Supercompliers[*]

Matthew L. Comey, Joint Committee on Taxation
Amanda R. Eng, U.S. Census Bureau
Zhuan Pei, Cornell University and IZA[†]

August 2023

Abstract

In a binary-treatment instrumental variable framework, we define supercompliers as the subpopulation whose treatment take-up positively responds to eligibility and whose outcome positively responds to take-up. Supercompliers are the only subpopulation to benefit from treatment eligibility and, hence, are of great policy interest. Given a set of jointly testable assumptions and a binary outcome, we can completely identify the characteristics of supercompliers. Specifically, we require the standard assumptions from the local average treatment effect literature along with an outcome monotonicity assumption (i.e., treatment is weakly beneficial). We can estimate and conduct inference on supercomplier characteristics using standard instrumental variable regression.

Key Words: Local Average Treatment Effect; Principal Stratification; Compliers; Supercompliers

JEL Codes: C10; C26.

1   Introduction

Seminal studies by Imbens and Angrist (1994), Angrist and Imbens (1995), and Angrist, Imbens and Rubin (1996) establish the now well-known local average treatment effect (LATE) interpretation of an estimand from an instrumental variable regression. In the canonical setting with a binary instrument and binary treatment, the LATE is the average treatment effect for individuals who comply with treatment assignment (i.e., compliers). Unsurprisingly, methodological interest in describing compliers followed (Imbens and Rubin 1997; Abadie 2003), and estimation of complier characteristics, popularized by Angrist and Pischke (2009), is now widely implemented in empirical studies.[1] Describing the characteristics of compliers provides a direct answer to the question: who are induced to take up treatment when eligible? The exercise is informative of the external validity of the identified treatment effect. As Angrist and Pischke (2009) note: "if the compliant subpopulation is similar to other populations of interest, the case for extrapolating estimated causal effects to these other populations is stronger."

In this paper, we extend the complier literature and devise tools to provide a direct answer to a different but related question: who benefit from gaining treatment eligibility? This population consists of those whose treatment take-up responds positively to treatment eligibility *and* whose outcome improves following treatment take-up. In other words, it is the subset of compliers for whom treatment improves the outcome. We term this subpopulation "supercompliers".

Supercompliers are a key building block of a LATE. Under a set of testable assumptions and given a binary outcome, we show that the supercomplier population share is the intent-to-treat effect. And since the LATE is the ratio of the intent-to-treat and the first stage effects, it is equivalent to the supercomplier population share as a fraction of the complier population share. Put simply, the prevalence of supercompliers drives the LATE as we know it. Therefore, learning about supercompliers can be even more informative than learning about the broader complier subpopulation. In the case where supercomplier characteristics differ from complier characteristics, it would be more effective to target treatment eligibility to an external population similar to the supercompliers.

---

[1]Recent examples include Borghans, Gielen and Luttmer (2014), Dahl, Kostøl and Mogstad (2014), Dobbie, Goldin and Yang (2018), Finkelstein and Notowidigdo (2019), Ouss and Stevenson (Forthcoming), and Agan, Doleac and Harvey (Forthcoming).

As with compliers, supercompliers cannot be directly observed. However, we show that their characteristics are identified with expressions analogous to those of complier characteristics. Furthermore, our identification result gives rise to estimators for supercomplier average and distributional characteristics that can be easily implemented with standard instrumental variable regressions. We also show that analogous complier characteristics estimators can be similarly implemented using instrumental variable regressions, which come with analytical standard errors. In contrast, previous studies that estimated complier characteristics either did not report standard errors or computed standard errors using bootstrap.

Our key results follow from the standard LATE assumptions along with an additional assumption that we call outcome monotonicity. Outcome monotonicity is the outcome counterpart of the treatment monotonicity assumption in the LATE framework. It states that receiving treatment either does not affect the outcome or changes it in a single direction. In other words, treatment does not improve the outcome for some while harming others.

While outcome monotonicity is not universally tenable, it is quite plausible in many situations. It is reasonable to assume, for example, that a deworming drug does not lower a child's weight (Andrews and Shapiro 2021). More broadly, Institutional Review Board approval for experiments is only granted conditional on demonstrated minimal risks of harm to participants. Other studies impose outcome monotonicity restrictions as well (e.g., Manski 1997, Manski and Pepper 2000, and Lee 2009), which we discuss in more detail in Section 2.3 below.

In other situations, there may be ambiguity as to whether outcome monotonicity holds. Hence, it is important that we be able to assess the validity of our identifying assumptions and understand the consequences of violation. We first develop tools that extend results by Kitagawa (2015) and Mourifié and Wan (2017) and provide an easy-to-implement joint test of the standard LATE assumptions and outcome monotonicity. Our test mirrors a result by Machado, Shaikh and Vytlacil (2019) on the identification of a positive average treatment effect under those same LATE assumptions. Then, to understand the consequences of outcome monotonicity violations, we show that the bias in our estimand is linear in the share of the complier subpopulation harmed by treatment take-up. This result implies that the bias is likely small when few individuals violate the outcome monotonicity assumption.

In the majority of our results, we also impose that the outcome be binary. However, we do not view this requirement as particularly restrictive. Binary outcomes are quite common in

economic research (e.g., employment, welfare receipt, graduation from college). Furthermore, as argued by Balke and Pearl (1997) who also focus on binary outcomes, we can easily transform a multi-valued or continuous outcome into a binary variable by categorizing its value as "high" or "low". Moreover, our estimand still identifies a weighted average of supercomplier characteristics in the case of a non-binary outcome, where the weight on each individual is proportional to her treatment effect.

We view analysis of supercomplier characteristics as complementary to subsample heterogeneity analysis of a treatment effect, just like describing complier characteristics is complementary to subsample heterogeneity analysis of treatment take-up. The methodological literature on heterogeneity analysis is advancing quickly, with the application of machine learning techniques at the forefront (e.g., Athey and Imbens 2016, Wager and Athey 2018, Chernozhukov et al. 2020, Knaus, Lechner and Strittmatter 2020; see Smith 2022 for a recent survey). Compared to these approaches, we acknowledge that estimating supercomplier characteristics has a drawback in the need for the outcome monotonicity and binary outcome assumptions. But our approach has several advantages too. First, it allows imperfect compliance with treatment, from which the recent machine learning methods abstract away. Second, it avoids any approximation of the conditional expectation of potential outcomes or treatment effects with covariates. As a result, our result does not depend on the choice of a particular machine learning algorithm or tuning parameters therein (Chernozhukov et al. 2020, for example, document that different machine learning techniques perform differently in their empirical application; see also Knaus, Lechner and Strittmatter 2021 for an extensive exercise comparing techniques across many data generating processes). Finally, our approach is easily implemented using existing statistical software commands, incurs low computational cost, and is transparent to understand.

We illustrate our proposed method using the experimental study of Ashraf et al. (2015), where U.S.-based migrants were randomly offered savings accounts in their home country with varying degrees of control over finances. We estimate mean characteristics of super-compliers, defined here as individuals induced to both sign up for an account and save in the account. Our results lead to new hypotheses to explain the substantial treatment effects found in the original paper.

2 Identification and Estimation

2.1 Theoretical Framework and Identification

We begin with an extension of the Rubin (1974) potential outcomes framework with a binary instrument $Z \in \{0, 1\}$, binary treatment $D \in \{0, 1\}$, and binary outcome $Y \in \{0, 1\}$. Let D_z represent the potential treatment status for an individual when she is assigned the instrument value z. Let Y_{zd} represent a potential outcome for $Z = z$ and $D = d$. With these notations, where each individual's D and Y do not depend on the Z and D of other individuals, we implicitly assume the stable unit treatment value assumption (e.g., Cox 1958).

Throughout this paper, we also maintain the other standard assumptions for the identification of a local average treatment effect (LATE) by Angrist, Imbens and Rubin (1996).

Assumption 1 (IV)

1. *Random Assignment:* $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, D_0, D_1) \perp\!\!\!\perp Z$ *and* $0 < \Pr(Z = 1) < 1$.

2. *Exclusion:* $\Pr(Y_{1d} = Y_{0d}) = 1$ *for* $d \in \{0, 1\}$.

3. *Treatment Monotonicity:* $\Pr(D_1 \geq D_0) = 1$.

4. *First Stage:* $\Pr(D_1 = 1) > \Pr(D_0 = 1)$.

With Assumptions 1.1 and 1.2 along with binary D and Y, we can partition the population into 16 unobserved subpopulations or groups based on potential treatment and potential outcome values. These 16 groups also appear in Balke and Pearl (1993) (the working paper version of Balke and Pearl 1997) and Chen and Flores (2015), and we refer to this partition as the "extended principal stratification". First, we categorize individuals by how treatment D responds to assignment Z. This corresponds to the well-known principal strata (Frangakis and Rubin 2002) in the LATE context, comprising "<u>a</u>lways takers", "<u>n</u>ever takers", "<u>c</u>ompliers", and "<u>de</u>fiers". Second, the exclusion assumption allows us to extend the principal stratification to incorporate how outcome Y responds to treatment D using the same taxonomy. Correspondingly, we define $Y_d \equiv Y_{zd}$ for $z \in \{0, 1\}$ to simplify notation. We index each of the 16 groups G by the pair to, where $t \in \{a, n, c, f\}$ indexes the four strata based on treatment response and $o \in \{a, n, c, f\}$ indexes the four strata based on outcome response. For each value of G, all individuals have the same vector (D_0, D_1, Y_0, Y_1). We fully

define the 16 groups in Table 1. Supercompliers correspond to the subpopulation $G = cc$, where $D_1 > D_0$ and $Y_1 > Y_0$. Assumptions 1.3 and 1.4 place restrictions on the shares of these subpopulations. Most importantly, Assumption 1.3 rules out all treatment defiers, i.e. $G \in \{fa, fn, fc, ff\}$ (for clarity, we use the word "treatment" as a qualifier when referring to the conventionally defined always takers, never takers, compliers, and defiers). Together, the two assumptions also require a nonzero share of treatment compliers. Next, we impose an additional assumption that further restricts the shares of certain subpopulations from Table 1.

Assumption 2 (Outcome Monotonicity and Reduced Form)

1. *Outcome Monotonicity:* $\Pr(Y_1 \geq Y_0) = 1$.

2. *Reduced Form:* $\Pr(G = cc) \equiv \Pr(Y_1 > Y_0, D_1 > D_0) > 0$.

Assumption 2 is the outcome analog of Assumptions 1.3 and 1.4. Assumption 2.1 rules out the existence of all outcome defiers (i.e., $G \in \{af, nf, cf, ff\}$). Together, Assumption 1.3 and Assumption 2.1 rule out 7 of the 16 groups, and the 9 remaining groups are bolded in Table 1 (in contrast, Balke and Pearl 1997 do not rule out groups with monotonicity assumptions but obtain partial identification of the average treatment effect instead). While Assumption 1.4 implies a nonzero population share of treatment compliers and consequently a nonzero first stage, Assumption 2.2 requires a nonzero population share of supercompliers, which implies a nonzero reduced form as shown in Lemma 1 below. (Along with Assumptions 1.1-1.3, Assumption 2.2 implies Assumption 1.4.)

Lemma 1. *Under Assumptions 1 and 2.1, the supercomplier share is identified by:*

$$\Pr(G = cc) = E[Y|Z = 1] - E[Y|Z = 0].$$

All proofs are in the Appendix.

Lemma 1 states that the share of the supercompliers is identified by the reduced form. This corresponds naturally to the well-known result that the share of treatment compliers is identified by the first stage. Since the reduced form is the product of the local average treatment effect and the first stage, the LATE (given a binary outcome) is simply the share of supercompliers as a fraction of the treatment compliers. This result is intuitive: only

the outcome compliers have a nonzero treatment effect (their treatment effect is equal to one), while all other admissible treatment compliers have a zero treatment effect. The more supercompliers there are, the higher the LATE.

In fact, the supercomplier group is the only subpopulation under Assumptions 1 and 2 whose outcome changes with the assignment Z. In other words, they are the only ones who benefit from being assigned to the treatment group. As such, learning about supercompliers should be of great importance to policy makers.

As with compliers, we cannot directly identify members of the supercomplier subpopulation. However, we can identify the distribution of their characteristics using observed data. Let X be a variable determined prior to random assignment (i.e., prior to the realization of Z), so it is reasonable to assume that jointly with (Y_1, Y_0, D_1, D_0), X is independent to Z (hereafter we use $X \perp\!\!\!\perp Z$ as a shorthand for this joint independence). For simplicity, we consider the case of a one-dimensional X, but many of our results generalize to a covariate vector of any dimension. Let h be a function such that $E[|h(X)|] < \infty$.

Proposition 1 *Under Assumptions 1 and 2 and provided that $X \perp\!\!\!\perp Z$, the supercomplier average of $h(X)$ is identified by:*

$$E[h(X)|G = cc] = \frac{1}{RF}E[\pi h(X)], \tag{1}$$

where $RF \equiv E[Y|Z = 1] - E[Y|Z = 0]$, and $\pi \equiv \kappa - (\kappa_0 Y + \kappa_1(1 - Y))$ with

$$\kappa \equiv 1 - \frac{D(1 - Z)}{\Pr(Z = 0)} - \frac{(1 - D)Z}{\Pr(Z = 1)}$$

$$\kappa_0 \equiv \frac{(1 - D)(1 - Z)}{\Pr(Z = 0)} - \frac{(1 - D)Z}{\Pr(Z = 1)}$$

$$\kappa_1 \equiv \frac{DZ}{\Pr(Z = 1)} - \frac{D(1 - Z)}{\Pr(Z = 0)}.$$

The supercomplier average can also be identified by a Wald-type estimand:

$$E[h(X)|G = cc] = \frac{E[h(X)Y|Z = 1] - E[h(X)Y|Z = 0]}{E[Y|Z = 1] - E[Y|Z = 0]}. \tag{2}$$

The weights κ, κ_0, and κ_1 are the unconditional counterpart of those defined by Abadie (2003) in his study of compliers. Lemma 2 in Appendix A.1 provides the analog of Theorem

3.1 from Abadie (2003) under our Assumption 1. While Abadie (2003) assumes the LATE assumptions hold conditional on X, our Lemma 2 shows that all three weights can still be used to identify the X distribution of the compliers when the assumptions hold unconditionally. Lemma 2 also shows that κ_0 (κ_1) can still be used to identify the Y_0 (Y_1) distribution of the compliers under our assumptions. In Remark 7 below, we consider identification under conditional independence, such as in a stratified experiment.

Remark 1. (**Compliers and Supercompliers**) There is a natural parallel between the identification results for supercomplier characteristics above and those for complier characteristics from the literature. While the κ weights from Abadie (2003) "find compliers" (Angrist and Pischke 2009), our π weights find supercompliers: heuristically, the κ weight starts from the population and subtracts the treatment always-takers and never-takers, while the π weight starts from the treatment compliers and subtracts the $G = ca$ and $G = cn$ groups. There is a complier analog to equation (2) as well: Kline and Walters (2016) and Marbach and Hangartner (2020) show that the complier characteristics can be identified by replacing Y with D in (2).

Remark 2. (**Perfect Treatment Compliance**) Our estimands from Lemma 1 and Proposition 1 are applicable to the case where $D = Z$. This arises when there is perfect compliance with treatment assignment.[2] Here, the supercomplier share from Lemma 1 is equal to the population average treatment effect. And the expressions for supercomplier characteristics identification under perfect treatment compliance are isomorphic (i.e., identical up to variable labels) to those for complier characteristics identification under treatment noncompliance.

Remark 3. (**Identification of the Distributions of Characteristics**) When h is the identity function, Proposition 1 says that we can identify the average characteristics of the supercompliers. We can also choose $h = 1_{[X \leq x]}$ for all $x \in \mathbb{R}$ and identify the entire c.d.f. of X among the supercompliers. More generally, when X is k-dimensional, letting $h = 1_{[X \leq x]}$ for $x \in \mathbb{R}^k$ identifies the joint distribution of the random vector X among the supercompliers.

Remark 4. (**Relaxing the Binary-Y Restriction**) For non-binary Y, we can generalize the definition of supercompliers to individuals with $D_1 > D_0$ and $Y_1 > Y_0$. In this case, the reduced-form estimand in Lemma 1 identifies the supercomplier population share scaled by the average treatment effect among the supercompliers. Similarly, the Wald estimand in

---

[2]This is the setting Kowalski (2020) considers, although she adheres to the conventional nomenclature and studies "defiers" where "take-up" is defined using an outcome variable.

Proposition 1 identifies a weighted average of supercomplier characteristics, where the weight of each individual is proportional to her treatment effect. See Appendix A.2 for details.

Remark 5. (**Share and Characteristics of Other Groups**) We can also identify the shares and characteristics of the other two groups within the unobserved treatment complier population: $G = ca$ and $G = cn$. See Appendix A.3 for details.

2.2   Estimation and Inference

The Wald-type equation (2) can be estimated via a two-stage least squares (2SLS) regression. For example, when $h(X)$ is the identity function—corresponding to the mean characteristics of supercompliers—this regression can be estimated in Stata by using the command

$$\texttt{ivregress 2sls XY (Y = Z) [, options]} \tag{3}$$

where `XY` is a variable defined as the product of X and Y.

However, equation (3) does not estimate an exact sample analog of the Abadie-style equation (1) from Proposition 1. As we show in Appendix A.4, the estimand from equation (1) also has a Wald-type representation similar to (2):

$$\frac{E[h(X)\{Y - (1 - \tau)\}|Z = 1] - E[h(X)\{Y - (1 - \tau)\}|Z = 0]}{E[Y - (1 - \tau)|Z = 1] - E[Y - (1 - \tau)|Z = 0]}, \tag{4}$$

where $\tau \equiv \Pr(Z = 1)$ denotes the proportion of units assigned to treatment. Therefore, the sample analog of (1) can also be implemented using 2SLS. But we need to first transform Y by subtracting from it the proportion of observations in the control group, and then use this transformed outcome variable in the 2SLS regression (it turns out that we can ignore the sampling variation in estimating τ when conducting inference). As discussed in Appendix A.4, of the two estimators based on (1) and (2), neither has an asymptotic variance that dominates the other in all data generating processes (DGPs). For convenience, we simply use the sample analog of (2) as the supercomplier characteristics estimator in Section 3.

Estimators based on complier analogs of both (1) and (2) have been used in existing studies, for which Y is replaced by D. Angrist and Pischke (2009) implement the complier analog of (1). Recent papers referenced in the introduction (e.g., Finkelstein and Notowidigdo 2019) use a different estimator, which is equivalent to the complier analog of (2) (see Appendix A.4

for details). However, these studies do not use 2SLS regression to estimate complier char-
acteristics. Instead, implementation involves assembling separately estimated quantities. In
addition, the studies either do not report standard errors on estimated complier character-
istics or report bootstrapped standard errors. Our results here imply that inference results
can be easily obtained for both estimators using existing Stata commands, including those
that account for a weak instrument.[3]

Remark 6. (Characteristics Distribution and Quantile Estimation) As we discussed
in Remark 3, we can identify the entire X distribution among supercompliers. It is straight-
forward to estimate the c.d.f. of X: we can replace the dependent variable in (3) with the
product of the indicator function $1_{[X \leq x]}$ and Y and estimate a series of 2SLS regressions
by varying x. To estimate the supercomplier quantiles of X, we can minimize a weighted
sum of the check function per Bassett and Koenker (1982). However, the same challenge
confronting the complier quantile estimation by Abadie, Angrist and Imbens (2002) is also
present here—since the individual π weight may be negative, the sample objective function
is usually non-convex and is therefore difficult to minimize. Abadie, Angrist and Imbens
(2002) overcome this challenge by using weights conditional on (D, Y, X). Directly applying
this strategy to the supercomplier setting does not lead to nonnegative weights, but applying
a modified strategy works, in which we use the π weights only conditional on (Y, X) or just
X. See Appendix A.5 for details.

Remark 7. (Conditional Independence and Stratified Randomization) Our iden-
tification and estimation results can be naturally generalized to accommodate cases where
independence (Assumption 1.1) holds conditionally on covariate set W — we just need to
add W into the conditioning set in Lemma 1 and Proposition 1. A common situation that
calls for conditional independence is stratified randomized experiments, in which researchers
typically include stratum fixed effects in treatment effect regressions (Bruhn and McKenzie
2009). It is then natural to also include the stratum fixed effects in the IV regression es-
timating supercomplier characteristics. In Appendix A.6, we follow Blandhol et al. (2022)
to show that the resulting population regression coefficient still identifies a non-negatively
weighted average of supercomplier characteristics across strata.

Remark 8. (Extension to Other Research Designs) Our work extends easily to regres-

---

[3]For estimating complier characteristics, a weak instrument has the standard meaning—Z fails to gen-
erate sufficient variation in D. For supercompliers, we have a weak instrument problem if Z fails to generate
sufficient variation in Y.

sion discontinuity (RD) designs. We can identify the characteristics of the supercompliers local to the policy cutoff and estimate it using the fuzzy RD counterpart of the 2SLS regression in (3). Our work may also be extended to difference-in-differences designs with a modified set of assumptions, including no anticipation (e.g., Roth and Sant'Anna 2022) and a stronger parallel trends assumption based on independence (e.g., Athey and Imbens 2006).

2.3   The Outcome Monotonicity Assumption

While Assumption 1 is standard in the RCT literature, Assumption 2.1 (outcome monotonicity) warrants more discussion. First, we point out that previous studies have maintained similar assumptions. For example, in their influential studies of treatment effect partial identification, outcome monotonicity is what Manski (1997) and Manski and Pepper (2000) refer to as "monotone treatment response" and what Lee (2009) refers to simply as "monotonicity". In their motivating examples, outcome monotonicity is taken to mean that the demand curve is weakly downward sloping (Manski 1997), that education does not decrease wages (Manski and Pepper 2000), or that participating in the Job Corps training program does not lower employment (Lee 2009).[4] Additionally, outcome monotonicity is implicitly assumed in the classic constant parameter endogenous treatment model of Heckman (1978). As with Assumption 1, the practical plausibility of Assumption 2 depends on the context. For example, it is quite plausible for the relationship between training program participation and subsequent employment or between health insurance coverage and doctor visits to be weakly positive. However, there is more ambiguity concerning the relationship between, say, health insurance coverage and out-of-pocket medical spending. Health insurance may lead to significant savings during emergency room visits, but it may also incentivize healthcare utilization and lead to higher spending.

Given uncertainty in the plausibility of this and the other identifying assumptions, researchers would be prudent to test them. We state such a test below.

---

[4]Chen and Flores (2015) extend Lee (2009) to bound treatment effects under imperfect compliance. Their monotonicity assumption is akin to Jobs Corps *take-up* not lowering employment for the treatment compliers, while Lee (2009) assumes *assignment* to Jobs Corps does not lower employment. Under treatment monotonicity, these assumptions are observationally equivalent.

2.3.1 Assumption Testing

We build on and extend results from Kitagawa (2015) and Mourifié and Wan (2017) and propose a sharp characterization of Assumptions 1.1-1.3 and 2.1.[5] The resulting test takes the form of a set of inequalities that must jointly hold.[6] Consistent with Kitagawa (2015), "sharp" here means that if the inequalities hold, then we can construct a data generating process which satisfies Assumptions 1 and 2 and rationalizes the observed data.

Formally,

Proposition 2 *Given the potential outcomes model described in Section 2.1, (i) under Assumptions 1.1-1.3 and 2.1, the following inequalities hold*

$$\Pr(Y = 0, D = 1 | Z = 1) - \Pr(Y = 0, D = 1 | Z = 0) \geq 0 \tag{5}$$

$$\Pr(Y = 1, D = 0 | Z = 0) - \Pr(Y = 1, D = 0 | Z = 1) \geq 0 \tag{6}$$

$$\Pr(Y = 1 | Z = 1) - \Pr(Y = 1 | Z = 0) \geq 0; \tag{7}$$

(ii) if inequalities (5)-(7) hold, there exists a joint distribution of $(Y_{11}, Y_{10}, Y_{01}, Y_{00}, D_1, D_0, Z)$ *that satisfies Assumptions 1.1-1.3 and 2.1 and induces the observed distribution of* (Y, D, Z).

Our identification results for the size of the treatment complier subpopulations (Lemma 1 and Proposition 5, introduced below) reveal an intuitive interpretation of inequalities (5)-(7). The left-hand-sides identify the population shares of the *cn*, *ca*, and *cc* groups, respectively, under our assumptions. Therefore, the testable implication of the identifying assumptions is simply that these quantities are weakly positive, which may not be the case if, for example, the disallowed subpopulation shares are positive.

Our test is related to a result presented in Machado, Shaikh and Vytlacil (2019), who study the identification of the sign of the average treatment effect in a LATE setting. Their Theorem 3.2(ii) establishes a set of inequalities that hold if and only if the standard LATE

---

[5]We exclude Assumption 1.4 (nonzero first-stage) and Assumption 2.2 (nonzero reduced-form) from the joint test below for two reasons. First, this exclusion is consistent with Kitagawa (2015) who only tests Assumptions 1.1-1.3 and not Assumption 1.4. Second, testing for nonzero first-stage and reduced-form is straightforward and a must-do in any empirical study.

[6]Note that testing treatment and outcome monotoncity (Assumptions 1.3 and 2.1) amounts to testing for the existence of treatment and outcome defiers, respectively. While we do not consider it here, Kowalski (2020) proposes a finite sample test of the existence of outcome defiers (Kowalski 2020 simply refers to them as defiers) in a perfect compliance framework. Extending her results to accommodate incomplete take-up is an avenue for future research.

assumptions are satisfied and the average treatment effect is nonnegative. In our binary outcome setting, the Machado, Shaikh and Vytlacil (2019) inequalities turn out to be identical to those in Proposition 2. While an assumption of a nonnegative average treatment effect is implied by and therefore weaker than our outcome monotonicity assumption, data cannot tell the two apart. Indeed, the proof of Proposition 2 reveals that all the information available for evaluating outcome monotonicity is contained within the reduced form estimate (Inequality 7), which translates to a nonnegative average treatment effect.

Testing the inequalities one-by-one is straightforward: Each requires a one-sided test based on a treatment-control comparison. To test all three inequalities jointly, we propose running a "stacked" regression, where each stack corresponds to a single inequality. That is: first, create three copies of the data; second, for each individual i, define $Y_i^1 = (1-Y_i)D_i$ $Y_i^2 = Y_i(1-D_i)$, and $Y_i^3 = Y_i$; and third, estimate the regression

$$Y_i^s = \phi^s + \theta_1 Z_i 1_{[s=1]} + \theta_2(1 - Z_i)1_{[s=2]} + \theta_3 Z_i 1_{[s=3]} + \varepsilon_i^s, \tag{8}$$

where s indexes each stack, ϕ^s is the stack specific constant, and standard errors are clustered at the individual level. The joint test amounts to testing whether $\theta \equiv \min(\theta_1, \theta_2, \theta_3)$ is nonnegative. Specifically, because the estimators $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$ are asymptotically multivariate normal with a covariance matrix we can consistently estimate, we can simulate the asymptotic distribution of $\hat{\theta}$ under the null. We reject the null hypothesis and conclude violations of the identifying assumptions when $\hat{\theta}$ is less than the 5th percentile in that distribution.

Remark 9. Inequality (7) corresponds to testing outcome monotonicity (Assumption 2.1). If we wish to test outcome monotonicity under the LATE Assumptions (Assumptions 1.1-1.3)—as opposed to testing all these assumptions jointly—we can simply rely on inequality (7) alone.

Remark 10. We can incorporate covariates to increase power in the outcome monotonicity test. Specifically, instead of using inequality (7) to test Assumption 2.1, we check whether the inequality holds when further conditioning on covariates. A simple way to implement this covariate-augmented test is to first discretize the covariate space (if necessary) and then jointly test the conditional analog of inequality (7) via a stacked regression, in which each stack corresponds to a value of the covariate vector. In Appendix B, we provide a proof of concept by showing that this test indeed rejects outcome monotonicity in a well known example by Bitler, Gelbach and Hoynes (2006, 2017) where Assumption 2.1 fails.

While the result from the exercise in Appendix B is reassuring, we acknowledge that our testing procedures cannot detect all violations of the identifying assumptions. By following the same construction as in the proof of Kitagawa (2015)'s Proposition 1.1(ii), we can show that for any observed (Y, D, Z) that satisfies inequalities (5)-(7), there exists an underlying joint distribution $(Y_{11}, Y_{10}, Y_{01}, Y_{00}, D_1, D_0, Z)$ that violates some of the identifying assumptions. If we use inequality (7) or its extension conditional on covariates as mentioned in Remark 10 to only test outcome monotonicity, we will fail to detect a violation if for every covariate value, the $G = cf$ group (or the "compfier" group) share is nonzero but is less than the $G = cc$ (supercomplier) share. Finally, like any other statistical validity test, we may lack the precision to detect a violation in a given sample. For all these reasons, we investigate the consequence of violating outcome monotonicity.

2.3.2 Relaxing Outcome Monotonicity

Proposition 3 below shows what the supercomplier characteristics estimand identifies when we relax outcome monotonicity (Assumption 2.1). It is analogous to Proposition 3 in Angrist, Imbens and Rubin (1996) on relaxing treatment monotonicity (our Assumption 1.3).

Proposition 3 *Under Assumptions 1 and 2.2 and provided that $X \perp\!\!\!\perp Z$,*

$$\frac{E[h(X)Y|Z=1] - E[h(X)Y|Z=0]}{E[Y|Z=1] - E[Y|Z=0]} = E[h(X)|G=cc] +$$
$$\xi \left\{ E[h(X)|G=cc] - E[h(X)|G=cf] \right\},$$

where we define

$$\xi \equiv \frac{\Pr(G=cf)}{E[Y|Z=1] - E[Y|Z=0]}.$$

Proposition 3 says that when outcome monotonicity does not hold, the estimand for supercomplier characteristics will be biased unless the supercompliers and the compfiers have the same average characteristics. The bias increases linearly with the compfier share, which can be interpreted as the degree of an outcome monotonicity violation. The bias is small when the degree of outcome monotonicity violation is low.

3   Empirical Application

To demonstrate the utility of the tools we have developed for characterizing supercompliers, we turn to the household savings experiment of Ashraf et al. (2015).[7] The study experimentally varies access to savings technology in order to investigate savings and remittance behavior among migrants to the United States. Based on previous research in development economics and economics of the household, Ashraf et al. (2015) hypothesize that migrants' inability to monitor or control the financial decisions of their household members in the home country results in low household savings. To test this hypothesis, the authors partnered with a bank in El Salvador and offered savings accounts with differing levels of control to migrants from El Salvador living in the Washington, D.C. area.

The experiment randomly assigned the promotion of two new savings accounts. The first account, called Cuenta Unidos, is a remittance account where the migrant can deposit savings to be accessed by a designated individual in El Salvador. In some treatment arms, treated migrants were given the ability to designate the Cuenta Unidos account as a joint savings account, accessible by both the migrant and the individual in El Salvador. The second account, called Ahorro Directo, resembles a traditional savings account, accessible to the migrant alone. In the experiment, individuals in the treatment group were given information about the accounts and offered assistance in opening them. Notably, both the joint Cuenta Unidos account and the Ahorro Directo account offer increased control over savings by migrants relative to the account for the remittance recipient only.

The experiment consisted of four treatment arms:

- Treatment 0 (comparison group): Migrants were visited by a marketer who encouraged them to make remittances to someone in El Salvador using a bank account. Migrants in this group could sign up for the accounts offered in the other treatments if they learned of the accounts from outside sources, but they did not receive any information or assistance in signing up.[8]

- Treatment 1: Marketers offered migrants assistance in setting up a Cuenta Unidos account. But the account could not be made into a joint account: migrants could only

---

[7]We obtain replication data from the Harvard Dataverse (Ashraf et al. 2014). We commend the authors of the original study for making their data available.

[8]We follow Ashraf et al. (2015) and use the term "comparison group", as opposed to "control group", to refer to Treatment 0. This is to avoid confusion, as "control" has a distinct meaning in the study.

deposit into the account and could not withdraw from it nor observe withdrawals.

- Treatment 2: Marketers offered migrants assistance in setting up a Cuenta Unidos account, and migrants could make the account a joint account fully accessible by both the migrant and the designated recipient.

- Treatment 3: Marketers offered migrants assistance in setting up a joint Cuenta Unidos account as in Treatment 2, and marketers also offered migrants assistance in setting up the migrant-only Ahorro Directo account.

Ashraf et al. (2015) find that all three treatments led to significantly higher savings in Cuenta Unidos accounts relative to the comparison group. In addition, individuals in Treatment 3 had significantly higher savings in Ahorro Directo accounts. However, only individuals in Treatment 3 had higher savings across all accounts held at the partner bank, which included preexisting accounts and accounts not offered as part of the experiment. To further study the role of control in savings decisions, the authors construct an indicator of demand for control based on migrants' answers to a baseline survey. The authors conclude that individuals who exhibited demand for control in the survey responded most to the offer of greater control over savings. In summarizing their findings, the authors write:

> [T]he treatment that offered migrants the greatest degree of control over El Salvador savings (offering both joint accounts and accounts in the migrant's name alone) led to substantial increases in savings at the partner bank. The increase in savings is likely due to enhanced control exerted by migrants; the effect of the treatment is significantly larger among migrants who report greater demand for such control in the baseline survey.

Our supercompliers analysis expands upon these findings.

To examine the characteristics of individuals who benefited from treatment assignment in the Ashraf et al. (2015) experiment, we define supercompliers as compliers who are induced to save in one of the new accounts. Note that households may have zero savings in their account, so take-up of an account is not equivalent to having savings in an account.[9] For brevity, we focus our attention on Treatment 2 (the group exposed to promotion of the joint Cuenta Unidos account) and Treatment 3 (the group exposed to promotion of both the joint

---

[9]We considered alternatively defining supercompliers as individuals induced to save over various dollar amounts. We find similar, although somewhat more muted, patterns of characteristics in these analyses.

account and the individual Ahorro Directo account). While not shown here, we implement our outcome monotonicity test and do not reject that treatment has a monotonically positive effect on the propensity to save in a Cuenta Unidos or Ahorro Directo account. This result is unsurprising as the accounts had not been offered to the public prior to the experiment, so there was no way for treated individuals to decrease their use of the accounts.

In our analysis, we deviate from the Ashraf et al. (2015) regression specification for practical reasons, but we show in Appendix Tables A.2 and A.3 that these deviations are not consequential for the conclusions they reach with respect to treatment effects and heterogeneous effects.[10] To describe the characteristics of supercompliers, we focus on the means of five variables: whether the individual has demand for control, whether the individual tracks spending, annual household income, whether the individual had any savings prior to the experiment, and individual baseline savings (transformed with a quartic root).

Table 2 corresponds to our analysis of Treatment 2 (the joint account treatment). Here, we compare the mean characteristics of supercompliers to the mean characteristics of compliers and the entire experimental population. Columns (1), (2), and (3) display these means for the population, compliers, and supercompliers, respectively, and columns (4), (5), and (6) show the pairwise differences across the first three columns. Heteroskedasticity-robust standard errors are in parentheses (we have also computed the Anderson-Rubin confidence intervals for the mean characteristics and found similar results). In the first row of the table, we see that compliers and supercompliers are more likely than the overall population to have demand for control. Compliers are 13 percentage points more likely than the overall population to have demand for control, while supercompliers are 40 percentage points more likely to have demand for control. While the differences are not statistically significant at the conventional level, the magnitudes are large and the pattern is broadly consistent with

---

[10]Specifically, Ashraf et al. (2015) estimate treatment effects in a single regression that includes all three treatment arms, and their heterogeneity analysis also pools all treatment arms. In contrast, we estimate supercomplier characteristics separately by treatment arm. We also consider binary savings outcomes instead of the continuous measures used in the original study.

Ashraf et al. (2015).[11] When we examine other characteristics we find additional differences between supercompliers and the overall population. In particular, presence and amount of baseline savings are substantially higher among supercompliers, as indicated in the final two rows of Table 2. Supercompliers are nearly three times as likely to have any savings at the partner bank prior to the experiment, relative to the overall population. This pattern, in addition to the fact that Treatment 2 did not significantly increase total savings, adds nuance to the hypothesis in Ashraf et al. (2015) that lack of ability to control and monitor savings discourages savings. While individuals who were induced to save in a Cuenta Unidos account had higher than average demand for control, they also had higher than average savings prior to being offered greater control over their savings.

We find similar patterns when we estimate the mean characteristics of individuals who took up and used Cuenta Unidos accounts in Treatment 3. Panel A of Table 3 displays these estimates. Again, we see that supercompliers are more likely to have demand for control than all individuals in Treatment 0 or Treatment 3. In addition, as with Treatment 2, supercompliers also have higher baseline savings.

Turning to the Ahorro Directo account, we find that individuals who took up and used these accounts have different characteristics than those who used the Cuenta Unidos accounts. In Panel B of Table 3, column (5) shows that individuals induced to save in an Ahorro Directo account have somewhat *lower* demand for control and baseline savings than the overall population, although none of these differences are statistically significant. In contrast, supercompliers have higher household income than the population and are significantly more likely to track their spending.

The findings in Panel B of Table 3 speak to another line of questioning in Ashraf et al. (2015): whether the same impacts could have been achieved by simply encouraging individuals to exert more control over their accounts or whether offering accounts and providing assistance to open those accounts was an integral part of the treatment. Through a follow up survey,

---

[11]Although the patterns we find for supercompliers' demand for control align with those found in Ashraf et al. (2015), readers may be surprised that they do not align more closely. Ashraf et al. (2015) find, for example, that individuals in Treatment 2 with demand for control were significantly more likely than those without demand for control to increase their savings in a Cuenta Unidos account, while we do not find significant differences in demand for control between supercompliers in Treatment 2 and the overall population. This difference may be partly driven by the fact that Ashraf et al. (2015) estimate reduced form regressions instead of instrumental variables regressions. As a result, their heterogeneity estimates reflect the relationship between demand for control and savings in addition to the relationship between demand for control and take-up of the accounts.

Ashraf et al. (2015) find suggestive evidence that providing assistance played an important role. The supercomplier characteristics estimates in Panel B offer additional suggestive evidence on this point: the combination of high income and low savings among supercompliers suggests that Treatment 3 was most effective for individuals who had enough income to be savers but faced other barriers to saving. Helping individuals set up a bank account is one way to overcome this barrier.

4   Conclusion

In this paper, we develop methods to characterize the supercomplier subpopulation in an instrumental variable framework with a binary instrument, treatment, and outcome. We show that under the standard LATE assumptions plus outcome monotonicity, we can completely identify characteristics of supercompliers. Because the plausibility of our identifying assumptions may depend on context, we propose a sharp joint test of their validity. In the presence of outcome monotonicity violations, we show that the average characteristics we identify are a linear combination of the characteristics of those benefiting from and those harmed by treatment receipt, implying that the bias will be small when few individuals are harmed by treatment receipt. Our identification results lead to natural estimators, which can be easily implemented using standard instrumental variable regression techniques via existing Stata commands. We also show that complier characteristics can be similarly estimated via instrumental variable regressions, without the need for bootstrapped standard errors. Finally, we note that analogous supercomplier characteristics estimators can be devised in other research designs, such as the regression discontinuity design.

Our methods complement existing tools in heterogeneity analysis and help to answer the question of who benefit from gaining treatment eligibility. We illustrate their utility in a randomized experiment by Ashraf et al. (2015), which intervened to incentivize migrant savings. A fruitful avenue for future research would be to derive treatment assignment rules based on supercomplier characteristics and study their efficacy.

References

Abadie, Alberto. 2003. "Semiparametric Instrumental Variable Estimation of Treatment Response Models." *Journal of Econometrics*, 113(2): 231–263.

Abadie, Alberto, Joshua Angrist, and Guido Imbens. 2002. "Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings." *Econometrica*, 70(1): 91–117.

Agan, Amanda Y., Jennifer L. Doleac, and Anna Harvey. Forthcoming. "Misdemeanor Prosecution." *Quarterly Journal of Economics*.

Andrews, Isaiah, and Jesse M. Shapiro. 2021. "A Model of Scientific Communication." *Econometrica*, 89(5): 2117–2142.

Angrist, Joshua D., and Guido W. Imbens. 1995. "Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity." *Journal of the American Statistical Association*, 90(430): 431–442.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics.* Princeton Univers. Press.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association*, 91(434): 444–455.

Ashraf, Nava, Diego Aycinena, Claudia Martínez A., and Dean Yang. 2014. "Replication data for: Savings in transnational households: A field experiment among migrants from El Salvador." Harvard Dataverse. https://doi.org/10.7910/DVN/27431.

Ashraf, Nava, Diego Aycinena, Claudia Martínez A., and Dean Yang. 2015. "Savings in transnational households: A field experiment among migrants from El Salvador." *Review of Economics and Statistics*, 97(2): 332–351.

Athey, Susan, and Guido Imbens. 2016. "Recursive partitioning for heterogeneous causal effects." *Proceedings of the National Academy of Sciences*, 113(27): 7353–7360.

Athey, Susan, and Guido W. Imbens. 2006. "Identification and Inference in Nonlinear Difference-in-Differences Models." *Econometrica*, 74(2): 431–497.

Balke, Alexander, and Judea Pearl. 1993. "Nonparametric Bounds on Causal Effects from Partial Compliance Data." Computer Science Department, University of California Los Angeles Technical Report R-199.

Balke, Alexander, and Judea Pearl. 1997. "Bounds on Treatment Effects from Studies with Imperfect Compliance." *Journal of the American Statistical Association*, 92(439): 1171–1176.

Bassett, Gilbert, and Roger Koenker. 1982. "An Empirical Quantile Function for Linear Models with iid Errors." *Journal of the American Statistical Association*, 77(378): 407.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2006. "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments." *American Economic Review*, 96(4): 988–1012.

Bitler, Marianne P., Jonah B. Gelbach, and Hilary W. Hoynes. 2017. "Can Variation in Subgroups' Average Treatment Effects Explain Treatment Effect Heterogeneity? Evidence from a Social Experiment." *Review of Economics and Statistics*, 99(4): 683–697.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky. 2022. "When is TSLS Actually LATE?" NBER Working Paper 29709 Working Paper.

Bloom, Dan, Susan Scrivener, Charles Michalopoulos, Pamela Morris, Richard Hendra, Diana Adams-Ciardullo, Johanna Walter, and Wanda Vargas. 2002. "Jobs First: Final Report on Connecticut's Welfare Reform Initiative." Manpower Demonstration Research Corporation.

Borghans, Lex, Anne C. Gielen, and Erzo F. P. Luttmer. 2014. "Social Support Substitution and the Earnings Rebound: Evidence from a Regression Discontinuity in Disability Insurance Reform." *American Economic Journal: Economic Policy*, 6(4): 34–70.

Bruhn, Miriam, and David McKenzie. 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*, 1(4): 200–232.

Chen, Xuan, and Carlos A. Flores. 2015. "Bounds on Treatment Effects in the Presence of Sample Selection and Noncompliance: The Wage Effects of Job Corps." *Journal of Business & Economic Statistics*, 33(4): 523–540.

Chernozhukov, Victor, Mert Demirer, Esther Duflo, and Iván Fernández-Val. 2020. "Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India." NBER Working Paper 24678.

Cox, David R. 1958. *Planning of Experiments.* Wiley, New York.

Dahl, Gordon B., Andreas Ravndal Kostøl, and Magne Mogstad. 2014. "Family Welfare Cultures." *Quarterly Journal of Economics*, 129(4): 1711–1752.

Dobbie, Will, Jacob Goldin, and Crystal S. Yang. 2018. "The Effects of Pre-Trial Detention on Conviction, Future Crime, and Employment: Evidence from Randomly Assigned Judges." *American Economic Review*, 108(2): 201–240.

Finkelstein, Amy, and Matthew J Notowidigdo. 2019. "Take-Up and Targeting: Experimental Evidence from SNAP." *Quarterly Journal of Economics*, 134(3): 1505–1556.

Frangakis, Constantine E., and Donald B. Rubin. 2002. "Principal Stratification in Causal Inference." *Biometrics*, 58(1): 21–29.

Heckman, James J. 1978. "Dummy Endogenous Variables in a Simultaneous Equation System." *Econometrica*, 46(4): 931.

Imbens, Guido W., and Donald B. Rubin. 1997. "Estimating Outcome Distributions for Compliers in Instrumental Variables Models." *Review of Economic Studies*, 64(4): 555–574.

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

Kitagawa, Toru. 2015. "A Test for Instrument Validity." *Econometrica*, 83(5): 2043–2063.

Kline, Patrick, and Christopher R. Walters. 2016. "Evaluating Public Programs with Close Substitutes: The Case of Head Start." *Quarterly Journal of Economics*, 131(4): 1795–1848.

Knaus, Michael C., Michael Lechner, and Anthony Strittmatter. 2020. "Heterogeneous Employment Effects of Job Search Programs: A Machine Learning Approach." *Journal of Human Resources*, 57(2): 597–636.

Knaus, Michael C., Michael Lechner, and Anthony Strittmatter. 2021. "Machine Learning Estimation of Heterogeneous Causal Effects: Empirical Monte Carlo Evidence." *Econometrics Journal*, 24(1): 134–161.

Kowalski, Amanda E. 2020. "A Model of a Randomized Experiment with an Application to the PROWESS Clinical Trial." NBER Working Paper 25670.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies*, 76(3): 1071–1102.

Machado, Cecilia, Azeem M. Shaikh, and Edward J. Vytlacil. 2019. "Instrumental Variables and the Sign of the Average Treatment Effect." *Journal of Econometrics*, 212(2): 522–555.

Manski, Charles F. 1997. "Monotone Treatment Response." *Econometrica*, 65(6): 1311–1334.

Manski, Charles F., and John V. Pepper. 2000. "Monotone Instrumental Variables: With an Application to the Returns to Schooling." *Econometrica*, 68(4): 977–1010.

Marbach, Moritz, and Dominik Hangartner. 2020. "Profiling Compliers and Noncompliers for Instrumental-Variable Analysis." *Political Analysis*, 28(3): 435–444.

Mourifié, Ismael, and Yuanyuan Wan. 2017. "Testing Local Average Treatment Effect Assumptions." *Review of Economics and Statistics*, 99(2): 305–313.

Ouss, Aurelie, and Megan Stevenson. Forthcoming. "Does Cash Bail Deter Misconduct?" *American Economic Journal: Applied Economics*.

Roth, Jonathan, and Pedro H. C. Sant'Anna. 2022. "When Is Parallel Trends Sensitive to Functional Form?" arXiv:2010.04814.

Rubin, Donald B. 1974. "Estimating the Causal Effects of Treatments in Randomized and Non-randomized Studies." *Journal of Educational Psychology*, 66(5): 688–701.

Smith, Jeffrey. 2022. "Treatment Effect Heterogeneity." *Evaluation Review*, 46(5): 652–677.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects using Random Forests." *Journal of the American Statistical Association*, 113(523): 1228–1242.

Table 1: Extended Principal Stratification

G-value	Treatment Type	Outcome Type	D_0	D_1	Y_0	Y_1
aa	**Always Taker**	**Always Taker**	**1**	**1**	**1**	**1**
an	**Always Taker**	**Never Taker**	**1**	**1**	**0**	**0**
ac	**Always Taker**	**Complier**	**1**	**1**	**0**	**1**
af	Always Taker	Defier	1	1	1	0
na	**Never Taker**	**Always Taker**	**0**	**0**	**1**	**1**
nn	**Never Taker**	**Never Taker**	**0**	**0**	**0**	**0**
nc	**Never Taker**	**Complier**	**0**	**0**	**0**	**1**
nf	Never Taker	Defier	0	0	1	0
ca	**Complier**	**Always Taker**	**0**	**1**	**1**	**1**
cn	**Complier**	**Never Taker**	**0**	**1**	**0**	**0**
cc	**Complier**	**Complier**	**0**	**1**	**0**	**1**
cf	Complier	Defier	0	1	1	0
fa	Defier	Always Taker	1	0	1	1
fn	Defier	Never Taker	1	0	0	0
fc	Defier	Complier	1	0	0	1
ff	Defier	Defier	1	0	1	0

Table 2: Characteristics of Compliers and Supercompliers: Joint Account Only Treatment

	Means			Differences		
	Population	Compliers	Supercompliers	Comp.-Pop.	Scomp. - Pop.	Scomp. - Comp.
	(1)	(2)	(3)	(4)=(2)-(1)	(5)=(3)-(1)	(6)=(3)-(2)
	A. Outcome: Any Savings in Joint Account					
Demand for control	0.51	0.63	0.91	0.13	0.40	0.27
	(0.02)	(0.13)	(0.30)	(0.13)	(0.30)	(0.23)
Tracks spending	0.45	0.54	0.58	0.10	0.13	0.04
	(0.02)	(0.12)	(0.22)	(0.12)	(0.22)	(0.15)
Annual household income	39,535	30,538	34,565	-8,997	-4,969	4,028
	(4,376)	(4,378)	(7,810)	(6,040)	(8,793)	(5,591)
Any baseline savings	0.26	0.58	0.73	0.31	0.47	0.15
	(0.02)	(0.13)	(0.28)	(0.13)	(0.28)	(0.20)
Avg. baseline savings, quartic	0.95	1.73	2.90	0.78	1.95	1.17
	(0.09)	(0.49)	(1.26)	(0.48)	(1.25)	(0.91)

Notes: Population means are estimated on all individuals in Treatment 2 and the comparison group (Treatment 0). Estimates and standard errors come from a stacked regression that simultaneously estimates characteristics for all three groups. Stacked regressions include fixed effects for marketer, stratification cell, and month of marketing visit when treatments were administered. Baseline savings are savings at the partner bank in the twelve months prior to treatment. The demand for control indicator is constructed from answers in the baseline survey. Household income and the indicator for tracking spending and budgeting expenses are also from the baseline survey. Household income is reported in US dollars. Baseline savings were originally reported in US dollars and are transformed using a quartic root.

Table 3: Characteristics of Compliers and Supercompliers: Joint and Migrant-Only Account Treatment

	Means			Differences		
	Population	Compliers	Supercompliers	Comp.-Pop.	Scomp. - Pop.	Scomp. - Comp.
	(1)	(2)	(3)	(4)=(2)-(1)	(5)=(3)-(1)	(6)=(3)-(2)
	A. Outcome: Any Savings in Joint Account					
Demand for control	0.52	0.70	0.81	0.18	0.28	0.10
	(0.02)	(0.10)	(0.14)	(0.10)	(0.14)	(0.08)
Tracks spending	0.46	0.40	0.27	-0.06	-0.19	-0.13
	(0.02)	(0.09)	(0.13)	(0.09)	(0.13)	(0.08)
Annual household income	36,843	48,891	35,185	12,048	-1,658	-13,706
	(1,429)	(8,398)	(5,006)	(7,739)	(4,961)	(7,930)
Any baseline savings	0.26	0.60	0.59	0.34	0.33	-0.01
	(0.02)	(0.10)	(0.13)	(0.09)	(0.13)	(0.07)
Avg. baseline savings, quartic	1.08	2.13	2.13	1.05	1.05	0.00
	(0.10)	(0.41)	(0.58)	(0.39)	(0.56)	(0.32)
	B. Outcome: Any Savings in Migrant-Only Account					
Demand for control	0.52	0.50	0.27	-0.02	-0.26	-0.24
	(0.02)	(0.09)	(0.18)	(0.08)	(0.18)	(0.14)
Tracks spending	0.46	0.61	0.99	0.15	0.53	0.38
	(0.02)	(0.09)	(0.19)	(0.08)	(0.19)	(0.17)
Annual household income	36,843	42,072	77,218	5,228	40,375	35,147
	(1,429)	(7,408)	(21,296)	(6,762)	(20,776)	(16,390)
Any baseline savings	0.26	0.33	0.10	0.07	-0.16	-0.23
	(0.02)	(0.08)	(0.16)	(0.08)	(0.16)	(0.13)
Avg. baseline savings, quartic	1.08	1.06	0.45	-0.02	-0.63	-0.61
	(0.10)	(0.40)	(0.97)	(0.39)	(0.97)	(0.72)

Notes: Population means are estimated on all individuals in Treatment 3 and the comparison group (Treatment 0). Estimates and standard errors come from a stacked regression that simultaneously estimates characteristics for all three groups. Stacked regressions include fixed effects for marketer, stratification cell, and month of marketing visit when treatments were administered. Baseline savings are savings at the partner bank in the twelve months prior to treatment. The demand for control indicator is constructed from answers in the baseline survey. Household income and the indicator for tracking spending and budgeting expenses are also from the baseline survey. Household income is reported in US dollars. Baseline savings were originally reported in US dollars and are transformed using a quartic root.

Appendix (For Online Publication Only)

A Additional Theoretical Results

A.1 Proofs of Lemma 1 and Propositions 1-3

Proof of Lemma 1: We start from the right hand side of the equation

$$
\begin{aligned}
&E[Y|Z = 1] - E[Y|Z = 0] \\
&= \Pr(Y = 1|Z = 1) - \Pr(Y = 1|Z = 0) \\
&= \Pr(G = aa, ac, na, ca, cc|Z = 1) - \Pr(G = aa, ac, na, ca|Z = 0) \\
&= \Pr(G = cc).
\end{aligned}
$$

The first equality follows from Y being binary. The second equality follows from Assumptions 1.2, 1.3, and 2.1. The last equality follows from Assumption 1.1 and the fact that the groups in Table 1 are mutually exclusive. *QED.*

To prove Proposition 1, we first state a lemma that provides the analog of Theorem 3.1 of Abadie (2003) under our Assumption 1.

Lemma 2. *Let $g(\cdot)$ be any function of (Y, D, X) such that $E[|g(Y, D, X)|] < \infty$. Under Assumption 1, provided that $X \perp\!\!\!\perp Z$, and with κ, κ_0, and κ_1 defined in Proposition 1,*

(a) $E[g(Y, D, X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} E[\kappa g(Y, D, X)]$. Also,

(b) $E[g(Y_0, X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} E[\kappa_0 g(Y, X)]$, and

(c) $E[g(Y_1, X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} E[\kappa_1 g(Y, X)]$.

We omit the proof of Lemma 2 as it follows the same line of reasoning as the proof of Theorem 3.1 in Abadie (2003). However, Lemma 2 is not a special case of Theorem 3.1 of Abadie (2003). Abadie (2003) assumes conditional independence, and only the covariates in the conditioning set appear in his Theorem 3.1. We assume unconditional independence, which means our conditioning set is empty, but we still have covariates in Lemma 2. We can generalize Lemma 2 to nest both our unconditional case and Abadie (2003) by allowing for an arbitrary conditioning set and additional covariates not in the conditioning set.

Proof of Proposition 1: To prove the identification result in equation (1), we note that under Assumption 2, the treatment complier population consists of three groups: $G = ca, cn, cc$. Thus,

$$E[h(X)|G = cc]\Pr(G = cc|D_1 > D_0)$$
$$= E[h(X)|D_1 > D_0] - E[h(X)|G = ca]\Pr(G = ca|D_1 > D_0)$$
$$- E[h(X)|G = cn]\Pr(G = cn|D_1 > D_0). \tag{A1}$$

We examine each of the three terms on the right-hand side of equation (A1). The first term is simply the average $h(X)$ for the treatment compliers. Per Lemma 2(a),

$$E[h(X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)}E[\kappa h(X)].$$

For the second term,

$$E[h(X)|G = ca]\Pr(G = ca|D_1 > D_0)$$
$$= E[h(X)|D_1 > D_0, Y_0 = Y_1 = 1]\Pr(Y_0 = Y_1 = 1|D_1 > D_0)$$
$$\overset{(i)}{=} E[h(X)|D_1 > D_0, Y_0 = 1]\Pr(Y_0 = 1|D_1 > D_0)$$
$$= E[h(X)Y_0|D_1 > D_0, Y_0 = 1]\Pr(Y_0 = 1|D_1 > D_0)$$
$$= E[h(X)Y_0|D_1 > D_0]$$
$$\overset{(ii)}{=} \frac{1}{\Pr(D_1 > D_0)}E[\kappa_0 Y h(X)],$$

where equality (i) follows from Assumption 2.1 and (ii) from Lemma 2(b).

Analogously, for the third term, we can show that

$$E[h(X)|G = cn]\Pr(G = cn|D_1 > D_0) = \frac{1}{\Pr(D_1 > D_0)}E[\kappa_1(1 - Y)h(X)].$$

Plugging these results back into equation (A1), we have

$$E[h(X)|G = cc] = \frac{E[\pi h(X)]}{\Pr(G = cc|D_1 > D_0)\Pr(D_1 > D_0)} = \frac{E[\pi h(X)]}{\Pr(G = cc)}.$$

Because Lemma 1 implies the denominator in the equation above to be the reduced form, RF, we have the desired result.

To prove equation (2), note that we can plug in the expressions of κ, κ_0, and κ_1 and rewrite the numerator in equation (1) as

$$
\begin{aligned}
&E[\pi h(X)]\\
=&E\left[h(X) - \frac{(1-D)Z}{\Pr(Z=1)}h(X) - \frac{D(1-Z)}{\Pr(Z=0)}h(X) - \frac{1-Z}{\Pr(Z=0)}Yh(X)\right.\\
&\left.+ \frac{D(1-Z)}{\Pr(Z=0)}Yh(X) + \frac{Z}{\Pr(Z=1)}Yh(X) - \frac{DZ}{\Pr(Z=1)}Yh(X)\right.\\
&\left.- \frac{DZ}{\Pr(Z=1)}h(X) + \frac{D(1-Z)}{\Pr(Z=0)}h(X) + \frac{DZ}{\Pr(Z=1)}Yh(X) - \frac{D(1-Z)}{\Pr(Z=0)}Yh(X)\right]\\
=&E\left[h(X) - \frac{Z}{\Pr(Z=1)}h(X) - \frac{1-Z}{\Pr(Z=0)}Yh(X) + \frac{Z}{\Pr(Z=1)}Yh(X)\right]\\
=&E[h(X)] - E[h(X)|Z=1] + E[h(X)Y|Z=1] - E[h(X)Y|Z=0]\\
=&E[h(X)Y|Z=1] - E[h(X)Y|Z=0],
\end{aligned}
$$

and equation (2) follows immediately. *QED*.

Proof of Proposition 2: To prove (i), we first introduce the notation $p_g \equiv \Pr(G=g)$ to denote the share of each of the 16 unobserved subpopulations in Table 1. Assumption 1.2 allows us to focus on these groups and use them as the fundamental building blocks. Under Assumptions 1.1, 1.3, and 2.1, we can translate these subpopulation shares into observed shares of subgroups based on the values of Y, D, and Z as stated in Table A.1 below.

Table A.1: Observed Subgroups and Makeup Under Assumptions

	Conditional on $Z=0$		Conditional on $Z=1$	
	$Y=0$	$Y=1$	$Y=0$	$Y=1$
$D=0$	$p_{nn}+p_{nc}+p_{cn}+p_{cc}$	$p_{na}+p_{ca}$	$p_{nn}+p_{nc}$	p_{na}
$D=1$	p_{an}	$p_{aa}+p_{ac}$	$p_{an}+p_{cn}$	$p_{aa}+p_{ac}+p_{ca}+p_{cc}$

Table A.1 makes clear six inequalities that must hold under Assumptions 1.1-1.3 and 2.1:

$$\Pr(Y=0, D=0|Z=1) \leq \Pr(Y=0, D=0|Z=0) \tag{A2}$$

$$\Pr(Y=0, D=1|Z=0) \leq \Pr(Y=0, D=1|Z=1) \tag{A3}$$

$$\Pr(Y=1, D=0|Z=1) \leq \Pr(Y=1, D=0|Z=0) \tag{A4}$$

$$\Pr(Y=1, D=1|Z=0) \leq \Pr(Y=1, D=1|Z=1) \tag{A5}$$

$$\Pr(Y=0|Z=1) \leq \Pr(Y=0|Z=0) \tag{A6}$$

$$\Pr(Y = 1|Z = 0) \le \Pr(Y = 1|Z = 1) \tag{A7}$$

The first four inequalities come from directly comparing cells, and the last two are made apparent by vertically adding up cells. The first four inequalities are implications of Assumptions 1.1-1.3 (they are the same as inequality 1.1 in Kitagawa 2015 and as inequalities 1 and 2 in Mourifié and Wan 2017 adapted to the binary-Y case). The last two inequalities come from Assumption 2.1: If only Assumptions 1.1-1.3 hold but not Assumption 2.1 (i.e. $p_{cf} \ne 0$), these two inequalities will fail whenever $p_{cf} > p_{cc}$.

Finally, three of these inequalities are redundant: they are implied by the other three. First, it is easy to see that inequalities (A6) and (A7) are equivalent by writing the probability of each event as one minus that of its complement. Second, (A3) and (A6) imply (A2):

$$\begin{aligned}
\Pr(Y = 0, D = 0|Z = 1) &= 1 - \Pr(Y = 1|Z = 1) - \Pr(Y = 0, D = 1|Z = 1) \\
&\le 1 - \Pr(Y = 1|Z = 0) - \Pr(Y = 0, D = 1|Z = 0) \\
&= \Pr(Y = 0, D = 0|Z = 0).
\end{aligned}$$

And third, an analogous argument shows that (A4) and (A7) imply (A5). Therefore, we omit inequalities (A2), (A5), and (A6), and the three remaining inequalities are inequalities (5), (6), and (7) in the statement of Proposition 2. This completes the proof of part (i).

The proof of part (ii) proceeds similarly to the corresponding proof of Proposition 1.1 in Kitagawa (2015). We construct the distribution of $(Y_{11}, Y_{10}, Y_{01}, Y_{00}, D_1, D_0, Z)$ as follows. First, we assign zero probability to the cases where Y_{1d} and Y_{0d} $(d = 0, 1)$ take on different values. This ensures that the resulting construction satisfies the exclusion restriction and allows us to specify the joint distribution by specifying the share of each group G within the treatment and control populations. We construct these group shares based on the observed (Y, D, Z) distribution as follows.

$$\begin{aligned}
\Pr(G = aa|Z = 0) &= \Pr(G = aa|Z = 1) \equiv \frac{1}{2}\Pr(Y = 1, D = 1|Z = 0) \\
\Pr(G = an|Z = 0) &= \Pr(G = an|Z = 1) \equiv \Pr(Y = 0, D = 1|Z = 0) \\
\Pr(G = ac|Z = 0) &= \Pr(G = ac|Z = 1) \equiv \frac{1}{2}\Pr(Y = 1, D = 1|Z = 0) \\
\Pr(G = na|Z = 0) &= \Pr(G = na|Z = 1) \equiv \Pr(Y = 1, D = 0|Z = 1) \\
\Pr(G = nn|Z = 0) &= \Pr(G = nn|Z = 1) \equiv \frac{1}{2}\Pr(Y = 0, D = 0|Z = 1) \\
\Pr(G = nc|Z = 0) &= \Pr(G = nc|Z = 1) \equiv \frac{1}{2}\Pr(Y = 0, D = 0|Z = 1) \\
\Pr(G = ca|Z = 0) &= \Pr(G = ca|Z = 1) \equiv \Pr(Y = 1, D = 0|Z = 0) - \Pr(Y = 1, D = 0|Z = 1)
\end{aligned}$$

4

$$\Pr(G = cn|Z = 0) = \Pr(G = cn|Z = 1) \equiv \Pr(Y = 0, D = 1|Z = 1) - \Pr(Y = 0, D = 1|Z = 0)$$
$$\Pr(G = cc|Z = 0) = \Pr(G = cc|Z = 1) \equiv \Pr(Y = 0|Z = 0) - \Pr(Y = 0|Z = 1)$$
$$\Pr(G = g|Z = 0) = \Pr(G = g|Z = 1) \equiv 0 \text{ for } g = af, nf, cf, fa, fn, fc, ff$$

The construction above satisfies Assumptions 1.1-1.3 and 2.1, and the shares are nonnegative because of inequalities (5)-(7). Next, we can check that this constructed joint distribution induces the observed (Y, D, Z) distribution. Specifically, if we assume that the specification is the true DGP and use these group shares as building blocks to put together the distributions of the observables, which we denote by $\tilde{Y}, \tilde{D}, Z$, we can show that the distribution of $\tilde{Y}, \tilde{D}, Z$ is the same as that of Y, D, Z. For brevity, we illustrate it for one case where all variables take the value of 1, and the other cases are analogous.

$$
\begin{aligned}
&\Pr(\tilde{Y} = 1, \tilde{D} = 1, Z = 1)\\
&= \big[\Pr(G = aa|Z = 1) + \Pr(G = ac|Z = 1)\\
&\quad + \Pr(G = ca|Z = 1) + \Pr(G = cc|Z = 1) \big] \Pr(Z = 1)\\
&= \big[\frac{1}{2} \Pr(Y = 1, D = 1|Z = 0) + \frac{1}{2} \Pr(Y = 1, D = 1|Z = 0)\\
&\quad + \Pr(Y = 1, D = 0|Z = 0) - \Pr(Y = 1, D = 0|Z = 1)\\
&\quad + \Pr(Y = 0|Z = 0) - \Pr(Y = 0|Z = 1) \big] \Pr(Z = 1)\\
&= \big[\Pr(Y = 1, D = 1|Z = 0) + \Pr(Y = 1, D = 0|Z = 0) - \Pr(Y = 1, D = 0|Z = 1)\\
&\quad + \Pr(Y = 0|Z = 0) - \Pr(Y = 0|Z = 1) \big] \Pr(Z = 1)\\
&= \big[\Pr(Y = 1|Z = 0) - \Pr(Y = 1, D = 0|Z = 1)\\
&\quad + \Pr(Y = 0|Z = 0) - \Pr(Y = 0|Z = 1) \big] \Pr(Z = 1)\\
&= \big[1 - \Pr(Y = 1, D = 0|Z = 1) - \Pr(Y = 0|Z = 1) \big] \Pr(Z = 1)\\
&= \big[\Pr(Y = 1|Z = 1) - \Pr(Y = 1, D = 0|Z = 1) \big] \Pr(Z = 1)\\
&= \Pr(Y = 1, D = 1, Z = 1)
\end{aligned}
$$

Finally, the joint distribution $(Y_{11}, Y_{10}, Y_{01}, Y_{00}, D_1, D_0, Z)$ that we specified is non-negative, additive, and sums to one. It is non-negative and additive by construction. It is easy to check that it sums up to one by simply plugging in their definitions in terms of the probabilities of (Y, D) given Z and rearranging. We omit the details here. QED.

Proof of Proposition 3: The numerator of the estimand is

$$
\begin{aligned}
&E[h(X)Y|Z = 1] - E[h(X)Y|Z = 0]\\
=&E[h(X)Y_1|Z = 1, D_0 = 1] \Pr(D_0 = 1|Z = 1) + E[h(X)Y_0|Z = 1, D_1 = 0] \Pr(D_1 = 0|Z = 1)+\\
&E[h(X)Y_1|Z = 1, D_1 > D_0] \Pr(D_1 > D_0|Z = 1) - E[h(X)Y_1|Z = 0, D_0 = 1] \Pr(D_0 = 1|Z = 0)-\\
&E[h(X)Y_0|Z = 0, D_1 = 0] \Pr(D_1 = 0|Z = 0) - E[h(X)Y_0|Z = 0, D_1 > D_0] \Pr(D_1 > D_0|Z = 0)\\
=&E[h(X)Y_1|D_1 > D_0] \Pr(D_1 > D_0) - E[h(X)Y_0|D_1 > D_0] \Pr(D_1 > D_0)
\end{aligned}
$$

$$=E[h(X)|D_1 > D_0, Y_1 = 1, Y_0 = 1] \Pr(Y_1 = 1, Y_0 = 1|D_1 > D_0) \Pr(D_1 > D_0)+$$
$$E[h(X)|D_1 > D_0, Y_1 = 1, Y_0 = 0] \Pr(Y_1 = 1, Y_0 = 0|D_1 > D_0) \Pr(D_1 > D_0)-$$
$$E[h(X)|D_1 > D_0, Y_1 = 1, Y_0 = 1] \Pr(Y_1 = 1, Y_0 = 1|D_1 > D_0) \Pr(D_1 > D_0)-$$
$$E[h(X)|D_1 > D_0, Y_1 = 0, Y_0 = 1] \Pr(Y_1 = 0, Y_0 = 1|D_1 > D_0) \Pr(D_1 > D_0)$$
$$=E[h(X)|G = cc] \Pr(G = cc) - E[h(X)|G = cf] \Pr(G = cf).$$

Similarly, we can show that the denominator is

$$E[Y|Z = 1] - E[Y|Z = 0] = \Pr(Y_1 > Y_0, D_1 > D_0) - \Pr(Y_1 < Y_0, D_1 > D_0).$$

Combining the numerator and denominator expressions and rearranging, we have our desired result. *QED.*

A.2   Relaxing the Binary Y Restriction

In this section, we provide the formal statement for Remark 4. It generalizes Lemma 1 and Proposition 1 and applies regardless of whether Y is binary or not.

Proposition 4 *Under Assumptions 1 and 2, the reduced form identifies the supercomplier share scaled by the supercomplier average treatment effect:*

$$\Pr(G = cc)E[Y_1 - Y_0|G = cc] = E[Y|Z = 1] - E[Y|Z = 0]. \tag{A8}$$

With the additional assumption that $X \perp\!\!\!\perp Z$, the Wald estimand identifies supercomplier characteristics weighted by treatment effect:

$$\frac{E[h(X)(Y_1 - Y_0)|G = cc]}{E[Y_1 - Y_0|G = cc]} = \frac{E[h(X)Y|Z = 1] - E[h(X)Y|Z = 0]}{E[Y|Z = 1] - E[Y|Z = 0]}. \tag{A9}$$

Proof of Proposition 4: For equation (A8),

$$E[Y|Z = 1] - E[Y|Z = 0]$$
$$=E[Y_1 - Y_0|D_1 > D_0] \Pr(D_1 > D_0)$$
$$=E[Y_1 - Y_0|D_1 > D_0, Y_1 > Y_0] \Pr(Y_1 > Y_0|D_1 > D_0) \Pr(D_1 > D_0)$$
$$=\Pr(G = cc)E[Y_1 - Y_0|G = cc].$$

6

For equation (A9), the numerator of the Wald estimand is

$$E[h(X)Y|Z = 1] - E[h(X)Y|Z = 0]$$
$$= \sum_g \left\{ E[h(X)Y|Z = 1, G = g] \Pr(G = g|Z = 1) \right.$$
$$\left. - E[h(X)Y|Z = 0, G = g] \Pr(G = g|Z = 0) \right\}$$
$$= \left\{ E[h(X)Y_1|G = cc] - E[h(X)Y_0|G = cc] \right\} \Pr(G = cc).$$

Plugging the result from equation (A8) into the denominator of the Wald estimand, we have the desired result. *QED.*

A.3   Shares and Characteristics of Other Groups

Lemma 1 and Proposition 1 show that we can identify the share and characteristics of the supercompliers. The proposition below establishes identification of the shares and characteristics of the other two groups within the stratum of compliers, $G = ca, cn$.

Proposition 5 *Under Assumptions 1 and 2 and provided that $X \perp\!\!\!\perp Z$, the shares of the two groups are*

$$\Pr(G = ca) = E[\kappa_0 Y] = E[(1 - D)Y|Z = 0] - E[(1 - D)Y|Z = 1]$$
$$\Pr(G = cn) = E[\kappa_1(1 - Y)] = E[D(1 - Y)|Z = 1] - E[D(1 - Y)|Z = 0],$$

and their characteristics can be identified as

$$E[h(X)|G = ca] = \frac{E[\kappa_0 Y h(X)]}{E[\kappa_0 Y]} = \frac{E[(1 - D)Y h(X)|Z = 1] - E[(1 - D)Y h(X)|Z = 0]}{E[(1 - D)Y|Z = 1] - E[(1 - D)Y|Z = 0]}$$
$$E[h(X)|G = cn] = \frac{E[\kappa_1(1 - Y)h(X)]}{E[\kappa_1(1 - Y)]} = \frac{E[D(1 - Y)h(X)|Z = 1] - E[D(1 - Y)h(X)|Z = 0]}{E[D(1 - Y)|Z = 1] - E[D(1 - Y)|Z = 0]}.$$

Proof of Proposition 5: We only prove the identification results for $G = ca$, as the proof for $G = cn$ is analogous. For the share of the $G = ca$ group, first note that

$$\Pr(G = ca) \equiv \Pr(Y_0 = 1, D_1 > D_0) = E[Y_0|D_1 > D_0] \Pr(D_1 > D_0) = E[\kappa_0 Y],$$

where the last equality follows from Lemma 2(b). This proves the first part of the equality, and the second part can be easily proved by plugging in the definition of κ_0 and simplifying.

7

For the characteristics result, note that

$$E[h(X)|G = ca] = E[h(X)|D_1 > D_0, Y_0 = 1] = \frac{E[Y_0 h(X)|D_1 > D_0]}{E[Y_0|D_1 > D_0]}$$

Since Lemma 2(b) also implies

$$E[Y_0 h(X)|D_1 > D_0] = \frac{1}{\Pr(D_1 > D_0)} E[\kappa_0 Y h(X)],$$

we have the first part of the identification result:

$$E[h(X)|G = ca] = \frac{E[\kappa_0 Y h(X)]}{E[\kappa_0 Y]}$$

We obtain the second part of the equality, i.e., identification by the Wald-type estimand, by plugging in the definition of κ_0 and simplifying. *QED*.

Like Proposition 1, the identification result in Proposition 5 leads to natural estimators. We can estimate the population share of the $G = ca$ group by regressing $(1 - D)Y$ on $(1 - Z)$ and the population share of the $G = cn$ group by regressing $D(1 - Y)$ on Z. With the Wald-type estimand for the characteristics, we can implement the corresponding 2SLS estimators like we do for the supercomplier characteristics.

Finally, we note that it is easy to directly identify the observations assigned to treatment (control) as belonging to the $G = na$ ($G = an$) group, namely those with $D = 0$ and $Y = 1$ ($D = 1$ and $Y = 0$). Therefore, we can identify the shares and characteristics of these two groups by directly using these observations. However, we cannot identify the shares and characteristics of the four remaining groups in Table 1, $G = aa, ac, nc, nn$. To see this, suppose for the sake of contradiction that we can identify the share of the aa group. Because we can identify the share of the treatment always takers, we will also be able to identify the share of the ac group (by subtraction) and, consequently, the probability of being in the ac group conditional on being a treatment always taker. However, the latter is simply the treatment effect for the treatment always takers:

$$\Pr(G = ac|D_0 = D_1 = 1) = E[Y_1 - Y_0|D_0 = D_1 = 1],$$

which we know cannot be identified. This contradiction shows the non-identifiability of the aa group share, and the proofs for the other groups are analogous.

A.4 Estimator Details

We prove several claims about complier and supercomplier characteristics estimators from Section 2.2. First, we show that the Angrist and Pischke (2009) plug-in estimator for complier characteristics and the analogous estimator for supercomplier characteristics have Wald-type representations. These plug-in estimators are defined, respectively, as:

$$\tilde{\chi}_c \equiv \frac{\sum_i \hat{\kappa}_i h(X_i)}{\sum_{Z_i=1} D_i - \sum_{Z_i=0} D_i}$$

$$\tilde{\chi}_{cc} \equiv \frac{\sum_i \hat{\pi}_i h(X_i)}{\sum_{Z_i=1} Y_i - \sum_{Z_i=0} Y_i},$$

where $\hat{\kappa}_i$ and $\hat{\pi}_i$ are sample analogs of the κ and π weighting functions evaluated at the D_i, Z_i, and Y_i for individual i.

Define $\hat{\chi}_{cc}$ as the Wald-type estimator associated with equation (2) and $\hat{\chi}_c$ as the analogous treatment complier characteristics estimator:

$$\hat{\chi}_c \equiv \frac{\sum_{Z_i=1} h(X_i) D_i - \sum_{Z_i=0} h(X_i) D_i}{\sum_{Z_i=1} D_i - \sum_{Z_i=0} D_i}$$

$$\hat{\chi}_{cc} \equiv \frac{\sum_{Z_i=1} h(X_i) Y_i - \sum_{Z_i=0} h(X_i) Y_i}{\sum_{Z_i=1} Y_i - \sum_{Z_i=0} Y_i}.$$

Our second claim is that $\hat{\chi}_c$ is equivalent to the complier characteristics estimator used by studies such as Finkelstein and Notowidigdo (2019).[1] Finally, we compare the asymptotic variances of the two candidate supercomplier estimators, $\tilde{\chi}_{cc}$ and $\hat{\chi}_{cc}$, and show that neither dominates the other (the same is true for the two complier characteristics estimators).

Proposition 6 *Let N_1 and N_0 denote the number of sample units assigned to the treatment and control groups, respectively, and $\hat{\tau} \equiv N_1/N$ the proportion in the treatment group. We can rewrite the plug-in estimators as*

$$\tilde{\chi}_c = \frac{\frac{1}{N_1}\sum_{Z_i=1}\{D_i - (1-\hat{\tau})\}h(X_i) - \frac{1}{N_0}\sum_{Z_i=0}\{D_i - (1-\hat{\tau})\}h(X_i)}{\frac{1}{N_1}\sum_{Z_i=1}\{D_i - (1-\hat{\tau})\} - \frac{1}{N_0}\sum_{Z_i=0}\{D_i - (1-\hat{\tau})\}}$$

$$\tilde{\chi}_{cc} = \frac{\frac{1}{N_1}\sum_{Z_i=1}\{Y_i - (1-\hat{\tau})\}h(X_i) - \frac{1}{N_0}\sum_{Z_i=0}\{Y_i - (1-\hat{\tau})\}h(X_i)}{\frac{1}{N_1}\sum_{Z_i=1}\{Y_i - (1-\hat{\tau})\} - \frac{1}{N_0}\sum_{Z_i=0}\{Y_i - (1-\hat{\tau})\}}$$

---

[1] It is easy to show that $\hat{\chi}_c$ is also the sample analog of $E[\kappa_1 h(X)]/FS$ where FS stands for the population first stage.

Proof of Proposition 6: We prove the result for $\tilde{\chi}_c$, and the proof for $\tilde{\chi}_{cc}$ is analogous. We begin by plugging into the numerator of $\tilde{\chi}_c$ the estimator of κ_i, for which $\Pr(Z = 1)$ is estimated with $\hat{\tau}$:

$$\frac{1}{N}\sum_i \hat{\kappa}_i h(X_i) = \frac{1}{N}\sum_i \left[h(X_i) - \frac{h(X_i)D_i(1 - Z_i)}{1 - \hat{\tau}} - \frac{h(X_i)(1 - D_i)Z_i}{\hat{\tau}} \right]$$

$$= \frac{1}{N}\sum_i h(X_i) - \frac{1}{N}\sum_i \frac{h(X_i)Z_i}{\hat{\tau}} + \frac{1}{N}\sum_i \frac{h(X_i)D_iZ_i}{\hat{\tau}} - \frac{1}{N}\sum_i \frac{h(X_i)D_i(1 - Z_i)}{1 - \hat{\tau}}.$$

For the last three terms,

$$-\frac{1}{N}\sum_i \frac{h(X_i)Z_i}{\hat{\tau}} = -\frac{1}{N_1}\sum_{Z_i=1} h(X_i)$$

$$\frac{1}{N}\sum_i \frac{h(X_i)D_iZ_i}{\hat{\tau}} = \frac{1}{N_1}\sum_{Z_i=1} D_i h(X_i)$$

$$-\frac{1}{N}\sum_i \frac{h(X_i)D_i(1 - Z_i)}{1 - \hat{\tau}} = -\frac{1}{N_0}\sum_{Z_i=0} D_i h(X_i),$$

and it follows that

$$\frac{1}{N}\sum_i \kappa_i h(X_i) = \frac{1}{N}\sum_i h(X_i) - \frac{1}{N_1}\sum_{Z_i=1} h(X_i) + \frac{1}{N_1}\sum_{Z_i=1} D_i h(X_i) - \frac{1}{N_0}\sum_{Z_i=0} D_i h(X_i).$$

With simple algebra, we can show that

$$\frac{1}{N}\sum_i h(X_i) - \frac{1}{N_1}\sum_{Z_i=1} h(X_i) = (1 - \hat{\tau})\left(\frac{1}{N_0}\sum_{Z_i=0} h(X_i) - \frac{1}{N_1}\sum_{Z_i=1} h(X_i) \right),$$

and the result for $\tilde{\chi}_c$ easily follows. *QED.*

As a corollary to Proposition 6, expression (4) is the estimand corresponding to the $\tilde{\chi}_{cc}$ estimator. Comparing (2) and (4) reveals that the wedge between the two estimators, $\hat{\chi}_{cc}$ and $\tilde{\chi}_{cc}$, is due to the difference in the average characteristics between the treatment and control groups. While independence between X and Z implies this difference to be zero in the population, it is generally nonzero in a given sample.

We can implement both $\hat{\chi}_{cc}$ and $\tilde{\chi}_{cc}$ using a standard 2SLS command in a statistical software. The only difference is that to implement the latter, we need to first transform Y by subtracting from it the proportion of sample units assigned to the control group (i.e., $1 - \hat{\tau}$).

As it turns out, because $\hat{\tau}$ is consistent for τ, we can proceed as if we know the true τ, and the inference results reported by the 2SLS command are valid.

For our second claim, we remind the reader that the estimand in Finkelstein and Notowidigdo (2019) is (the expression is provided in the Online Appendix F of their 2018 NBER working paper; note that the roles of D and Z are reversed in their notation):

$$E[h(X)|D_1 > D_0] = \frac{(s_{AT} + s_C)\mu_T - s_{AT}\mu_{AT}}{s_c} \tag{A10}$$

where the various quantities are identified with

$$s_{AT} = \Pr(D = 1|Z = 0)$$
$$s_C = \Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0)$$
$$\mu_{AT} = E[h(X)|D = 1, Z = 0]$$
$$\mu_T = E[h(X)|D = 1, Z = 1].$$

The resulting estimator, which we denote by $\hat{\chi}_{FN}$, is obtained by plugging in the sample analogs of s_{AT}, s_C, μ_{AT}, and μ_T. The following proposition establishes that $\hat{\chi}_{FN}$ can be alternatively implemented using a 2SLS regression with the product $h(X)D$ as the dependent variable, D the endogenous variable, and Z the instrument.

Proposition 7 *The two estimators $\hat{\chi}_{FN}$ and $\hat{\chi}_c$ are equal.*

Proof of Proposition 7: First notice that the estimand in (A10) is equivalent to

$$E[h(X)|D_1 > D_0] = \frac{\Pr(D = 1|Z = 1)\mu_T - \Pr(D = 1|Z = 0)\mu_{AT}}{\Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0)}. \tag{A11}$$

Since the denominator of (A11) is the same as (2), we only need to show that the sample analogs of the numerators are equal. The sample analog of the first term in the numerator of (A11) is

$$\left(\frac{1}{N_1}\sum_{Z_i=1} D_i\right)\left(\frac{1}{\sum_{Z_i=1} D_i}\sum_{D_i=1,Z_i=1} h(X_i)\right) = \frac{1}{N_1}\sum_{D_i=1,Z_i=1} h(X_i) = \frac{1}{N_1}\sum_{Z_i=1} h(X_i)D_i,$$

which is the sample analog of $E[h(X)D|Z = 1]$. Similarly, we can show that the second term $\Pr(D = 1|Z = 0)\mu_{AT}$ and $E[h(X)D|Z = 0]$ have the same sample analog. *QED.*

The final claim of this section is that the relative asymptotic efficiency of the two estimators, $\hat{\chi}_{cc}$ and $\tilde{\chi}_{cc}$, depends on the underlying DGP. To see this, first focus on the numerators of the estimators. For simplicity, consider the case where h is the identity function. Following standard arguments, the numerators of $\hat{\chi}_{cc}$ and $\tilde{\chi}_{cc}$ have asymptotic variances

$$\text{avar of } \hat{\chi}_{cc} \text{ numerator}: \quad \frac{1}{\tau}var(XY|Z=1) + \frac{1}{1-\tau}var(XY|Z=0)$$

$$\text{avar of } \tilde{\chi}_{cc} \text{ numerator}: \quad \frac{1}{\tau}var(X(Y-1+\tau)|Z=1) + \frac{1}{1-\tau}var(X(Y-1+\tau)|Z=0).$$

We further narrow in on the first term in each of the expressions above (the variance conditional on $Z=1$). Consider the simple (though likely unrealistic) case where $\tau = 0.5$ and X is binary, and where conditional on $Z=1$, X and Y are independent with $\Pr(X=1) = 0.5$ and $\Pr(Y=1|Z=1) = \mu_Y$. The independence allows us to express the two variances in terms of just the means and variances of X and Y conditional on $Z=1$. As a result, we can derive the difference between the two terms as

$$var(XY|Z=1) - var(X(Y-1+\tau)|Z=1) = 0.25\mu_Y - 0.0625$$

The implication is that in this simple setting, the first term of $\hat{\chi}_{cc}$'s numerator has a smaller variance than that of $\tilde{\chi}_{cc}$ if and only if $\mu_Y < 0.25$. It is easy to construct similar scenarios under which one numerator or one estimator has an asymptotic variance that is larger or smaller than the other, so we omit the details here. For simplicity and practical consistency, we implement $\hat{\chi}_{cc}$ in this paper.

A.5 Estimating Supercomplier Characteristics Quantiles

Abadie, Angrist and Imbens (2002) (henceforth, AAI) propose an estimation and inference procedure for the complier quantile treatment effect. They start from the observation that the θ-quantile treatment effect parameter solves the population minimization problem

$$(\alpha_\theta, \beta_\theta) = \underset{\alpha,\beta}{\text{argmin}}\, E[\kappa\rho_\theta(Y - \alpha D - X\beta)], \tag{A12}$$

where $\rho_\theta(\lambda) \equiv (\theta - 1_{[\lambda<0]})\lambda$ is the check function per Bassett and Koenker (1982). AAI note that while the minimization problem (A12) is convex in the parameters, its sample analog is not because the κ weight can be negative. Their solution comes from the insight that, by

the law of iterated expectations, (A12) is equivalent to

$$(\alpha_\theta, \beta_\theta) = \underset{\alpha,\beta}{\operatorname{argmin}} E[\kappa_\nu \rho_\theta(Y - \alpha D - X\beta)], \tag{A13}$$

where $\kappa_\nu \equiv E[\kappa|U]$ with $U \equiv (D, Y, X)$. AAI show that $\kappa_\nu = \Pr(D_1 > D_0|U)$ has a probability interpretation and is therefore nonnegative. They proceed to estimate the quantile effect by plugging in the estimated κ_ν in the sample analog of (A13).

Our problem is analogous, but we need to adjust the AAI approach. The supercomplier characteristics θ-quantile solves the problem

$$\xi_\theta = \underset{\xi}{\operatorname{argmin}} E[\pi \rho_\theta(X - \xi)],$$

and we face the same challenge that π may be negative. Like AAI, we can apply the law of iterated expectations and use a conditional version of π instead. However, it turns out that when conditioning on the same triplet $U = (D, Y, X)$, $E[\pi|U]$ does not have a probability interpretation and may still be negative. In Lemma 3 below, we show that we need to instead condition on just $V \equiv (Y, X)$ or simply $V \equiv X$. With the result in Lemma 3, we can adapt the estimation approach of AAI. Specifically, we can let $\pi_\nu \equiv E[\pi|V]$ and solve the sample analog of the minimization problem

$$\xi_\theta = \underset{\xi}{\operatorname{argmin}} E[\pi_\nu \rho_\theta(X - \xi)].$$

Lemma 3. *Under Assumptions 1.1-1.3 and 2.1,*

$$E[\pi|V] = \Pr(G = cc|V).$$

Proof of Lemma 3: We examine the conditional expectation of each of the three terms that make up π:

$$E[\pi|V] = E[\kappa|V] - E[\kappa_0 Y|V] - E[\kappa_1(1 - Y)|V] \tag{A14}$$

For the first term on the right hand side of (A14), we can simply replace U by V in the proof of Lemma 3.2 by AAI and still have the analogous probability interpretation: $E[\kappa|V] = \Pr(D_1 > D_0|V)$.

For the second term of (A14),

$$E[\kappa_0 Y | V] = E\left[\left.\frac{(1-D)(1-Z)Y}{\Pr(Z=0)}\right| V\right] - E\left[\left.\frac{(1-D)ZY}{\Pr(Z=1)}\right| V\right]$$

Because

$$
\begin{aligned}
&E[(1-D)(1-Z)Y|V] \\
&= \Pr(D=0, Z=0, Y=1|V) \\
&= \Pr(D_0=0, Z=0, Y_0=1|V) \\
&= \sum_{k=0,1} \Pr(D_0=0, D_1=k, Z=0, Y_0=1|V) \\
&= \sum_{k=0,1} \Pr(D_0=0, D_1=k, Y_0=1|V) \Pr(Z=0|D_0=0, D_1=k, Y_0=1, V) \\
&= \Pr(G=na|V)\Pr(Z=0) + \Pr(G=ca|V)\Pr(Z=0),
\end{aligned}
$$

we have
$$E\left[\left.\frac{(1-D)(1-Z)Y}{\Pr(Z=0)}\right| V\right] = \Pr(G=na|V) + \Pr(G=ca|V).$$

At the same time,

$$
\begin{aligned}
&E[(1-D)ZY|V] \\
&= \Pr(D=0, Z=1, Y=1|V) \\
&= \Pr(D_1=0, Y_0=1|V)\Pr(Z=1|D_1=0, Y_0=1, V) \\
&= \Pr(G=na|V)\Pr(Z=1),
\end{aligned}
$$

and therefore,
$$E\left[\left.\frac{(1-D)ZY}{\Pr(Z=1)}\right| V\right] = \Pr(G=na|V).$$

It follows that
$$E[\kappa_0 Y|V] = \Pr(G=ca|V).$$

Similarly, we can show that the third term of (A14) is

$$E[\kappa_1(1-Y)|V] = \Pr(G=cn|V).$$

Since the three groups $G = ca, cn, cc$ partition the set of treatment compliers, we prove

Lemma 3 by putting the three terms together. *QED.*

A.6 Implications under Conditional Independence

In this section, we analyze the supercomplier characteristics estimand under the conditional independence assumption. Specifically, we investigate identification under the corresponding 2SLS regression of (3) with covariate controls, and we focus on the case of stratified experiments where the controls are the set of stratum dummies. We show that the population 2SLS estimator identifies a non-negatively weighted average of supercomplier characteristics, as opposed to an expression where other unobserved subpopulations receive positive weight. This result is analogous to the Blandhol et al. (2022) finding on "weak causality"—under certain conditions, the LATE estimand with covariate controls identifies a non-negatively weighted average of complier treatment effects.

Assumption 3 (Stratified Experiment)

1. *Stratified Random Assignment:* $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, D_0, D_1) \perp\!\!\!\perp Z$ *conditional on* W *and* $0 < \Pr(Z = 1|W) < 1$.

2. *Conditional Reduced Form: The population coefficient on* Z *from the reduced form regression of* Y *on* Z *and* W *is nonzero.*

3. *Saturation:* $E[Z|W] = \mathbb{L}(Z|W)$, *where* $\mathbb{L}(Z|W)$ *is the linear projection of* Z *on* W.

Assumptions 3.1 and 3.2 are the covariate-control counterparts to Assumptions 1.1 and 2.2, respectively. Assumption 3.3 stipulates that the true conditional probability of treatment eligibility assignment can be recovered with linear regression. This condition is satisfied when the controls are saturated stratum fixed effects in a stratified randomized experiment.

Proposition 8 *Denote the population coefficient from a supercomplier characteristics regression with covariate control* W *by* β_{2SLS}. *Under Assumptions 1.2, 1.3, 2.2 and 3, and provided that* $X \perp\!\!\!\perp Z|W$,

$$\beta_{2SLS} = E\big[\omega_W E[h(X)|G = cc, W]\big], \tag{A15}$$

where the weight ω_W *is nonnegative for all values of* W.

Proof of Proposition 8: We define $\tilde{Z} = Z - \mathbb{L}(Z|W)$ and use Proposition 6 of Blandhol et al. (2022) to write

$$\beta_{2SLS} = \frac{E[h(X)Y\tilde{Z}]}{E[Y\tilde{Z}]}.$$

Decompose the numerator of β_{2SLS} as

$$
\begin{aligned}
E[h(X)Y\tilde{Z}] &= E[E[h(X)Y\tilde{Z}|W]] \\
&= E\big[E[cov(h(X)Y, \tilde{Z}|W)] + E[E[h(X)Y|W]E[\tilde{Z}|W]] \\
&= E\big[\underbrace{E[cov(h(X)Y, Z|W)]}_{\text{(i)}} + \underbrace{E[E[h(X)Y|W]E[\tilde{Z}|W]]}_{\text{(ii)}} \big],
\end{aligned}
\qquad \text{(A16)}
$$

where the last equality follows from the fact that conditional on W, $\mathbb{L}(Z|W)$ is a constant and is therefore uncorrelated with $h(X)Y$.

We can further decompose term (i) of equation (A16) as:

$$
\begin{aligned}
cov(h(X)Y, Z|W) &= E[h(X)Y(Z - E[Z|W])|W] \\
&= E[h(X)Y|Z = 1, W](1 - E[Z|W])E[Z|W] \\
&\quad - E[h(X)Y|Z = 0, W](E[Z|W])(1 - E[Z|W]) \\
&= (E[h(X)Y|Z = 1, W] - E[h(X)Y|Z = 0, W])(1 - E[Z|W])E[Z|W] \\
&= E[h(X)|G = cc, W] \times RF_W \times (1 - E[Z|W])E[Z|W]
\end{aligned}
\qquad \text{(A17)}
$$

where the last equality follows from the conditional variation of Proposition 1 with $RF_W \equiv E[Y|Z = 1, W] - E[Y|Z = 0, W]$ being the reduced form conditional on W.

For term (ii) of equation (A16), note that

$$E[\tilde{Z}|W] = E[Z|W] - \mathbb{L}(Z|W) = 0$$

if $E[Z|W] = \mathbb{L}(Z|W)$. Thus, term (ii) is zero whenever the expectation of Z in linear in W, which is true in stratified experiments with W being the full set of stratum dummies.

Putting together equations (A17) and (A18), we have that the numerator of β_{2SLS} is

$$E[h(X)Y\tilde{Z}] = E\big[E[h(X)|G = cc, W] \times RF_W \times (1 - E[Z|W])E[Z|W]\big]. \qquad \text{(A18)}$$

We can simiarly show that the denominator of β_{2SLS} is:

$$E[h(X)Y\tilde{Z}] = E\big[RF_W \times (1 - E[Z|W])E[Z|W]\big],$$

and the result of Proposition 8 follows with weight $\omega_W = RF_W \times (1 - E[Z|W])E[Z|W]$. Since $RF_W \geq 0$ by Assumption 3.2 and Z is binary, $\omega_W \geq 0$ for all values of W. *QED.*

B Testing Outcome Monotonicity in Bitler, Gelbach and Hoynes (2006, 2017)

As a proof of concept, we implement our outcome monotonicity test on the Jobs First evaluation data. Briefly, the Jobs First evaluation, implemented by the Manpower Demonstration Research Corporation (MDRC) beginning in 1997, was an experimental evaluation of Connecticut's Jobs First welfare reform initiative (see Bloom et al. 2002 for a description). The evaluation compared outcomes like earnings, employment, and welfare use among individuals randomly assigned to the Jobs First program to those in the traditional Connecticut welfare program. The Jobs First program imposed a time limit on welfare receipt and allowed participants to earn up to the federal poverty line without any decrease to their benefits.

As discussed in Bitler, Gelbach and Hoynes (2006, 2017), based on theory of labor supply, the predicted effect of the Jobs First earnings disregard relative to the traditional welfare benefit schedule should vary across the earnings distribution. For women with previously low earnings, the intervention is predicted to increase earnings. However, for women with relatively high earnings, the intervention could have a negative effect. Thus, Jobs First is an experiment where we would expect the outcome monotonicity assumption to fail and where it has indeed been shown to fail by Figure 3 of Bitler, Gelbach and Hoynes (2006). It therefore provides a useful test case for our proposed outcome monotonicty test.

To examine the viability of our test, we assess whether the Jobs First treatment had a monotonically positive effect on participants' average quarterly earnings in the first seven quarters after random assignment. In the experiment, there is no non-compliance, so we limit our focus to inequality (7). Our approach follows the subsample analyses in Bitler, Gelbach and Hoynes (2017). We split participants into three groups, $r = \{1, 2, 3\}$, based on their earnings seven quarters prior to random assignment—no earnings, earnings below median, and earnings at or above median earnings—where the median is measured only

among participants with non-zero earnings.

We then create three copies of the data and estimate the following stacked regression:

$$Y_i^r = \phi^r + \theta_1 Z_i 1_{[r=1]} + \theta_2 Z_i 1_{[r=2]} + \theta_3 Z_i 1_{[r=3]} + \varepsilon_i^r$$

Similar to equation (8), r indexes each stack, ϕ^r is the stack specific constant, and standard errors are clustered at the individual level. However, in this case, Y_i^r is simply equal to average quarterly earnings in the first seven quarters of Jobs First for all stacks. Note that the choice of three stacks here corresponds to the number of subsamples examined, not to the inequalities in Proposition 2.

We test whether $\theta \equiv \min(\theta_1, \theta_2, \theta_3)$ is nonnegative by taking draws from a joint normal distribution with the estimated covariance matrix from the stacked regression. Our estimate of θ is -347.257 while the 5-percent critical value from the simulated normal distribution is -303.591. Thus, the test rejects outcome monotonicity at the 5-percent level.

We note, however, that our ability to reject outcome monotonicity varies based on the subsamples we examine. This pattern is similar to the results in Bitler, Gelbach and Hoynes (2017), where the authors find that using alternative subsamples does not adequately reflect treatment effect heterogeneity. We performed similar tests with all of the subsamples in Table 1 of Bitler, Gelbach and Hoynes (2017), which includes subsamples by education of the participant, by whether the youngest child in the case is under 6 years old, by number of children in the case, by marital status of the participant, by number of quarters with any earnings before random assignment, and by whether the participant was receiving welfare benefits seven quarters before random assignment. We cannot reject outcome monotonicity at standard significance levels in any of these tests.

We conclude from this proof of concept that our proposed outcome monotonicity test can detect the presence of subsamples for which the outcome monotonicity assumption fails. However, the choice of subsamples to examine matters.

Table A.2: Ashraf et al. Table 3 Comparison

	Cuenta Unidos Accounts (a)		Ahorro Directo Accounts (b)		Other Accounts (c)		Total Across All Accounts (d)=(a)+(b)+(c)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Quartic Root								
Treatment 3 (joint + migrant-only account)	0.354***	0.417***	0.305***	0.338***	0.206	0.181	0.639***	0.703***
	(0.114)	(0.129)	(0.094)	(0.116)	(0.192)	(0.194)	(0.212)	(0.225)
Treatment 2 (joint account)	0.231**	0.244**	-0.057	-0.101**	-0.006	0.072	0.102	0.138
	(0.109)	(0.123)	(0.047)	(0.051)	(0.178)	(0.193)	(0.198)	(0.219)
Treatment 1 (recipient-only account)	0.162*	0.152*	-0.004	-0.006	-0.122	-0.043	0.001	0.081
	(0.088)	(0.092)	(0.051)	(0.063)	(0.180)	(0.189)	(0.195)	(0.208)
Pooled Regression	X		X		X		X	
Demand for Control Indicator	X		X		X		X	
p-value of F-test of equality of								
Treatment 3 and 2 coefficients	0.334	0.258	0.000	0.000	0.305	0.595	0.017	0.018
Treatment 3 and 1 coefficients	0.101	0.043	0.000	0.001	0.105	0.277	0.004	0.008
Treatment 2 and 1 coefficients	0.555	0.467	0.091	0.016	0.531	0.560	0.624	0.797
B. Any Savings								
Treatment 3 (joint + migrant-only account)	0.112***	0.131***	0.071***	0.079***	0.008	0.006	0.112**	0.130***
	(0.030)	(0.033)	(0.022)	(0.026)	(0.041)	(0.044)	(0.044)	(0.048)
Treatment 2 (joint account)	0.058**	0.060**	-0.017	-0.028**	0.014	0.027	0.025	0.032
	(0.027)	(0.028)	(0.011)	(0.012)	(0.040)	(0.045)	(0.043)	(0.048)
Treatment 1 (recipient-only account)	0.069**	0.076**	0.001	0.002	-0.023	-0.013	0.027	0.051
	(0.027)	(0.030)	(0.013)	(0.015)	(0.040)	(0.044)	(0.044)	(0.049)
Pooled Regression	X		X		X		X	
Demand for Control Indicator	X		X		X		X	
p-value of F-test of equality of								
Treatment 3 and 2 coefficients	0.112	0.046	0.000	0.000	0.898	0.649	0.050	0.050
Treatment 3 and 1 coefficients	0.214	0.142	0.001	0.001	0.451	0.664	0.064	0.123
Treatment 2 and 1 coefficients	0.720	0.623	0.043	0.009	0.367	0.364	0.959	0.703

Notes: Panel A, Columns (1), (3), (5), and (7) replicate the estimates in Panel A of Table 3 in Ashraf et al. (2015). The estimates come from regressions that pool all treatment groups and that include fixed effects for marketer, stratification cell, and month of marketing visit, in addition to the demand for control indicator. Columns (2), (4), (6), and (8) of Panel A show estimated coefficients from regressions estimated on only the relevant treatment group and the comparison group. Estimates in even columns include the same fixed effects as used in the odd columns but omit the demand for control indicator. Estimates in Panel B repeat this pattern but switch the outcome variable from the quartic root of savings in the relvant account to an indicator for having any savings in the relevant account.

Table A.3: Ashraf et al. Table 5 Comparison

	Cuenta Unidos Accounts		Ahorro Directo Accounts		Other Accounts		Total Across All Accounts	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Quartic Root								
Treatment 3 × Demand for Control	0.572***	0.630***	0.113	0.128	0.331	0.460	0.768***	0.932***
	(0.171)	(0.164)	(0.110)	(0.112)	(0.261)	(0.285)	(0.293)	(0.309)
Treatment 3 × No Demand for Control	0.117	0.121	0.511***	0.449***	0.070	0.090	0.497*	0.458
	(0.135)	(0.136)	(0.146)	(0.141)	(0.279)	(0.279)	(0.302)	(0.304)
Treatment 2 × Demand for Control	0.478***	0.464***	-0.119*	-0.129*	0.047	0.100	0.283	0.310
	(0.161)	(0.154)	(0.072)	(0.071)	(0.254)	(0.254)	(0.286)	(0.282)
Treatment 2 × No Demand for Control	-0.022	-0.005	0.009	-0.022	-0.054	-0.021	-0.074	-0.081
	(0.134)	(0.133)	(0.044)	(0.022)	(0.256)	(0.255)	(0.275)	(0.273)
Treatment 1 × Demand for Control	0.121	0.170	-0.114	-0.088	-0.355	-0.270	-0.340	-0.193
	(0.110)	(0.106)	(0.076)	(0.082)	(0.231)	(0.222)	(0.253)	(0.245)
Treatment 1 × No Demand for Control	0.186	0.183	0.114*	0.069	0.099	0.101	0.326	0.271
	(0.140)	(0.152)	(0.066)	(0.057)	(0.277)	(0.284)	(0.298)	(0.302)
Pooled Regression	X		X		X		X	
p-value of F-test of equality of interactions with								
Treatment 3	0.031	0.017	0.025	0.075	0.490	0.354	0.516	0.274
Treatment 2	0.014	0.022	0.100	0.147	0.783	0.736	0.372	0.320
Treatment 1	0.722	0.944	0.024	0.117	0.209	0.304	0.091	0.233
B. Any Savings								
Treatment 3 × Demand for Control	0.162***	0.205***	0.040	0.042	-0.008	0.006	0.097	0.142**
	(0.041)	(0.040)	(0.028)	(0.030)	(0.058)	(0.060)	(0.061)	(0.064)
Treatment 3 × No Demand for Control	0.058	0.070	0.105***	0.092***	0.027	0.027	0.128**	0.125*
	(0.044)	(0.046)	(0.032)	(0.031)	(0.057)	(0.058)	(0.063)	(0.066)
Treatment 2 × Demand for Control	0.131***	0.151***	-0.029*	-0.034**	-0.046	-0.041	0.013	0.025
	(0.036)	(0.037)	(0.017)	(0.017)	(0.057)	(0.058)	(0.061)	(0.062)
Treatment 2 × No Demand for Control	-0.018	-0.001	-0.004	-0.010	0.076	0.058	0.040	0.025
	(0.039)	(0.039)	(0.013)	(0.010)	(0.057)	(0.057)	(0.062)	(0.062)
Treatment 1 × Demand for Control	0.069**	0.099***	-0.027	-0.025	-0.069	-0.060	-0.042	-0.007
	(0.033)	(0.034)	(0.017)	(0.020)	(0.057)	(0.060)	(0.062)	(0.064)
Treatment 1 × No Demand for Control	0.066	0.071	0.030	0.018	0.025	0.020	0.097	0.082
	(0.044)	(0.047)	(0.020)	(0.019)	(0.057)	(0.058)	(0.065)	(0.066)
Pooled Regression	X		X		X		X	
p-value of F-test of equality of interactions with								
Treatment 3	0.083	0.028	0.111	0.249	0.668	0.800	0.724	0.845
Treatment 2	0.004	0.005	0.215	0.207	0.132	0.223	0.760	1.000
Treatment 1	0.953	0.626	0.033	0.113	0.251	0.335	0.124	0.332

Notes: Panel A, Columns (1), (3), (5), and (7) replicate the estimates in Panel A of Table 5 in Ashraf et al. (2015). The estimates come from regressions that pool all treatment groups and that include fixed effects for marketer, stratification cell, and month of marketing visit, in addition to the demand for control indicator. Columns (2), (4), (6), and (8) of Panel A show estimated coefficients from regressions estimated on only the relevant treatment group and the comparison group. Estimates in even columns include the same controls as the odd columns. Estimates in Panel B repeat this pattern but switch the outcome variable from the quartic root of savings in the relvant account to an indicator for having any savings in the relevant account.