken over the support of U common to both treated and untreated groups because values of U without variation in treatment cannot generate a conditional treatment effect. A stronger form of the common support idea is $P(T_i = 1 \mid U_i = u) < 1$, which states that for every value of U where treatment occurs, untreated observations are also available. If this does not hold, we are only estimating the ATT for a subset of the treated (Heckman et al., 1998).

[3]It is often left unsaid that the ITT is only causally-interpretable if Z has an effect on T, and not the other way around. This may be intuitive, but the above assumptions do not make this explicit. If this didn't hold, Equation 4 only represents a spurious correlation between Z and Y (Donovan, 2024).

sumption, monotonicity, which states that the instrument weakly influences take-up of treatment for all individuals. This is needed to maintain a causal interpretation amid heterogeneous treatment effects (Angrist and Pischke, 2009).

Valid instruments only exist in particular circumstances. It is often the case that a candidate Z fails to play the exact role needed for identification of the ITT and LATE. However, a class of variables that fail only the exclusion restriction may provide identification of a causal effect under a different assumption. If a variable M *mediates* the causal effect of T on Y and is unconfounded with either variable, then we can estimate the effect T has on Y that is facilitated by M. This is done in two [unbiased] stages. The first stage estimates the effect of T on M, and the second stage estimates the effect of M on Y. With this approach, the exogeneity of the mediator "filters" the endogenous variation in T generated by U before it reaches Y.[4]

Which of the above treatment effects are identified by the random filter? Even when data are subject to selection into treatment, the random filter can uncover either the ATE or the ATT.[5] To facilitate a proof, I will start with a formalization of the random filter identification assumptions. I will work with the simplest case with binary treatment and mediator variables. The assumptions are as follows:

$$
\begin{aligned}
&\text{(A1)}\ \ Y^1_{Mi} = Y^0_{Mi} = Y_{Mi} && (M \text{ intercepts } T \to Y)\\
&\text{(A2)}\ \ M_{1i}, M_{0i} \perp T_i && (M \text{ unconfounded with } T)\\
&\text{(A3)}\ \ Y_{1i}, Y_{0i} \perp M_i \mid T_i && (M \text{ unconfounded with } Y, \text{ conditional on } T)\\
&\text{(A4)}\ \ 0 < P(M_{Ti} = 1) < 1, T \in \{0,1\} && (\text{common support, ATE})\\
&\text{(A4')}\ \ P(M_{Ti} = 1) < 1, T \in \{0,1\} && (\text{common support, ATT})
\end{aligned}
$$

The subscripts on Y and M denote the potential outcome and mediator under $M \in \{0,1\}$ and $T \in \{0,1\}$, respectively. Two additional definitions decompose M_i and Y_i into their potential outcomes:

$$
\begin{aligned}
&\text{(D1)}\ \ M_i = M_{0i} + (M_{1i} - M_{0i}) \cdot T_i\\
&\text{(D2)}\ \ Y_i\ = Y_{0i} + (Y_{1i} - Y_{0i}) \cdot M_i \qquad (\text{given (A1)}).
\end{aligned}
$$

[4] M is effectively a lottery, and T determines the lottery in which an observation takes part. With this analogy, it becomes clear that both stages are estimable without bias, because T does not have any influence on the lottery result once the lottery is determined, and M is as good as randomly assigned with respect to the potential outcomes of Y, conditional on the lottery that T selects.

[5] In a less policy-relevant case, the random filter can also potentially identify the average treatment on the untreated (ATU). This result turns out to be a trivial corollary to the main proof after relaxing the second inequality in (A4) rather than the first.

We first assume that treatment can only impact the outcome via the mediator, and (A1) demonstrates that the value of T is immaterial to Y once M is fixed. If there is an additional causal channel between T and Y that M does not intercept, this partial effect will be missed by the estimation strategy. This is ultimately due to (A3), which requires us to control for T when estimating the impact of the M on the Y. This will remove any correlation between M and another mechanism stemming from T.

The independence assumptions (A2) and (A3) state that the potential outcome and mediator distributions do not depend on the realized values of M and T, respectively.[6] First, there cannot be any confounding factors driving a spurious relationship between treatment and mediator. This allows us to interpret a difference in M across the groups delineated by T as causal. Second, there cannot be any confounding factors driving a spurious relationship between the mediator and the outcome, after controlling for treatment. Clearly, this second assumption doesn't hold unconditionally, since the confounder will impact the mediator through its effect on treatment. However, this assumption is met after conditioning on treatment, since the value of U is immaterial to M after T is fixed.[7]

An estimator can only exploit variation in M where both $T_i = 1$ and $T_i = 0$ exist. Assumptions (A4) and (A4') provide the common support statements necessary to label the treatment effect identified by the random filter. Something similar has been proposed—but not proven—in Pearl (2000) and the literature following, and is supported by simulation in Bellemare et al. (2024). This alternative, $P(T_i = t \mid M_i = m) > 0 \ \forall t, m$, is merely a data requirement which states that for each value of the mediator, there must be a non-zero probability of treatment and non-treatment. The intuition works better once we apply Bayes' Theorem—this condition is equivalent to $P(M_i = m \mid T_i = t) > 0 \ \forall t, m$. If variation in the mediator is only accessible to those who are treated, then there is no way to identify the effect of M on Y for those untreated. In this case, we can relax this constraint if we only want to measure the ATT, which just requires individuals with $M_i = 0$ for any condition under which there are others with $M_i = 1$, and so $P(M_i = m \mid T_i = 1) > 0 \ \forall m$ suffices.

However, what is actually desired is a statement about the possibility of each potential outcome of M realizing 0 or 1. If, for example, $P(M_{0i} = 1) = 0$, then it is clear that this cannot happen within sample either—but the converse is not true. If $P(M_i = 1 \mid T_i = 0) = 0$, we must assume $P(M_{0i} = 1) = 0$ to be true anyway in order to estimate the ATT,

[6] These two independence assumptions can be relaxed to [conditional] mean-independence for the sake of identification if one finds the strong forms of these assumptions implausible.

[7] In the case that there is some structural relationship between U and M, the random filter approach may still be admissible. If, for instance, there is another variable responsible for this link, and this variable is observable, we can control for this in both stages of the estimation. Donovan (2024) discuses this idea in the context of analyzing the effectiveness of crop insurance programs, and Bellemare et al. (2024) use this approach with their identification of the ride-share effect on tipping behavior.

because if $T_i = 1$ reduced the likelihood of $M_i = 1$ for some individuals, running the random filter on data without this archetype will fail to estimate the full ATT.[8] This subtle difference between (A4) and (A4') versus the previous data requirement is absent from the literature, and is shown to be the correct identification assumption in the proof below.

2.2 The Random Filter Theorem

Random Filter Theorem. *Provided (A1)-(A4) hold, then the random filter estimand,*

$$\begin{aligned}\beta_{RF} = &\{E[M_i \mid T_i = 1] - E[M_i \mid T_i = 0]\} \\ &\cdot E_{T|M}\left[E[Y_i \mid M_i = 1, T_i = T] - E[Y_i \mid M_i = 0, T_i = T]\right]\end{aligned} \tag{5}$$

is equivalent to the ATE. If instead (A1)-(A3) and (A4') hold, then the it is equivalent to the ATT.

The $E_{T|M}[\cdot]$ notation clarifies that the outer expectation is taken over the common support of T. My proof uses the familiar potential outcomes framework (Rubin, 1974, 1977; Holland, 1986).[9] I first prove three lemmas, which support proof of the above theorem.

Lemma 1. *Provided (A2) and (A4) hold, the first stage estimand of the random filter is*

$$E[M_i \mid T_i = 1] - E[M_i \mid T_i = 0] = P(M_{1i} > M_{0i}) - P(M_{1i} < M_{0i}). \tag{6}$$

Proof.

$$\begin{aligned}&E[M_i \mid T_i = 1] - E[M_i \mid T_i = 0] \\ &= E[M_{1i} \mid T_i = 1] - E[M_{i0} \mid T_i = 0] && (D1) \\ &= E[M_{1i} - M_{0i}] && (A2) \\ &= P(M_{1i} > M_{0i}) - P(M_{1i} < M_{0i}) \ \square && (A4)\end{aligned}$$

Corollary. *If (A4) does not hold, but (A4') does, then the estimand becomes*

$$E[M_i \mid T_i = 1] - E[M_i \mid T_i = 0] = P(M_{1i} > M_{0i}) = P(M_{1i} = 1). \tag{7}$$

[8] The saving grace of the data requirement is that it would be rare for (A4) to hold while all of the admissible mediator values are not observed within sample, and examples thus far have been compatible with the unconsciously-assumed common support assumptions stated here.

[9] The usual stable unit treatment value assumption (SUTVA) is implicit in this derivation. In the random filter setting, this assumption states that there are no treatment and mediator externalities between observations (otherwise recorded $T_i = 0$ or $M_i = 0$ may be false) and that the treatment/mediator doseage is identical for those with $T_i = 1$ or $M_i = 1$ (Duflo et al., 2006; Cunningham, 2021).

Lemma 1 provides the rather intuitive result that the treatment must have some impact on the mediator if it is to impact the outcome of interest. Archetypes whose M isn't driven by T pull the first stage estimate towards zero because their Y will also be unmotivated by T. Note how these individuals do not receive zero weighting, in contrast to the LATE estimator for instrumental variables. That method assigns zero weight to those not affected by Z, which removes the ability to estimate the ATE or ATT.

Lemma 2. *Provided (A1), (A3), and (A4) hold, the second stage of the random filter is*

$$\begin{aligned} &E_{T|M}\left[E[Y_i \mid M_i = 1, T_i = T] - E[Y_i \mid M_i = 0, T_i = T]\right] \\ &= P(T_i = 1) \cdot E[Y_{1i} - Y_{0i} \mid T_i = 1] + P(T_i = 0) \cdot E[Y_{1i} - Y_{0i} \mid T_i = 0]. \end{aligned} \tag{8}$$

Proof.

$$\begin{aligned} &E_{T|M}\left[E[Y_i \mid M_i = 1, T_i = T] - E[Y_i \mid M_i = 0, T_i = T]\right] \\ &= P(T_i = 1) \cdot \{E[Y_i \mid M_i = 1, T_i = 1] - E[Y_i \mid M_i = 0, T_i = 1]\} \\ &+ P(T_i = 0) \cdot \{E[Y_i \mid M_i = 1, T_i = 0] - E[Y_i \mid M_i = 0, T_i = 0]\} && (A4) \\ &= P(T_i = 1) \cdot \{E[Y_{1i} \mid M_i = 1, T_i = 1] - E[Y_{0i} \mid M_i = 0, T_i = 1]\} \\ &+ P(T_i = 0) \cdot \{E[Y_{0i} \mid M_i = 1, T_i = 0] - E[Y_{0i} \mid M_i = 0, T_i = 0]\} && (A1), (D2) \\ &= P(T_i = 1) \cdot E[Y_{1i} - Y_{0i} \mid T_i = 1] + P(T_i = 0) \cdot E[Y_{1i} - Y_{0i} \mid T_i = 0] \ \square && (A3) \end{aligned}$$

Corollary. *If (A4) does not hold, but (A4') does, then the estimand becomes*

$$E_{T|M}\left[E[Y_i \mid M_i = 1, T_i = T] - E[Y_i \mid M_i = 0, T_i = T]\right] = E[Y_{1i} - Y_{0i} \mid T_i = 1]. \tag{9}$$

Lemma 2 provides the average mediator effect, which is [conceivably] a weighted average of the average mediator effect on the treated and the average mediator effect on the untreated. If part of the support of T does not include observations where $M_i = 1$ and $M_i = 0$, all of the density in $P(T_i = T \mid S_M(T_i))$ shifts to the region of common support and trivializes the $P(T_i = T)$ distribution. For example, if there is no variation in M for the untreated observations, $P(T_i = 1) = 1$.

Lemma 3. *(a) Provided (A1)-(A4) hold, the ATT estimand can be decomposed into*

$$E\left[Y_i^1 - Y_i^0 \mid T_i = 1\right] = \{P(M_{1i} > M_{0i}) - P(M_{1i} < M_{0i})\} \cdot E\left[Y_{1i} - Y_{0i} \mid T_i = 1\right], \quad (10)$$

and (b) under the same assumptions, the ATE can be decomposed into

$$\begin{aligned} E\left[Y_i^1 - Y_i^0\right] = &\{P(M_{1i} > M_{0i}) - P(M_{1i} < M_{0i})\} \\ &\cdot \{P(T_i = 1) \cdot E\left[Y_{1i} - Y_{0i} \mid T_i = 1\right] + P(T_i = 0) \cdot E\left[Y_{1i} - Y_{0i} \mid T_i = 0\right]\} \end{aligned} \quad (11)$$

Proof of 3(a).

$$\begin{aligned} &E\left[Y_i^1 - Y_i^0 \mid T_i = 1\right] \\ &= P(M_{1i} > M_{0i} \mid T_i = 1) \cdot E\left[Y_i^1 - Y_i^0 \mid T_i = 1, M_{1i} > M_{0i}\right] \\ &+ P(M_{1i} < M_{0i} \mid T_i = 1) \cdot E\left[Y_i^1 - Y_i^0 \mid T_i = 1, M_{1i} < M_{0i}\right] \\ &+ P(M_{1i} = M_{0i} = 1 \mid T_i = 1) \cdot E\left[Y_i^1 - Y_i^0 \mid T_i = 1, M_{1i} = M_{0i} = 1\right] \\ &+ P(M_{1i} = M_{0i} = 0 \mid T_i = 1) \cdot E\left[Y_i^1 - Y_i^0 \mid T_i = 1, M_{1i} = M_{0i} = 0\right] && (A4) \\ &= P(M_{1i} > M_{0i} \mid T_i = 1) \cdot E\left[Y_{1i} - Y_{0i} \mid T_i = 1, M_{1i} > M_{0i}\right] \\ &+ P(M_{1i} < M_{0i} \mid T_i = 1) \cdot E\left[Y_{0i} - Y_{1i} \mid T_i = 1, M_{1i} < M_{0i}\right] && (A1), (D2) \\ &= \{P(M_{1i} > M_{0i}) - P(M_{1i} < M_{0i})\} \cdot E\left[Y_{1i} - Y_{0i} \mid T_i = 1\right] \ \square && (A2), (A3), (D1) \end{aligned}$$

Corollary for 3(b). *Under the same assumptions, the ATU can be decomposed into*

$$E\left[Y_i^1 - Y_i^0 \mid T_i = 0\right] = \{P(M_{1i} > M_{0i}) - P(M_{1i} < M_{0i})\} \cdot E\left[Y_{1i} - Y_{0i} \mid T_i = 0\right], \quad (12)$$

and therefore the ATE,

$$E\left[Y_i^1 - Y_i^0\right] = P(T_i = 1) \cdot E\left[Y_i^1 - Y_i^0 \mid T_i = 1\right] + P(T_i = 0) \cdot E\left[Y_i^1 - Y_i^0 \mid T_i = 0\right],$$

matches the form in 3(b).

Corollary. *If instead (A1)-(A3) and (A4′) hold, then the ATT is*

$$E\left[Y_i^1 - Y_i^0 \mid T_i = 1\right] = P(M_{1i} = 1) \cdot E\left[Y_{1i} - Y_{0i} \mid T_i = 1\right], \quad (13)$$

and the ATU and ATE are unidentifiable.

In the second line of the proof of 3(a), I list the archetypes in the data created by the

combinations of T and M permitted by (A4). In the third line, I have implemented the potential outcomes of Y consistent with the conditional M statements in each expectation. In the two cases where M is the same for either value of T, Y does not respond to T when T fails to manipulate M, and the corresponding expectations contribute a zero. In the final line, (A3) applies to the potential outcomes of M with the aid of (D1).

An interesting result of Lemma 3 is that the ATT estimand changes depending on the region of common support for T. Equations 10 and 13 are not the same because the former case admits the possibility for T to decrease M. In cases where (A4) holds, this subpopulation of the treated should appear in the sample (via observed $M_{0i} = 1$) and thus be incorporated into the first stage of the random filter in order to identify the full ATT. If only (A4') holds, there is no discrepancy between the estimand and estimators meeting Pearl's data requirement and unconscious assumption.

The central result of this section is easily-determined by combining the relevant conclusions from the three lemmas:

Proof of the Random Filter Theorem

Provided (A1)-(A4) hold, multiplying the resulting estimands of Equation 6 and Equation 8 yields the estimand in Equation 11. If instead (A1)-(A3) and (A4') hold, then multiplying the resulting estimands of Equation 7 and 9 yields the estimand in Equation 13. □

While this section has framed the bias to be overcome as something purely due to selection into treatment, the random filter also combats spurious correlations driven by a sampling process or other treatments impacting the outcome of interest over the same time period. In the next section, I apply the random filter to measure the effect of a widely-adopted financial instrument in El Salvador on financial well-being—and remove the influence of selection bias and bias generated by a volunteer sample. In this application, I utilize both a parametric and non-parametric estimator of the random filter estimand.

3 Chivo Wallet

This section details the key events and motivations of the Chivo Wallet rollout, develops a rationale for expecting benefits from the intervention, and explains why it failed.[10]

[10]This section builds on ground-truths in Alvarez et al. (2024) via interviews with individuals at PADE-COSMS, Credicampo, SPTF, and ASEI—four organizations providing financial services to disadvantaged communities in El Salvador—and our survey about the technical issues Chivo Wallet users faced.

3.1 The boom and bust of Chivo Wallet

In September 2021, the El Salvadoran government deployed "Chivo Wallet," a mobile banking application that allowed citizens to transact with businesses, save money securely, and send money to others. Those who enrolled received $30—about 8.5% of the median monthly salary in El Salvador—for creating an account. Account management and actions such as in-network transactions, transfers, and currency conversions carried no fees.[11,12] 53% of the adult population of El Salvador attempted to create an account by the end of 2021, with 40% of downloads occurring within the first month (Alvarez et al., 2024).

The launch of Chivo Wallet was sudden, and users had many issues when interacting with the network. Numerous programming errors in the phone application, ATM software, and server code prevented users from claiming the $30 sign-up bonus, interacting with ATMs, sending funds to other users, or paying for goods with Chivo Wallet. Claims of identity theft were also common as new users occasionally discovered that an account had already been created under their name. These programming and security issues were eventually fixed in March 2022, but by this time the majority of users had stopped using the app entirely. Interest in Chivo Wallet had declined due to the chaotic implementation, unease generated by political opposition and the volatility of bitcoin, and a lack of incentive beyond the initial sign-up bonus (*ASEI*; *Credicampo*; *PADECOSMS*; *SPTF*).

A second issue stemming from the hastiness of the launch was that many individuals and businesses had little to no instruction on the benefits of using Chivo Wallet unless they purposefully sought out this information.[13] Crucially, many individuals reasonably conflated Chivo Wallet with cryptocurrency and did not realize that an account was able to hold and transact in both dollars and bitcoin (*Credicampo*). Firms and would-be remittance-receivers largely shunned Chivo Wallet because they did not want to be exposed to the volatility of bitcoin, even though received bitcoin could be—and in practice, were—immediately converted to dollars (*Credicampo*). Had the government focused on delineating the functionality of Chivo Wallet and the network's [rather tenuous] relation to cryptocurrency, much of this confusion could have been avoided. A relaunch may be more effective with sufficient financial education, but it is unlikely that the government will attempt this due to the cost of the initial intervention (*ASEI*; *SPTF*).

[11]Chivo Wallet is compatible with the El Salvadoran tax authority, bank accounts, and certain decentralized finance applications, but transferring out of the Chivo network results in transaction fees.

[12]Users also received an 8% discount on gas bought using Chivo Wallet.

[13]This is not for a lack of support infrastructure, however. Government officials distributed materials to help users with the Chivo Wallet app and employees were stationed around Chivo ATMs to help users navigate the machines for at least a year after the launch. Additionally, the President of El Salvador, Nayib Bukele, was found providing technical support via Twitter during the initial launch of Chivo Wallet.

3.2 Chivo Wallet's wasted potential

The intervention's emphasis on cryptocurrency adoption hampered a push for financial inclusion with significant potential. The El Salvadoran government had created Chivo Wallet to facilitate and promote bitcoin usage under the 2021 "Bitcoin Law"—a play to attract outside investment during a period of exceptionally-low creditworthiness. The law established bitcoin as an alternative to the U.S. dollar for paying taxes and required businesses to accept bitcoin as legal tender (Alvarez et al., 2024).

This conflation of cryptocurrency propaganda and increased access to affordable banking made any real progress towards financial inclusion unlikely. The benefits from having a bank account would not come from cryptocurrency adoption because the structure of decentralized finance actively disenfranchises smaller players (Cong et al., 2023).[14] Access to decentralized finance has had no positive impact for users except in niche cases where individuals aim to escape hyperinflation—a problem which El Salvador has not faced since its adoption of the U.S. dollar in 2001—or perpetrate scams or fraud.[15]

Nevertheless, Chivo Wallet brought a high amount of attention to virtual currencies and banking in a short period of time (*Credicampo*). In a country where 64% of adults had access to a mobile phone but 70% of adults were unbanked prior to Chivo Wallet (Alvarez et al., 2024), the potential for improving financial inclusion was great. Access would reduce the risk of carrying cash and remove the need for rural customers to travel long distances for physical cash transfers or micro-financing (*PADECOSMS*). The reduction in remittance fees should have also posed a massive benefit. Remittances constitute nearly a quarter of El Salvadoran GDP, and El Salvadorans outside of the country were able to access the network and send [subsidized] funds to family members (Alvarez et al., 2024).

Chivo Wallet bears a resemblance to mobile money applications that have increased financial inclusion through access to peer-to-peer transfers via mobile phone accounts (Batista and Vicente, 2020). M-Pesa—the most recognizable mobile money system—had a similarly explosive start in 2007, with over 1.1 million Kenyans enrolling within eight months of its launch (Mbiti and Weil, 2011). M-Pesa was particularly successful in developing a resilient informal credit system (Jack et al., 2013; Jack and Suri, 2014), and digital traces of economic behavior allowed formal banking institutions to assess creditworthi-

[14] In a study of the Ethereum platform, Cong et al. (2023) shows that transactions, mining, and wealth are concentrated among a few large players and that the bidding structure that determines transaction costs forces disproportionate fees on smaller players. High percentage fees, congestion-induced fluctuations in transaction costs, misunderstandings regarding reserve prices, and volatility in the Ether token all contribute to diminished consumer surplus. These issues are present with all other popular cryptocurrencies as well.

[15] The general erosion of societal welfare due to decentralized finance makes the whole movement objectively reprehensible. For examples of the failed decentralized finance experiment, see the many investigations by Stephen Findeisen, Dan Olsen, or Molly White.

ness (Björkegren and Grissen, 2018). These gains in access to credit could be expected of Chivo Wallet as well, given that transaction costs were even lower than that of M-Pesa and all activity could be observed by the government.

There are many other reasons that the rollout of Chivo Wallet could have offered a viable pathway out of poverty. Expanding access to formal financial products has had a remarkable impact on asset accumulation, the ability to protect against income shocks, and the relaxation of credit constraints—relative to informal mechanisms such as storing cash at home or buying durable, but illiquid assets (Demirgüç-Kunt et al., 2015; Demirgüç-Kunt and Singer, 2017). The introduction of high-value savings products has led to improvements in financial well-being (Prina, 2015), higher income through enabled entrepreneurship (Dupas and Robinson, 2013; Schaner, 2018), increased workforce participation in response to higher returns on capital (Callen et al., 2019), and greater trust and engagement with financial institutions Bachas et al. (2021). Interpersonal financial relationships can generate additional positive spillovers to those connected to someone who is formally banked (Dupas et al., 2019). However, low-cost savings accounts are not necessarily sufficient for improving financial well-being, and positive results are very much context and mechanism-dependent (Dupas et al., 2018; de Mel et al., 2022). Chivo Wallet provides another case study to support this last point.

3.3 Personal finance survey and random filter data generating process

The Chivo Wallet story provides an excellent opportunity to demonstrate the effectiveness of the random filter. The threat to identification is that downloading the Chivo Wallet application was a voluntary choice and ostensibly driven by several factors that would also have some effect of financial well-being, thus it is likely that a naïve comparison of those who engaged with Chivo Wallet and those who did not would reveal a large positive treatment effect. But the impact of access to an institution that did not function effectively in the first six months of use should intuitively be much more muted than this. This paper ultimately demonstrates that this impact is very small—and likely zero—by filtering out the variation in treatment that contributes no causal interpretation.

In September 2022, on the anniversary of the Chivo Wallet launch, I collected data on Chivo Wallet enrollment, the barriers faced while using the application, and coarse measures of financial well-being with the help of an enthusiastic undergraduate assistant and CID Gallup, a Latin American enumerator and research company previously utilized by Alvarez et al. (2024). The survey generated a nationally-representative sample of 700 El Salvadoran residents that were eligible to download Chivo Wallet (i.e. had a phone

and were an adult) through randomized dialing of a large set of active phone numbers. The interviews occurred throughout a single week in September, with calls attempted throughout the day from 8:30 AM to 5:30 PM. Each number was tried up to three times, at different times of day. Respondents were asked to engage in an anonymous, three to five minute survey on "personal finances" that later pivoted towards questions on Chivo Wallet if the individual reported having attempted to create an account.

The partner enumerators first conducted a 100 person pilot sample and provided feedback on questions before the final survey. The most relevant finding in the pilot with respect to survey design was that most individuals could not provide a reliable dollar amount for a year-over-year change in savings or income since the initial launch, so we opted for Likert-scale questions to capture changing financial well-being in order to fulfil the primary empirical goal of demonstrating the mitigation of bias by the random filter. This would muddy the interpretation of the measured effect size in most cases, but presently the purpose is to show the data are consistent with a null hypothesis of no effect.

Table 1 shows that those who download Chivo Wallet are systematically different from those who do not. This should not be surprising given the previous demographic survey results from Alvarez et al. (2024). To illustrate, individuals who downloaded Chivo Wallet were more likely to already interact with formal banking institutions, use digital forms of payment, be young and male, and complete high school but not college.

The majority of those who did not download Chivo Wallet prefer to use cash. Alvarez et al. (2024) determines that this preference is due to privacy and security concerns. Most transactions in El Salvador are made with cash—an anonymous form of payment—and half of El Salvadorans use cash exclusively according to both surveys. In contrast, Chivo Wallet account transactions are not private or anonymous, as the accounts are linked to El Salvadoran identification and phone numbers. I find that these concerns, as well as a lack of trust in the application and perceived difficulty of using it were the three main reasons (cited by 75% of individuals) that someone chose not to download Chivo Wallet.

These systematic differences between treatment and control groups generate selection biases when considering differences in financial outcomes. Experience with formal financial institutions and interest in cryptocurrency are likely positively correlated with financial well-being (and improvements in well-being), as the former signals higher wealth and the latter signals higher discretionary income. Thus a naïve regression of the change in well-being on downloading Chivo Wallet will provide a positive, but biased result.

The design of the sample presents additional "collider biases" that must be overcome as well. Collider bias occurs when a *lack* of variation in a variable—either through explicit control or the structure of the data generating process—generates a spurious correlation

Table 1: Imbalance across treatment groups, balance across mediator groups

	Downloaded Chivo			Experienced no Chivo issue		
Variable	$T = 0$	$T = 1$	Difference	$M = 0$	$M = 1$	Difference
no savings	0.347	0.321	-0.026	0.301	0.330	0.029
	(0.477)	(0.467)	(0.041)	(0.460)	(0.471)	(0.044)
money in bank	0.318	0.392	0.075*	0.380	0.398	0.017
	(0.467)	(0.489)	(0.043)	(0.487)	(0.490)	(0.046)
money in house	0.324	0.245	-0.078**	0.276	0.232	-0.044
	(0.469)	(0.431)	(0.039)	(0.448)	(0.422)	(0.041)
unbanked	0.512	0.389	-0.123***	0.429	0.371	-0.059
	(0.501)	(0.488)	(0.043)	(0.497)	(0.484)	(0.046)
bank account	0.424	0.511	0.088**	0.442	0.542	0.101**
	(0.496)	(0.500)	(0.044)	(0.498)	(0.499)	(0.047)
bank trips/month	1.244	1.949	0.705***	1.908	1.967	0.059
	(1.728)	(2.144)	(0.181)	(2.390)	(2.029)	(0.202)
bank travel time	43.556	36.954	-6.602	39.443	35.941	-3.502
	(39.343)	(42.611)	(4.223)	(48.681)	(39.917)	(4.418)
credit card	0.112	0.145	0.034	0.172	0.134	-0.038
	(0.316)	(0.353)	(0.030)	(0.378)	(0.341)	(0.033)
remittance change	2.435	2.527	0.092	2.506	2.535	0.029
	(0.866)	(1.107)	(0.142)	(1.193)	(1.075)	(0.143)
cash vs digital	1.854	2.517	0.663***	2.426	2.557	0.131
	(1.235)	(1.476)	(0.132)	(1.419)	(1.501)	(0.142)
only uses cash	0.609	0.404	-0.205***	0.419	0.397	-0.022
	(0.490)	(0.491)	(0.046)	(0.495)	(0.490)	(0.047)
dollars vs bitcoin	1.066	1.391	0.325***	1.283	1.438	0.155**
	(0.275)	(0.769)	(0.064)	(0.665)	(0.806)	(0.074)
female	0.524	0.368	-0.156***	0.423	0.343	-0.080*
	(0.501)	(0.483)	(0.043)	(0.496)	(0.475)	(0.045)
$18 \leq$ age ≤ 39	0.382	0.564	0.182***	0.528	0.580	0.053
	(0.487)	(0.496)	(0.044)	(0.501)	(0.494)	(0.047)
$40 \leq$ age ≤ 62	0.482	0.355	-0.128***	0.337	0.362	0.025
	(0.501)	(0.479)	(0.043)	(0.474)	(0.481)	(0.045)
age ≥ 63	0.135	0.081	-0.054**	0.135	0.057	-0.078***
	(0.343)	(0.273)	(0.026)	(0.343)	(0.233)	(0.026)
primary school	0.447	0.325	-0.123***	0.288	0.341	0.052
	(0.499)	(0.469)	(0.042)	(0.454)	(0.475)	(0.044)
high school	0.241	0.360	0.119***	0.387	0.349	-0.038
	(0.429)	(0.481)	(0.041)	(0.488)	(0.477)	(0.045)
college	0.312	0.315	0.003	0.325	0.311	-0.015
	(0.465)	(0.465)	(0.041)	(0.470)	(0.463)	(0.044)
Observations	170	530	700	163	367	530

Notes: The mediator panel makes its comparison conditional on being treated. The remittance and currency variables have a Likert scale from one to five. A remittance value of three implies remittances received did not change in the year since the launch, and lower/higher numbers imply a decrease/increase. A cash vs digital score of three implies cash is used as frequently as digital payments, with a score of one meaning that the individual only uses cash; this logic applies to the dollars vs bitcoin score as well. Bank trips per month is a count, and travel time to a bank/ATM is in minutes. All other variables are binary. Not shown: there were no observed differences in treatment/mediator groups across geographical location (department).

between treatment and outcome (Donovan, 2024).[16] To think about collider bias properly, it pays to first consider the external validity of the results to come: since our sample contains no individuals who refused to be surveyed nor those who do not own a mobile phone, we are estimating a treatment effect only for those with phones, some free time when they were called, and higher-than-average interest in answering a survey about personal finance (i.e. some "conditional" treatment effect, as is implicit in most studies).[17]

However, collider bias is not an external validity concern, and in this setting it works like this: having a recent change in financial well-being may make someone more likely to volunteer to take the survey. But so does having more free time. Conditional on volunteering, these two factors become substitute reasons for appearing in the sample. Further, free time during the work week may imply something about income, education, or any of the other aforementioned issues that led to the differential download rates of Chivo Wallet across treatment and control groups. Thus a systematic correlation between treatment and outcome is born that is driven by the sampling process, rather than a treatment effect—and this could could ruin the internal validity of a result. Collisions like these are usually unannounced, yet unconsciously-managed successfully via control of other relevant confounding factors. In the current example, the random filter will handle this task.

The hastiness of the Chivo Wallet launch provides a unique way to mitigate these selection and collider biases. As mentioned previously, many users found out that—due to poor validation procedure within the application—their identity had already been used to collect the $30 incentive. For those who did receive the incentive, their were numerous programming errors that plagued the application with inoperable barriers for end-users. Theese complications were widespread, with a third of users revealing difficulties with withdrawing money from ATMs, making purchases, sending money to others, receiving remittances, and other technical glitches. Any potential impact of Chivo Wallet on financial well-being is therefore mediated by these early issues, satisfying (A1).

Table 1 shows that while there is a lack of balance in demographic and financial characteristics across treatment vs control groups, the realized value of the mediator is as good as randomly assigned with respect to these factors. This is because many of these issues had nothing to do with user error.[18] The exogeneity of these "barriers" is therefore evident,

[16]These are Angrist and Pischke's "bad controls" (Angrist and Pischke, 2009).

[17]So the survey may yield a slightly different treatment effect relative to that of the general adult population with cell phones. But this doesn't seem likely given that the demographic results here are consistent with the in-person Alvarez et al. (2024) sample. With the experience of hundreds of past surveys under their belt, CID Gallup simply claimed that El Salvadorans generally liked participating in phone surveys.

[18]Those who faced issues were slightly less likely to have banking or bitcoin familiarity and more likely to be older, but the odds of 3/19 rejections in group-level similarity given no true difference in these characteristics is 62%. Nonetheless, it could be expected that some issues could have been due to user error,

given that problems with Chivo Wallet cannot be confounded with the download decision or a change in finances. The volunteer decision is also unrelated to facing a barrier due to the generic framing of the survey. Thus (A2)-(A3) are also satisfied.

Since those who choose not to download Chivo Wallet will never face a problem with its use, the weaker support assumption (A4') holds, and thus the data generating process created by this simple single-wave phone survey facilitates the estimation of the ATT.[19]

3.4 Random filter estimates of the impact of Chivo Wallet

The null hypothesis tested in this paper is that downloading Chivo Wallet did not increase financial well-being one year after its launch.[20] From Section 2, the causal relationship between T (downloading Chivo Wallet) and Y (a change in financial well-being score) can be decomposed into two separately identifiable estimands representing the effect of T on M (access to a "functional" Chivo Wallet application) and the effect of M on Y (conditional on T). Following Bellemare et al. (2024), I use seemingly unrelated regressions to recover and combine estimates of these effects:[21,22]

$$M_i = \gamma + \delta \cdot T_i + \varepsilon_i \tag{14}$$

$$Y_i = \theta + \lambda \cdot M_i + \phi \cdot T_i + \nu_i \, . \tag{15}$$

From (A1)-(A3) and (A4'), the point-estimate of the effect of T on Y, ($\hat{\beta}_{RF} = \hat{\delta} \cdot \hat{\lambda}$), is identified. In small samples, bootstrapping allows us to recover the distribution of this point-estimate, and with larger samples, the delta-method approximation is permissible. I use the former approach here. Table 2 provides the random filter estimate of the Chivo Wallet effect alongside the [biased] estimate resulting from a naïve regression.

---

so controlling for these variables in the random filter regressions below would demonstrate robustness of results, since the identification assumptions can be conditionally met.

[19] For the same stylistic reason, Bellemare et al. (2024) identify the ATT in their setting as well.

[20] There is no reason to think the application would actively harm individuals, hence the one-sided test.

[21] A matching estimator provides an identical measurement of the ATT. The reduction in usable observations (since no individuals with $T_i = 0$ were impacted by the mediator) would typically decrease the precision of this estimate relative to regression, but in this setting, the variation in Y_i that is removed is orthogonal to what can be explained by M_i. This means that the matching estimator could theoretically be *more* precise in cases where treatment is predominantly driven by confounding factors. In this setting, the sample size and orthogonal variation effects are roughly equal in magnitude and the standard errors from either method are nearly identical, so this output was suppressed.

[22] Non-linear estimators can be used here as well if one wants to make the stronger independence assumptions in Section 2 rather than the weaker mean-independence versions. However, in the present ATT estimation, something like logit or probit cannot estimate the first stage since $M = 0$ whenever $T = 0$. Ultimately, the linear probability model specification will not make any predictions outside of the M and Y domains within the T and M domains since T and M are binary, and the sample size is large enough that $\hat{\beta}$ is t-distributed, so neither of the downsides typically associated with the linear model apply here.

Table 2: Random filter (via SUR) and OLS Chivo Wallet effect estimates

	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	

Notes: "finance score" refers to the Likert-scale question asking about improvements in financial well-being at the time of the survey, one year after the launch of Chivo Wallet. Bolded estimates are the naïve and random filter estimates of the effect of Chivo Wallet on financial well-being. Italicized estimates represent the relevant first and second stage SUR estimates that are multiplied to create the random filter estimate. The random filter estimate includes bootstrapped standard errors.

Table 2 presents little evidence that downloading Chivo Wallet made a lasting positive impact on financial well-being. A naïve observation (the OLS estimate) would suggest a significant positive impact, but this is driven by selection into treatment and selection into the sample. The random filter instead measures a precise zero for the treatment effect, owing to the fact that having access to a "functional" Chivo Wallet application did not have any meaningful impact on the financial well-being measure in the second stage of the SUR estimation.

The random filter decomposition allows us to determine the size of the selection bias because the spurious correlations driven by unobserved factors are captured by the second stage regression. Since the mediator intercepts all of the exogenous variation in treatment, the second stage regression only has the residual endogenous variation to assign to the "effect" of treatment. The non-causal relationship between treatment and outcome (controlling for the mediator) is unsurprisingly close to the naïve treatment effect estimate.

The financial score variable ranges from "significant decrease" (one) to "significant increase" (five) in perceived financial well-being between September 2021 and September 2022, with a score of three signalling no meaningful change. To put the magnitude of the two point estimates in context, an "effect" of 0.015 on the Likert scale is akin to claiming that ten individuals in the sample experienced a one-point improvement in this scale due to Chivo Wallet, while the naïve estimate of 0.301 equates to 210 people reporting a one-point improvement—an estimate that is off by a factor of 20. Applying the random filter estimate to the general population, we would reject any positive effect size greater than "one-point for 11% of the population" with a one-sided hypothesis test and 5% false positive rate. While this scale is somewhat coarse, we can rule out effects with the scope

needed to call Chivo Wallet a financial inclusion success because those most likely to use the application are not those that stood to gain the most from adopting it.[23]

From Alvarez et al. (2024), issues with Chivo Wallet did not discourage continued use, so we know that Chivo Wallet's impact was not reduced due to error-induced attrition. Instead, the behavioral response to the launch explains the mechanism behind this null result. The greatest stated motivator to downloading the application was the $30 incentive (Alvarez et al., 2024). My data show that 12% of all downloads were by individuals who downloaded solely to acquire this bonus—making this policy equivalent to an undersized unconditional cash transfer—which wasn't likely to have a lasting impact on its own.

The Chivo Wallet launch was not explicitly a push for financial inclusion—it was a push for cryptocurrency adoption, which complicated the rollout and eroded any potential long-term benefits. My data suggests that 76% of the eligible population downloaded Chivo Wallet, and 31% of those who downloaded the application encountered some sort of issue with it. An actionable policy implication of my findings suggests that, had El Salvadorans been subject to a pure "traditional" finance institution with the sole purpose of education and financial inclusion, there could have been significant successes with respect to continued use and eventually financial well-being.[24]

4 Discussion

In this paper, I provide a rigorous introduction to the random filter identification strategy and demonstrate its usefulness in settings where the identification assumptions of other popular approaches are not met. Even in settings with selection into treatment, the random filter can uncover the ATE or the ATT when an exogenous mediator facilitates the treatment effect on an outcome of interest. The random filter is a valuable contribution that can become a regular feature in applied econometrics instruction and expand the range of causal inferences and questions asked in empirical research.

The random filter expands the scope of "natural experiment" opportunities for causal inference. In the El Salvador example, there was no way to control who was treated and who was not, however, the "effective" treatment was determined by a random process outside of individual control. This allowed for the measurement of a treatment effect that

[23] Additionally, this study is able to detect an effect size of "one-point for 19% of the population" with 80% power against the null hypothesis of "no positive effect." If we apply the quasi-Bayesian approach in Lang (2023), the likelihood that the null is true (instead of the alternative 19%) given a test statistic of -0.2 is 81%, even with a prior of 10% (representing a strong belief in the effectiveness of Chivo Wallet).

[24] For example, if the government had simply applied subsidies to remittances via traditional bank transfers (a much simpler and cheaper policy), El Salvadorans could have experienced a significant and repeatable wealth shock in response to creating a savings account.

may have been thought to be unrecoverable, and it is likely that this data structure exists—or could be engineered—in many other settings. Any potential setting where researchers learned of an intervention after the fact but an exogenous factor modified the effectiveness of that intervention would yield a causal estimate with a simple retrospective study like the one in this paper. Even if researchers have relatively small research budgets and an inability to randomize treatment, observational studies with data that fit this motif can still meet the prerequisite assumptions for estimating treatment effects.

For researchers pursuing field experiments, one promising opportunity for the random filter is to measure "ambition effects." For example, in over-subscription designs for randomized controlled trials, it is ultimately a selected group of individuals upon which randomization is done. In this case, everyone in the treated and control groups have some level of ambition to receive the treatment that is different from the rest of the population. This would imply that the sample ATE measured by this design would be higher than the ATE for the general population. This is typically acceptable, as it often isn't desirable from a policy perspective to provide treatment to those who do not want it.

The over-subscription design can be used to estimate the impact of wanting to receive the treatment, which is distinct from the selection effect described above. In the case where data can be collected for those who did not enroll to be treated, we can compare outcomes for those who enrolled to those who did not, and use the random assignment to treatment—the random filter in this case—to remove the bias associated with selection. This ambition effect—which applies to everyone who enrolled for treatment regardless if they were treated—provides additional behavioral insight complementary to the ATE generated by the study. Summing the ATE and ambition effects—since they are independent—reveals the full impact of the treatment if one were to change an individual's mind about enrolling in the program of interest.

The random filter identification strategy has not yet been integrated into econometrics research or instruction. This is partly due to a dependence on prerequisite knowledge of a niche literature in computer science and a lack of empirical examples. But another more subtle setback is the incompleteness (à la Gödel) of the theory that generated early writings on a precursor to the random filter. DAGs and do-calculus are tools designed to determine the causal interpretability of an estimand given an assumed data generating process. This provides a claim of internal validity, but does not establish the population to which the causal effect applies. In response to these complications, I derive the random filter from scratch using language familiar to economists. This provides the prerequisite knowledge for widespread dissemination and employment in applied economics.

In empirical studies, causal inference typically considers variations on one of five es-

tablished identification strategies—randomization, matching, instrumental variables, regression discontinuity, and differences in differences (Angrist and Pischke, 2015). This paper introduces a sixth strategy—the *random filter*. But the random filter may be one of many methodological discoveries made possible by embracing a transdisciplinary approach to causality (Donovan, 2024). Model paradigms like DAGs and do-calculus make causal discovery—the determination of an estimation strategy given assumptions about a data generating process—fairly straightforward. The potential outcomes framework is not proficient here, but it can reveal crucial complementary information about the estimand of interest that DAGs and do-calculus are not capable of explaining. Using both of these modeling techniques together may empower researchers to answer new questions previously thought to be inaccessible.

References

Alvarez, F., Argente, D., and Van Patten, D. (2024). Are Cryptocurrencies Currencies? Bitcoin as Legal Tender in El Salvador. *Science*, Forthcoming.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press, Princeton.

Angrist, J. D. and Pischke, J.-S. (2015). *Mastering 'metrics: the path from cause to effect*. Princeton University Press, Princeton.

Bachas, P., Gertler, P., Higgins, S., and Seira, E. (2021). How Debit Cards Enable the Poor to Save More. *The Journal of Finance*, 76(4):1913–1957.

Batista, C. and Vicente, P. C. (2020). Adopting Mobile Money: Evidence from an Experiment in Rural Africa. *AEA Papers and Proceedings*, 110:594–598.

Bellemare, M. F., Bloem, J. R., and Wexler, N. (2024). The Paper of How: Estimating Treatment Effects Using the Front-Door Criterion. *Oxford Bulletin of Economics and Statistics*, Forthcoming.

Björkegren, D. and Grissen, D. (2018). The Potential of Digital Credit to Bank the Poor. *AEA Papers and Proceedings*, 108:68–71.

Callen, M., de Mel, S., McIntosh, C., and Woodruff, C. (2019). What Are the Headwaters of Formal Savings? Experimental Evidence from Sri Lanka. *The Review of Economic Studies*, 86(6):2491–2529.

Cong, L. W., Tang, K., Wang, Y., and Zhao, X. (2023). Inclusion and Democratization Through Web3 and DeFi? Initial Evidence from the Ethereum Ecosystem.

Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press.

de Mel, S., McIntosh, C., Sheth, K., and Woodruff, C. (2022). Can Mobile-Linked Bank Accounts Bolster Savings? Evidence from a Randomized Controlled Trial in Sri Lanka. *The Review of Economics and Statistics*, 104(2):306–320.

Demirgüç-Kunt, A., Klapper, L. F., Singer, D., and van Oudheusden, P. (2015). The Global Findex Database 2014: Measuring Financial Inclusion Around the World.

Demirgüç-Kunt, A. and Singer, D. (2017). Financial Inclusion and Inclusive Growth: A Review of Recent Empirical Evidence.

Donovan, P. (2024). Visualizing Causal Hypotheses in Environmental Economics. *Review of Environmental Economics and Policy*, Forthcoming.

Duflo, E., Glennerster, R., and Kremer, M. (2006). Using Randomization in Development Economics Research: A Toolkit.

Dupas, P., Karlan, D., Robinson, J., and Ubfal, D. (2018). Banking the Unbanked? Evidence from Three Countries. *American Economic Journal: Applied Economics*, 10(2):257–297.

Dupas, P., Keats, A., and Robinson, J. (2019). The Effect of Savings Accounts on Interpersonal Financial Relationships: Evidence from a Field Experiment in Rural Kenya. *The Economic Journal*, 129(617):273–310.

Dupas, P. and Robinson, J. (2013). Savings Constraints and Microenterprise Development: Evidence from a Field Experiment in Kenya. *American Economic Journal: Applied Economics*, 5(1):163–192.

Heckman, J. J., Ichimura, H., and Todd, P. (1998). Matching as an Econometric Evaluation Estimator. *The Review of Economic Studies*, 65(2):261–294.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies*, 64(4):605–654.

Heckman, J. J. and Pinto, R. (2022). The Econometric Model for Causal Policy Analysis. *Annual Review of Economics*, 14(1):893–923.

Holland, P. W. (1986). Statistics and Causal Inference. *Journal of the American Statistical Association*, 81(396):945–960.

Huntington-Klein, N. (2022a). *The effect: an introduction to research design and causality*. CRC Press, Taylor & Francis Group, Boca Raton.

Huntington-Klein, N. (2022b). Pearl before economists: the book of why and empirical economics. *Journal of Economic Methodology*, 29(4):326–334.

Imbens, G. W. (2020). Potential Outcome and Directed Acyclic Graph Approaches to Causality: Relevance for Empirical Practice in Economics. *Journal of Economic Literature*, 58(4):1129–1179.

Imbens, G. W. and Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–475.

Jack, W., Ray, A., and Suri, T. (2013). Transaction Networks: Evidence from Mobile Money in Kenya. *American Economic Review*, 103(3):356–361.

Jack, W. and Suri, T. (2014). Risk Sharing and Transactions Costs: Evidence from Kenya's Mobile Money Revolution. *American Economic Review*, 104(1):183–223.

Lang, K. (2023). How Credible is the Credibility Revolution?

Mbiti, I. and Weil, D. N. (2011). Mobile Banking: The Impact of M-Pesa in Kenya.

Pearl, J. (1995). Causal diagrams for empirical research. *Biometrika*, 82(4):669–688.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge University Press, Cambridge, U.K. ; New York.

Prina, S. (2015). Banking the poor via savings accounts: Evidence from a field experiment. *Journal of Development Economics*, 115:16–31.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701.

Rubin, D. B. (1977). Assignment to Treatment Group on the Basis of a Covariate. *Journal of Educational Statistics*, 2(1):1–26.

Schaner, S. (2018). The Persistent Power of Behavioral Change: Long-Run Impacts of Temporary Savings Subsidies for the Poor. *American Economic Journal: Applied Economics*, 10(3):67–100.