Discussion Paper Series

IZA DP No. 18598

April 2026

Supercompliers

Matthew L. Comey
Joint Committee on Taxation

Amanda R. Eng
Internal Revenue Service

Pauline Leung
Cornell University

Zhuan Pei
Cornell University and IZA@LISER

IZA@LISER NETWORK
ISSN: 2365-9793

IZA DP No. 18598 April 2026

Supercompliers*

Abstract

In an 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 evaluation. We propose conditions for characterizing supercompliers, and show how 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 analyses.

JEL classification

C10, H00, J24

Keywords

compliers, supercompliers, marginal value of public funds

Corresponding author

Pauline Leung

pleung@cornell.edu

* We are grateful for the insightful suggestions from the co-editor, Manasi Deshpande, and three anonymous reviewers. We also thank Josh Angrist, Eric Auerbach, David Card, Amy Finkelstein, Toru Kitagawa, David S. Lee, Lars Lefgren, Linden McBride, Doug Miller, Jon Roth, Steve Pischke, Andres Santos, Jeff Smith, and seminar and conference participants at Binghamton University, Cornell University, the Princeton University Industrial Relations Section Centennial Symposium, the Stata Virtual Symposium, WEAI, and the SOLE World Labor Conference for helpful comments. Fiona Qiu and Kathryn McGinnis 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.

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 is induced to take up treatment when eligible? In this paper, we extend the complier literature and devise tools to provide a direct answer to a different but related question: who benefits 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."

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.

Our analysis of supercomplier characteristics is complementary to subsample heterogeneity analysis of a treatment effect (see Smith 2022 for a recent review). While conventional heterogeneity analysis—whether it is based on simple subsampling or machine learning methods (e.g., Athey and Imbens 2016; Wager and Athey 2018; Knaus, Lechner, and Strittmatter 2021)—asks which subgroups benefit more, our supercomplier analysis collects all beneficiaries and profiles them. Characterizing supercompliers mirrors the increasingly common practice of describing compliers, which researchers have accepted as a natural complement

[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).

to subsample heterogeneity analysis of treatment take-up. Complier analysis is typically used to assess how representative the complier subpopulation is of the broader experimental population, informing whether the local average treatment effect generalizes beyond the experiment. Supercomplier analysis serves an analogous purpose on the outcome side: by profiling those who actually benefit from the intervention, it speaks to whether the benefits are concentrated in a narrow subgroup or distributed more broadly—information that is directly relevant for distributional and welfare assessments of the kind we develop in this paper.

For ease of exposition, we first focus on the case of a binary instrument, a binary treatment, and a binary outcome. In this setting, we maintain the standard LATE assumptions along with an additional assumption we call outcome monotonicity: treatment either does not affect the outcome or changes it in a single direction. Outcome monotonicity is the outcome counterpart of the treatment monotonicity assumption in the LATE framework and has been imposed in other studies (e.g., Manski 1997, Manski and Pepper 2000, Lee 2009, and Celislami, Kastoryano, and Mastrobuoni 2023). These assumptions partition the population into groups defined by their potential responses to eligibility and treatment, giving rise to the supercomplier concept. We show that the supercomplier average of baseline characteristics X is identified by a Wald estimand: the intent-to-treat effect on the product of X and outcome Y, divided by the intent-to-treat effect on Y. It can be estimated with a standard two-stage least squares regression.

Our results generalize beyond the restrictions in the focal case. When the instrumental variable is continuous, an analogous estimand identifies "marginal supercomplier characteristics," connecting our framework to the marginal treatment effects literature that started with Heckman and Vytlacil (1999). When the treatment variable is non-binary, we can still interpret the identified characteristics as the average among those whose outcome responds to a change in eligibility (because the treatment variable does not appear in our estimand). When the outcome variable is non-binary, we can either "binarize" it by setting thresholds (Balke and Pearl 1997), or apply our estimand regardless of the outcome distribution and interpret the identified parameter as a weighted average of supercomplier characteristics, with weights proportional to individual treatment effects. Finally, without outcome monotonicity, the treatment effects and therefore the weights can be negative, which could be appropriate in certain contexts.

One such context is the analysis of the marginal value of public funds (MVPF; Hendren and Sprung-Keyser 2020), which has become commonplace in empirical 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 redistributive preferences, it is not feasible when only the overall LATE is reported. We show, however, that when welfare weights depend linearly on observed characteristics, a weighted MVPF can be computed by combining supercomplier characteristics with the reported LATE. A key advantage of this approach is that access to the study's microdata may not be necessary for recomputing an MVPF with different social welfare functions. Further, this reweighting exercise requires only the LATE assumptions, highlighting the utility of our estimand in the absence of outcome monotonicity.

We illustrate our proposed method using data from two well-known randomized experiments on job training, the National Job Corps Study (Schochet, Burghardt, and McConnell 2008) and the National Job Training Partnership Act (JTPA) Study (Bloom et al. 1997). For labor market outcomes, we find that the Job Corps beneficiaries are more advantaged relative to the study population. In contrast, among the group that experienced the largest impact from JTPA—adult women—beneficiaries are relatively less 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.

While our exposition focuses on randomized experiments, our framework extends to other research designs. In regression discontinuity designs, we can identify supercomplier characteristics at the policy threshold under smoothness conditions. In difference-in-differences designs, researchers (e.g., Ouss and Stevenson 2023; Tadjfar and Vira 2025) have conducted complier analysis by invoking LATE-type assumptions, beyond the standard parallel trends. Our framework extends to these settings as well, though additional assumptions (e.g., outcome monotonicity) may be needed depending on the parameter of interest.

Related theoretical results were independently developed in exemplary research by Yu (2025).[2] Yu (2025) and our baseline result rely on the same set of assumptions, propose the same estimand, and derive essentially the same key identification for a binary Y (Our Propositions 1 and 4 and Yu 2025's Theorem 3.2). Our paper goes further in several directions. First, we

[2] We became aware of Yu (2025) in November 2024, which was first circulated in November 2022 as a job market paper. We first posted ours on arXiv in December 2022.

show that the estimand remains interpretable when the treatment or outcome is non-binary, and we relate our framework to the marginal treatment effects literature when the instrument is continuous. Second, we connect our estimand to the increasingly influential MVPF literature and illustrate how they can be used to compute MVPF under different social welfare weights. As mentioned above, this connection requires only the standard LATE assumptions, and not outcome monotonicity, substantially broadening the scope of applicability. Third, we address statistical issues of direct interest to applied researchers, including the interpretation of estimands under stratified randomization, the precision of supercomplier characteristics estimates, and the relationships among complier characteristics estimators used in the literature. Finally, while Yu (2025) is oriented toward the political persuasion literature, our paper is aimed at a general audience in economics.

2 Identification and Estimation

2.1 Theoretical Framework and Identification

We begin with an extension of the 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 $D = d$. (These notations implicitly incorporate the stable unit treatment value assumption, or SUTVA.)

Throughout this paper, we maintain the standard assumptions for the identification of LATE by Angrist, Imbens, and Rubin (1996).

Assumption 1 (IV)

1. *Random Assignment:* $(Y_{00}, Y_{01}, Y_{10}, Y_{11}, D_0, D_1) \perp\!\!\!\perp Z$ *and* $0 < \Pr(Z = 1) < 1$.
2. *Exclusion:* $\Pr(Y_{1d} = Y_{0d}) = 1$ *for* $d \in \{0, 1\}$.
3. *Treatment Monotonicity:* $\Pr(D_1 \geq D_0) = 1$.
4. *First Stage:* $\Pr(D_1 = 1) > \Pr(D_0 = 1)$.

With Assumptions 1.1 and 1.2 along with binary D and Y, we can partition the population into 16 unobserved groups (e.g., Balke and Pearl 1993), which we refer to as "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. In Table 1, 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. Supercompliers correspond to the group $G = cc$, where $D_1 > D_0$ and $Y_1 > Y_0$.

Assumptions 1.3 and 1.4 place restrictions on the shares of these groups. 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 group 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.[3] While Assumption 1.4 implies a nonzero share of treatment compliers and consequently a nonzero first stage, Assumption 2.2 requires a nonzero 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.)

While Assumption 1 is standard in the randomized controlled trial (RCT) literature, Assumption 2.1 (outcome monotonicity) warrants more discussion. Previous studies have maintained similar assumptions. Outcome monotonicity corresponds to "monotone treatment

[3] Balke and Pearl (1997) do not eliminate groups with monotonicity assumptions but partially identify the average treatment effect instead. Celislami, Kastoryano, and Mastrobuoni (2023) consider groups jointly defined by two outcomes, both subject to outcome monotonicity.

response" in Manski (1997) and Manski and Pepper (2000) or "monotonicity" in Lee (2009). In the motivating examples of these seminal studies, outcome monotonicity is taken to mean that the demand curve is weakly downward sloping, that education does not decrease wages, or that participating in the Job Corps training program does not lower employment. Outcome monotonicity is also implicit in the classic 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 job training and labor market outcomes 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. In this case, researchers may not want to assume outcome monotonicity.

Even in cases where the identifying assumptions are plausible, it is prudent to test them and probe the consequences of their violation. In previous versions of this paper (Comey et al. 2024), we proposed a joint test of the identifying assumptions by building on Kitagawa (2015). For a binary Y, our test is based on the same inequalities as Yu (2025)'s; we extend the test to a general Y. In Appendix A.8, we investigate the consequence of violating outcome monotonicity and show that the bias is small when the degree of violation is low. Finally, as we discuss below, our estimand remains useful even in the absence of outcome monotonicity, and we illustrate a situation—incorporating social welfare weights into marginal value of public funds analyses—in which outcome monotonicity is not required.

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. That is, they are the only ones who benefit from being assigned to the treatment group. Thus, learning about supercompliers should be of great importance to policymakers and evaluators.

This result is related to a broader literature that partially identifies the distribution of treatment effects (e.g., Fan and Park 2010) or the share of units affected by treatment (e.g., Borusyak 2015; Huang et al. 2017; Huang et al. 2019). The share of units with a strictly positive treatment effect is generally only partially identified, but Borusyak (2015) notes that point identification is achieved when the outcome Y is binary and under outcome monotonicity (termed "non-negative treatment effect" by Borusyak 2015), as in our setting. While this prior literature is primarily interested in counting the size of this beneficiary subpopulation, we place more emphasis on describing its characteristics.

As with compliers, we cannot pinpoint members of the supercomplier subpopulation, but we can identify the distribution of their characteristics. Let X be a variable determined prior to random assignment 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 covariate vectors of any dimension. Let h be a function satisfying $E[|h(X)|] < \infty$.

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

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

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

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

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

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

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

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

The weights κ, κ_0, and κ_1 are the unconditional counterparts of those defined by Abadie (2003) in his study of compliers. While Abadie (2003) assumes the LATE assumptions hold conditional on X, our Lemma 2 in Appendix A.1 shows that these weights identify compliers' X distribution and that κ_0 (κ_1) identifies their Y_0 (Y_1) distribution under our (unconditional) assumptions. That is, Lemma 2 provides the unconditional analog of Theorem 3.1 of Abadie (2003).

Remark 1. (**Compliers and Supercompliers**) There is a 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, κ starts from the population and subtracts the treatment always-takers and never-takers, while π starts from the treatment compliers and subtracts the $G=ca$ and $G=cn$ groups. There is also a complier analog to equation (2): complier characteristics are identified by replacing Y with D in (2) (Kline and Walters 2016 and Marbach and Hangartner 2020).[4] This correspondence reflects a fundamental parallel between compliers and supercompliers. To see this, define alternative potential outcomes $\tilde{Y}_z \equiv Y_{D_z}$. Assumptions 1 and 2 jointly imply the standard LATE assumptions for the relationship between Z and $\tilde{Y}$: random assignment, monotonicity ($\tilde{Y}_1 \geq \tilde{Y}_0$), and a nonzero "first stage" ($\Pr(\tilde{Y}_1=1) > \Pr(\tilde{Y}_0=1)$). Under this relabeling, supercompliers are the "compliers" in the reduced form, and equation (2) holds under standard complier reasoning. Intuitively, supercompliers are the only group whose outcome changes with Z, so the reduced-form effect of Z on $h(X)Y$ isolates them (just as the effect of Z on $h(X)D$ isolates the treatment compliers), and scaling by the share of supercompliers identifies their average characteristics. This interpretation is agnostic to the distribution of D and applies whether the treatment variable is binary, multi-valued, or continuous.

Remark 2. (**Supercomplier Characteristics and Subsample LATEs**) The supercom-

[4]One can rearrange the terms in π and show that it can be similarly obtained from the Abadie κ by replacing D with Y: $\pi = 1 - \frac{Y(1-Z)}{\Pr(Z=0)} - \frac{(1-Y)Z}{\Pr(Z=1)}$.

plier distribution of characteristics X and the LATEs within subsamples defined by the value of X (we focus on a discrete X for ease of exposition) are related by the identity:

$$\Pr(X = x \mid G = cc) = \frac{E[Y_1 - Y_0 \mid X = x,\, D_1 > D_0]}{E[Y_1 - Y_0 \mid D_1 > D_0]} \cdot \Pr(X = x \mid D_1 > D_0).$$

This equality, which follows from Bayes' Rule, shows that the supercomplier X-distribution equals the complier X-distribution reweighted by the ratio of the subsample LATE to the overall LATE. In particular, a subgroup can have a large supercomplier share either because it is prevalent among treatment compliers or because it has a large conditional treatment effect—supercomplier characteristics reflect both margins. This feature is useful for welfare analysis: a social planner assessing the total benefit of an intervention to a preferred subgroup needs to account for both the subgroup's prevalence among those induced into treatment and its treatment effect. Supercomplier characteristics combine these two factors into a single object, whereas subsample LATEs, by design, abstract from the subgroup's prevalence among compliers.

Remark 3. (**Identification of Characteristics Distributions**) When h is the identity function, Proposition 1 identifies average supercomplier characteristics. For a k-dimensional X, setting $h = 1_{[X \leq x]}$ for $x \in \mathbb{R}^k$ identifies the joint distribution of X among supercompliers.

Remark 4. (**Share and Characteristics of Other Groups**) We can also identify the shares and characteristics of the other two groups within the treatment compliers: $G = ca$ and $G = cn$. See Appendix A.2.

Remark 5. (**Connections to Marginal Treatment Effects**) Our identification results connect to the literature on marginal treatment effects (MTE) that started with Heckman and Vytlacil (1999). Specifically, we can interpret the estimand on the right hand side of (2) as identifying an MTE-weighted average of supercomplier characteristics. Going beyond our focal framework, with a continuous instrument that traces out the propensity score, we can formulate an analogous estimand that identifies the "marginal supercomplier characteristics" at each value of the propensity score. We provide details in Appendix A.3.

Remark 6. (**Binarizing a Non-Binary** Y) As mentioned in the introduction, we can recode a non-binary Y into a binary variable ("high" or "low") by setting thresholds, allowing us to identify supercomplier characteristics for each "binarized" Y. Whereas using a binarized *treatment* variable can violate the exclusion restriction and invalidate the Wald estimand (Angrist and Imbens 1995; Andresen and Huber 2021; Rose and Shem-Tov forthcoming), binarizing Y does not violate our IV assumptions or outcome monotonicity.

Although 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 values (0, 1, and 2), in which case there are three supercomplier groups whose Y changes from 0 to 1, from 0 to 2, and from 1 to 2 because of treatment. See Appendix A.4.[5]

When Y is continuous, it is impossible and impractical to characterize infinitely many supercomplier types. Instead, we extend the identification results in Lemma 1 and Proposition 1 to accommodate non-binary Y. The proposition below shows that regardless of the Y-distribution, 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 weights are proportional to treatment effects.

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 explore below in Sections 3 and 4, welfare analysis may require weighted treatment effects as inputs, where weights correspond to social planner preferences. A consequence of equation (4) is that the weighted treatment effect for compliers, where the weights are a function of observable characteristics, is equal to the unweighted treatment effect, "scaled"

[5] Relatedly, while $Y_0 = 0$ and $Y_1 = 1$ among the supercompliers when Y is binary, the marginal distributions of Y_0 and Y_1 among the supercompliers are not identified in general for a non-binary Y. See the last paragraph of Appendix A.4.

by our estimand for supercomplier characteristics (see Appendix A.1 for derivation):

$$E[h(X_i)(Y_{1i} - Y_{0i})|\text{complier}] = \frac{E[h(X)Y|Z=1] - E[h(X)Y|Z=0]}{E[Y|Z=1] - E[Y|Z=0]}\text{LATE}, \qquad (5)$$

where LATE $= E[Y_1 - Y_0|\text{complier}]$ is the (unweighted) average treatment effect for treatment compliers. It is important to note that while equation (5) can be derived from equation (4), it holds under weaker assumptions:

Proposition 3 *Under Assumption 1 and provided that $X \perp\!\!\!\perp Z$ and that the reduced form $E[Y|Z=1] - E[Y|Z=0]$ is nonzero, equation (5) holds.*

That is, equation (5) does not require outcome monotonicity but only a nonzero reduced form. The weaker assumptions that are required substantially broaden the scope of applicability of the MVPF analysis discussed in Section 3.

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 supercomplier characteristics—we can obtain point estimates and conduct inference in Stata with

```
ivregress 2sls XY (Y = Z) [, options]        (6)
```

where `XY` is defined as the product of X and Y. In the remainder of this paper, we rely on (6) for estimation and inference because the estimator naturally accommodates arbitrary distributions of Y—it is akin to the complier IV estimator used by Alsan et al. (2025) for a non-binary treatment. In Appendix A.5, we survey alternative complier characteristics estimators and discuss their supercomplier analogs.

Remark 7. (**2SLS First Stage and Precision**) The 2SLS regression (6) instruments Y with Z. Consequently, the relevant "first stage" for profiling supercompliers is the reduced-form effect of Z on Y. For a binary Y, Lemma 1 shows that this reduced form estimates the share of supercompliers, which cannot exceed the share of treatment compliers. Since the latter is the relevant first stage for profiling treatment compliers (Appendix A.5), supercomplier characteristics estimates tend to be less precise than complier characteristics estimates. Appendix A.6 provides details for this result and also relates the variance of the (super)complier characteristics estimator to the F-statistic from the relevant first stage for a general Y. The bottom line is that it is advisable to limit characterization of supercompliers to settings with a strong reduced-form effect (i.e., compelling evidence that supercompliers exist).

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. (forthcoming) to show that the resulting population regression coefficient still identifies a non-negatively weighted average of supercomplier characteristics across strata.

3 Relationship to Welfare Analysis

In this section, we discuss how supercomplier characteristics can enrich welfare analyses. In particular, we consider the case where a social planner places more weight on individuals with certain characteristics who may differentially benefit from a particular intervention or policy. We first describe how social weights are (and are not) used in the widely adopted framework of Hendren and Sprung-Keyser (2020), and then discuss how the result in Proposition 3 allows us 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.[6] A small policy change, denoted by dp, impacts social welfare by:

$$\frac{dW}{dp} = \sum_i \eta_i \frac{dU_i}{dp} \equiv \sum_i \eta_i WTP_i, \tag{7}$$

where WTP_i is individual i's "willingness to pay" (WTP) for the policy. Equation (7) encapsulates the benefits of the policy change. The cost of the policy, G, equals its impact on the government's budget. Hendren and Sprung-Keyser (2020) formally define the MVPF as

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

or, the ratio of the (unweighted) aggregate WTP 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 implement in practice. Hendren and Sprung-Keyser (2020) suggest, therefore, to use MVPFs to compare policies that benefit similar beneficiaries to "reduce the role of social preferences in driving conclusions." However, we show that we can use supercomplier characteristics to incorporate weights into the MVPF to reflect a social planner's redistributive 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 the WTP (numerator). We focus our discussion on the numerator because that is the relevant quantity for incorporating social weights. In a typical job training application, for example, the WTP is equal to its causal effect on beneficiaries' after-tax earnings net of transfers from government safety net programs. If the policy also affects other outcomes (e.g., improved health), those effects are also translated into a dollar amount and enter the WTP.[7] 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 an RCT with imperfect

[6] Any utility function can be normalized to a money-metric by dividing by the marginal utility of income. Since U_i is already a money-metric, the weights η_i are the "generalized weights" from Saez and Stantcheva (2016).

[7] The denominator in the job training example will contain the upfront cost of the intervention to the government, the causal effects on tax revenue, and any effects on government transfers. In other settings, the denominator may also incorporate spillover effects on government spending, such as changes in government-subsidized healthcare utilization, changes in public school enrollment, or changes in criminal justice system involvement. Since it is typically assumed that these costs are borne by an "average" taxpayer, there is no need to apply social weights to the costs.

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' WTP, even within the complier group. Suppose that the WTP is measured by the effect of the policy on an outcome Y (e.g., earnings), so $WTP_i = Y_{1i} - Y_{0i}$. Proposition 3 shows that if we have the (unweighted) local average treatment effect, we can obtain the weighted WTP for the complier group by scaling with the proposed estimand for supercomplier characteristics, provided that the social weights η_i are a linear function of observable characteristics X_i. 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 WTP directly by estimating the treatment effect on weighted outcomes, the formulation in (5) shows that it can be done without microdata. If supercomplier characteristics are documented along with the LATE, readers 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. While it is true that our proposed method is slightly more restrictive than if one were to calculate the weighted MVPF directly with microdata, a major advantage of our approach is that it allows for readers to use different social weights than those chosen by the researchers who estimate the relevant treatment effect. Since social weights are quite subjective, researchers typically report MVPFs using unweighted causal effects.

4 Empirical Applications

To demonstrate the utility of our tools for characterizing supercompliers, we consider experimental evaluations of the Job Corps and Job Training Partnership Act programs (JTPA). We show estimated supercomplier characteristics when the outcome of interest is binary (educational attainment in Job Corps) and when it is continuous (measures of earnings in Job Corps and JTPA). Finally, we use the framework from Section 3 to compare weighted MVPFs with respect to the earnings impacts in both programs.

4.1 Job Corps

Job Corps is a federally sponsored job training program for disadvantaged youths in the United States that offers academic education, vocational training, and job search assistance. In the 1990s eligible applicants ages 16 to 24 were randomly assigned access to Job Corps in the National Job Corps Study (Schochet, Burghardt, and Glazerman 2001).

Job Corps had large educational impacts during the 48-month follow-up period: 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) and 15.0 percentage points more likely to obtain a GED (41.6 vs. 26.6 percent; t-stat: 14.3; N: 8,579; among those without high school credentials). Labor market impacts were more moderate. 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 that supercompliers were statistically 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. But even for this earnings outcome, its reduced-form t-stat is smaller than those of the education outcomes underlying

Table 2, and therefore—consistent with Remark 7— the standard errors of the supercomplier characteristics are larger here.[8] Nevertheless, the patterns in Table 3 share many similarities with Table 2: individuals whose earnings increased as a result of treatment 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 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, the tests proposed in Comey et al. (2024) cannot reject our identifying assumptions at the 5 percent level, indicating their plausibility (as stated in Comey et al. 2024, our tests only report whether the null hypothesis is rejected at a given level, which is consistent with Mourifié and Wan 2017).

Overall, our findings 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 evaluated job training programs targeting disadvantaged adults and out-of-school youths across 16 US sites in the late 1980s. Eligible applicants were randomly assigned to control or treatment groups that offered classroom training, subsidized on-the-job training, or other services. Following Bloom et al. (1997), we focus on 30-month earnings impacts. Hendren and Sprung-Keyser (2020) also use this outcome to calculate the MVPF.

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:

[8]Similarly, because the first-stage effect of treatment on Job Corps take-up is much stronger (F-statistic of 15027.53) than the reduced-form effects underlying Tables 2 and 3, the complier characteristics reported in both tables are more precisely estimated.

[9]Treatment take-up was around 60 percent for all four groups.

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 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 (Remark 7) 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, we also cannot reject our identifying assumptions in the JTPA data for the earnings outcome, using the test proposed in Comey et al. (2024).

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

[10]The "first-stage" F-statistics for Table 2 are 205.72 for receiving a GED and 731.56 for receiving a vocational certificate. The F-statistics for earnings outcomes are smaller but still above 10: 22.63 for quarter-16 earnings in Table 3 and 13.15 for 30-month earnings in Table 4. Lee et al. (2022) suggest that a much higher F-stat threshold of 104.7 is needed to ensure valid inference with a standard t-test (though Angrist and Kolesár 2024 argue that conventional inference in a just-identified IV setting is typically reliable). As a robustness check, we compute Anderson-Rubin (AR) confidence intervals, which are valid regardless of instrument strength. While the AR confidence intervals are 9 to 13 percent (17 to 25 percent) wider for the characteristics reported in Table 3 (Table 4) than conventional counterparts, our qualitative conclusions remain the same: The statistical significance (at the 5 percent level) of the differences for white and older youths in Table 3 is unchanged by applying a conservative test based on the Bonferroni inequality.

[11]Note that estimated characteristics shares of supercompliers (or compliers) are not guaranteed to lie between zero and one, though none of our point estimates in Tables 2-4 are outside this interval.

[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. The reduced form earnings effect for men and the AFDC effects for both men and women are not strong enough to compute supercomplier weights. AFDC effects contribute minimally to the MVPF: without accounting for it, the unweighted MVPF is 1.35 (vs. 1.38).

$\Psi(u) = \frac{u^{1-\phi}}{1-\phi}$ (Salanié, 2011), where ϕ denotes the extent to which the social planner desires redistribution and u is income.[13] 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 welfare 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.[14]

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 at 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 (5), 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

[13] We apply this social welfare function directly to income rather than utility to avoid needing to make assumptions about the utility function. If we instead define $u(I)$ as utility, where I is income, the weight in step 1 below would be $u^{-\phi}u'(I)$ instead of $u^{-\phi}$.

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

do not have the data to estimate a weighted MVPF for Job Corps that directly corresponds to Hendren and Sprung-Keyser (2020), our results from survey-based public-use data indicate that incorporating social welfare weights would decrease the MVPF for Job Corps youths by up to 43 percent depending on the value of ϕ.[15] 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 supercompliers—those who comply with treatment assignment and benefit from treatment. Under standard LATE assumptions plus outcome monotonicity, we can consistently estimate supercomplier characteristics using simple IV regressions. Under the LATE assumptions alone, we can use our estimand to incorporate social weights into MVPF calculations.

We illustrate the utility of our methods using 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.

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. 2025. "'Something Works' in U.S. Jails: Misconduct and Recidivism Effects of the IGNITE Programs." *Quarterly Journal of Economics* 140(2): 1367–1415.

[15] Hendren and Sprung-Keyser (2020) estimate an MVPF of 0.18 for Job Corps based on 20-year impacts from Schochet (2018), which uses restricted tax data. The impacts on earnings from the restricted data are more muted even in the short run, which Schochet, Burghardt, and McConnell (2008) attribute to the exclusion of informal earnings in tax data.

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.

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

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., 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. 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. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press.

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." Technical Report R-199, Computer Science Department, University of California Los Angeles.

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.

Blandhol, Christine, John Bonney, Magne Mogstad, and Alexander Torgovitsky. forthcoming. "When is TSLS Actually LATE?" *Review of Economic Studies*.

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.

Borusyak, Kiril. 2015. "Bounding the Population Shares Affected by Treatments." SSRN Working Paper 2473827.

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.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil. 2011. "Estimating Marginal Returns to Education." *American Economic Review* 101(6): 2754–2781.

Celislami, Elda, Stephen Kastoryano, and Giovanni Mastrobuoni. 2023. "Strategic Bureaucratic Opacity: Evidence from Death Investigation Laws and Police Killings." IZA Discussion Paper 16609.

Comey, Matthew L., Amanda R. Eng, Pauline Leung, and Zhuan Pei. 2024. "Supercompliers." arXiv:2212.14105v3.

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

Cornelissen, Thomas, Christian Dustmann, Anna Raute, and Uta Schönberg. 2016. "From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions." *Labour Economics* 41: 47–60.

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.

Fan, Yanqin and Sang Soo Park. 2010. "Sharp Bounds on the Distribution of Treatment Effects and Their Statistical Inference." *Econometric Theory* 26(3): 931–951.

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.

Heckman, James J, Sergio Urzua, and Edward Vytlacil. 2006. "Understanding Instrumental Variables in Models with Essential Heterogeneity." *Review of Economics and Statistics* 88(3): 389–432.

Heckman, James J. and Edward J. Vytlacil. 1999. "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects." *Proceedings of the National Academy of Sciences* 96(8): 4730–4734.

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

Huang, Emily J., Ethan X. Fang, Daniel F. Hanley, and Michael Rosenblum. 2017. "Inequality in Treatment Benefits: Can We Determine if a New Treatment Benefits the Many of the Few?" *Biostatistics* 18(2): 308–324.

Huang, Emily J., Ethan X. Fang, Daniel F. Hanley, and Michael Rosenblum. 2019. "Constructing a Confidence Interval for the Fraction Who Benefit from Treatment, Using Randomized Trial Data." *Biometrics* 75: 1228–1239.

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

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.

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.

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.

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

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

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

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

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

Rose, Evan K. and Yotam Shem-Tov. forthcoming. "On Recoding Ordered Treatments as Binary Indicators." *Review of Economics and Statistics*.

Saez, Emmanuel and Stefanie Stantcheva. 2016. "Generalized social marginal welfare weights for optimal tax theory." *American Economic Review* 106(1): 24–45.

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

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

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.

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

Shea, John. 1997. "Instrument Relevance in Multivariate Linear Models: A Simple Measure." *Review of Economics and Statistics* 79(2): 348–352.

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

Tadjfar, Nagisa and Kartik Vira. 2025. "Prediction or Prejudice? Standardized Testing and University Access." Unpublished Manuscript.

Vytlacil, Edward. 2002. "Independence, Monotonicity, and Latent Index Models: An Equivalence Result." *Econometrica* 70(1): 331–341.

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. 2025. "A Binary IV Model for Persuasion: Profiling Persuasion Types among Compliers." *Econometrics Journal* 28(3): 462–481.

Declaration of Gen-AI Technologies in the Writing Process

In preparation for the original submission, we used Claude 4 to streamline the paper and improve clarity. During the revision process, we iterated with AI (ChatGPT 5.2/5.4; Claude 4.5/4.6) on response strategies and used it for editing. All substantive content is our own, and we have reviewed and edited all AI-assisted output. We take full responsibility for this article.

Table 1: Extended Principal Stratification

G-value	Treatment Type	Outcome Type	D_0	D_1	Y_0	Y_1
$\boldsymbol{aa}$	**Always Taker**	**Always Taker**	**1**	**1**	**1**	**1**
$\boldsymbol{an}$	**Always Taker**	**Never Taker**	**1**	**1**	**0**	**0**
$\boldsymbol{ac}$	**Always Taker**	**Complier**	**1**	**1**	**0**	**1**
af	Always Taker	Defier	1	1	1	0
$\boldsymbol{na}$	**Never Taker**	**Always Taker**	**0**	**0**	**1**	**1**
$\boldsymbol{nn}$	**Never Taker**	**Never Taker**	**0**	**0**	**0**	**0**
$\boldsymbol{nc}$	**Never Taker**	**Complier**	**0**	**0**	**0**	**1**
nf	Never Taker	Defier	0	0	1	0
$\boldsymbol{ca}$	**Complier**	**Always Taker**	**0**	**1**	**1**	**1**
$\boldsymbol{cn}$	**Complier**	**Never Taker**	**0**	**1**	**0**	**0**
$\boldsymbol{cc}$	**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	SC-Pop	SC-C
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	SC-Pop	SC-C
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 SC-Pop and SC-C differences. As with Schochet, Burghardt, and Glazerman (2001) and Schochet, Burghardt, and McConnell (2008), we use weights included in the public use data to adjust for sample and survey designs. Standard errors are in parentheses. The F-statistics for the reduced form effects on receiving a GED and receiving a vocational certificate (i.e., the "first-stage" F-stats for the supercompliers estimates) are 205.72 and 731.56, respectively.

Table 3: Job Corps Supercomplier Characteristics for Earnings

	Means			Differences	
	Population	Complier	Supercomplier	SC-Pop	SC-C
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 SC-Pop and SC-C 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), we use weights included in the public use data to adjust for sample and survey designs. Standard errors are in parentheses. The F-statistic for the reduced form effect on earnings in the 16th quarter (i.e., the "first-stage" F-stat for the supercompliers estimates) is 22.63.

Table 4: National JTPA Study Supercomplier Characteristics for Earnings

	Means			Differences	
	Population	Complier	Supercomplier	SC-Pop	SC-C
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 SC-Pop and SC-C 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. The F-statistic for the reduced form effect on 30-month earnings (i.e., the "first-stage" F-stat for the supercompliers estimates) is 13.15.

Appendix (For Online Publication Only)

A Additional Theoretical Results

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

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

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

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

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

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

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

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

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

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

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

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

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

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

For the second term,

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

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: We show that equation (5) holds under two sets of assumptions. We first show that it follows from equation (4) under the assumptions of Proposition 2. We then show that equation (5) holds under the weaker assumptions of Proposition 3.

Equation (4) implies equation (5) under the assumptions in Proposition 2 because

$$E[h(X_i)(Y_{1i} - Y_{0i})|\text{complier}]$$

$$\overset{\text{(i)}}{=} E[h(X_i)(Y_{1i} - Y_{0i})|\text{supercomplier}] \Pr(Y_{1i} - Y_{0i} > 0|\text{complier})$$
$$\overset{\text{(ii)}}{=} \frac{E[h(X_i)Y_i|Z_i = 1] - E[h(X_i)Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} E[Y_{1i} - Y_{0i}|\text{supercomplier}] \Pr(Y_{1i} - Y_{0i} > 0|\text{complier})$$
$$\overset{\text{(iii)}}{=} \frac{E[h(X_i)Y_i|Z_i = 1] - E[h(X_i)Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} E[Y_{1i} - Y_{0i}|\text{complier}]$$
$$\overset{\text{(iv)}}{=} \frac{E[h(X_i)Y_i|Z_i = 1] - E[h(X_i)Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} \text{LATE},$$

where equalities (i) and (iii) follow from the definition of supercompliers, equality (ii) from Proposition 2, and equality (iv) from the definition of LATE.

To show that equation (5) holds under the assumptions in Proposition 3, i.e., to prove Proposition 3, note that

$$\begin{aligned} & E[h(X_i)(Y_{1i} - Y_{0i})|\text{complier}] \\ =& \frac{E[h(X_i)Y_i|Z_i = 1] - E[h(X_i)Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} \\ =& \frac{E[h(X_i)Y_i|Z_i = 1] - E[h(X_i)Y_i|Z_i = 0]}{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]} \text{LATE}, \end{aligned}$$

where the first equality follows from standard local-average-treatment-effect identification reasoning, and the second equality follows from dividing and multiplying the (non-zero) reduced form. *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 4 *Under Assumptions 1 and 2 and provided that $X \perp\!\!\!\perp Z$, the shares of the two groups are*

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

and their characteristics can be identified as

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

Proof of Proposition 4: 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 4 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 Connections to Marginal Treatment Effects

Remark 5 notes the connections between our identification results and the marginal treatment effect literature. We provide details in this section.

Vytlacil (2002) shows that the LATE framework characterized by Assumption 1 is equivalent to a latent index threshold crossing formulation: $D = 1_{[p(Z) \geq U]}$ where $p(z)$ denotes the (non-constant) propensity score function $\Pr(D = 1|Z = z)$ and "resistance" U a uniform random variable with $Z \perp\!\!\!\perp (Y_1, Y_0, U)$. It follows that the treatment compliers are those with $U \in [p(0), p(1)]$.

Focusing on a binary Y, we define the marginal supercomplier characteristics as

$$\text{SC}(u) \equiv E[h(X)|U = u, Y_1 > Y_0]$$

for $u \in [p(0), p(1)]$.[1] $\text{SC}(u)$ represents the average characteristics among the supercompliers with $U = u$. The next proposition shows that our estimand identifies an average of supercomplier characteristics weighted by MTEs.

[1]Technically, $\text{SC}(u)$ is only well-defined when $\Pr(Y_1 > Y_0|U = u) > 0$. When $\Pr(Y_1 > Y_0|U = u) = 0$, we can define $\text{SC}(u)$ arbitrarily since it receives zero weight in the integrals in Proposition 5.

Proposition 5 *Under Assumptions 1 and 2 and provided that* $X \perp\!\!\!\perp Z$*,*

$$\frac{\int_{p(0)}^{p(1)} \mathrm{SC}(u)\mathrm{MTE}(u)du}{\int_{p(0)}^{p(1)} \mathrm{MTE}(u)du} = \frac{E[h(X)Y|Z=1] - E[h(X)Y|Z=0]}{E[Y|Z=1] - E[Y|Z=0]},$$

where $\mathrm{MTE}(u)$ *is the marginal treatment effect* $E[Y_1 - Y_0|U=u]$*.*

Proof of Proposition 5: Under Assumptions 1 and 2 and with $X \perp\!\!\!\perp Z$,

$$\begin{aligned}
&E[h(X)Y|Z=1] - E[h(X)Y|Z=0] \\
=&E[h(X)(Y_1 - Y_0)|p(0) \leq U \leq p(1)] \Pr(p(0) \leq U \leq p(1)) \\
=&\int_{p(0)}^{p(1)} E[h(X)(Y_1 - Y_0)|U=u]du \\
=&\int_{p(0)}^{p(1)} E[h(X)|U=u, Y_1 > Y_0]E[Y_1 - Y_0|U=u]du \\
=&\int_{p(0)}^{p(1)} \mathrm{SC}(u)\mathrm{MTE}(u)du,
\end{aligned}$$

where the first equality follows from Assumption 1, $X \perp\!\!\!\perp Z$, and the definition of U, the second equality from the definition of conditional expectation, the third equality from Assumption 2 and the observation that $\Pr(Y_1 > Y_0|U=u) = E[Y_1 - Y_0|U=u]$, and the final equality from the definitions of $\mathrm{SC}(u)$ and $\mathrm{MTE}(u)$. Similarly,

$$E[Y|Z=1] - E[Y|Z=0] = \int_{p(0)}^{p(1)} \mathrm{MTE}(u)du,$$

and the desired result follows. *QED.*

As mentioned in Remark 5, with a continuous instrument that traces out the propensity score from zero to one, we can formulate the corresponding estimand that identifies $\mathrm{SC}(u)$ at any value of u between 0 and 1:

$$\mathrm{SC}(u) = \frac{\frac{d}{du}E[h(X)Y|p(Z)=u]}{\frac{d}{du}E[Y|p(Z)=u]}. \tag{A2}$$

Our estimand in equation (A2) is the continuous analog of that in Proposition 5: It is constructed from derivatives—as opposed to differences—of the relevant conditional expectation functions. Previous studies have established that the denominator of the right hand side of

(A2) identifies MTE(u) (e.g., Heckman, Urzua, and Vytlacil 2006; Carneiro, Heckman, and Vytlacil 2011; Cornelissen et al. 2016). Similar reasoning reveals that the corresponding numerator identifies the product SC(u) $\cdot$ MTE(u), and (A2) follows.

MTE(u) and SC(u) describe the supercomplier population at the margin as the instrument shifts treatment take-up among individuals with different resistance levels u. MTE(u) tells us the average gains at each u; in the binary-Y, monotone-outcome case it equals the marginal fraction of beneficiaries. SC(u) additionally characterizes the composition of these beneficiaries in terms of pre-treatment covariates.

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 from 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 takes on three values in $\{0, 1, 2\}$ and assume perfect compliance ($D = Z$). We can categorize each unit as one of six types: $a_0, a_1, a_2, c_{01}, c_{02}, c_{12}$, where

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

The (unknown) shares of each type 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

$$\begin{aligned} \Pr(Y = 0 \mid Z = 0) &= \Pr(a_0) + \Pr(c_{01}) + \Pr(c_{02}) \\ \Pr(Y = 1 \mid Z = 0) &= \Pr(a_1) + \Pr(c_{12}) \\ \Pr(Y = 2 \mid Z = 0) &= \Pr(a_2) \\ \Pr(Y = 0 \mid Z = 1) &= \Pr(a_0) \\ \Pr(Y = 1 \mid Z = 1) &= \Pr(a_1) + \Pr(c_{01}) \\ \Pr(Y = 2 \mid Z = 1) &= \Pr(a_2) + \Pr(c_{02}) + \Pr(c_{12}). \end{aligned} \tag{A3}$$

These six equations form a linear mapping

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

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 related observation from this non-identification result is that the distributions of Y_0 and Y_1 among the supercompliers are also not identified in general. To see this in the simple example above, suppose the observed data satisfy

$$\Pr(Y=0 \mid Z=0)=0.5;\ \Pr(Y=1 \mid Z=0)=0.5;\ \Pr(Y=2 \mid Z=0)=0$$
$$\Pr(Y=0 \mid Z=1)=0;\ \Pr(Y=1 \mid Z=1)=0.5;\ \Pr(Y=2 \mid Z=1)=0.5.$$

Then (A3) admits an infinite number of solutions for the shares of the latent types. We focus on two of these solutions; the first one is

$$\big(\Pr(a_0), \Pr(a_1), \Pr(a_2), \Pr(c_{01}), \Pr(c_{02}), \Pr(c_{12})\big) = \big(0, 0, 0, 0.5, 0, 0.5\big),$$

in which all units are supercompliers, so among the supercompliers Y_0 takes values 0 and 1 with probability 0.5 each and Y_1 takes values 1 and 2 with probability 0.5 each. The other solution is

$$\big(\Pr(a_0), \Pr(a_1), \Pr(a_2), \Pr(c_{01}), \Pr(c_{02}), \Pr(c_{12})\big) = \big(0, 0.5, 0, 0, 0.5, 0\big),$$

in which only the c_{02} units are supercompliers, so among the supercompliers $Y_0 = 0$ and $Y_1 = 2$ with probability one. Because these two latent solutions (or data generating processes) yield the same observed distribution of (Y, Z) (and hence of (Y, D, Z), since $D = Z$) but imply different marginal distributions of Y_0 and Y_1 among the supercompliers, those distributions are not identified.

A.5 Estimation and Inference: Additional Discussion

In this section, we discuss alternative supercomplier characteristics estimators by focusing on those with a direct analog in the existing complier characteristics literature. First, we note that the Stata command in (6) does not estimate an exact sample analog of the Abadie-style equation (1) from Proposition 1. As we show in Comey, Eng, and Pei (2023), the estimand from equation (1) also has a Wald-type representation similar to (2):

$$\frac{E[h(X)\{Y-(1-\tau)\}|Z=1]-E[h(X)\{Y-(1-\tau)\}|Z=0]}{E[Y-(1-\tau)|Z=1]-E[Y-(1-\tau)|Z=0]}, \tag{A4}$$

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). We also show in Comey, Eng, and Pei (2023) that 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 Comey, Eng, and Pei 2023 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.

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, as stated in Section 2.2, 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. (2025) when the treatment variable is non-binary.

A.6 Details on Remark 7

In this section, we provide more details on Remark 7 and discuss the factors that influence the precision of (super)complier characteristics estimates. We focus on the case of estimating supercomplier characteristics without controls, but our reasoning easily extends to estimation with controls after proper residualizing.

Let γ_Y denote the Wald estimand of the mean supercomplier characteristics underlying (6), i.e., the right hand side of equations (2) and (4) when setting $h(X) = X$. Using standard asymptotics with the delta method, it is easy to show that the asymptotic variance of the 2SLS estimator $\hat{\gamma}_Y$ is the constant

$$avar(\hat{\gamma}_Y) = \frac{1}{\delta_Y^2}\left[\left(\frac{\sigma^2_{XY|1}}{\tau} + \frac{\sigma^2_{XY|0}}{1-\tau}\right) + \gamma_Y^2\left(\frac{\sigma^2_{Y|1}}{\tau} + \frac{\sigma^2_{Y|0}}{1-\tau}\right)\right.$$
$$\left. - 2\gamma_Y\left(\frac{\sigma_{XY,Y|1}}{\tau} + \frac{\sigma_{XY,Y|0}}{1-\tau}\right)\right], \quad \text{(A5)}$$

where $\delta_Y \equiv E[Y|Z=1] - E[Y|Z=0]$ denotes the reduced-form effect, $\sigma^2_{XY|z} \equiv var(XY|Z = z)$, $\sigma^2_{Y|z} \equiv var(Y|Z = z)$, and $\sigma_{XY,Y|z} \equiv cov(XY, Y|Z = z)$ denote second moments within the treatment ($z = 1$) and control ($z = 0$) groups, and τ represents the probability of being assigned to the treatment group as defined in Appendix A.5. Note that we can define $u \equiv XY - \gamma_Y Y$ and $\sigma^2_{u|z} \equiv var(u|Z = z)$, and the quantity inside the square bracket of equation (A5) is simply $\sigma^2_{u|1}/\tau + \sigma^2_{u|0}/(1-\tau)$.

Three observations emerge from equation (A5). First, the same derivation applies to the analogous 2SLS estimator of mean complier characteristics $\hat{\gamma}_D$, and its asymptotic variance is given by the right hand side of (A5) with Y replaced by D where δ_D represents the first-stage effect of Z on D. Second, the asymptotic variance of $\hat{\gamma}_Y$ ($\hat{\gamma}_D$) is inversely proportional to δ_Y^2 (δ_D^2). When Y is binary, δ_Y and δ_D are the shares of supercompliers and treatment compliers, respectively, under our identifying assumptions, and therefore, $\delta_Y \leq \delta_D$. Consequently, all else equal, we might expect the standard error of $\hat{\gamma}_Y$ to be larger than $\hat{\gamma}_D$. And third, we can rewrite equation (A5) as

$$avar(\hat{\gamma}_Y) = \frac{N}{F_Y} \cdot \frac{\sigma^2_{u|1}/\tau + \sigma^2_{u|0}/(1-\tau)}{\sigma^2_{Y|1}/\tau + \sigma^2_{Y|0}/(1-\tau)}, \quad \text{(A6)}$$

where N is the sample size, and $F_Y \equiv N \cdot \delta_Y^2 / \big(\sigma^2_{Y|1}/\tau + \sigma^2_{Y|0}/(1-\tau)\big)$ is the asymptotic F-

statistic from the reduced form. Therefore, regardless of the distribution of Y, $avar(\hat{\gamma}_Y)$ is inversely proportional to F_Y.[2] Just like the first observation above, equation (A6) also applies to $avar(\hat{\gamma}_D)$ with Y replaced by D, and it follows that $avar(\hat{\gamma}_D)$ is inversely proportional to F_D, the asymptotic F-statistic from the first stage. It is common in empirical applications to have $F_Y < F_D$ and often substantially so; as a result, it would be harder to precisely estimate supercomplier characteristics than complier characteristics.

As we emphasize in Remark 7, it makes sense to apply our machinery in settings where the reduced-form effect is strong. That is, it will be meaningful to characterize supercompliers only if there is compelling evidence that they exist.

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 (6) 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. (forthcoming) finding on "weak causality"—under certain conditions, the LATE estimand with covariate controls identifies a non-negatively weighted average of complier treatment effects.

Assumption 3 (Stratified Experiment)

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

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

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

Assumptions 3.1 and 3.2 are the covariate-control counterparts to Assumptions 1.1 and 2.2, respectively. Assumption 3.3 stipulates that the true conditional probability of treatment

[2] The result is similar in spirit to that by Shea (1997) who connects the asymptotic variance of a 2SLS estimator with the R^2 of the first stage.

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

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

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

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

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

Decompose the numerator of β_{2SLS} as

$$\begin{aligned} E[h(X)Y\tilde{Z}] &= E[E[h(X)Y\tilde{Z}|W]] \\ &= E\big[cov(h(X)Y, \tilde{Z}|W) + E[h(X)Y|W]E[\tilde{Z}|W]\big] \\ &= E\big[\underbrace{cov(h(X)Y, Z|W)}_{\text{(i)}} + \underbrace{E[h(X)Y|W]E[\tilde{Z}|W]}_{\text{(ii)}}\big], \end{aligned} \tag{A7}$$

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 (A7) as:

$$\begin{aligned} cov(h(X)Y, Z|W) &= E[h(X)Y(Z - E[Z|W])|W] \\ &= E[h(X)Y|Z = 1, W](1 - E[Z|W])E[Z|W] \\ &\quad - E[h(X)Y|Z = 0, W](E[Z|W])(1 - E[Z|W]) \\ &= (E[h(X)Y|Z = 1, W] - E[h(X)Y|Z = 0, W])(1 - E[Z|W])E[Z|W] \\ &= E[h(X)|G = cc, W] \times RF_W \times (1 - E[Z|W])E[Z|W] \end{aligned} \tag{A8}$$

where the last equality follows from the conditional version 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 (A7), 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 is linear in W, which is true in stratified experiments with W being the full set of stratum dummies.

Hence, the numerator of β_{2SLS} is

$$E[h(X)Y\tilde{Z}] = E\big[E[h(X)|G = cc, W] \times RF_W \times (1 - E[Z|W])E[Z|W]\big]. \tag{A9}$$

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

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

and the result of Proposition 6 follows with weight $\omega_W = RF_W \times (1 - E[Z|W])E[Z|W]$. Under Assumptions 1.2, 1.3, 2.1, and 3.1, the conditional version of Lemma 1 holds and implies $RF_W = \Pr(G = cc \mid W) \geq 0$; since Z is binary, $\omega_W \geq 0$ for all values of W. *QED.*

A.8 Relaxing Outcome Monotonicity

Focusing on the case of a binary Y, Proposition 7 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 7 *Under Assumptions 1 and 2.2, and provided that $X \perp\!\!\!\perp Z$ and $E[Y|Z = 1] - E[Y|Z = 0] \neq 0$,*

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

where we define

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

Proof of Proposition 7: The numerator of the estimand is

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

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.*

Proposition 7 says that when outcome monotonicity does not hold, the estimand for supercomplier characteristics will be biased unless the supercompliers and the cf group have the same average characteristics. The bias increases linearly with the cf 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.