Supercompliers*

Matthew L. Comey, Joint Committee on Taxation
Amanda R. Eng, Internal Revenue Service
Pauline Leung, Cornell University
Zhuan Pei, Cornell University and IZA†§

February 2025

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 important for policy. We provide tools to characterize supercompliers under a set of jointly testable assumptions. Specifically, we require standard assumptions from the local average treatment effect literature plus an outcome monotonicity assumption. Estimation and inference can be conducted with instrumental variable regression. In two job-training experiments, we demonstrate our machinery's utility, particularly in incorporating social welfare weights into marginal-value-of-public-funds analysis.

Key Words: Local Average Treatment Effect; Principal Stratification; Compliers; Supercompliers; Marginal Value of Public Funds

JEL Codes: C10; H00; J24.

*We thank Josh Angrist, Eric Auerbach, David Card, Amy Finkelstein, Toru Kitagawa, David S. Lee, Doug Miller, Jon Roth, Steve Pischke, Andres Santos, and participants in the Princeton University Industrial Relations Section Centennial Symposium and the Stata Virtual Symposium for helpful discussions. Fiona Qiu provided excellent research assistance. Disclaimers: Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the Joint Committee on Taxation or any Member of Congress. Eng performed this work prior to joining the Internal Revenue Service. All views and opinions expressed herein do not represent the Internal Revenue Service.

†Comey: matthew.comey@jct.gov; Eng: amanda.r.eng@irs.gov; Leung: pleung@cornell.edu; Pei: zhuan.pei@cornell.edu

§The authors received no specific funding for this project and have no conflicts of interest to disclose.

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 a LATE is the ratio of intent-to-treat and 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 (2023), and Agan, Doleac and Harvey (2023).

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. Additionally, in our exploration of various estimators, we show that the plug-in complier characteristics estimators used in the literature can be equivalently implemented using simple instrumental variable regression. Finally, we provide results for the identification and estimation of characteristics for compliers whose outcomes are unaffected by treatment.

Our key results follow from 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 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, supporting the assumption of nonnegative treatment effects. Other studies have likewise imposed outcome monotonicity restrictions (e.g., Manski 1997, Manski and Pepper 2000, and Lee 2009), highlighting their applicability across a range of settings.

In other situations, there is 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 consequence of violation. To that end, we first extend the seminal work by Kitagawa (2015) to characterize sharp testable implications under the standard LATE assumptions and outcome monotonicity, and we develop an easy-to-implement joint test. Then, to understand the consequence 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.

Much of our econometric discussion centers around a binary outcome. However, we do not view this focus as restrictive. First, binary outcomes such as employment, welfare receipt,

and education credential attainment are quite common in economic research. Second, 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." Finally, our framework is still interpretable even when the outcome is not binary. Regardless of the distribution of the outcome variable, our estimand identifies a weighted average of supercomplier characteristics, 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. While conventional heterogeneity analysis asks which subgroups benefit more, our supercomplier analysis collects all beneficiaries and profiles them. Characterizing supercompliers also mirrors the increasingly common practice of describing compliers, which researchers have accepted as a natural complement to subsample heterogeneity analysis of treatment take-up.

In addition to describing beneficiaries, the supercomplier framework can also be used to enrich marginal value of public funds (MVPF; Hendren and Sprung-Keyser 2020) analyses, which have recently become commonplace in empirical economic and policy studies. An input into the MVPF is the social willingness to pay (WTP) for a policy, which is often quantified by the LATE of the policy on some outcome of interest (e.g., earnings). Although social WTP (and therefore MVPF) should be calculated by taking a weighted sum of treatment effects, with weights reflecting the planner's redistributive preferences, it is not feasible in practice when only one LATE is reported. We show, however, that when social welfare weights depend linearly on observed characteristics, a weighted MVPF can be computed using only information on supercomplier characteristics and the reported LATE. A key advantage of this approach is that access to the microdata may not be necessary for recomputing an MVPF with different social welfare functions.

We illustrate our proposed method using data from two well-known randomized experiments on job training. The first is the National Job Corps Study, where eligible youths were randomly assigned access to an intensive residential education and job training program (Schochet, Burghardt and McConnell 2008*a*). The second is the National Job Training Partnership Act (JTPA) Study, where economically disadvantaged adults were randomly assigned access to Title II-A programs under the JTPA (Bloom et al. 1997). We find that, for labor market outcomes, the Job Corps supercompliers are more advantaged relative to

the study population. In contrast, among the group that experienced the largest impact from JTPA—adult women—supercompliers are less relatively advantaged. For a social planner who favors redistribution and places greater welfare weight on the more disadvantaged, our results suggest that the MVPF is lower for Job Corps and higher for JTPA.

Related theoretical results were independently developed in exemplary research by Yu (forthcoming).[2] Both papers rely on the same set of assumptions, propose the same estimand, and derive essentially the same key identifying result for a binary Y (Our Propositions 1 and 5 and Yu forthcoming's Theorem 3.2). However, unlike Yu (forthcoming), we extend our framework to accommodate non-binary outcomes, which significantly broadens its applicability. We then connect our results to the increasingly influential MVPF literature and illustrate how they can be used to compute MVPF under different social welfare weights. Additionally, we discuss various statistical issues not considered by Yu (forthcoming), such as the interpretation of estimands under stratified randomization and the relationships among various complier characteristics estimators used in the literature, that are of direct interest to applied researchers. Finally, while Yu (forthcoming) is oriented primarily toward the political persuasion literature, our paper is aimed at a general audience across disciplines.

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)

[2]We became aware of this paper in November 2024. The first version of Yu (forthcoming) was circulated in November 2022 as a job market paper. We first posted ours on arXiv in December 2022.

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 "always takers," "never takers," "compliers," and "defiers." 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 subpopulation shares 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 intent-to-treat effect (or 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 of 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

$$\begin{aligned}
\kappa &\equiv 1 - \frac{D(1-Z)}{\Pr(Z=0)} - \frac{(1-D)Z}{\Pr(Z=1)} \\
\kappa_0 &\equiv \frac{(1-D)(1-Z)}{\Pr(Z=0)} - \frac{(1-D)Z}{\Pr(Z=1)} \\
\kappa_1 &\equiv \frac{DZ}{\Pr(Z=1)} - \frac{D(1-Z)}{\Pr(Z=0)}.
\end{aligned}$$

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.

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 Propo-

sition 1 are applicable to the case where $D = Z$, i.e., the case of perfect compliance with treatment assignment.[3] 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. (**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.2 for details.

Remark 5. (**Adding a Mediator**) The tools we have developed can also shed light on how a binary mediator, M, operates on different populations (the causal chain is $Z \to D \to M \to Y$). If we extend the exclusion restriction and monotonicity assumption to cover M (that is, Y_{zdm}=Y_m and $M_1 > M_0$), our supercomplier estimand simply identifies the characteristics of those with $D_1 > D_0, M_1 > M_0, Y_1 > Y_0$ (the "superdupercompliers"). In addition, we can identify the shares and characteristics of those with $D_1 > D_0, M_1 = M_0 = m$ and $D_1 > D_0, M_1 > M_0, Y_1 = Y_0 = y$ for $m, y = 0, 1$. See Appendix A.3 for details.

Remark 6. (**Binarizing a Non-Binary** Y) As mentioned in the Introduction, we can recode a non-binary Y into a binary variable by categorizing its value as "high" or "low" depending on whether Y exceeds a threshold (Balke and Pearl 1997). This allows us to identify the characteristics of supercompliers corresponding to each "binarized" Y. While binarizing a *treatment* variable can lead to violations of the exclusion restriction and thus invalidate the Wald estimand (Angrist and Imbens 1995, Andresen and Huber 2021, Rose and Shem-Tov 2024), binarizing Y does not violate our IV assumptions or outcome monotonicity.

While we can characterize the supercompliers for any given binarized outcome, we cannot, in general, separately identify the shares and characteristics of different supercomplier types when Y is non-binary. Identification fails even in a simple example where Y takes only three

[3]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.

values (0, 1, and 2) and we have perfect compliance ($D = Z$). Specifically, we consider the case where there are three always taker groups whose Y is always 0, 1, or 2 regardless of treatment, and three complier groups whose Y changes from 0 to 1, from 0 to 2, and from 1 to 2 because of treatment. In Appendix A.4, we show that we cannot identify the shares (and hence the characteristics) of these three supercomplier types in this simple case.

When Y is continuous, it is impossible and impractical to characterize the infinitely many supercomplier types. Instead, we take a different approach by extending the identification results in Lemma 1 and Proposition 1 to accommodate non-binary Y. The proposition below shows that regardless of the distribution of Y, the reduced-form estimand in Lemma 1 identifies the supercomplier population share scaled by the average treatment effect among the supercompliers (supercompliers are still defined as those with $D_1 > D_0$ and $Y_1 > Y_0$ and referred to with the shorthand $G = cc$). Similarly, the Wald estimand in Proposition 1 identifies a weighted average of supercomplier characteristics, where the weight of each individual is proportional to her treatment effect.

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

$$\Pr(G = cc)E[Y_1 - Y_0|G = cc] = E[Y|Z = 1] - E[Y|Z = 0]. \tag{3}$$

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

$$\frac{E[h(X)(Y_1 - Y_0)|G = cc]}{E[Y_1 - Y_0|G = cc]} = \frac{E[h(X)Y|Z = 1] - E[h(X)Y|Z = 0]}{E[Y|Z = 1] - E[Y|Z = 0]}. \tag{4}$$

As we demonstrate in Sections 3 and 4, the weights in (4) connect naturally to the MVPF analysis, and Proposition 2 substantially widens the applicability of our machinery to different types of outcomes.

More broadly, while our discussion here has focused on randomized experiments, our framework can be applied to quasi-experimental research designs by making appropriate adjustments to the underlying assumptions. In a regression discontinuity design, for instance, we can extend our logic and identify supercompliers at the policy threshold under smoothness conditions. We now turn to estimation and inference.

2.2 Estimation and Inference

The Wald-type estimand in (2) and (4) can be implemented via a two-stage least squares (2SLS) regression. For example, when $h(X)$ is the identity function—corresponding to the mean characteristics of supercompliers—we can run this regression in Stata as

$$\texttt{ivregress 2sls XY (Y = Z) [, options]} \tag{5}$$

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

However, equation (5) does not estimate an exact sample analog of the Abadie-style equation (1) from Proposition 1. As we show in Appendix A.5, the estimand from equation (1) also has a Wald-type representation similar to (2):

$$\frac{E[h(X)\{Y-(1-\tau)\}|Z=1]-E[h(X)\{Y-(1-\tau)\}|Z=0]}{E[Y-(1-\tau)|Z=1]-E[Y-(1-\tau)|Z=0]}, \tag{6}$$

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.5, neither of the two estimators based on (1) and (2) has an asymptotic variance that dominates the other in all data generating processes (DGPs).

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). Other empirical studies 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.5 for details). However, these studies do not use 2SLS regression to estimate complier characteristics. Instead, implementation involves assembling separately estimated quantities. In addition, these studies either do not report standard errors on estimated complier characteristics 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.[4]

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

Recent work by Angrist, Hull and Walters (2023) follows and extends the arguments by Abadie (2002) and uses additional estimators to characterize compliers. In addition to the complier analog of (2), which amounts to using κ_1 weights, they also consider weighting covariates by κ_0, equivalent to replacing Y by $1-D$ in (2). Angrist, Hull and Walters (2023) also propose a "pooled" estimator, which is the average of the κ_1 and κ_0 weighted estimators and can be implemented via a stacked regression. While we can easily use the supercomplier analogs of these additional estimators, replacing Y with $1-Y$ seems unnatural when Y is non-binary. Therefore, we simply use the sample analog of the Wald estimand in (2) and (4) to estimate supercomplier characteristics, which is akin to the complier IV estimators used by Alsan et al. (2024) when the treatment variable is non-binary.

Remark 7. (**Characteristics Distribution and Quantile Estimation**) As we discussed in Remark 3, we can identify the entire X distribution among supercompliers. It is straightforward to estimate the c.d.f. of X: we can replace the dependent variable in (5) 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 encountered by Abadie, Angrist and Imbens (2002) in complier quantile estimation 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.6 for details.

Remark 8. (**Conditional Independence and Stratified Randomization**) Our identification 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 estimating supercomplier characteristics. In Appendix A.7, 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.

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).[5] 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.

2.3.1 Assumption Testing

We build on and extend results from Kitagawa (2015) and propose a sharp characterization of Assumptions 1.1-1.3 and 2.1.[6] The resulting test takes the form of a set of inequalities

[5] 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 equivalent.

[6] 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.

that must jointly hold.[7] 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 3 *Given the potential outcomes model described in Section 2.1, (i) under Assumptions 1.1-1.3 and 2.1, the following inequalities hold*

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

$$\Pr(Y=1, D=0|Z=0) - \Pr(Y=1, D=0|Z=1) \geq 0 \tag{8}$$

$$\Pr(Y=1|Z=1) - \Pr(Y=1|Z=0) \geq 0; \tag{9}$$

(ii) if inequalities (7)-(9) 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 in the Appendix) reveal an intuitive interpretation of inequalities (7)-(9). 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 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 3. 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 3 reveals that all the information available for evaluating outcome monotonicity is contained within the reduced form estimate (Inequality 9), which translates to a nonnegative average treatment effect.

[7]Note that testing treatment and outcome monotonicity (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.

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

$$Y_i^s = \phi^s + \theta_1 Z_i 1_{[s=1]} + \theta_2 (1 - Z_i) 1_{[s=2]} + \theta_3 Z_i 1_{[s=3]} + \varepsilon_i^s, \tag{10}$$

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 Σ which can be consistently estimated by estimator $\hat{\Sigma}$, we can easily generate the critical value of the test via simulation. That is, we draw S 3-dimensional random vectors from the distribution $\mathrm{N}(0,\hat{\Sigma})$, take the minimum of each vector within each draw, and the critical value is the 5th percentile of the S minimums. We reject the null hypothesis and conclude violations of the identifying assumptions when $\hat{\theta}$ is less than this critical value.

Remark 9. We can generalize the test in Proposition 1 by removing the binary restriction on Y. In Appendix A.8, we state a sharp joint test of our identifying assumptions, which allows Y to have an arbitrary distribution. We implement this generalized test in our empirical applications where Y represents earnings, an important outcome in MVPF analyses.

Remark 10. Inequality (9) 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 (9) alone. However, it is a very weak test—while a significantly negative reduced form effect contradicts the null of $Y_1 \geq Y_0$, a positive reduced form effect can mask underlying violations of outcome monotonicity. To that end, we can incorporate covariates to increase the power of the test. Specifically, instead of relying only on inequality (9), 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 (9) 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 (7)-(9), 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 (9) 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 outcome monotonicity violations next.

2.3.2 Relaxing Outcome Monotonicity

Proposition 4 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 4 *Under Assumptions 1 and 2.2 and provided that* $X \perp\!\!\!\perp Z$*,*

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

where we define

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

Proposition 4 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.

In spite of our discussion here and in Section 2.3, we acknowledge that needing outcome monotonicity is a drawback of our supercompliers framework relative to subsample hetero-

geneity analysis, which increasingly relies on machine learning techniques (e.g, Athey and Imbens 2016, Wager and Athey 2018; see Smith 2022 for a recent review). But our approach also has several advantages. First, it allows imperfect compliance with treatment, from which machine learning methods often abstract away. Second, it avoids any approximation of the conditional expectation of potential outcomes or treatment effects with covariates. Consequently, our result does not depend on the choice of a particular machine learning algorithm or tuning parameters therein (see, for example, Knaus, Lechner and Strittmatter 2021 for an extensive exercise comparing techniques across many data generating processes). Finally, our approach is transparent to understand, easy to implement, and cheap to compute with existing statistical software commands.

3 Relationship to Welfare Analysis

In this section, we discuss how supercomplier characteristics can be used to enrich welfare analyses. Specifically, documenting the differences between supercompliers and compliers informs how social welfare would change if a social planner put different "weights" on individuals of different circumstances. More concretely, we describe how social weights are (and are not) used in the widely adopted framework of Hendren and Sprung-Keyser (2020), and elaborate on how to appropriately weight treatment effects in this framework using supercomplier characteristics.

Hendren and Sprung-Keyser (2020) advocate for the systematic reporting of the marginal value of public funds, or MVPF, to facilitate comparisons of public policies. The starting point for the MVPF is that the government is interested in the impact of a policy on social welfare, defined as the weighted sum of individual utilities:

$$W = \sum_i \eta_i U_i.$$

U_i is expressed as a money-metric, and η_i is the social welfare impact of transferring \$1 to individual i.[8] A small policy change, denoted by dp, impacts social welfare by:

$$\frac{dW}{dp} = \sum_i \eta_i \frac{dU_i}{dp} \equiv \sum_i \eta_i WTP_i, \tag{11}$$

[8] Any utility function can be normalized to a money-metric by dividing by the marginal utility of income.

where WTP_i is interpreted as individual i's "willingness to pay" (WTP) for the policy. Equation (11) encapsulates the benefits of the policy change. The cost of the policy equals its impact on the government's budget, denoted by G. Hendren and Sprung-Keyser (2020) define the MVPF formally as

$$MVPF = \frac{\sum_i WTP_i}{G},$$

or, the ratio of the (unweighted) societal willingness to pay to the government net costs. Note that while the benefits of a policy change should incorporate social weights, the formal definition of the MVPF leaves the weights out, presumably because it is difficult to practically implement. Hendren and Sprung-Keyser (2020) suggest, therefore, to use MVPFs to compare policies that benefit similar beneficiaries, so as to "reduce the role of social preferences in driving conclusions." As shown below, we can use supercomplier characteristics to incorporate weights into the MVPF to reflect a social planner's redistributional preferences, so long as the weights are functions of the characteristics reported.

The MVPF of a policy is calculated by plugging in estimates of its effect on the government budget (denominator) and, in many cases, the willingness to pay (numerator). In a typical application, the policy's effect on beneficiaries' (and possibly their children's) earnings are translated into an estimate of changes in tax revenue collected. The denominator may also incorporate spillover effects on government spending, such as changes in healthcare utilization among Medicare and Medicaid beneficiaries, changes in public school enrollment, or changes in individual contact with the criminal justice system. The willingness to pay is the dollar value that is transferred to individuals (before behavioral responses, per the envelope theorem) for policies that include a transfer or tax. However, if a policy affects later-life outcomes (e.g., additional education or improved health), causal estimates of these effects could better capture social benefits and therefore enter the numerator.

The use of causal estimates as "plug-in" components of an MVPF calculation means that it corresponds to the subpopulation for which the causal estimates are identified. In particular, when using estimates from a randomized controlled trial with imperfect compliance as inputs, the resulting MVPF is the value of the policy for the (treatment) complier group.

A social planner may desire to place different weights on individuals' willingness to pay, even within the complier group. Suppose that the WTP is measured by the effect of the policy on an outcome Y, so that $WTP_i = Y_{1i} - Y_{0i}$. When Y is binary, it is easy to see that the

weighted WTP for the complier group is

$$E[\eta_i(Y_{1i} - Y_{0i})|\text{complier}] = E[\eta_i|\text{supercomplier}]\text{LATE}_Y \tag{12}$$

In other words, the weighted WTP is simply the (unweighted) local average treatment effect on the outcome of interest (LATE_Y), multiplied by the average social welfare weight of the supercompliers. We show in Appendix C that this logic extends to the case where Y is not binary. An important implication of this result is that if the social weights η_i are a linear function of observable characteristics X_i, one can calculate the appropriately weighted WTP (and corresponding MVPF) from reported supercomplier characteristics. Linearity need not be restrictive in practice. In our empirical applications below, for example, baseline family income was recorded in predefined bins. By including an indicator variable for each bin (i.e., a fully saturated set of dummies), we effectively allow the welfare weights to vary flexibly across income categories, and the specification remains linear.

While it is possible to simply estimate the weighted willingness to pay directly by estimating the treatment effect on weighted outcomes, the formulation in (12) shows that it can be done without microdata. If supercomplier characteristics are documented along with the LATE, a reader can construct a customized weighted WTP and MVPF themselves, to the extent that the desired social weights are a linear function of those reported characteristics. A major advantage of this approach is that it allows for readers to compute the MVPF using different social weights, which may not necessarily be the same as the weights chosen by the researchers who estimate the relevant treatment effect.

4 Empirical Applications

To demonstrate the utility our tools for characterizing supercompliers, we consider experimental evaluations of the Job Corps and Job Training Partnership Act programs.

4.1 Job Corps

Job Corps is a federally sponsored job training program for disadvantaged youths in the United States. The program offers a suite of services, including academic education, vocational training, and job search assistance. In the 1990s, Department of Labor sponsored the National Job Corps Study, where eligible Job Corps applicants between ages 16 and 24 were randomly assigned access to Job Corps. Researchers have used this experiment to study the

effects of Job Corps on various outcomes (Schochet, Burghardt and Glazerman 2001).

Consistent with Job Corps's mission, there were large impacts on educational attainment during the 48-month follow-up period in the public use data (Schochet, Burghardt and McConnell 2008*b*). Members of the treatment group were 22.3 percentage points more likely to obtain a vocational certificate (37.5 percent compared to 15.2 percent; *t*-stat: 27.1; *N*: 11,151). Among students without a high school credential at random assignment, members of the treatment group were 15.0 percentage points more likely to obtain a GED (41.6 vs. 26.6 percent; *t*-stat: 14.3; *N*: 8,579). Job Corps also improved labor market outcomes, though the effects are more moderate. Specifically, the intent-to-treat effects on employment and earnings in the 16th quarter after random assignment—the final quarter of study by Schochet, Burghardt and Glazerman (2001)—were 2.4 percentage points (71.1 vs. 68.7 percent; *t*-stat: 2.59; *N*: 10,872) and $266, or 11 percent of the control mean, (*t*-stat: 4.76; *N*: 10,872), respectively.

In Table 2, we present supercomplier and complier characteristics with respect to educational attainment, as well as differences in characteristics between supercompliers and both compliers and the full experimental population (standard errors for the differences are estimated via stacked regressions). In Panel A, we examine characteristics for those attaining a GED, restricting to the sample without high school credentials prior to random assignment. In Panel B, we examine characteristics for those attaining a vocational credential. In both panels, we define (treatment) compliers as those who ultimately enrolled in Job Corps. They made up 73 percent and 72 percent of the respective samples. Panel A shows that supercompliers with respect to GED attainment were more likely to be female and white, and also to have had employment prior to random assignment, all relative to both the full population and compliers. Although not statistically significant at the 5 percent level, supercompliers also appeared to be slightly more advantaged with respect to family income. Similar patterns hold for supercompliers with respect to vocational certificate attainment (Panel B), and here we also find significant differences in arrest history, where supercompliers were less likely to have had a prior arrest.

In Table 3, we present supercomplier characteristics with respect to a labor market outcome. Because of the relatively weak intent-to-treat effect on employment reported above, we focus on earnings in the 16th quarter after random assignment. The patterns in Table 3 share many similarities with Table 2: supercompliers were more likely to be white, older, less likely to

have been arrested, and more likely to have had an employment history. But the magnitude differences are much more pronounced for two of these characteristics. Compared to Table 2, supercompliers were substantially more likely to be white and age 20 or older. While not statistically significant at the 5 percent level, point estimates suggest that supercompliers were also more likely to come from higher-income families. Finally, for all outcomes in Tables 2 and 3, our tests based on Propositions 3 and 10 cannot reject our identifying assumptions at the 5 percent level, indicating their plausibility.

Overall, our labor market findings combined with our results on educational attainment suggest that participants who benefited most from Job Corps tended to be those who were less marginalized upon application: they were less likely to have had an arrest history, more likely to have been white, more likely to have had an employment history, and tended to be from families with higher income. We return to this point and its normative implications at the end of Section 4.2.

4.2 Job Training Partnership Act

The National Job Training Partnership Act (JTPA) Study was a randomized controlled trial evaluating job training programs in 16 sites across the US in the late 1980s. The JTPA programs primarily target economically disadvantaged adults and out-of-school youths. Eligible applicants to each site's JTPA program were randomly assigned to a control group or a treatment group, which was offered one of three service types: classroom training, subsidized on-the-job training, or other services. Following Bloom et al. (1997), we focus on earnings in the 30-month follow-up period as the main outcome of interest. Hendren and Sprung-Keyser (2020) also rely on this outcome for the MVPF calculation of JTPA.

While Bloom et al. (1997) estimated program effects for four distinct groups—adult men, adult women, male youths, and female youths—we restrict our analysis to adult women.[9] Among the four groups, adult women saw the largest impact of the intervention with an intent-to-treat effect of \$1,190, or 10 percent of the control mean (t-stat: 3.63; N: 6,102). Adult men saw an intent-to-treat effect of \$1,051, or 6 percent of the control mean (t-stat: 2.01; N: 5,102). The intent-to-treat effects for the youth groups were insignificant at the 5 percent level. Since the intent-to-treat effect is the supercomplier "first stage," it is only

[9]The treatment take-up effects were around 60 percent for all four groups.

strong for adult women by the conventional rule of thumb of F-stat > 10.[10]

Following the same format as Tables 2 and 3, Table 4 reports the characteristics of the adult women in the study. We find that compliers were similar to the experimental population, but that supercompliers were somewhat different. With the caveat that standard errors are large and that none of the differences are statistically significant at the 5 percent level, we see that supercompliers were less attached to the labor force at baseline: they had lower prior earnings and weeks worked, were more likely to be on Aid to Families with Dependent Children (AFDC, i.e., cash assistance), and had lower family income.[11] As with Job Corps, the test based on Proposition 10 also cannot reject our identifying assumptions in the JTPA data for the earnings outcome.

We now turn to the normative implications of these results. Following the framework from Section 3, we estimate the JTPA program's socially weighted MVPF. Hendren and Sprung-Keyser (2020) estimate that the *unweighted* MVPF of the JTPA program for adults (pooling men and women) is 1.38. This is calculated by assuming that participants value the program by their post-tax earnings impacts, net of AFDC benefits. Since the earnings impacts for women appear to be driven by less advantaged participants, and to the extent that one places a larger social weight on these participants, the weighted MVPF will be higher.[12]

To compute the weighted WTP (the numerator of the weighted MVPF), we first estimate the average social weight of the supercompliers. We assume a textbook social welfare function $\Psi(u) = \frac{u^{1-\phi}}{1-\phi}$ (Salanié, 2011), where ϕ denotes the extent to which the social planner desires

[10]Recent studies by Lee et al. (2022) and Angrist and Kolesár (2024) reexamine the reliability of inference procedures for just identified single-instrumental-variable models. We first note that the "first-stage" F-stats (square of the t-stats of the intent-to-treat effects) for the two panels in Table 2 are sufficiently strong (above the threshold of 142.6 according to Lee et al. 2022) to dispel inference concerns. In addition to the statistics reported in Tables 3 and 4, we have also examined the Anderson-Rubin (AR) confidence intervals for supercomplier characteristics: the 95 percent AR confidence intervals are 9 to 13 percent (17 to 25 percent) wider for the characteristics reported in Table 3 (Table 4) than their conventional counterparts. However, we can show that the statistical significance (at the 5 percent level) of the differences for white and older youths in Table 3 is unchanged with AR intervals by applying a conservative test that results from the Bonferroni inequality. Angrist and Kolesár (2024) advocate for screening for the sign of the "first stage" coefficient, and all of our intent-to-treat effects have the expected signs.

[11]Note that estimated characteristics shares of supercompliers are not guaranteed to lie between zero and one (in fact, neither are complier characteristics estimates). It is reassuring that none of our point estimates are negative or larger than one in Tables 2-4.

[12]While the WTP calculated by Hendren and Sprung-Keyser (2020) contains earnings and AFDC impacts for both men and women, we only weight the earnings impacts for women. As mentioned above, the reduced form earnings effect for men is not strong enough to compute supercomplier weights. The same is true of AFDC impacts for both men and women. Furthermore, JTPA impacts on AFDC contribute minimally to the MVPF calculation: without accounting for it, the unweighted MVPF is 1.35 (vs. 1.38).

redistribution and u is income. A value of $\phi = 0$ means that a dollar transfer has the same impact on social welfare regardless of income, while larger values of ϕ imply greater preference for redistribution. For a given ϕ, the estimation of the average supercomplier social weight proceeds in three steps:

1. Calculate the marginal social welfare of transferring \$1 to someone at the midpoint of each of the five income bins. That is, calculate $u^{-\phi}$ by setting u equal to each midpoint. We truncate the top income bin (greater than \$12K) to (\$12K,\$15K) for the purpose of calculating marginal utilities.[13]

2. Calculate the social weights by normalizing the marginal welfare so that they average to one across the five income bins. For example, when $\phi = 0.5$, the weights are 1.83, 1.05, 0.82, 0.69, and 0.61, respectively.

3. Multiply each weight with the corresponding supercomplier share from Table 4 and sum the products to arrive the average supercomplier social weight. When $\phi = 0.5$, the average weight estimate is 1.29.

After these three steps, we multiply the average supercomplier weight with the reported LATE as per equation (12), and plug the result into the companion program by Hendren and Sprung-Keyser (2020) to arrive at the weighted MVPF. We find that when $\phi = 0.5$, the weighted MVPF is 1.63, while $\phi = 1$ implies a weighted MVPF of 1.97.

As discussed above, this MVPF exercise depends on the normalization of the weights. Had we included additional higher income bins, supercompliers would receive an even greater weight, driving up the MVPF. The bottom line is that, whatever the choice of social welfare weights, it is possible to compute the weighted MVPF using only the reported supercomplier characteristics, provided that the weights are a linear function of those characteristics.

Finally, we note the contrast between the results for the JTPA and Job Corps studies. The earnings supercompliers (among women) in the JTPA study are less advantaged compared to the experimental population while the opposite is true in the Job Corps study. Although we do not have the data to estimate a weighted MVPF for Job Corps that directly corresponds to Hendren and Sprung-Keyser (2020), our results based on publicly available survey data indicate that incorporating social welfare weights would decrease the MVPF for Job Corps

[13]This truncation artificially lowers the assumed income for individuals in the top bin, thereby inflating their marginal utility, resulting in a higher social weight for this income bin, and biasing against our results.

youths by up to 43 percent depending on the value of ϕ.[14] In contrast, the MVPF with social welfare weights increases by up to 43 percent for the JTPA adults as shown above.

5 Conclusion

In this paper, we develop methods to characterize the supercomplier subpopulation in a canonical instrumental variable framework with a binary instrument and treatment. We show that under the standard LATE assumptions plus outcome monotonicity, we can identify the characteristics of supercompliers. Because the plausibility of our identifying assumptions may depend on context, we develop sharp joint tests 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 instrumental variable regressions via existing Stata commands. Finally, we show that supercomplier characteristics can be used to incorporate social weights in a welfare analysis of the marginal value of public funds.

We illustrate the utility of our tools using data from two job-training experiments, the National Job Corps Study and the National Job Training Partnership Act Study. We find that participants whose labor market outcomes were improved by Job Corps had relatively higher baseline incomes, while the adult women who benefited from JTPA had relatively lower baseline incomes. To the extent that a social planner values redistribution, our results imply a lower MVPF for Job Corps and a higher MVPF for JTPA.

References

Abadie, Alberto. 2002. "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models." *Journal of the American Statistical Association*, 97(457): 284–292.

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

[14]Hendren and Sprung-Keyser (2020) estimate an MVPF of 0.18 for Job Corps based on 20-year follow-up impacts from Schochet (2018), which come from tax data that are not publicly available. The impacts on earnings from the tax data are more muted even in the short run, and one explanation by Schochet, Burghardt and McConnell (2008*a*) is that survey data includes informal earnings not captured by tax data.

Abadie, Alberto, Joshua D. Angrist, and Guido W. 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. 2023. "Misdemeanor Prosecution." *Quarterly Journal of Economics*, 138(3): 1453–1505.

Alsan, Marcella, Arkey Barnett, Peter Hull, and Crystal Yang. 2024. "'Something Works' in U.S. Jails: Misconduct and Recidivism Effects of the IGNITE Programs." mimeograph.

Andresen, Martin E., and Martin Huber. 2021. "Instrument-Based Estimation with Binarised Treatments: Issues and Tests for the Exclusion Restriction." *Econometrics Journal*, 24(3): 536–558.

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 University Press.

Angrist, Joshua D., and Michal Kolesár. 2024. "One Instrument to Rule Them All: The Bias and Coverage of Just-ID IV." *Journal of Econometrics*, 240(2): 105398.

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.

Angrist, Joshua D., Peter Hull, and Christopher R. Walters. 2023. "Methods for Measuring School Effectiveness." In *Handbook of the Economics of Education* Vol. 7, ed. Eric A. Hanushek, Stephen Machin and Ludger Woessmann, Chapter 1, 1–60. Elsevier.

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

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.

Bloom, Dan, and MDRC. 2022. "Connecticut's Jobs First Program Analysis Data, 1996-2001."

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.

Bloom, Howard S., Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. 1997. "The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study." *Journal of Human Resources*, 32(3): 549–576.

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, Sokbae Lee, and Adam M. Rosen. 2013. "Intersection Bounds: Estimation and Inference." *Econometrica*, 81(2): 667–737.

Chernozhukov, Victor, Wooyoung Kim, Sokbae Lee, and Adam M. Rosen. 2015. "Implementing Intersection Bounds in Stata." *Stata Journal*, 15(1): 21–44.

Comey, Matthew L., Amanda R. Eng, and Zhuan Pei. 2023. "Supercompliers." arXiv:2212.14105v2.

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.

Hendren, Nathaniel, and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *Quarterly Journal of Economics*, 135(3): 1209–1318.

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. 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.

Lee, David S., Justin McCrary, Marcelo J. Moreira, and Jack Porter. 2022. "Valid t-ratio Inference for IV." *American Economic Review*, 112(10): 3260–3290.

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.

Orr, Larry L., Howard S. Bloom, Stephen H. Bell, Winston Lin, George Cave, and Fred Doolittle. 2003. "Data for: The National JTPA Study: Impacts, Benefits, And Costs of Title II-A."

Ouss, Aurelie, and Megan Stevenson. 2023. "Does Cash Bail Deter Misconduct?" *American Economic Journal: Applied Economics*, 15(3): 150–182.

Rose, Evan K., and Yotam Shem-Tov. 2024. "On Recoding Ordered Treatments as Binary Indicators." NBER Working Paper 32234.

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

Salanié, Bernard. 2011. *The Economics of Taxation.* MIT press.

Schochet, Peter Z. 2018. "National Job Corps Study: 20-YearFollow-Up Study Using Tax Data." Mathematica Policy Research.

Schochet, Peter Z., John Burghardt, and Sheena McConnell. 2008*a*. "Does Job Corps Work? Impact Findings from the National Job Corps Study." *American Economic Review*, 98(5): 1864–1886.

Schochet, Peter Z., John Burghardt, and Sheena McConnell. 2008*b*. "Replication data for: Does Job Corps Work? Impact Findings from the National Job Corps Study."

Schochet, Peter Z., John Burghardt, and Steven Glazerman. 2001. "National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes." Mathematica Policy Research.

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.

Yu, Zeyang. forthcoming. "A Binary IV Model for Persuasion: Profiling Persuasion Types among Compliers." *Econometrics Journal*.

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: Job Corps Supercomplier Characteristics for Education Outcomes

Panel A: Outcome: Receiving GED

	Means			Differences	
	Population	Complier	Supercomplier	Pop-SC	C-SC
Female	0.38	0.38	0.48	0.10	0.11
	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
White	0.25	0.25	0.34	0.08	0.09
	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
Age >=20	0.17	0.16	0.21	0.04	0.05
	(0.00)	(0.01)	(0.03)	(0.03)	(0.03)
Never Arrested	0.69	0.71	0.78	0.09	0.06
	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
Previously Employed	0.60	0.61	0.75	0.15	0.14
	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)
Family Income <$3K	0.16	0.15	0.11	-0.05	-0.04
	(0.00)	(0.01)	(0.03)	(0.03)	(0.03)
Family Income $3-6K	0.13	0.12	0.09	-0.04	-0.04
	(0.00)	(0.01)	(0.03)	(0.03)	(0.03)
Family Income $6-9K	0.06	0.07	0.05	-0.02	-0.02
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)
Family Income $9-12K	0.06	0.06	0.09	0.03	0.03
	(0.00)	(0.00)	(0.02)	(0.02)	(0.02)
Family Income >$12K	0.59	0.60	0.67	0.08	0.07
	(0.01)	(0.01)	(0.04)	(0.04)	(0.04)

Panel B: Outcome: Receiving Vocational Certificate

	Means			Differences	
	Population	Complier	Supercomplier	Pop-SC	C-SC
Female	0.41	0.39	0.44	0.03	0.04
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)
White	0.27	0.27	0.31	0.04	0.05
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)
High School	0.23	0.21	0.24	0.01	0.03
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)
Age >=20	0.27	0.25	0.28	0.01	0.03
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)
Never Arrested	0.71	0.73	0.78	0.07	0.05
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)
Previously Employed	0.64	0.64	0.69	0.05	0.05
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)
Family Income <$3K	0.16	0.15	0.16	0.00	0.01
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Family Income $3-6K	0.13	0.12	0.12	0.00	0.00
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Family Income $6-9K	0.07	0.07	0.09	0.02	0.01
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Family Income $9-12K	0.06	0.06	0.06	0.00	0.00
	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
Family Income >$12K	0.59	0.59	0.57	-0.01	-0.02
	(0.00)	(0.01)	(0.02)	(0.02)	(0.02)

Notes: This table reports the population (Pop), complier (C), and supercomplier (SC) averages in baseline characteristics as well as the Pop-SC and C-SC differences. As with Schochet, Burghardt and Glazerman (2001) and Schochet, Burghardt and McConnell (2008*a*), we use weights included in the public use data to adjust for sample and survey designs. Standard errors are in parentheses.

Table 3: Job Corps Supercomplier Characteristics for Earnings

	Means			Differences	
	Population	Complier	Supercomplier	Pop-SC	C-SC
Female	0.41	0.39	0.38	-0.03	-0.02
	(0.01)	(0.01)	(0.13)	(0.13)	(0.12)
White	0.27	0.27	0.65	0.38	0.39
	(0.00)	(0.01)	(0.15)	(0.15)	(0.14)
High School	0.23	0.22	0.25	0.02	0.04
	(0.00)	(0.01)	(0.13)	(0.13)	(0.12)
Age >=20	0.27	0.25	0.55	0.28	0.30
	(0.00)	(0.01)	(0.14)	(0.14)	(0.14)
Never Arrested	0.72	0.74	0.79	0.08	0.05
	(0.00)	(0.01)	(0.13)	(0.13)	(0.13)
Previously Employed	0.64	0.64	0.75	0.12	0.12
	(0.00)	(0.01)	(0.12)	(0.12)	(0.12)
Family Income <$3K	0.16	0.15	0.10	-0.06	-0.05
	(0.00)	(0.01)	(0.09)	(0.09)	(0.09)
Family Income $3-6K	0.13	0.12	0.01	-0.11	-0.11
	(0.00)	(0.00)	(0.09)	(0.09)	(0.09)
Family Income $6-9K	0.07	0.07	0.01	-0.06	-0.06
	(0.00)	(0.00)	(0.08)	(0.08)	(0.07)
Family Income $9-12K	0.06	0.06	0.07	0.00	0.00
	(0.00)	(0.00)	(0.06)	(0.06)	(0.06)
Family Income >$12K	0.58	0.59	0.81	0.23	0.22
	(0.00)	(0.01)	(0.14)	(0.14)	(0.14)

Notes: This table reports the population (Pop), complier (C), and supercomplier (SC) averages in baseline characteristics as well as the Pop-SC and C-SC differences for the outcome of earnings during the 16th quarter after random assignment (quarter 16 is the final quarter of study for the initial Job Corps assessment by Schochet, Burghardt and Glazerman 2001). As with Schochet, Burghardt and Glazerman (2001) and Schochet, Burghardt and McConnell (2008*a*), we use weights included in the public use data to adjust for sample and survey designs. Standard errors are in parentheses.

Table 4: National JTPA Study Supercomplier Characteristics for Earnings

	Means			Differences	
	Population	Complier	Supercomplier	Pop-SC	C-SC
Black	0.26	0.24	0.37	0.11	0.13
	(0.01)	(0.01)	(0.13)	(0.13)	(0.13)
Hispanic	0.12	0.13	0.02	-0.09	-0.10
	(0.00)	(0.01)	(0.09)	(0.09)	(0.09)
Has High School/GED Credential	0.68	0.69	0.70	0.03	0.02
	(0.01)	(0.01)	(0.14)	(0.14)	(0.13)
Ever Received Vocational Training	0.45	0.45	0.48	0.03	0.03
	(0.01)	(0.01)	(0.15)	(0.15)	(0.15)
Earnings in Year before RA	2489	2461	1773	-717	-688
	(40)	(63)	(1282)	(1279)	(1284)
Worked 1-12 Weeks in Year before RA	0.16	0.16	0.21	0.04	0.05
	(0.00)	(0.01)	(0.10)	(0.10)	(0.10)
Worked 13-52 Weeks in Year before RA	0.43	0.45	0.13	-0.29	-0.32
	(0.01)	(0.01)	(0.18)	(0.18)	(0.18)
Received AFDC in Year before RA	0.38	0.38	0.49	0.11	0.10
	(0.01)	(0.01)	(0.16)	(0.16)	(0.16)
Married	0.22	0.22	0.29	0.07	0.07
	(0.01)	(0.01)	(0.12)	(0.12)	(0.12)
Number of Children	1.75	1.76	2.51	0.76	0.74
	(0.02)	(0.03)	(0.47)	(0.47)	(0.47)
Family Income <$3K	0.31	0.29	0.46	0.15	0.17
	(0.01)	(0.01)	(0.13)	(0.13)	(0.13)
Family Income $3-6K	0.34	0.35	0.15	-0.18	-0.20
	(0.01)	(0.01)	(0.15)	(0.15)	(0.15)
Family Income $6-9K	0.16	0.16	0.26	0.10	0.10
	(0.00)	(0.01)	(0.11)	(0.11)	(0.11)
Family Income $9-12K	0.09	0.09	0.04	-0.05	-0.05
	(0.00)	(0.01)	(0.10)	(0.10)	(0.10)
Family Income >$12K	0.09	0.10	0.08	-0.02	-0.03
	(0.00)	(0.01)	(0.10)	(0.10)	(0.10)

Notes: This table reports the population (Pop), complier (C), and supercomplier (SC) averages in baseline characteristics as well as the Pop-SC and C-SC differences. Our regressions control for the same covariates as in the impact analysis by Bloom et al. (1997). Standard errors are in parentheses. "RA" is short for random assignment. Data source: Orr et al. (2003).

Appendix (For Online Publication Only)

A Additional Theoretical Results

A.1 Proofs of Lemma 1 and Propositions 1-4

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,

$$\begin{aligned}
&E[h(X)|G = cc] \Pr(G = cc|D_1 > D_0) \\
=&E[h(X)|D_1 > D_0] - E[h(X)|G = ca] \Pr(G = ca|D_1 > D_0) \\
&- E[h(X)|G = cn] \Pr(G = cn|D_1 > D_0). \qquad \text{(A1)}
\end{aligned}$$

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,

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

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.

For equation (2), we note that it is a special case of Proposition 2, which we prove next. Specifically, when Y is binary, we can plug $Y_1 - Y_0 = 1$ into equation (4) and obtain the desired result. *QED.*

Proof of Proposition 2: For equation (3),

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

For equation (4), the first term in the numerator of the Wald estimand is

$$\begin{aligned}
&E[h(X)Y|Z=1] \\
=&E[h(X)Y|Z=1, (D_1 = D_0 \text{ or } Y_1 = Y_0)] \Pr(D_1 = D_0 \text{ or } Y_1 = Y_0|Z=1) + \\
&E[h(X)Y|Z=1, G=cc] \Pr(G=cc|Z=1).
\end{aligned}$$

We can similarly expand the second term in the numerator. The difference between the two terms is nonzero only for the supercomplier group. Therefore,

$$\begin{aligned}
&E[h(X)Y|Z=1] - E[h(X)Y|Z=0] \\
=&E[h(X)Y|Z=1, G=cc] \Pr(G=cc|Z=1) - \\
&E[h(X)Y|Z=0, G=cc] \Pr(G=cc|Z=0) \\
=&E[h(X)(Y_1 - Y_0)|G=cc] \Pr(G=cc).
\end{aligned}$$

Plugging the left hand side of equation (3) into the denominator of the Wald estimand, we have the desired result. *QED.*

Proof of Proposition 3: Since Proposition 3 is a special case of Proposition 10 in Appendix A.8 below, we omit its proof here. Interested readers can still find a proof specifically for the simpler Proposition 3 in a previous version of our paper Comey, Eng and Pei (2023).

Proof of Proposition 4: The numerator of the estimand is

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

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

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 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*

$$
\begin{aligned}
\Pr(G=ca) &= E[\kappa_0 Y] = E[(1-D)Y|Z=0] - E[(1-D)Y|Z=1] \\
\Pr(G=cn) &= E[\kappa_1(1-Y)] = E[D(1-Y)|Z=1] - E[D(1-Y)|Z=0],
\end{aligned}
$$

and their characteristics can be identified as

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

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. 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.3 Adding a Mediator

In this section, we provide details for Remark 5. To incorporate a mediator M, we first expand on Assumptions 1 and 2.

Assumption 3 (Mediator)

1. *Random Assignment:* $(\{Y_{zdm}\}, \{M_{zd}\}, \{D_z\}) \perp\!\!\!\perp Z$ *for* $z, d, m = 0, 1$ *and* $0 < \Pr(Z = 1) < 1$.

2. *Exclusion:* $\Pr(M_{1d} = M_{0d}) = 1$ *for* $d \in \{0, 1\}$ *and* $\Pr(Y_{zdm} = Y_{z'd'm}) = 1$ *for* $\{zd\} \neq \{z'd'\}$ *and* $m \in \{0, 1\}$.

3. *Monotonicity:* $\Pr(D_1 \geq D_0) = 1$, $\Pr(M_1 \geq M_0) = 1$, *and* $\Pr(Y_1 \geq Y_0) = 1$.

4. *Causal Chain:* $\Pr(Y_1 > Y_0, M_1 > M_0, D_1 > D_0) > 0$.

The inclusion of mediator M expands the dimensionality of the principal stratum partitioning from Table 1. With the monotonicity conditions in Assumption 3.3, we now have 27 admissible sub-groups, as opposed to nine admissible sub-groups under Assumptions 1 and 2 alone. We denote each of the groups in the expanded stratification as $\tilde{G} = dmy$ for $d, m, y = a, c, n$.

Proposition 6 *Under Assumption 3 and provided that* $X \perp\!\!\!\perp Z$,

(A) The $\tilde{G} = ccc$ *group ("super-duper complier") share is*

$$\Pr(\tilde{G} = ccc) = E[Y|Z = 1] - E[Y|Z = 0];$$

and its mean characteristics

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

(B) The $\tilde{G} = cca$ group share is

$$\Pr(\tilde{G} = cca) = E[(1-M)Y|Z=0] - E[(1-M)Y|Z=1],$$

and its mean characteristics

$$E[h(X)|\tilde{G} = cca] = \frac{E[(1-M)Yh(X)|Z=1] - E[(1-M)Yh(X)|Z=0]}{E[(1-M)Y|Z=1] - E[(1-M)Y|Z=0]}$$

(C) The $\tilde{G} = ccn$ group share is

$$\Pr(\tilde{G} = ccn) = E[M(1-Y)|Z=1] - E[M(1-Y)|Z=0],$$

and its mean characteristics

$$E[h(X)|\tilde{G} = ccn] = \frac{E[M(1-Y)h(X)|Z=1] - E[M(1-Y)h(X)|Z=0]}{E[M(1-Y)|Z=1] - E[M(1-Y)|Z=0]}.$$

Proof of Proposition 6:

(A) The key insight for this part is that $\tilde{G} = ccc$ the only group whose Y changes with Z. The formal proof follows the same logic as that in the proof of Proposition 2 and is therefore omitted.

(B) For the share $\Pr(\tilde{G} = cca)$, note that

$$\begin{aligned} &E[(1-M)Y|Z=0] \\ =&\Pr(D=1, M=0, Y=1|Z=0) + \Pr(D=0, M=0, Y=1|Z=0) \\ =&\Pr(\tilde{G} = ana) + \Pr(\tilde{G} = cca, cna, nca, nna). \end{aligned} \tag{A2}$$

Similarly, we have

$$E[(1-M)Y|Z=1] = \Pr(\tilde{G} = ana, cna) + \Pr(\tilde{G} = nca, nna). \tag{A3}$$

Taking the difference of (A2) and (A3), we have the desired result.

For average characteristics, the first term in the numerator of the Wald estimand is

$$\begin{aligned}
&E[(1-M)Yh(X)|Z=1] \\
=&E[D(1-M)Yh(X)|Z=1]+E[(1-D)(1-M)Yh(X)|Z=1] \\
=&E[h(X)|\tilde{G}=ana,cna]\Pr(\tilde{G}=ana,cna)+ \\
&E[h(X)|\tilde{G}=nca,nna]\Pr(\tilde{G}=nca,nna) \\
=&\sum_{\tilde{g}=ana,cna,nca,nna} E[h(X)|\tilde{G}=\tilde{g}]\Pr(\tilde{G}=\tilde{g}). \qquad \text{(A4)}
\end{aligned}$$

Similarly, we can show that

$$E[(1-M)Yh(X)|Z=0]=\sum_{\tilde{g}=ana,cna,nca,nna,cca} E[h(X)|\tilde{G}=\tilde{g}]\Pr(\tilde{G}=\tilde{g}). \qquad \text{(A5)}$$

We have the desired result after taking the difference of (A4) and (A5) and replacing the denominator of the Wald estimand by $\Pr(\tilde{G}=cca)$.

(C) The proof proceeds analogously to part (B) by noticing:

$$\begin{aligned}
E[M(1-Y)|Z=1]=&\Pr(\tilde{G}=aan,acn,can,ccn)+\Pr(\tilde{G}=nan) \\
E[M(1-Y)|Z=0]=&\Pr(\tilde{G}=aan,acn)+\Pr(\tilde{G}=nan,can).\ QED.
\end{aligned}$$

A.4 (Non-)Identification of Separate Supercomplier Shares and Characteristics for a Non-Binary Y

In this section, we provide details on the simple example mentioned in Section 2.1 to illustrate the general non-identifiability of the various supercomplier shares and characteristics associated with a non-binary Y. We consider the simple case where Y can take three values in $\{0,1,2\}$ and there is perfect compliance ($D=Z$). Each unit is one of six types under our identifying assumptions: $a_0, a_1, a_2, c_{01}, c_{02}, c_{12}$, where

$$\begin{aligned}
&a_y : Y_0=Y_1=y \quad \text{for } y\in\{0,1,2\}, \\
&c_{y_0y_1} : (Y_0,Y_1)=(y_0,y_1) \quad \text{for } y_0,y_1\in\{0,1,2\} \text{ and } y_0<y_1.
\end{aligned}$$

The shares of each type are the six unknown probabilities, and they relate to the six observed probabilities of $\Pr(Y = y \mid Z = z)$ for $y \in \{0, 1, 2\}$ and $z \in \{0, 1\}$ by

$$\begin{aligned}
&\Pr(Y = 0 \mid Z = 0) = \Pr(a_0) + \Pr(c_{01}) + \Pr(c_{02}) \\
&\Pr(Y = 1 \mid Z = 0) = \Pr(a_1) + \Pr(c_{12}) \\
&\Pr(Y = 2 \mid Z = 0) = \Pr(a_2) \\
&\Pr(Y = 0 \mid Z = 1) = \Pr(a_0) \\
&\Pr(Y = 1 \mid Z = 1) = \Pr(a_1) + \Pr(c_{01}) \\
&\Pr(Y = 2 \mid Z = 1) = \Pr(a_2) + \Pr(c_{02}) + \Pr(c_{12})
\end{aligned} \tag{A6}$$

These six equations form a linear mapping

$$(a_0, a_1, a_2, c_{01}, c_{02}, c_{12}) \;\longmapsto\; \big(\Pr(Y = 0 \mid Z = 0), \ldots, \Pr(Y = 2 \mid Z = 1)\big),$$

but the corresponding 6×6 coefficient matrix is singular. Consequently, while we directly observe $\Pr(a_0)$ and $\Pr(a_2)$ in the data, we cannot separately recover $\Pr(a_1)$, $\Pr(c_{01})$, $\Pr(c_{02})$, and $\Pr(c_{12})$. Hence, shares of the different supercomplier types are not identified, and, following similar reasoning, the corresponding characteristics are also not identified.

A.5 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:

$$\begin{aligned}
\tilde{\chi}_c &\equiv \frac{\sum_i \hat{\kappa}_i h(X_i)}{\sum_{Z_i=1} D_i - \sum_{Z_i=0} D_i} \\
\tilde{\chi}_{cc} &\equiv \frac{\sum_i \hat{\pi}_i h(X_i)}{\sum_{Z_i=1} Y_i - \sum_{Z_i=0} Y_i},
\end{aligned}$$

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 7 *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*

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

Proof of Proposition 7: 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 $\hat{\tau}$:

$$\begin{aligned}\frac{1}{N}\sum_i \hat{\kappa}_i h(X_i) &= \frac{1}{N}\sum_i \left[h(X_i) - \frac{h(X_i)D_i(1-Z_i)}{1-\hat{\tau}} - \frac{h(X_i)(1-D_i)Z_i}{\hat{\tau}}\right] \\ &= \frac{1}{N}\sum_i h(X_i) - \frac{1}{N}\sum_i \frac{h(X_i)Z_i}{\hat{\tau}} + \frac{1}{N}\sum_i \frac{h(X_i)D_iZ_i}{\hat{\tau}} - \frac{1}{N}\sum_i \frac{h(X_i)D_i(1-Z_i)}{1-\hat{\tau}}.\end{aligned}$$

For the last three terms,

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

[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.

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

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 7, expression (6) is the estimand corresponding to the $\tilde{\chi}_{cc}$ estimator. Comparing (2) and (6) 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):

$$E[h(X)|D_1 > D_0] = \frac{(s_{AT}+s_C)\mu_T - s_{AT}\mu_{AT}}{s_c} \tag{A7}$$

where the various quantities are identified with

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

$$\mu_T = E[h(X)|D = 1, Z = 1].$$

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

Proof of Proposition 8: First notice that the estimand in (A7) is equivalent to

$$E[h(X)|D_1 > D_0] = \frac{\Pr(D = 1|Z = 1)\mu_T - \Pr(D = 1|Z = 0)\mu_{AT}}{\Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0)}. \tag{A8}$$

Since the denominator of (A8) 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 (A8) 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

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

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.6 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

$$(\alpha_\theta, \beta_\theta) = \underset{\alpha,\beta}{\operatorname{argmin}}\, E[\kappa\rho_\theta(Y - \alpha D - X\beta)], \tag{A9}$$

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 (A9) 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, (A9) is equivalent to

$$(\alpha_\theta, \beta_\theta) = \underset{\alpha,\beta}{\operatorname{argmin}}\, E[\kappa_\nu\rho_\theta(Y - \alpha D - X\beta)], \tag{A10}$$

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 (A10).

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

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

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

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

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:

$$E[\pi|V] = E[\kappa|V] - E[\kappa_0 Y|V] - E[\kappa_1(1-Y)|V] \tag{A11}$$

For the first term on the right hand side of (A11), 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 (A11),

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

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

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

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,

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

It follows that

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

Similarly, we can show that the third term of (A11) 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.7 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 (5) 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 4 (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 4.1 and 4.2 are the covariate-control counterparts to Assumptions 1.1 and 2.2, respectively. Assumption 4.3 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 9 *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* $X \perp\!\!\!\perp Z|W$*,*

$$\beta_{2SLS} = E\big[\omega_W E[h(X)|G = cc, W]\big], \tag{A12}$$

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

Proof of Proposition 9: 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

$$\begin{aligned} E[h(X)Y\tilde{Z}] &= E[E[h(X)Y\tilde{Z}|W]] \\ &= E\big[E[cov(h(X)Y, \tilde{Z}|W)] + E[E[h(X)Y|W]E[\tilde{Z}|W]\big] \\ &= E\big[\underbrace{E[cov(h(X)Y, Z|W)]}_{\text{(i)}} + \underbrace{E[E[h(X)Y|W]E[\tilde{Z}|W]]}_{\text{(ii)}}\big], \end{aligned} \tag{A13}$$

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 (A13) as:

$$\begin{aligned} cov(h(X)Y, Z|W) &= E[h(X)Y(Z - E[Z|W])|W] \\ &= E[h(X)Y|Z = 1, W](1 - E[Z|W])E[Z|W] \\ &\quad - E[h(X)Y|Z = 0, W](E[Z|W])(1 - E[Z|W]) \\ &= (E[h(X)Y|Z = 1, W] - E[h(X)Y|Z = 0, W])(1 - E[Z|W])E[Z|W] \\ &= E[h(X)|G = cc, W] \times RF_W \times (1 - E[Z|W])E[Z|W] \end{aligned} \tag{A14}$$

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 (A13), 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 (A14) and (A15), we have that the numerator of β_{2SLS} is

$$E[h(X)Y\tilde{Z}] = E\big[E[h(X)|G = cc, W] \times RF_W \times (1 - E[Z|W])E[Z|W]\big]. \tag{A15}$$

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 9 follows with weight $\omega_W = RF_W \times (1 - E[Z|W])E[Z|W]$. Since $RF_W \geq 0$ by Assumption 4.2 and Z is binary, $\omega_W \geq 0$ for all values of W. *QED.*

A.8 Joint Test of Identifying Assumptions for a General Y

The formulation of the joint test of identifying assumptions in Proposition 3 requires the outcome variable Y be binary. In this section, we provide a generalized test that does not impose any restriction on the distribution of Y. Our arguments follow those of Kitagawa (2015) but incorporate outcome monotonicity of Assumption 2.

Proposition 10 *Define the conditional probability distributions*

$$P(B,d) \equiv \Pr(Y \in B, D = d|Z = 1)$$
$$Q(B,d) \equiv \Pr(Y \in B, D = d|Z = 0),$$

where B is a Borel set of $\mathbb{R}$. (i) Under Assumptions 1.1-1.3 and 2.1, these inequalities hold:

$$P(B,1) - Q(B,1) \geq 0 \tag{A16}$$
$$Q(B,0) - P(B,0) \geq 0 \tag{A17}$$
$$Q(A,1) - P(A,1) \geq P(A,0) - Q(A,0), \tag{A18}$$

where B is any Borel set and A is any lower Borel interval of the form $A = (-\infty, y]$. (ii) If P and Q satisfy inequalities (A16)-(A18), then there exists a joint probability law $(Y_{11}, Y_{10}, Y_{01}, Y_{00}, D_1, D_0, Z)$ that satisfies Assumptions 1.1-1.3 and 2.1 and induces P and Q.

Proof. (i) For this part, Kitagawa (2015) already shows that Assumptions 1.1-1.3 imply (A16) and (A17). We only need to show that our identifying assumptions imply (A18), or equivalently, for any $y \in \mathbb{R}$:

$$\Pr(Y \leq y|Z = 1) \leq \Pr(Y \leq y|Z = 0). \tag{A19}$$

Inequality (A19) holds under Assumptions 1.1-1.3 and 2.1 because 1) the Y distribution is the same among always takers and the same among never takers across the treatment and control groups and 2) the Y distribution among treated compliers first-order stochastically dominates the Y distribution among untreated compliers:[2]

$$\begin{aligned}
&\Pr(Y \leq y|Z = 1) \\
=&\Pr(Y \leq y|Z = 1, D_1 = D_0) \Pr(D_1 = D_0|Z = 1) + \\
&\Pr(Y_1 \leq y|Z = 1, D_1 > D_0) \Pr(D_1 > D_0|Z = 1) \\
\leq&\Pr(Y \leq y|Z = 0, D_1 = D_0) \Pr(D_1 = D_0|Z = 0) + \\
&\Pr(Y_0 \leq y|Z = 0, D_1 > D_0) \Pr(D_1 > D_0|Z = 0) \\
=&\Pr(Y \leq y|Z = 0).
\end{aligned}$$

[2]Always takers, never takers, compliers, and defiers in this proof refer to the treatment always takers, treatment never takers, treatment compliers, and treatment defiers.

(ii) We prove this part of the proposition by slightly modifying the construction of Kitagawa (2015) in the proof of his Proposition 1.1. Kitagawa (2015) assumes instrument exclusion (Assumption 1.2) and constructs a joint distribution of (Y_1, Y_0, D_1, D_0, Z), which satisfies random assignment and treatment monotonicity (Assumptions 1.1 and 1.3), and which induces the observed P and Q satisfying the first two inequalities (A16) and (A17).

We proceed similarly, with the difference being that the joint distribution also satisfies outcome monotonicity (Assumption 2.1) and that the induced P and Q satisfy all three inequalities (A16)-(A18). Our modification of Kitagawa (2015)'s construction concerns the unobserved distributions of Y_0 among always takers and of Y_1 among never takers. Specifically, Kitagawa (2015) lets $\Pr(Y_0 \in B|\text{always taker})$ and $\Pr(Y_1 \in B|\text{never taker})$—the unobserved distribution of Y_0 among always takers and the unobserved distribution of Y_1 among never takers respectively—be arbitrary as long as $\Pr(Y_0 \in \mathbb{R}|\text{always taker}) = \Pr(D = 1|Z = 0)$ and $\Pr(Y_1 \in \mathbb{R}|\text{never taker}) = \Pr(D = 0|Z = 1)$. Because of our outcome monotonicity assumption, we additionally require Y_1 to have weak first-order stochastic dominance over Y_0 within each of the two populations. For ease of exposition, we use the simple construction where Y_1 and Y_0 have the same distribution among always takers and among never takers, respectively. That is, $\Pr(Y_0 \in B_0, D = 1|Z = 0) = Q(B_0, 1)$ and $\Pr(Y_1 \in B_1, D = 0|Z = 1) = P(B_1, 0)$ for all Borel sets B_0 and B_1.

The rest of the construction of a joint distribution of (Y_1, Y_0, D_1, D_0, Z) is the same as that in Kitagawa (2015). For any Borel sets B_1 and B_0,

$$\begin{aligned}
&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 1, D_0 = 0|Z = 1) \\
=&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 1, D_0 = 0|Z = 0) \\
\equiv&\frac{P(B_1, 1) - Q(B_1, 1)}{P(\mathbb{R}, 1) - Q(\mathbb{R}, 1)} \cdot \frac{Q(B_0, 0) - P(B_0, 0)}{Q(\mathbb{R}, 0) - P(\mathbb{R}, 0)} \cdot \\
&[P(\mathbb{R}, 1) - Q(\mathbb{R}, 1)] \cdot 1_{[P(\mathbb{R},1)-Q(\mathbb{R},1)>0]}
\end{aligned} \tag{A20}$$

$$\begin{aligned}
&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 0, D_0 = 0|Z = 1) \\
=&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 0, D_0 = 0|Z = 0) \\
\equiv&\frac{P(B_1, 0)}{P(\mathbb{R}, 0)} \cdot \frac{P(B_0, 0)}{P(\mathbb{R}, 0)} \cdot P(\mathbb{R}, 0) \cdot 1_{[P(\mathbb{R},0)>0]}
\end{aligned}$$

$$
\begin{aligned}
&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 1, D_0 = 1 | Z = 1) \\
=&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 1, D_0 = 1 | Z = 0) \\
\equiv&\frac{Q(B_1, 1)}{Q(\mathbb{R}, 1)} \cdot \frac{Q(B_0, 1)}{Q(\mathbb{R}, 1)} \cdot Q(\mathbb{R}, 1) \cdot 1_{[Q(\mathbb{R},1)>0]}
\end{aligned}
$$

$$
\begin{aligned}
&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 0, D_0 = 1 | Z = 1) \\
=&\Pr(Y_1 \in B_1, Y_0 \in B_0, D_1 = 0, D_0 = 1 | Z = 0) \\
\equiv&0,
\end{aligned}
$$

where a product is defined to be zero if the indicator function inside equals zero. Following the arguments in Kitagawa (2015), this constructed joint distribution is a probability measure, satisfies Assumptions 1.1 and 1.3, and induces the observed distributions of P and Q.

We just need to show that the joint distribution also satisfies outcome monotonicity. Since outcome monotonicity holds for the always taker, never taker, and defier populations by construction, we only need to examine the compliers. For any lower Borel interval A,

$$
\begin{aligned}
&\Pr(Y_1 \in A, D_1 = 1, D_0 = 0) \\
=&P(A, 1) - Q(A, 1) \\
\leq&Q(A, 0) - P(A, 0) \\
=&\Pr(Y_0 \in A, D_1 = 1, D_0 = 0).
\end{aligned}
$$

We obtain the first (last) equality by replacing B_1 with A and B_0 with $\mathbb{R}$ (B_1 with $\mathbb{R}$ and B_1 with A) in (A20), and by using the identities of $P(\mathbb{R}, 0) = 1 - P(\mathbb{R}, 1)$ and $Q(\mathbb{R}, 0) = 1 - Q(\mathbb{R}, 1)$ based on the definitions of P and Q. The inequality in the middle is simply (A18), which is assumed to hold. *QED.* □

In the specific case of a binary Y, we already have a simple test as discussed in Section 2.3. To implement the joint test based on Proposition 10 for a non-binary Y, we note that the first two inequalities (A16) and (A17) are exactly the same as those that underlie the tests by Kitagawa (2015) and Mourifié and Wan (2017). Kitagawa (2015) proposes a variance-weighted Kolmogorov-Smirnov test statistic for these inequalities and provides an implementation in R. Meanwhile, Mourifié and Wan (2017) transform (A16) and (A17) into moment inequalities where the moments are conditional on the value of Y, allowing their

test to be implemented using a Stata command by Chernozhukov et al. (2015) based on Chernozhukov, Lee and Rosen (2013). Since most of our intended audience uses Stata, we build on Mourifié and Wan (2017) for convenience, though R users could easily implement a similar joint test based on the Kitagawa (2015) procedure.

Because inequality (A18) only holds for lower Borel intervals and not just any Borel set, it cannot be turned into the same type of moment condition and incorporated into the Chernozhukov et al. (2015) command. However, it is straightforward to test (A18) with a standard one-sided two-sample Kolmogorov-Smirnov test via the `ksmirnov` command in Stata. Consequently, our procedure entails testing (A16) and (A17) separately from (A18), and we will conclude that one or more of our identifying assumptions are violated if either test rejects the null at the 5 percent level.

It is well known that conducting two tests separately like this generally fails to control the overall size, and a common remedy is to apply a multiple hypothesis correction like Bonferroni. However, a Bonferroni procedure is conservative, which can be a virtue in treatment effect testing but a vice in assumption or specification testing—a conservative test tends to lack power, which means that we would often fail to reject incorrect assumptions. Therefore, we choose not to correct for multiple hypotheses and note that the actual (asymptotic) size of our procedure is between 5 and 10 percent. In our empirical applications, we use the test from Section 2.3 (which controls size at 5 percent) for binary outcomes and the test described here for continuous outcomes.

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 use data by Bloom and MDRC (2022) to 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 (9). 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 (10), 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 3.

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.

C Weighted MVPF and Supercompliers

In this section, we show that weighted willingness to pay (the denominator of MVPF) equals unweighted WTP (the local average treatment effect) multiplied by our identified supercomplier characteristics for any distribution of the outcome variable. Our result here nests equation (12) as a special case for binary Y.

The weighted willingness to pay for the complier group is

$$
\begin{aligned}
&E[\eta_i(Y_{1i} - Y_{0i})|\text{complier}] \\
=&E[\eta_i(Y_{1i} - Y_{0i})|\text{complier}, Y_{1i} > Y_{0i}] \Pr(Y_{1i} - Y_{0i} > 0|\text{complier}) \\
=&E[\eta_i(Y_{1i} - Y_{0i})|\text{supercomplier}] \Pr(Y_{1i} - Y_{0i} > 0|\text{complier}) \\
=&\frac{E[\eta_i Y_i|Z_i = 1] - E[\eta_i Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} \cdot E[Y_{1i} - Y_{0i}|\text{supercomplier}] \Pr(Y_{1i} - Y_{0i} > 0|\text{complier}) \\
=&\frac{E[\eta_i Y_i|Z_i = 1] - E[\eta_i Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} \cdot E[Y_{1i} - Y_{0i}|\text{complier}, Y_{1i} > Y_{0i}] \Pr(Y_{1i} - Y_{0i} > 0|\text{complier}) \\
=&\frac{E[\eta_i Y_i|Z_i = 1] - E[\eta_i Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} \cdot E[Y_{1i} - Y_{0i}|\text{complier}] \\
=&\frac{E[\eta_i Y_i|Z_i = 1] - E[\eta_i Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} \text{LATE}_Y.
\end{aligned}
$$

It follows by Proposition 2 that the first term represents the supercomplier average of the weights under our identifying assumptions.