Estimating Treatment Complementarity

Hyewon Kim*

November 8, 2025

Abstract

How can we estimate the complementarity between two treatments when assignment is not fully random, such as in randomized experiments with imperfect compliance or in quasi-experimental settings? The first part of this paper shows that the commonly used two-stage least squares (2SLS) method—with instruments for each treatment and their interaction—is often not suitable for estimating treatment interaction effects. Specifically, 2SLS requires strong assumptions about (1) treatment effect heterogeneity or (2) types of compliers. I show that these assumptions have testable implications on first stage patterns, and these often fail in published empirical studies on complementarity. The second part of the paper proposes an alternative estimation strategy for cases where these assumptions for 2SLS are unlikely to hold. Building on the marginal treatment effect literature, this approach models potential outcomes as a linear function of individuals' unobserved resistance to treatment and offers a clearer connection to the intended estimand of treatment interaction. Lastly, the paper revisits Angelucci and Bennett (2024), an experimental study of complementarity under imperfect compliance, to illustrate how the proposed diagnostics and alternative estimator can enhance empirical analysis of interactions between two treatments.

JEL classification codes: C26, C36.

Keywords: Instrumental variables; Treatment complementarity.

*Department of Economics, Cornell University (email: hk935@cornell.edu)

I am deeply grateful to Doug Miller, Zhuan Pei, and Evan Riehl for their guidance and support. I would also like to thank Francesca Molinari, Joerg Stoye, Neil Cholli, Leonard Goff, Pauline Leung, and Seojung Oh for their helpful feedback and discussions. I am grateful to the participants of the Cornell Labor Work in Progress Seminar, Cornell Econometrics Reading Group, and Econometric Society World Congress for their constructive feedback.

1 Introduction

Treatment interactions are central to many questions in social science. Is an early-childhood intervention more effective when followed by a later intervention? Do skills beget skills (Cunha and Heckman, 2007)? Is a combination of policies more effective at reducing poverty than individual policies in isolation? Why is a policy effective in one context but ineffective in another?

Researchers often test for such complementarities using instrumental variable (IV) designs with interaction terms. Let $Y(T_1, T_2)$ denote a potential outcome as a function of two distinct binary treatments, T_1 and T_2. A treatment complementarity exists when the sum of the separate treatment effects differs from the combined treatment effect on average:

$$\mathbb{E}[Y(1,0) - Y(0,0)] + \mathbb{E}[Y(0,1) - Y(0,0)] \neq \mathbb{E}[Y(1,1) - Y(0,0)]. \tag{1}$$

If both treatments are randomly assigned, a valid test for complementarity is to estimate the OLS regression

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_c (T_1 \times T_2) + \epsilon, \tag{2}$$

and test whether $\beta_c = 0$. But treatment assignment is often not random, so researchers commonly use a two-stage least squares (2SLS) approach in which T_1, T_2 and $T_1 \times T_2$ are instrumented with the corresponding instruments, Z_1, Z_2, and $Z_1 \times Z_2$. Such IV regressions are used in both quasi-experimental research designs[1] (e.g. Johnson and Jackson 2019) and in randomized controlled trials where compliance with treatment assignment is imperfect (e.g. Angrist, Lang and Oreopoulos 2009; Fang et al. 2023)[2].

[1] To examine treatment interactions using quasi-experiments, researchers have also combined other research designs, such as difference-in-differences or regression discontinuity designs, into a single regression that includes an interacted treatment variable (Neumark and Wascher, 2011; Johnson and Jackson, 2019; Rossin-Slater and Wüst, 2020; Kerwin and Thornton, 2021; Gilligan et al., 2022; Goff et al., 2023). While this paper does not directly address these combined designs, the concern about nonconvex weighting of treatment effects under heterogeneity raised here may also call for a careful examination of the assumptions underlying such approaches.

[2] Some experimental papers do not use 2SLS but rely on reduced-form estimates instead (Angelucci and Bennett, 2024; Duflo, Dupas and Kremer, 2015). As explained in a later section, reduced-form estimates can provide misleading information about treatment complementarity under imperfect compliance. Some research questions can therefore benefit from exploring treatment complementarity beyond reduced-form evidence, depending on the target estimand of the research question.

This paper shows that this common approach—2SLS with instruments for each treatment and their interaction—can frequently yield biased estimates of treatment complementarity. Bias can arise when two plausible conditions hold: 1) treatment effects are heterogeneous and 2) there is imperfect compliance between the instruments and treatment assignment. Under these conditions, the assumptions for 2SLS to identify causal treatment interactions require unrealistic restrictions on individuals' compliance behavior. This motivates an alternative design that allows for more plausible compliance patterns. I propose an alternative estimation strategy that extends the marginal treatment effects framework to the estimation of treatment interactions. Using Monte Carlo simulation, I show the new approach performs better than 2SLS. I also introduce diagnostic tools that can help researchers assess when the identification assumptions for 2SLS are likely to fail in practice. I apply the findings to Angelucci and Bennett (2024), an experimental design that examines treatment complementarity under imperfect compliance, to illustrate how the proposed diagnostics and alternative estimator can enhance empirical analysis of interactions between two treatments.

The first part of this paper discusses the pitfalls of using 2SLS to estimate treatment interactions. Using a potential outcomes framework, I show that researchers must assume either homogeneous treatment effects or impose strong restrictions on complier types for 2SLS to correctly identify treatment complementarity. A key requirement for 2SLS to be valid is that each instrument affects only its intended treatment and has no influence on the uptake of the other treatment, e.g., Z_2 must not affect take-up of T_1. Yet such patterns of treatment compliance are implausible in many settings where treatment interaction is of interest. For example, if families are considering whether to move to a better neighborhood (T_1) and send their child to a better school (T_2), a voucher that reduces the cost of moving (Z_1) may impact the choice of schools (T_2).

My framework shows that violations of these conditions generate contamination bias, where 2SLS estimands reflect not only the intended treatment effect but also spillovers from the other treatment effect. In this case, estimates of β_c using 2SLS may capture not only the treatment complementarity effect but also unrelated treatment effects, leading to misleading conclusions. Notably, contamination bias for interaction terms differs from that in the general context of multiple treatments (Goldsmith-Pinkham, Hull and Kolesár, 2024; Bhuller and Sigstad, 2024) because the parameters of interest often involve comparisons across treatment effects. For example, research on treatment complementarity commonly examines not only the interaction effect itself but also the individual treatment effects and the combined effect (which is the sum of the individual effects and the complementarity effect)[1]. To identify

[1] Alternatively, researchers often estimate the separate treatment effects and the combined effect in a

the combined effect using the 2SLS estimates, researchers must impose an even stronger assumption: that all 2SLS coefficients, β_1,β_2 and β_c, are identified from the same set of compliers. This requirement further restricts the allowable complier patterns compared to standard multiple-treatment settings.

I develop testable implications for first-stage regressions to assess whether the assumptions required for identifying treatment complementarity hold in empirical applications. These diagnostics clarify how each instrument should affect only its corresponding treatment and reveal that such ideal first-stage patterns—necessary for 2SLS to yield causal estimates—are rarely satisfied in practice. Specifically, the assumption that each instrument affects only its intended treatment translates into a requirement that the coefficient on the non-targeted instrument be zero. For example, regressing T_1 on Z_1, Z_2 and $(Z_1 \times Z_2)$ should yield zero coefficients for Z_2 and $(Z_1 \times Z_2)$. The same is true for the first-stage regressions of T_2 and $(T_1 \times T_2)$. The common compliers assumption further requires that each relevant first-stage coefficient be equal and positive. For example, the coefficient on Z_1 in the T_1 regression should be equal to the coefficient on Z_2 in the T_2 regression and the coefficient on $(Z_1 \times Z_2)$ in the $(T_1 \times T_2)$ regression. As I show in the replication part of my paper below, such first-stage patterns are rarely observed in published empirical studies of treatment interactions.

Furthermore, my framework clarifies that reduced-form estimates of instrument interactions can also be misleading when the same assumptions about complier types do not hold. For example, the coefficient on $(Z_1 \times Z_2)$ in the reduced-form regression may not only capture the treatment complementarity effect but also unrelated treatment effects, through the take-up of other treatments not intended by this instrument itself. This may be acceptable when the sole interest lies in the effect of the policy instrument itself regardless of the channel. However, researchers often care about the underlying interaction between treatments to better understand the policy effect. These misleading conclusions can arise even when treatment effects are homogeneous. When the assumptions underpinning 2SLS are unlikely to hold, an alternative estimation strategy may be necessary to credibly assess treatment complementarity.

In the second part of the paper, I extend the marginal treatment effects (MTE) literature by developing an alternative estimation strategy to 2SLS that recovers interaction effects. This approach models potential outcomes as a linear function of individuals' underlying resistance to treatment. It uses average outcomes conditional on combinations of instruments

regression, then compute treatment complementarity by subtracting the sum of the two separate effects from the combined effect. The argument in this paragraph applies to that specification as well, since it also requires assuming that all 2SLS coefficients are identified from the same set of compliers.

and treatments to estimate the linear parameters in the outcome function, where the outcome function is modeled using two-dimensional resistance to each treatment. Based on the extrapolated outcome function, the researcher can construct either an average treatment interaction estimate or an estimate conditional on complier types. Although this method relies on strong assumptions including linearity, it accommodates a broader and more realistic set of compliers than the 2SLS framework and offers a clearer connection to the intended estimand. In a simulated example where 2SLS produces biased estimates, the linear extrapolation approach yields estimates that are closer to the true interaction effect, provided that the data-generating process satisfies the identifying assumptions.

In addition, this section introduces a diagnostic and a practical measure for researchers who continue to rely on 2SLS, thereby improving the paper's applicability to empirical practices. The first part of this subsection discusses how to detect underlying potential outcome heterogeneity—the core threat to 2SLS validity—by comparing outcomes among individuals with the same treatment take-up but different instrument assignments. The second part explores how covariates related to treatment effect heterogeneity can mitigate bias in 2SLS estimation. Simulation results show that simply including covariates linearly does little to reduce bias, even when those covariates are informative about treatment effect heterogeneity. To fully exploit the role of covariates in addressing bias in 2SLS, the regression specification should be saturated, consistent with the findings of Blandhol et al. (2022).

The final section applies the findings from the first two sections to an empirical study on the complementarity between psychiatric treatment and economic assistance in a randomized controlled trial with non-compliance (Angelucci and Bennett, 2024). The analysis suggests that the coefficient comparisons in their paper may have limitations, as indicated by the first-stage patterns. In particular, differences in coefficients may reflect variation in treatment effects across complier groups rather than genuine complementarity between treatments. To complement Angelucci and Bennett (2024)'s reduced-form evidence, I provide additional analysis examining treatment complementarities using the proposed diagnostics and alternative estimation approach. Applying the potential-outcome-heterogeneity diagnostics and the estimator introduced in Section 4, I find that both 2SLS and the alternative method yield limited insights. One of the key challenges for the alternative method is the small sample size, highlighting the empirical challenge of identifying interaction effects under imperfect compliance.

To my knowledge, this is the first paper in economics to focus specifically on the methodological challenges of identifying interaction effects between treatments.[2] Earlier work has

[2] Two related papers also address treatment interactions, one in the context of a political experiment and

examined settings in which treatments act as substitutes by design. For example, Goldsmith-Pinkham, Hull and Kolesár (2024) show that regressions generally fail to estimate convex averages of heterogeneous treatment effects when multiple treatments are included alongside covariates. Bhuller and Sigstad (2024) demonstrate the presence of contamination bias in two-stage least squares (2SLS) estimation and identify conditions under which 2SLS recovers convex averages of treatment effects. Their insights on the contamination bias present in 2SLS overlaps with my paper, but their analysis does not discuss potential interactions between multiple treatments. I show that the assumptions required for causal identification of interaction effects using 2SLS can be more restrictive than those needed to estimate the effects of each treatment separately. This is because the parameters of interest often involve comparisons between Local Average Treatment Effects (LATEs) identified by different regression coefficients. In other words, when researchers are interested in interaction effects, it is important to recognize that each coefficient may correspond to a different set of compliers, a concern that could be less relevant when the goal is to estimate separate effects of each treatment. Papers with specific empirical applications such as schooling choice, college major choices or Moving to Opportunity experiment also contain insights on how 2SLS estimands average across complier types (Kline and Walters, 2016; Kirkeboen, Leuven and Mogstad, 2016; Mountjoy, 2022; Pinto, 2022), but none of these empirical contexts directly address the methodological implications for estimating treatment interactions.

Second, this paper builds on the literature on instrumental variables and MTE estimation to develop a new approach for estimating treatment interactions. While the existing research has primarily focused on single-treatment setups (Mogstad and Torgovitsky, 2024; Brinch, Mogstad and Wiswall, 2017; Kowalski, 2023*a*,*b*), only limited work has extended the framework to multiple treatments. My contribution is to provide a use case motivating the need to extend the framework to multiple treatments and multiple resistance parameters based on treatment interactions. I also clarify the set of assumptions researchers should consider when extending this method to estimate treatment interaction effects.

Third, this paper extends recent work on the interpretation of causal estimands under treatment effect heterogeneity to the case of treatment complementarity. This literature

the other within a broader discussion of instrumental-variable identification. Blackwell (2017) shows the limitations of 2SLS in identifying treatment interactions in a political science experiment. While sharing a similar motivation, my analysis generalizes Blackwell's insight by clarifying how contamination bias arises when the treatment exclusion condition assumed in Blackwell (2017) is violated. I also develop diagnostics and alternative estimation strategies to guide empirical researchers when such conditions are not satisfied. Goff (2025) discusses complementarity as one application within a general framework of instrumental variables under unrestricted treatment effect heterogeneity, but Goff (2025)'s focus differs from mine in that Goff does not focus on estimation strategies when identification fails.

has explored the consequences of treatment effect heterogeneity for the interpretation of 2SLS estimates (Blandhol et al., 2022; Mogstad, Torgovitsky and Walters, 2021; Bhuller and Sigstad, 2024) and difference-in-differences designs (De Chaisemartin and D'Haultfœuille, 2020; Goodman-Bacon, 2021). This paper contributes to this work by showing that researchers who study treatment interactions by combining different sources of quasi-experimental variation in a single regression may also obtain estimands that are distorted by unintended weighting or inappropriate comparisons under treatment effect heterogeneity.

Lastly, this paper lays a foundation for empirical research on treatment complementarity by clarifying when and why standard tools may fail, what diagnostic analyses can be used, and what alternative methods may be more appropriate. Despite many compelling research questions in this area, empirical work on treatment complementarity remains limited (Neumark and Wascher, 2011; Johnson and Jackson, 2019; Rossin-Slater and Wüst, 2020; Kerwin and Thornton, 2021; Gilligan et al., 2022; Goff et al., 2023). A key reason for this gap is the lack of methodological guidance for obtaining credible estimates of treatment interactions. This paper fills that gap by connecting recent methodological advances on multiple treatments to empirical studies of treatment complementarity.

The remainder of the paper is organized as follows. Section 2 presents a motivating simulation that illustrates why 2SLS can fail and why reduced-form estimates may be misleading. Section 3 generalizes these insights and introduces first-stage diagnostics. Section 4 proposes an alternative estimation strategy and introduces additional diagnostics and practical tools for applied researchers using 2SLS. Section 5 applies these insights to Angelucci and Bennett (2024), and Section 6 concludes.

2 Motivating Example

This section explains why 2SLS can fail to recover unbiased estimates of treatment complementarity when treatment effects are heterogeneous. I first present a simulation in which all individuals have homogeneous treatment effects and the econometrician recovers the complementarity estimate. I then introduce heterogeneous treatment effects and show that the econometrician fails to recover the true complementarity estimate. The section concludes by explaining why reduced-form estimates can also be misleading in such settings.

Consider a randomized experiment with imperfect compliance designed to estimate the following regression:

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_c (T_1 \times T_2) + \epsilon$$

Here, Y is an outcome such as income, $T_1 \in \{0,1\}$ indicates living in a good neighborhood, $T_2 \in \{0,1\}$ indicates attending a high-quality school, and $(T_1 \times T_2) \in \{0,1\}$ indicates doing both at the same time. The econometrician is interested in the effect of neighborhood quality, the effect of school quality, and whether there is a complementarity effect. However, ϵ may be correlated with the housing and schooling decisions. This correlation prevents the econometrician from obtaining unbiased estimates of β_1, β_2, and β_c using OLS.

Suppose there is a randomly assigned housing voucher $Z_1 \in \{0,1\}$, a randomly assigned schooling voucher $Z_2 \in \{0,1\}$, and a randomly assigned voucher for both housing and schooling $(Z_1 \times Z_2) \in \{0,1\}$. Each instrument encourages take-up of the corresponding treatment. Assume a cross-randomization design in which assignment of Z_1 and Z_2 is independent: $\epsilon \perp (Z_1, Z_2)$ and $Z_1 \perp Z_2$. Let $\Pr(Z_1 = 1) = \Pr(Z_2 = 1) = 1/2$.

Now introduce imperfect compliance represented by three distinct complier types. One third are never takers, who never take up any treatment under any voucher assignment. Another third are "dutiful compliers," who take up whichever treatment they are assigned. The remaining third are "reluctant compliers," who take up both treatments only when they receive both vouchers and take up no treatment otherwise. This pattern is realistic when reluctant compliers face high costs of taking up either treatment individually, but the joint vouchers provide sufficient incentive to adopt both. The table below summarizes these compliance patterns.

	Receives no vouchers $(Z_1 = 0, Z_2 = 0)$	Receives housing voucher only $(Z_1 = 1, Z_2 = 0)$	Receives schooling voucher only $(Z_1 = 0, Z_2 = 1)$	Receives both vouchers $(Z_1 = 1, Z_2 = 1)$
Never-takers	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$
Dutiful-compliers	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 0, T_2 = 1$	$T_1 = 1, T_2 = 1$
Reluctant-compliers	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 1$

Table 1: Types of compliers in the motivating simulation

Assume homogeneous treatment effects for all individuals, regardless of complier type. Suppose growing up in a good neighborhood increases later wages by \$2 and attending a high-quality school also increases later wages by \$2. Assume there is no complementarity between neighborhood and school quality. Let the true data-generating process be:

$$Y = 2T_1 + 2T_2 + 0(T_1 \times T_2) + \epsilon$$

Under homogeneous effects, if the econometrician estimates 2SLS with three endogenous variables T_1, T_2, and $(T_1 \times T_2)$ instrumented by the randomly assigned Z_1, Z_2, and $(Z_1 \times Z_2)$, the estimates recover the two individual treatment effects and the complementarity effect. The equation below reports the mean 2SLS estimates from 1,000 Monte Carlo simulations (each with $N = 100{,}000$). Standard deviations of the estimates are shown in parentheses:

$$Y = \underset{(0.03)}{2.00}T_1 + \underset{(0.03)}{2.00}T_2 + \underset{(0.04)}{0.00}(T_1 \times T_2) + \epsilon$$

Now introduce treatment effect heterogeneity. Consider the following data-generating process:

$$\begin{aligned}
Y &= 1T_1 + 1T_2 + 0(T_1 \times T_2) + \epsilon \quad \text{if the individual is a never taker}\\
Y &= 2T_1 + 2T_2 + 0(T_1 \times T_2) + \epsilon \quad \text{if the individual is a dutiful complier}\\
Y &= 3T_1 + 3T_2 + 0(T_1 \times T_2) + \epsilon \quad \text{if the individual is a reluctant complier}
\end{aligned}$$

In this setting, the same 2SLS specification yields biased estimates of treatment complementarity. The equation below shows the mean 2SLS estimates from 1,000 Monte Carlo simulations (again with $N = 100{,}000$), with standard deviations in parentheses:

$$Y = \underset{(0.03)}{2.00}T_1 + \underset{(0.03)}{2.00}T_2 + \underset{(0.04)}{1.00}(T_1 \times T_2) + \epsilon$$

Let Δ_1 denote the effect of the first treatment, Δ_2 the effect of the second treatment, and Δ_c the complementarity effect. Even though Δ_c equals zero for every individual, the 2SLS estimate of the complementarity effect is positive. Denoting dutiful compliers by DC and reluctant compliers by RC, the expected value of the third 2SLS coefficient is the following. This expression follows directly from the general 2SLS expectation formula derived in Section 3.1, after substituting in the complier types assumed in this simulation:

$$\mathbb{E}[\hat{\beta}_c] = \mathbb{E}[\Delta_c \mid DC, RC] + \tfrac{1}{2}\big(\mathbb{E}[\Delta_1 \mid RC] - \mathbb{E}[\Delta_1 \mid DC]\big) + \tfrac{1}{2}\big(\mathbb{E}[\Delta_2 \mid RC] - \mathbb{E}[\Delta_2 \mid DC]\big)$$

Intuitively, the biased interaction coefficient arises because reluctant compliers respond only when both vouchers are offered, which induces correlation between the second instrument and the first treatment (and vice versa). Their first- and second-treatment effects are therefore misattributed to the interaction term, generating a spurious estimate of complementarity. This issue does not arise in standard IV settings with a single treatment but emerges in multi-treatment settings where the take-up of one treatment can depend on the assignment of the other instrument.

This example illustrates that the 2SLS estimand for treatment interaction reflects not only the intended interaction effect but also differences in the main effects across complier types. Note that the three complier patterns satisfy standard IV compliance assumptions. Each voucher increases take-up of its corresponding treatment for all complier types, and there are no defiers. This setting therefore calls for a more careful account of the complier structure when studying treatment interactions.

Finally, I demonstrate that reduced-form estimates can also be misleading under the same data-generating process. The following equations report the mean reduced-form estimates from 1,000 Monte Carlo simulations (each with $N = 100{,}000$) for both the homogeneous- and heterogeneous-effects settings. Standard deviations are shown in parentheses:

$$\text{Homogeneous Effect setting: } Y = \underset{(0.01)}{0.67} Z_1 + \underset{(0.01)}{0.67} Z_2 + \underset{(0.02)}{1.33} (Z_1 \times Z_2)$$

$$\text{Heterogeneous Effect setting: } Y = \underset{(0.01)}{0.67} Z_1 + \underset{(0.01)}{0.67} Z_2 + \underset{(0.02)}{2.00} (Z_1 \times Z_2)$$

The third coefficient in the reduced-form specification has the following expectation, where dutiful compliers are denoted by DC and reluctant compliers by RC:

$$\mathbb{E}[\hat{\gamma}_c] \;=\; \mathbb{E}[\Delta_c \mid DC, RC] \cdot \Pr(DC, RC) + \mathbb{E}[\Delta_1 \mid RC] \cdot \Pr(RC) + \mathbb{E}[\Delta_2 \mid RC] \cdot \Pr(RC)$$

This expression shows that the treatment effects associated with reluctant compliers spill over into the interaction coefficient in the reduced-form specification as well. This happens because the reluctant complier's take-up of either treatment depends on receiving both vouchers, causing their single-treatment effects to load onto the instrument interaction term. While this may be acceptable when the goal is to estimate the effect of the policy instrument itself—regardless of the behavioral channel—it becomes problematic when the estimand of interest is the causal effect of treatment take-up. In such cases, reduced-form estimates are not informative about the underlying treatment complementarity, even under homogeneous treatment effects, unless types such as reluctant compliers can be credibly ruled out.

3 General Problem

This section formalizes the challenges of estimating treatment complementarity with 2SLS. Section 3.1 sets up the problem and characterizes the 2SLS estimands. Section 3.2 provides the corresponding results for reduced form. Section 3.3 derives testable implications for the first-stage regressions and shows that they are often violated in empirical applications.

3.1 2SLS Estimation

There are four possible treatments, $(T_1, T_2) \in (0,0), (1,0), (0,1), (1,1)$, and individuals can be assigned to four possible instruments, $(Z_1, Z_2) \in (0,0), (1,0), (0,1), (1,1)$. This section provides an interpretation of the 2SLS estimate from the following equation with Z_1, Z_2, and $(Z_1 \times Z_2)$ as instruments:

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_c (T_1 \times T_2) + \epsilon \tag{3}$$

The 2SLS estimation is characterized by the moment conditions

$$\mathbb{E}[\epsilon] = \mathbb{E}[\epsilon Z_1] = \mathbb{E}[\epsilon Z_2] = \mathbb{E}[\epsilon Z_1 Z_2] = 0 \tag{4}$$

I denote the potential outcome as $Y(T_1, T_2, Z_1, Z_2)$ or $Y(T_1, T_2)$ and the potential treatment choice as $T_1(Z_1, Z_2)$ and $T_2(Z_1, Z_2)$, which take the value 1 if the individual takes up treatment T_k when given the instrument pair (Z_1, Z_2). Note that $(T_1 \times T_2)(Z_1, Z_2) \equiv T_1(Z_1, Z_2) \times T_2(Z_1, Z_2)$. I maintain the following assumptions throughout the paper.

Assumption 1. ***(Exclusion restriction).*** *$Y(T_1, T_2, Z_1, Z_2) = Y(T_1, T_2)$ for all values of Z_1, Z_2, T_1, and T_2.*

The first assumption implies that the instruments affect the outcome only through their influence on the treatment take-up.

Assumption 2. ***(Independence).*** *$(Y(T_1, T_2), T_1(Z_1, Z_2), T_2(Z_1, Z_2)) \perp (Z_1, Z_2)$, and $Z_1 \perp Z_2$.*

The second assumption is that the instruments are as good as randomly assigned and thus uncorrelated with potential outcomes and potential choices. In addition, the two instruments are assumed to be assigned independently of each other, which is the case in cross-randomized experiments.

Assumption 3. ***(Relevance).***

$$\mathbb{E}\big([1\ Z_1\ Z_2\ (Z_1 \times Z_2)]^T[1\ T_1\ T_2\ (T_1 \times T_2)]\big)\ \textit{has full rank.}$$

The third assumption ensures that the instruments provide sufficient variation to identify the treatment effects.

Assumption 4. ***(Monotonicity).*** *$T_1(1,0) \geq T_1(0,0)$, $T_2(0,1) \geq T_2(0,0)$, and $(T_1 \times T_2)(1,1) \geq (T_1 \times T_2)(0,0)$.*

The fourth assumption ensures that each instrument does not discourage take up of the corresponding treatment, which parallels the commonly assumed "No defiers" assumption in the standard binary treatment and binary instrument setup (Imbens and Angrist, 1994).

Note that the simulation example in the previous section also satisfies Assumptions 1–4 yet does not achieve identification of treatment complementarity. The following proposition formalizes this result.

Proposition 1. *Suppose Assumptions 1– 4 hold. Solving the moment condition equations for β_1, β_2, and β_c shows that each coefficient is a linear combination of all three treatment effects:*

$$\beta_1 = \mathbb{E}[w_1^1\Delta_1 + w_2^1\Delta_2 + w_c^1\Delta_c],$$
$$\beta_2 = \mathbb{E}[w_1^2\Delta_1 + w_2^2\Delta_2 + w_c^2\Delta_c],$$
$$\beta_c = \mathbb{E}[w_1^c\Delta_1 + w_2^c\Delta_2 + w_c^c\Delta_c],$$

where Δ denotes the individual-level treatment effects:

(i) Δ_1: the effect of taking up the first treatment, $Y(1,0) - Y(0,0)$,

(ii) Δ_2: the effect of taking up the second treatment, $Y(0,1) - Y(0,0)$,

(iii) Δ_c: the complementarity effect, $[Y(1,1) - Y(1,0)] - [Y(0,1) - Y(0,0)]$.

Furthermore, the own-weights satisfy $\mathbb{E}[w_1^1] = \mathbb{E}[w_2^2] = \mathbb{E}[w_c^c] = 1$, and all cross-weights satisfy $\mathbb{E}[w_2^1] = \mathbb{E}[w_c^1] = \mathbb{E}[w_1^2] = \mathbb{E}[w_c^2] = \mathbb{E}[w_1^c] = \mathbb{E}[w_2^c] = 0$. All weights depend on potential treatment choices.

Proof: See Appendix A.1.

Proposition 1 implies that with homogeneous treatment effects each 2SLS coefficient equals its target Δ. Under heterogeneity, however, individual weights can be negative or exceed one—even though their expectations are 1 (own-weights) or 0 (cross-weights). The extent of this contamination is determined by the covariance between the off-diagonal weights w_i^j and the corresponding treatment effects: because $\mathbb{E}[w_i^j] = 0$, we have $\mathbb{E}[w_i^j \Delta_i] = \text{cov}(w_i^j, \Delta_i)$. If potential choices are independent of treatment effects, contamination may not be of a concern even with some heterogeneity.

Consequently, Assumptions 1–4 are not sufficient to guarantee a causal interpretation of the 2SLS coefficients in (3). For example, β_1 generally reflects not only Δ_1 but also Δ_2 and Δ_c. To restore a causal interpretation, a stronger monotonicity assumption is needed. Bhuller and Sigstad (2024) propose such a condition (their Assumption 4) in a general multiple-treatments and instruments setting; here, I adapt it to treatment complementarity setup in a way that is more directly tied to potential treatment choices (instead of residualized predicted treatments), which helps clarify how compliance behavior drives the contamination in 2SLS.

Assumption 5. ***(No cross effects).*** *For $Z_1 \in \{0,1\}$ and $Z_2 \in \{0,1\}$,*

$$T_1(Z_1, 1) = T_1(Z_1, 0), \quad T_2(1, Z_2) = T_2(0, Z_2),$$

and

$$(T_1 \times T_2)(1,0) = (T_1 \times T_2)(0,1) = (T_1 \times T_2)(0,0).$$

The first part of Assumption 5 requires that each instrument affects only its intended treatment and not the other. The second part requires that joint take-up is encouraged only when both instruments are received; receiving a single instrument has no effect on joint take-up. The "reluctant complier" type from the previous example violates the first part of Assumption 5, since $T_1(1,1) \neq T_1(1,0)$ and $T_2(1,1) \neq T_2(0,1)$ for such individuals.

Assumption 5, together with Assumptions 1–4, ensures that each 2SLS coefficient is a convexly weighted average of its corresponding treatment effect. These restrictions, however, substantially limit permissible compliance behavior. With no restrictions, the 4 instrument combinations and 4 treatment combinations imply $4^4 = 256$ possible complier types. Imposing Assumption 4 (monotonicity) rules out discouragement patterns, reducing the set from 256 to 132 types, and Assumption 5 (No cross effects) further narrows the admissible set to just seven types. These seven types—precisely those consistent with Assumptions 4 and 5—are listed below.

Type	**Z00**	**Z10**	**Z01**	**Z11**
1. Never-taker	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$
2. T2-complier	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$
3. T1-complier	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$
4. Dutiful-complier	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 0, T_2 = 1$	$T_1 = 1, T_2 = 1$
5. Always-T1-taker	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$
6. Always-T2-taker	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$
7. Always-both-taker	$T_1 = 1, T_2 = 1$	$T_1 = 1, T_2 = 1$	$T_1 = 1, T_2 = 1$	$T_1 = 1, T_2 = 1$

Table 2: Admissible complier types under Assumptions 1–5.

Proposition 2 formalizes that valid 2SLS estimation of treatment complementarity (i.e., a causal interpretation of each coefficient in (3)) is possible if only these seven complier types are allowed, which is equivalent to imposing Assumptions 4 (monotonicity) and 5 (No cross effects).

Proposition 2. *Suppose Assumptions 1–5 hold. Then 2SLS coefficients have a causal interpretation and identify the following:*

$$\beta_1 = \mathbb{E}[\Delta_1 \mid T_1 \textit{ complier } \textbf{\textit{or}} \textit{ Dutiful complier}]$$

$$\beta_2 = \mathbb{E}[\Delta_2 \mid T_2 \textit{ complier } \textbf{\textit{or}} \textit{ Dutiful complier}]$$

$$\beta_c = \mathbb{E}[\Delta_c \mid \textit{Dutiful complier}].$$

Proof: See Appendix A.2.

Proposition 2 highlights that even after eliminating contamination bias, the 2SLS coefficients pertain to *different* complier groups. This complicates common practices of summing or comparing coefficients. For instance, the "combined effect," $\mathbb{E}[\Delta_1+\Delta_2+\Delta_c] = \mathbb{E}[Y(1,1)-Y(0,0)]$, is often reported in the complementarity literature, but the sum $\beta_1 + \beta_2 + \beta_c$ aggregates effects across distinct complier populations and therefore does not equal the combined effect under heterogeneous treatment effects. Additional restrictions on complier types are thus required—namely, ruling out the T_2-complier and T_1-complier types in Table 2.

Type	**Z00**	**Z10**	**Z01**	**Z11**
1. Never-taker	$T_1 = 0$, $T_2 = 0$	$T_1 = 0$, $T_2 = 0$	$T_1 = 0$, $T_2 = 0$	$T_1 = 0$, $T_2 = 0$
2. Dutiful-complier	$T_1 = 0$, $T_2 = 0$	$T_1 = 1$, $T_2 = 0$	$T_1 = 0$, $T_2 = 1$	$T_1 = 1$, $T_2 = 1$
3. Always-T1-taker	$T_1 = 1$, $T_2 = 0$	$T_1 = 1$, $T_2 = 0$	$T_1 = 1$, $T_2 = 0$	$T_1 = 1$, $T_2 = 0$
4. Always-T2-taker	$T_1 = 0$, $T_2 = 1$	$T_1 = 0$, $T_2 = 1$	$T_1 = 0$, $T_2 = 1$	$T_1 = 0$, $T_2 = 1$
5. Always-both-taker	$T_1 = 1$, $T_2 = 1$	$T_1 = 1$, $T_2 = 1$	$T_1 = 1$, $T_2 = 1$	$T_1 = 1$, $T_2 = 1$

Table 3: Admissible complier types under Assumptions 1–5, and to ensure 2SLS coefficients reflect the same complier group.

Corollary 1. *Suppose Assumptions 1–5 hold and we further rule out T2-complier and T1 complier types. Then 2SLS coefficients have a causal interpretation and all coefficients reflect the same underlying complier group, and:*

$$\beta_1 = \mathbb{E}[\Delta_1 \mid \textit{Dutiful complier}]$$
$$\beta_2 = \mathbb{E}[\Delta_2 \mid \textit{Dutiful complier}]$$
$$\beta_c = \mathbb{E}[\Delta_c \mid \textit{Dutiful complier}].$$

The same logic applies when estimating treatment complementarity via the mutually exclusive treatment-arm specification common in experimental studies:

$$Y = \beta_0' + \beta_1' \cdot \mathbf{1}[T_1 = 1, T_2 = 0] + \beta_2' \cdot \mathbf{1}[T_1 = 0, T_2 = 1] + \beta_3' \cdot \mathbf{1}[T_1 = 1, T_2 = 1] + \epsilon, \qquad (5)$$

and testing $\beta_1' + \beta_2' = \beta_3'$ for complementarity. Corollary 2 shows that the same assumptions are required for this specification to identify the causal estimands of interest.

Corollary 2. *Suppose Assumptions 1–5 hold and we further rule out T2-complier and T1 complier types. Then the 2SLS coefficients in (5) have a causal interpretation, all coefficients reflect the same underlying complier group, and:*

$$\beta_1' = \mathbb{E}[\Delta_1 \mid \textit{Dutiful complier}]$$
$$\beta_2' = \mathbb{E}[\Delta_2 \mid \textit{Dutiful complier}]$$
$$\beta_3' = \mathbb{E}[\Delta_1 + \Delta_2 + \Delta_c \mid \textit{Dutiful complier}]$$
$$\beta_3' - \beta_2' - \beta_1' = \mathbb{E}[\Delta_c \mid \textit{Dutiful complier}]$$

3.2 Reduced Form Estimation

Researchers often regress outcomes directly on the randomized instruments, which is commonly referred to as reduced-form estimation. However, without restrictions on the underlying complier types, reduced-form estimates can provide misleading evidence of treatment complementarity, as the simulation example in the previous section illustrates. The next proposition formalizes this point.

Proposition 3. *Suppose Assumptions 1–4 hold (as in Proposition 1). The reduced-form estimates from*

$$Y = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_c (Z_1 \times Z_2) + \epsilon \tag{6}$$

are linear combinations of all three treatment effects:

$$\begin{aligned} \gamma_1 &= \mathbb{E}[\delta_1^1 \Delta_1 + \delta_2^1 \Delta_2 + \delta_c^1 \Delta_c] \\ \gamma_2 &= \mathbb{E}[\delta_1^2 \Delta_1 + \delta_2^2 \Delta_2 + \delta_c^2 \Delta_c] \\ \gamma_c &= \mathbb{E}[\delta_1^c \Delta_1 + \delta_2^c \Delta_2 + \delta_c^c \Delta_c]. \end{aligned}$$

Proof: See Appendix A.3.

These expressions show that each reduced-form coefficient is a linear combination of the three individual-level effects $(\Delta_1, \Delta_2, \Delta_c)$ with weights determined by potential treatment choices. Therefore, reduced-form estimates may not be informative if researchers' interests are in Δ. Additional restrictions on potential choices—such as Assumption 5 (No cross effects)—are required to eliminate other Δ terms.

Next, I state the reduced-form analogue of Proposition 2.

Proposition 4. *Suppose Assumptions 1–5 hold. Then the reduced-form coefficients identify:*

$$\begin{aligned} \gamma_1 &= \Pr(T_1\textit{-complier or Dutiful complier}) \cdot \mathbb{E}[\Delta_1 \mid T_1\textit{-complier or Dutiful complier}], \\ \gamma_2 &= \Pr(T_2\textit{-complier or Dutiful complier}) \cdot \mathbb{E}[\Delta_2 \mid T_2\textit{-complier or Dutiful complier}], \\ \gamma_c &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_c \mid \textit{Dutiful complier}]. \end{aligned}$$

Proof: See Appendix A.4.

Proposition 4 shows that the common practice of combining or comparing reduced-form coefficients can be misleading even when seven complier types are assumed: each coefficient pertains to different compliers and is multiplied by the corresponding complier probability. Consequently, coefficient comparisons do not distinguish between variation in the treatment effects (Δ) and differences in the composition of complier types ($\Pr(\text{type})$). To facilitate meaningful comparisons or combinations of coefficients, a stronger assumption is required.

Corollary 3. *Suppose Assumptions 1–5 hold and we further rule out T2-complier and T1 complier types. Then all reduced-form coefficients reflect the same underlying complier group, and identify:*

$$\begin{aligned}
\gamma_1 &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_1 \mid \textit{Dutiful complier}], \\
\gamma_2 &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_2 \mid \textit{Dutiful complier}], \\
\gamma_c &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_c \mid \textit{Dutiful complier}].
\end{aligned}$$

The same logic applies to the mutually exclusive treatment-arm specification common in experimental studies:

$$Y = \gamma_0' + \gamma_1' \cdot \mathbf{1}[Z_1 = 1, Z_2 = 0] + \gamma_2' \cdot \mathbf{1}[Z_1 = 0, Z_2 = 1] + \gamma_3' \cdot \mathbf{1}[Z_1 = 1, Z_2 = 1] + \epsilon, \qquad (7)$$

where researchers test $\gamma_1' + \gamma_2' = \gamma_3'$ for complementarity. The same assumptions are required for this specification to identify the causal estimand of interest.

Corollary 4. *Suppose Assumptions 1–5 hold and we further rule out T2-complier and T1 complier types. Then the reduced-form coefficients in (7) reflect the same underlying complier group and identify:*

$$\begin{aligned}
\gamma_1' &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_1 \mid \textit{Dutiful complier}] \\
\gamma_2' &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_2 \mid \textit{Dutiful complier}] \\
\gamma_3' &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_1 + \Delta_2 + \Delta_c \mid \textit{Dutiful complier}] \\
\gamma_3' - \gamma_2' - \gamma_1' &= \Pr(\textit{Dutiful complier}) \cdot \mathbb{E}[\Delta_c \mid \textit{Dutiful complier}]
\end{aligned}$$

3.3 First stage regressions

The restrictions on complier behaviors have testable implications for the first-stage regressions. The first-stage regressions corresponding to the 2SLS estimation of equation (3) are:

$$\begin{aligned} T_1 &= \alpha_1^0 + \alpha_1^1 Z_1 + \alpha_1^2 Z_2 + \alpha_1^c (Z_1 \times Z_2) + \eta_1 \\ T_2 &= \alpha_2^0 + \alpha_2^1 Z_1 + \alpha_2^2 Z_2 + \alpha_2^c (Z_1 \times Z_2) + \eta_2 \\ (T_1 \times T_2) &= \alpha_c^0 + \alpha_c^1 Z_1 + \alpha_c^2 Z_2 + \alpha_c^c (Z_1 \times Z_2) + \eta_c \end{aligned}$$

If Assumptions 1–5 hold, ensuring that each 2SLS coefficient has a causal interpretation, then each first-stage coefficient corresponds to the proportion of compliers affected by the relevant instrument, as shown below.

Proposition 5. *Suppose Assumptions 1–5 hold. Then:*

$$\alpha_1^1 = \Pr(T_1\text{-}\textit{complier or Dutiful complier}), \alpha_1^2 = 0, \alpha_1^c = 0;$$

$$\alpha_2^1 = 0, \alpha_2^2 = \Pr(T_2\text{-}\textit{complier or Dutiful complier}), \alpha_2^c = 0;$$

$$\alpha_c^1 = 0, \alpha_c^2 = 0, \alpha_c^c = \Pr(\textit{Dutiful complier}).$$

Proof: See Appendix A.5.

Proposition 5 implies that, under Assumptions 1–5, all cross-coefficients in the first-stage regressions must be zero. Any nonzero estimate of α_1^2, α_2^1, α_1^c, α_2^c, α_c^1, or α_c^2 therefore provides direct evidence that Assumption 5 (No cross effects) is violated.

Proposition 6. *Suppose Assumptions 1–5 hold. For the 2SLS coefficients to have a causal interpretation and reflect the same set of compliers, each first-stage coefficient must correspond to the following complier proportions:*

$$\alpha_1^1 = \Pr(\textit{Dutiful complier}), \alpha_1^2 = 0, \alpha_1^c = 0;$$

$$\alpha_2^1 = 0, \alpha_2^2 = \Pr(\textit{Dutiful complier}), \alpha_2^c = 0;$$

$$\alpha_c^1 = 0, \alpha_c^2 = 0, \alpha_c^c = \Pr(\textit{Dutiful complier}).$$

Proof: See Appendix A.5.

Proposition 6 implies that, when the 2SLS coefficients are to be interpreted as causal effects for the same complier population, two conditions must hold: (i) all cross-coefficients in the first-stage regressions must be zero, and (ii) the relevant coefficients must be identical across the three first-stage equations. That is, the instrument's effect on T_1, T_2, and $(T_1 \times T_2)$

must be driven by the same complier group—Dutiful compliers.

In practice, these stringent implications are rarely satisfied. For instance, Angrist, Lang and Oreopoulos (2009) examine an experiment that randomly assigns students to two educational inputs. Replicating their first-stage regressions reveals that Assumption 5 (No cross effects) is not supported by the data. Another example by Angelucci and Bennett (2024), who cross-randomize psychiatric treatment and economic assistance, find first-stage patterns consistent with Assumption 5, but not with the additional restrictions required for each coefficient to reflect a common underlying complier group. Consequently, comparing 2SLS coefficients (or reduced-form coefficients) may yield misleading evidence of treatment complementarity when treatment effects are heterogeneous.

	(1)	(2)	(3)
	T1	T2	T1T2
Z1	0.60**	0.00	-0.00
	(0.02)	(0.02)	(0.01)
Z2	0.00	0.91**	-0.00
	(0.02)	(0.02)	(0.01)
Z1Z2	0.23**	-0.08**	0.84**
	(0.04)	(0.03)	(0.02)
_cons	-0.00	-0.00	0.00
	(0.01)	(0.01)	(0.00)
N	837	837	837

Standard errors in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

(a) Angrist, Lang and Oreopoulos (2009)

	(1)	(2)	(3)
	T1	T2	T1T2
Z1	0.46**	0.00	-0.00
	(0.03)	(0.03)	(0.02)
Z2	0.00	0.71**	-0.00
	(0.03)	(0.03)	(0.02)
Z1Z2	-0.03	-0.05	0.31**
	(0.04)	(0.04)	(0.03)
_cons	-0.00	-0.00	0.00
	(0.02)	(0.01)	(0.01)
N	1000	1000	1000

Standard errors in parentheses
+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$

(b) Angelucci and Bennett (2024)

Table 4: First-stage estimates from two cross-randomized experimental studies.

4 Solution

In this section, I propose an alternative estimation strategy for settings in which Assumptions 1 through 4 hold but Assumption 5 is unlikely to be satisfied, while allowing for a limited form of treatment effect heterogeneity. I extend a linear extrapolation approach in the marginal treatment effects (MTE) framework (Brinch, Mogstad and Wiswall, 2017; Kowalski, 2023*a*,*b*) to a setting with two jointly chosen treatments. Using simulated data, I show that this method accommodates a richer set of complier types and performs better than two stage least squares (2SLS), while remaining more interpretable. Its limitations are that it still requires some restrictions on choice behavior and it assumes linearity in the potential outcome function.

4.1 Assumptions on Choice model and Potential Outcomes

Let

$$
\begin{aligned}
V_1 &= \mu_1(Z_1, Z_2) - \varepsilon_1, \\
V_2 &= \mu_2(Z_1, Z_2) - \varepsilon_2,
\end{aligned}
$$

where V_1 is the latent utility of taking the first treatment and V_2 is the latent utility of taking the second treatment. Normalize the utility of taking neither treatment to zero, $V_0 = 0$. Assume separability between the observed components μ_1, μ_2 (suppressing observable covariates X) and the unobserved resistance to treatments $\varepsilon_1, \varepsilon_2$, as is standard in the MTE literature (Carneiro and Lee, 2009; Kowalski, 2023*a*,*b*). Assume further that the utility of taking both treatments is separable in V_1 and V_2, which yields the selection rule:

$$
\begin{aligned}
\mathbb{1}\{T_1 = 0, T_2 = 0\} &= \mathbb{1}\{V_1 < 0,\, V_2 < 0\}, \\
\mathbb{1}\{T_1 = 1, T_2 = 0\} &= \mathbb{1}\{V_1 > 0,\, V_2 < 0\}, \\
\mathbb{1}\{T_1 = 0, T_2 = 1\} &= \mathbb{1}\{V_1 < 0,\, V_2 > 0\}, \\
\mathbb{1}\{T_1 = 1, T_2 = 1\} &= \mathbb{1}\{V_1 > 0,\, V_2 > 0\}.
\end{aligned}
$$

I now introduce two additional assumptions on the choice model that facilitate the linear extrapolation approach, provided that the standard IV Assumptions 1 through 4 and the choice rules outlined above hold.

Assumption 6. *$\mu_1(Z_1, Z_2)$ is weakly increasing in Z_1 and in Z_1Z_2, and $\mu_2(Z_1, Z_2)$ is weakly increasing in Z_2 and in Z_1Z_2.*

Assumption 7. *$\varepsilon_1 \perp \varepsilon_2$.*

Assumptions 6 and 7 imply a richer, 16-type compliers (See Table 5), including the "reluctant compliers" from the previous simulation section. This expands the set of admissible choice behaviors relative to Assumption 5, which was required to give 2SLS a causal interpretation (See Table 2 for comparison). Therefore, the linear extrapolation approach can accommodate behaviors that violate Assumption 5 (No cross effects), at the cost of imposing functional structure on both the choice process and the potential outcome function.

Type	**Z00**	**Z10**	**Z01**	**Z11**
1. Never-taker	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$
2. [Type 2]	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 1$
3. [Type 3]	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$
4. Reluctant-complier	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 1$
5. T2-complier	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$
6. [Type 6]	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 0, T_2 = 1$	$T_1 = 1, T_2 = 1$
7. T1-complier	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$
8. [Type 8]	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 1$
9. Dutiful-complier	$T_1 = 0, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 0, T_2 = 1$	$T_1 = 1, T_2 = 1$
10. Always-T1-taker	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$
11. [Type 11]	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 1$
12. [Type 12]	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 0$	$T_1 = 1, T_2 = 1$	$T_1 = 1, T_2 = 1$
13. Always-T2-taker	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$
14. [Type 14]	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$	$T_1 = 0, T_2 = 1$	$T_1 = 1, T_2 = 1$
15. [Type 15]	$T_1 = 0, T_2 = 1$	$T_1 = 1, T_2 = 1$	$T_1 = 0, T_2 = 1$	$T_1 = 1, T_2 = 1$
16. Always-both-taker	$T_1 = 1, T_2 = 1$	$T_1 = 1, T_2 = 1$	$T_1 = 1, T_2 = 1$	$T_1 = 1, T_2 = 1$

Table 5: Types of compliers admissible under Linear Extrapolation Approach

Define $U_j = F_{\varepsilon_j}(\varepsilon_j)$ for $j \in \{1, 2\}$, so that $U_1, U_2 \sim \text{Unif}[0, 1]$. Interpret U_j as a normalized unobserved cost of treatment j. Next, I impose a linearity assumption on potential outcome heterogeneity, and therefore permit limited pattern of treatment effect heterogeneity.

Assumption 8. ***(linearity in potential outcomes).***

$$\mathbb{E}[Y(0,0) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{00} + \tilde{\beta}_{00} p_1 + \tilde{\gamma}_{00} p_2,$$
$$\mathbb{E}[Y(1,0) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{10} + \tilde{\beta}_{10} p_1 + \tilde{\gamma}_{10} p_2,$$
$$\mathbb{E}[Y(0,1) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{01} + \tilde{\beta}_{01} p_1 + \tilde{\gamma}_{01} p_2,$$
$$\mathbb{E}[Y(1,1) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{11} + \tilde{\beta}_{11} p_1 + \tilde{\gamma}_{11} p_2.$$

Assumption 8 allows for heterogeneous treatment effects that vary linearly in (U_1, U_2). It nests homogeneous treatment effects as special case, when all slopes are zero.

4.2 Estimation Steps

The estimation follows a standard MTE procedure, extended to joint choice over two treatments. First estimate the propensity scores $\hat{p}_1, \hat{p}_2$ by estimating the following with linear regression and predicting the treatment take-up.

$$
\begin{aligned}
T_1 &= \theta_{10} + \theta_{11} Z_1 + \theta_{12} Z_2 + \theta_{13} Z_1 Z_2, \\
T_2 &= \theta_{20} + \theta_{21} Z_1 + \theta_{22} Z_2 + \theta_{23} Z_1 Z_2.
\end{aligned}
$$

Next, define the average outcome conditional on the treatment take-up as

$$
\mathrm{AO}(t_1, t_2) \equiv \mathbb{E}[Y \mid T_1 = t_1,\ T_2 = t_2].
$$

Under the assumptions on the choice model, Assumption 8, and $U_j \sim \mathrm{Unif}[0,1]$, the following equalities hold:

$$
\begin{aligned}
\mathrm{AO}(0,0) &= \mathbb{E}[Y(0,0) \mid p_1 \le U_1 \le 1,\ p_2 \le U_2 \le 1] = \alpha_{00} + \beta_{00} p_1 + \gamma_{00} p_2, \\
\mathrm{AO}(1,0) &= \mathbb{E}[Y(1,0) \mid 0 \le U_1 \le p_1,\ p_2 \le U_2 \le 1] = \alpha_{10} + \beta_{10} p_1 + \gamma_{10} p_2, \\
\mathrm{AO}(0,1) &= \mathbb{E}[Y(0,1) \mid p_1 \le U_1 \le 1,\ 0 \le U_2 \le p_2] = \alpha_{01} + \beta_{01} p_1 + \gamma_{01} p_2, \\
\mathrm{AO}(1,1) &= \mathbb{E}[Y(1,1) \mid 0 \le U_1 \le p_1,\ 0 \le U_2 \le p_2] = \alpha_{11} + \beta_{11} p_1 + \gamma_{11} p_2.
\end{aligned}
$$

Then the parameters $(\alpha_{tt'}, \beta_{tt'}, \gamma_{tt'})$ can be estimated by regressing Y on $\hat{p}_1$ and $\hat{p}_2$ using observations in each treatment pair $(T_1 = t_1, T_2 = t_2)$. The following relationships map the estimated $(\alpha_{tt'}, \beta_{tt'}, \gamma_{tt'})$ to the parameters in the linear potential outcome functions:

$$
\begin{aligned}
\frac{\partial^2\{(1-p_1)(1-p_2)\mathrm{AO}(0,0)\}}{\partial(1-p_1)\,\partial(1-p_2)} &= \mathbb{E}[Y(0,0) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{00} + \tilde{\beta}_{00} p_1 + \tilde{\gamma}_{00} p_2, \\
\frac{\partial^2\{p_1(1-p_2)\mathrm{AO}(1,0)\}}{\partial p_1\,\partial(1-p_2)} &= \mathbb{E}[Y(1,0) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{10} + \tilde{\beta}_{10} p_1 + \tilde{\gamma}_{10} p_2, \\
\frac{\partial^2\{(1-p_1)p_2\mathrm{AO}(0,1)\}}{\partial(1-p_1)\,\partial p_2} &= \mathbb{E}[Y(0,1) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{01} + \tilde{\beta}_{01} p_1 + \tilde{\gamma}_{01} p_2, \\
\frac{\partial^2\{p_1 p_2\mathrm{AO}(1,1)\}}{\partial p_1\,\partial p_2} &= \mathbb{E}[Y(1,1) \mid U_1 = p_1, U_2 = p_2] = \tilde{\alpha}_{11} + \tilde{\beta}_{11} p_1 + \tilde{\gamma}_{11} p_2.
\end{aligned}
$$

Evaluating these derivatives yields the mapping from (α, β, γ) to $(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma})$:

$$\begin{aligned}
&\tilde{\alpha}_{00} = \alpha_{00} - \beta_{00} - \gamma_{00}, && \tilde{\beta}_{00} = 2\beta_{00}, && \tilde{\gamma}_{00} = 2\gamma_{00},\\
&\tilde{\alpha}_{10} = \alpha_{10} - \gamma_{10}, && \tilde{\beta}_{10} = 2\beta_{10}, && \tilde{\gamma}_{10} = 2\gamma_{10},\\
&\tilde{\alpha}_{01} = \alpha_{01} - \beta_{01}, && \tilde{\beta}_{01} = 2\beta_{01}, && \tilde{\gamma}_{01} = 2\gamma_{01},\\
&\tilde{\alpha}_{11} = \alpha_{11}, && \tilde{\beta}_{11} = 2\beta_{11}, && \tilde{\gamma}_{11} = 2\gamma_{11}.
\end{aligned}$$

Finally, the marginal treatment effects $(\mathrm{MTE}_1, \mathrm{MTE}_2)$, the marginal combined effect (MTE_3), and marginal treatment complementarity (MTE_c) are derived by taking differences between these linear potential outcomes:

$$\begin{aligned}
\mathrm{MTE}_1(p_1,p_2) &\equiv \mathbb{E}[Y(1,0) \mid U_1 = p_1, U_2 = p_2] - \mathbb{E}[Y(0,0) \mid U_1 = p_1, U_2 = p_2],\\
\mathrm{MTE}_2(p_1,p_2) &\equiv \mathbb{E}[Y(0,1) \mid U_1 = p_1, U_2 = p_2] - \mathbb{E}[Y(0,0) \mid U_1 = p_1, U_2 = p_2],\\
\mathrm{MTE}_3(p_1,p_2) &\equiv \mathbb{E}[Y(1,1) \mid U_1 = p_1, U_2 = p_2] - \mathbb{E}[Y(0,0) \mid U_1 = p_1, U_2 = p_2],\\
\mathrm{MTE}_c(p_1,p_2) &\equiv \mathbb{E}[Y(1,1) \mid U_1 = p_1, U_2 = p_2] - \mathbb{E}[Y(1,0) \mid U_1 = p_1, U_2 = p_2]\\
&\qquad - \mathbb{E}[Y(0,1) \mid U_1 = p_1, U_2 = p_2] + \mathbb{E}[Y(0,0) \mid U_1 = p_1, U_2 = p_2].
\end{aligned}$$

Average treatment effects follow by integrating these objects over the desired range of (U_1, U_2). To obtain an average treatment effect for receiving the first treatment, integrate $\mathrm{MTE}_1(p_1, p_2)$ over $[0,1] \times [0,1]$. To obtain an average treatment complementarity effect for the full population, integrate $\mathrm{MTE}_\mathrm{c}(p_1, p_2)$ over $[0,1] \times [0,1]$.

4.3 Simulation

This subsection evaluates how the proposed method performs relative to 2SLS and the reduced form using simulated data. I consider two latent subpopulations—Group 1 and Group 2—that are unobserved to the econometrician and differ in treatment effects. The population is split evenly between the two groups. In both groups, treatment take-up follows the same choice model:

$$\begin{aligned}
T_1 &= \mathbb{1}\{Z_1 + Z_1 Z_2 - \varepsilon_1 > 0.5\},\\
T_2 &= \mathbb{1}\{Z_2 + Z_1 Z_2 - \varepsilon_2 > 0.5\}.
\end{aligned}$$

Group 1 has a higher mean for ε_1 (higher resistance) and is therefore less likely to take up T_1. For Group 1, $\varepsilon_1 \sim \mathcal{N}(1,1)$; for Group 2, $\varepsilon_1 \sim \mathcal{N}(0,1)$. On the other hand, $\varepsilon_2 \sim \mathcal{N}(0,1)$

for both groups. Outcomes are determined by the following data generation process:

$$\begin{aligned} Y &= 1 \cdot T_1 + 2 \cdot T_2 + 0 \cdot (T_1 \times T_2) + \eta_Y \quad \text{for group 1,} \\ Y &= 2 \cdot T_1 + 2 \cdot T_2 + 0 \cdot (T_1 \times T_2) + \eta_Y \quad \text{for group 2.} \end{aligned}$$

where $\eta_Y \sim \mathcal{N}(0,1)$ for both groups. This data generating process reflects a Roy model argument: Group 2 has a higher treatment effect for the first treatment and therefore a lower resistance to select it. I simulate $N = 100{,}000$ observations for 1,000 replications and report the mean estimate with the standard deviations across replications in parentheses. I compare 2SLS, the reduced form, and the linear extrapolation method.

Table 6: Simulation results

	Truth	2SLS	Reduced Form	Linear Extrapolation
First treatment effect	1.5	1.670 (0.047)	0.504 (0.014)	1.481 (0.045)
Second treatment effect	2.0	2.035 (0.033)	0.766 (0.014)	2.000 (0.027)
Complementarity	0.0	-0.183 (0.063)	0.916 (0.019)	-0.018 (0.049)
Constant		0.028 (0.020)	0.959 (0.010)	

The figure below illustrates how the linear extrapolation method estimates the potential outcome function. Based on the data generating process described above, I plot the true average potential outcome for each of the sixteen complier types in the left panel. I plot four different planes—one plane for potential outcome of each of the four treatment statuses. These appear as step functions, since each complier type includes varying proportion of individuals from Group 2. The linear extrapolation approach approximates these potential outcomes with a plane, using observed outcomes across combinations of treatment take-up and instrument assignments. Treatment effects conditional on a given resistance level can then be obtained by taking differences in heights between the treated planes and the untreated plane at the bottom, while treatment complementarity can be obtained by taking differences across all four planes. Average treatment effects, reported in Table 6, follow by integrating these differences over the support of $[0,1] \times [0,1]$.

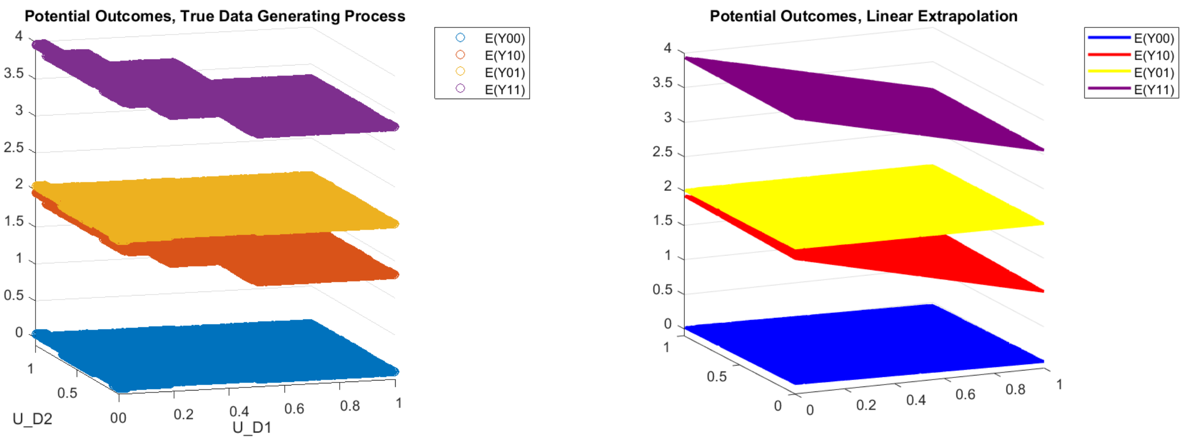

Figure 1: True potential outcomes (left) and extrapolated potential outcomes (right)

4.4 2SLS—Diagnostics and Practical Measures

Depending on the empirical setting, researchers may reasonably doubt that the assumptions required for the linear extrapolation method hold. When these assumptions are violated, the linear extrapolation method offers no guarantee of outperforming 2SLS.

In such cases, 2SLS remains the most accessible and widely used estimator. However, it becomes essential to demonstrate why the resulting estimates can still be considered credible and why potential bias from treatment effect heterogeneity is unlikely to be severe. Researchers can draw on the first-stage diagnostics presented in Section 3.3 to assess the possibility of contamination bias in 2SLS, and complement this with the diagnostics introduced in this section.

This subsection provides two complementary diagnostics for that purpose. The first examines how to detect unobserved heterogeneity in potential outcomes—the central threat to 2SLS validity—by comparing outcome means among individuals with identical treatment take-up but different instrument assignments. The second explores how covariates correlated with treatment effect heterogeneity can be leveraged to reduce bias in 2SLS estimation. This part echoes Blandhol et al. (2022) that the regression specification should be fully saturated.

4.4.1 Testing Potential Outcome Heterogeneity

A key threat to the validity of 2SLS is the presence of unobserved treatment effect heterogeneity. When treatment effects differ across complier types, 2SLS may give biased estimates depending on the composition of compliers, as discussed in Proposition 1. Detecting such heterogeneity can help researchers to assess whether 2SLS provides a credible estimate.

A simple diagnostic exploits the implication that, under homogeneous potential outcomes, individuals with the same realized treatment status (T_1, T_2) should have identical expected outcomes regardless of their instrument assignment (Z_1, Z_2). Any systematic difference in mean outcomes across instrument values—within the same treatment cell—implies that potential outcomes vary across complier groups, indicating unobserved heterogeneity in treatment effects. For each treatment combination (t_1, t_2), researchers can test

$$H_0 : \mathbb{E}[Y \mid T_1 = t_1,\, T_2 = t_2,\, Z_1 = z_1,\, Z_2 = z_2] = \mathbb{E}[Y \mid T_1 = t_1,\, T_2 = t_2,\, Z_1 = z_1',\, Z_2 = z_2']$$

for all (z_1, z_2) and (z_1', z_2'). Rejecting H_0 indicates the presence of unobserved heterogeneity in potential outcomes, and hence in treatment effects.

For instance, if the empirical setting admits the sixteen complier types in Table 5, testing whether

$$H_0 : \mathbb{E}[Y \mid T_1 = 1,\, T_2 = 0,\, Z_1 = 1,\, Z_2 = 0] = \mathbb{E}[Y \mid T_1 = 1,\, T_2 = 0,\, Z_1 = 0,\, Z_2 = 0]$$

is equivalent to comparing

$$H_0 : \mathbb{E}[Y(1,0) \mid \text{Complier type} \in \{7,8,9,10,11,12\}] = \mathbb{E}[Y(1,0) \mid \text{Complier type} \in \{10,11,12\}].$$

A rejection of this equality suggests that potential outcome Y(1,0) (and therefore treatment effects) differ across complier types, undermining the validity of 2SLS estimates.

4.4.2 Controlling for Covariates

If part of the treatment effect heterogeneity can be explained by observed characteristics X, conditioning on X may reduce bias in 2SLS estimates. However, even when X captures some of the variation in treatment effects, a simple linear inclusion (without saturating the specification) does little to improve 2SLS performance, consistent with findings in Blandhol et al. (2022). To illustrate this, I use the same data-generating process as in Section 4.3.

I introduce a binary covariate X taking value 1 for 30% of individuals in Group 1 and 70% in Group 2. This covariate is correlated with both complier type—captured by different means of ε across groups in the T_1 choice equation—and underlying treatment effect heterogeneity.

I estimate two 2SLS specifications. The first specification linearly includes X as an exogenous control variable, using Z_1, Z_2, and $(Z_1 \times Z_2)$ as instruments:

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_c (T_1 \times T_2) + X + \varepsilon.$$

The second specification fully interacts the treatment variables with X, allowing treatment effects to vary by the covariate. It uses Z_1, Z_2, $(Z_1 \times Z_2)$, $Z_1 \cdot X$, $Z_2 \cdot X$, and $(Z_1 \times Z_2) \cdot X$ as instruments:

$$Y = \beta_0 + (\beta_1 + \beta_1' X) T_1 + (\beta_2 + \beta_2' X) T_2 + (\beta_c + \beta_c' X)(T_1 \times T_2) + X + \varepsilon,$$

I simulate $N = 100{,}000$ observations for 1,000 replications to examine how controlling for X—linearly or through saturation—affects 2SLS estimates. Table 7 reports mean estimates with standard deviations in parentheses.

Table 7: Simulation results

	Truth	2SLS-not using X	2SLS-linear inclusion of X	2SLS-saturated
β_1	1.3	1.66629 (0.048)	1.66636 (0.048)	1.45010 (0.079)
β_2	2.0	2.03193 (0.035)	2.03188 (0.035)	2.01968(0.048)
β_c	0.0	-0.17749 (0.066)	-0.17751 (0.066)	-0.15904 (0.102)
$\beta_1 + \beta_1'$	1.7			1.83018 (0.061)
$\beta_2 + \beta_2'$	2.0			2.03268 (0.052)
$\beta_c + \beta_c'$	0.0			-0.13995 (0.086)

The results indicate that although 2SLS still fails to recover the true estimand for treatment complementarity, the bias diminishes when the covariate X captures relevant variation in treatment effect heterogeneity. This improvement, however, occurs only under a fully saturated specification; linear inclusion of X in the 2SLS estimation does not reduce bias.

5 Empirical Application

I apply the framework developed in the previous sections to the randomized controlled trial of Angelucci and Bennett (2024). The study cross-randomizes pharmacotherapy and livelihoods assistance among about 1,000 adults with depression, finding that the combined treatment substantially reduces depression severity, whereas pharmacotherapy alone has a weaker and less persistent effect.

$$\begin{aligned} Y_{ijt} = {} & \beta_1 (PC_j \cdot D_t) + \beta_2 (LA_j \cdot D_t) + \beta_3 (PC/LA_j \cdot D_t) \\ & + \beta_4 (PC_j \cdot A_t) + \beta_5 (LA_j \cdot A_t) + \beta_6 (PC/LA_j \cdot A_t) + X_{ijt}' \beta_7 + \epsilon_{ijt}, \end{aligned} \tag{8}$$

The authors estimate a reduced-form regression (8), where Y_{ijt} denotes standardized measure of depression severity (PHQ-9), PC_j denotes assignment to pharmacotherapy only, LA_j denotes assignment to livelihoods assistance only, and PC/LA_j denotes assignment to both treatments. The variables D_t and A_t indicate the "during" and "after" phases of the intervention, respectively. The vector X_{ijt} includes stratification indicators (as randomization was stratified by district and terciles of a locality socioeconomic index) and the baseline outcome of depression severity. Standard errors are clustered at the village level.

This reduced-form specification is valid for studying *instrument complementarity*—how assignment to combined treatments affects outcomes—but it can conflate differences in compliance with true interaction effects in treatment responses. The study reports substantial noncompliance: forty-five percent of participants complied with pharmacotherapy and sixty-eight percent with livelihoods assistance. With the unequal compliance and the first-stage patterns reported in Table 4, tests such as $H_0 : \beta_1 + \beta_2 = \beta_3$ (during phase) and $H_0 : \beta_4 + \beta_5 = \beta_6$ (after phase) can be challenging to interpret.

This section uses the replication data from Angelucci and Bennett (2024) to examine whether the data can be pushed further to estimate treatment complementarity itself ($H_0 : \mathbb{E}[\Delta_1]+\mathbb{E}[\Delta_2] = \mathbb{E}[\Delta_3]$), rather than relying solely on reduced-form evidence. A natural extension is to estimate 2SLS, which is not reported in the original study. However, comparing 2SLS coefficients is generally invalid for learning about complementarity if underlying treatment effect heterogeneity exists, as shown in Section 3.1. As discussed in Proposition 6, different coefficients in the first-stage regressions of this study imply that each 2SLS coefficient averages over distinct complier groups. Consequently, differences across 2SLS coefficients may reflect variation in complier composition and treatment effect heterogeneity rather than genuine complementarity.

To assess whether such heterogeneity in treatment effects is present and potentially biases the 2SLS estimates, I apply the diagnostic from Section 4.4.1 to the replication data. The diagnostic compares mean outcomes among individuals with the same realized treatment status but different instrument assignments, thereby testing for potential outcome heterogeneity across complier groups. Specifically, I compare $\mathbb{E}[Y \mid T_1 = 1, T_2 = 0, Z_1 = 1, Z_2 = 0]$ with $\mathbb{E}[Y \mid T_1 = 1, T_2 = 0, Z_1 = 1, Z_2 = 1]$, and, by the same logic, $\mathbb{E}[Y \mid T_1 = 0, T_2 = 1, Z_1 = 0, Z_2 = 1]$ with $\mathbb{E}[Y \mid T_1 = 0, T_2 = 1, Z_1 = 1, Z_2 = 1]$. Pooling during and after periods, equality of means is rejected at the five percent level in both comparisons. When analyzed separately, equality is not rejected during the intervention but is rejected afterward. This pattern indicates potential-outcome heterogeneity across complier types and cautions against interpreting 2SLS coefficient differences as evidence of treatment complementarity.

Given evidence from both the first-stage regressions and the diagnostic tests, I next implement the linear extrapolation method from Section 4 to study whether there is evidence of interaction in treatment effects. Two challenges complicate this application. The first is *one-sided compliance*, which is typical in randomized trials. Among those who take both treatments, there is no instrument-driven variation in propensities because the only individuals who take both are those assigned to both; the design rules out always-takers, limiting support for $(\hat{p}_1, \hat{p}_2)$ among those receiving both treatments. The second challenge is the small sample size, which limits statistical power.

To address the first challenge and increase variation in propensity scores, I proceed in two steps. First, I estimate treatment propensities using interactions between the instruments and baseline covariates X_{ijt}, including baseline depression severity and all stratification indicators. Second, in the regressions of outcomes on the estimated propensities within each treatment cell, I include only baseline depression as a covariate. This strategy expands variation in the estimated propensities sufficiently to identify the planes of potential outcomes, and it remains valid under the assumption that the excluded covariates (the stratification indicators) affect outcomes only through treatment.

For comparison, I report estimates for the during period. Table 8 presents the reduced-form, 2SLS, and linear extrapolation estimates computed from the replication data. Standard errors for the linear extrapolation are obtained by clustered bootstrap with 1,000 replications.

Table 8: Empirical application: Angelucci and Bennett (2024)
During period only

	Reduced Form	Two Stage Least Squares	Linear Extrapolation
First treatment (PC) effect	-0.151 (0.081)	-0.317 (0.173)	0.233 (0.390)
Second treatment (LA) effect	-0.082 (0.085)	-0.116 (0.120)	-0.143 (0.123)
Complementarity	-0.027 (0.125)	-0.141 (0.378)	-0.427 (1.123)

The linear extrapolation estimates exhibit large standard errors, limiting statistical significance. Nevertheless, their pattern differs notably from both the reduced-form and 2SLS results. During the intervention, the reduced-form estimates suggest little complementarity, and the 2SLS estimates indicate a smaller interaction effect than the main pharmacotherapy effect. By contrast, the linear extrapolation method yields estimates in which complementarity stands out relative to the individual treatment effects.

The second challenge—small sample size—further constrains inference. To gauge how large a sample would be needed to generate more precise evidence using the linear extrapolation method, I conduct a simple scaling exercise. I increase the dataset by a factor of ten and re-estimate the same models without clustering adjustments, to isolate the impact of sample size. The dataset is scaled by cloning observations without adding noise. The results underscore the empirical difficulty of detecting treatment complementarity: under imperfect compliance, uncovering interaction effects requires substantially larger samples than those typically sufficient for reduced-form analysis.

Table 9: Empirical application: Angelucci and Bennett (2024)
During period only

	Reduced Form (N×10)	Two Stage Least Squares (N×10)	Linear Extrapolation (N×10)
First treatment (PC) effect	-0.151 (0.020)	-0.317 (0.042)	0.528 (0.144)
Second treatment (LA) effect	-0.082 (0.020)	-0.116 (0.028)	-0.160 (0.029)
Complementarity	-0.027 (0.030)	-0.141 (0.091)	-0.953 (0.413)

6 Conclusion

This paper studies how to estimate complementarities between two treatments when assignment is not fully random. I show that the standard approach, two stage least squares with instruments for each treatment and their interaction, generally fails to recover causal interaction effects unless one of two strong conditions holds. First, treatment effects must be homogeneous. Second, instruments must shift only their own treatments and must generate a common complier population across coefficients. These conditions imply first stage restrictions that are testable and rarely satisfied in practice. This clarifies that learning about causal interactions requires attention not only to usual instrument validity but also to complier composition that underlies each estimand.

I develop an alternative strategy that extends the marginal treatment effects framework to two joint treatments. The method models potential outcomes as linear in unobserved resistance to each treatment and accommodates a richer, more realistic set of complier types than the set that delivers a clean 2SLS interpretation. In simulations where 2SLS is biased, the linear–extrapolation approach recovers the interaction effect more accurately when its identifying assumptions hold.

I apply these ideas to Angelucci and Bennett (2024), a randomized trial with substantial noncompliance that cross randomizes pharmacotherapy and livelihoods assistance. The first stage patterns suggest that the assumptions needed for interpretable coefficient comparisons are unlikely to hold. Reduced form tests detect instrument complementarity but cannot separate compliance differences from treatment interactions. Two stage least squares does not resolve this problem because coefficients represent different complier groups. The linear extrapolation approach, implemented with propensities enriched by interactions between instruments and baseline covariates to resolve limited support in propensities, yields estimates that map more directly to average interaction effects. However, the limitation is that the standard error is too large compared to the reduced form or 2SLS estimators.

The results in this paper offer some guidance for applied work. First, report first stage regressions that can speak to the complier restrictions: cross coefficients should be near zero if each instrument shifts only its targeted treatment, and relevant first stage coefficients should be equal if one wishes to sum or compare 2SLS coefficients. Second, avoid interpreting sums or comparisons of 2SLS or reduced form coefficients unless the evidence supports a common complier population. Third, if 2SLS must be used, saturate covariates to mitigate bias when observables are informative about treatment effect heterogeneity. When assumptions are met and broader complier sets are plausible, linear extrapolation could be one alternative.

The linear extrapolation approach has limits. It assumes linear potential outcomes and independence across unobserved resistance terms, and it requires support in the estimated propensities—often thin in randomized trials with one–sided compliance. Flexible first–stage specifications with baseline covariates can expand support, but validity requires that some covariates affect outcomes only through treatment. The empirical application also suggests that larger samples may be needed for precise inference.

There are two directions for future research. One is to study combinations of other quasi experimental strategies, such as difference in differences and regression discontinuity, in order to clarify what additional assumptions are needed to interpret interaction estimates as causal when treatment effects are heterogeneous. The other is to extend the analysis to continuous instruments or continuous treatments.

References

Angelucci, Manuela, and Daniel Bennett. 2024. “The Economic Impact of Depression Treatment in India: Evidence from Community-Based Provision of Pharmacotherapy.” *American Economic Review*, 114(1): 169–198.

Angrist, Joshua, Daniel Lang, and Philip Oreopoulos. 2009. “Incentives and Services for College Achievement: Evidence from a Randomized Trial.” *American Economic Journal: Applied Economics*, 1(1): 136–163.

Behaghel, Luc, Bruno Crepon, and Marc Gurgand. 2013. “Robustness of the Encouragement Design in a Two-Treatment Randomized Control Trial.” *SSRN Electronic Journal*.

Bhuller, Manudeep, and Henrik Sigstad. 2024. “2SLS with multiple treatments.” *Journal of Econometrics*, 242(1): 105785.

Blackwell, Matthew. 2017. “Instrumental Variable Methods for Conditional Effects and Causal Interaction in Voter Mobilization Experiments.” *Journal of the American Statistical Association*, 112(518): 590–599.

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

Brinch, Christian N., Magne Mogstad, and Matthew Wiswall. 2017. “Beyond LATE with a Discrete Instrument.” *Journal of Political Economy*, 125(4): 985–1039. Publisher: The University of Chicago Press.

Carneiro, Pedro, and Sokbae Lee. 2009. “Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality.” *Journal of Econometrics*, 149(2): 191–208.

Cunha, Flavio, and James Heckman. 2007. “The Technology of Skill Formation.” *American Economic Review*, 97(2).

De Chaisemartin, Clément, and Xavier D’Haultfœuille. 2020. “Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects.” *American Economic Review*, 110(9): 2964–2996.

Duflo, Esther, Pascaline Dupas, and Michael Kremer. 2015. “Education, HIV, and Early Fertility: Experimental Evidence from Kenya.” *American Economic Review*, 105(9): 2757–2797.

Fang, Ximeng, Lorenz Goette, Bettina Rockenbach, Matthias Sutter, Verena Tiefenbeck, Samuel Schoeb, and Thorsten Staake. 2023. "Complementarities in behavioral interventions: Evidence from a field experiment on resource conservation." *Journal of Public Economics*, 228: 105028.

Gilligan, Daniel O., Naureen Karachiwalla, Ibrahim Kasirye, Adrienne M. Lucas, and Derek Neal. 2022. "Educator Incentives and Educational Triage in Rural Primary Schools." *Journal of Human Resources*, 57(1): 79–111. Publisher: University of Wisconsin Press Section: Articles.

Goff, Leonard. 2025. "When does IV identification not restrict outcomes?" arXiv:2406.02835 [econ].

Goff, Leonard, Ofer Malamud, Cristian Pop-Eleches, and Miguel Urquiola. 2023. "Interactions Between Family and School Environments: Access to Abortion and Selective Schools." *Journal of Human Resources.* Publisher: University of Wisconsin Press Section: Articles.

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár. 2024. "Contamination Bias in Linear Regressions." *American Economic Review*, 114(12): 4015–4051.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of Econometrics*, 225(2): 254–277.

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

Johnson, Rucker C., and C. Kirabo Jackson. 2019. "Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending." *American Economic Journal: Economic Policy*, 11(4): 310–349.

Kerwin, Jason T., and Rebecca L. Thornton. 2021. "Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures." *The Review of Economics and Statistics*, 103(2): 251–264.

Kirkeboen, Lars J., Edwin Leuven, and Magne Mogstad. 2016. "Field of Study, Earnings, and Self-Selection." *The Quarterly Journal of Economics*, 131(3): 1057–1111.

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

Kowalski, Amanda E. 2023*a*. "Behaviour within a Clinical Trial and Implications for Mammography Guidelines." *The Review of Economic Studies*, 90(1): 432–462.

Kowalski, Amanda E. 2023*b*. "Reconciling Seemingly Contradictory Results from the Oregon Health Insurance Experiment and the Massachusetts Health Reform." *Review of Economics and Statistics*.

Mogstad, Magne, Alexander Torgovitsky, and Christopher R. Walters. 2021. "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables." *American Economic Review*, 111(11): 3663–3698.

Mogstad, Magne, and Alexander Torgovitsky. 2024. "Instrumental Variables with Unobserved Heterogeneity in Treatment Effects." *NBER Working Paper*.

Mountjoy, Jack. 2022. "Community Colleges and Upward Mobility." *American Economic Review*, 112(8): 2580–2630.

Neumark, David, and William Wascher. 2011. "Does a Higher Minimum Wage Enhance the Effectiveness of the Earned Income Tax Credit?" *ILR Review*, 64(4): 712–746. Publisher: SAGE Publications Inc.

Pinto, Rodrigo. 2022. "Beyond Intention-to-Treat: Using the Incentives of Moving to Opportunity to Identify Neighborhood Eects." *NBER Working Paper*.

Rossin-Slater, Maya, and Miriam Wüst. 2020. "What Is the Added Value of Preschool for Poor Children? Long-Term and Intergenerational Impacts and Interactions with an Infant Health Intervention." *American Economic Journal: Applied Economics*, 12(3): 255–286.

Appendix

A Proofs

A.1 Proof of Proposition 1

Rewrite[3] equation (3) as

$$Y = \beta_0 + \beta_1 T_1 + \beta_2 T_2 + \beta_c T_3 + \epsilon,$$

where $T_3 = T_1 \times T_2$. The 2SLS estimator is characterized by the moment conditions

$$\mathbb{E}[\epsilon] = \mathbb{E}[\epsilon Z_1] = \mathbb{E}[\epsilon Z_2] = \mathbb{E}[\epsilon Z_3] = 0,$$

with $Z_3 = Z_1 \times Z_2$.

Define the treatment effects as

$$\begin{aligned}
\Delta_1 &= Y(1,0) - Y(0,0),\\
\Delta_2 &= Y(0,1) - Y(0,0),\\
\Delta_c &= \big(Y(1,1) - Y(1,0)\big) - \big(Y(0,1) - Y(0,0)\big),
\end{aligned}$$

where Δ_c denotes the complementarity effect.

Using the potential treatment notation introduced in the main text, the error term in equation (3) can be expressed as

$$\begin{aligned}
\epsilon = {} & (Y(0,0) - \beta_0) + (\Delta_1 - \beta_1)T_1(0,0) + (\Delta_2 - \beta_2)T_2(0,0) + (\Delta_c - \beta_c)T_3(0,0)\\
& + Z_1\big[(\Delta_1 - \beta_1)\big(T_1(1,0) - T_1(0,0)\big) + (\Delta_2 - \beta_2)\big(T_2(1,0) - T_2(0,0)\big) + (\Delta_c - \beta_c)\big(T_3(1,0) - T_3(0,0)\big)\big]\\
& + Z_2\big[(\Delta_1 - \beta_1)\big(T_1(0,1) - T_1(0,0)\big) + (\Delta_2 - \beta_2)\big(T_2(0,1) - T_2(0,0)\big) + (\Delta_c - \beta_c)\big(T_3(0,1) - T_3(0,0)\big)\big]\\
& + Z_3\big[(\Delta_1 - \beta_1)\big(T_1(1,1) - T_1(1,0) - T_1(0,1) + T_1(0,0)\big)\\
& \qquad + (\Delta_2 - \beta_2)\big(T_2(1,1) - T_2(1,0) - T_2(0,1) + T_2(0,0)\big)\\
& \qquad + (\Delta_c - \beta_c)\big(T_3(1,1) - T_3(1,0) - T_3(0,1) + T_3(0,0)\big)\big].
\end{aligned}$$

Substitute this expression into the moment conditions. By Assumption 2, $(Y(\cdot), T_1(\cdot), T_2(\cdot)) \perp (Z_1, Z_2)$ and $Z_1 \perp Z_2$. Using these assumptions as well as $Z_1^2 = Z_1$ and $Z_2^2 = Z_2$ since each instrument are binary indicators, the moment conditions can be written as

[3] This section extends Kirkeboen, Leuven and Mogstad (2016) and Behaghel, Crepon and Gurgand (2013) to the case of treatment complementarities.

$$\begin{pmatrix} 1 & 0 & \mathbb{E}[Z_2] \\ 0 & 1 & \mathbb{E}[Z_1] \\ 1-\mathbb{E}[Z_1] & 1-\mathbb{E}[Z_2] & 1-\mathbb{E}[Z_3] \end{pmatrix} \begin{pmatrix} \text{condition 1} \\ \text{condition 2} \\ \text{condition 3} \end{pmatrix} = 0. \tag{9}$$

The determinant of the first matrix equals $(1-\mathbb{E}[Z_1])(1-\mathbb{E}[Z_2])$, which is nonzero as long as neither Z_1 nor Z_2 is almost surely equal to 1.

The three conditions are:

$$\begin{aligned}
&\mathbb{E}\big[(\Delta_1-\beta_1)\big(T_1(1,0)-T_1(0,0)\big)+(\Delta_2-\beta_2)\big(T_2(1,0)-T_2(0,0)\big)+(\Delta_c-\beta_c)\big(T_3(1,0)-T_3(0,0)\big)\big]=0,\\
&\mathbb{E}\big[(\Delta_1-\beta_1)\big(T_1(0,1)-T_1(0,0)\big)+(\Delta_2-\beta_2)\big(T_2(0,1)-T_2(0,0)\big)+(\Delta_c-\beta_c)\big(T_3(0,1)-T_3(0,0)\big)\big]=0,\\
&\mathbb{E}\big[(\Delta_1-\beta_1)\big(T_1(1,1)-T_1(1,0)-T_1(0,1)+T_1(0,0)\big)\\
&\qquad+(\Delta_2-\beta_2)\big(T_2(1,1)-T_2(1,0)-T_2(0,1)+T_2(0,0)\big)\\
&\qquad+(\Delta_c-\beta_c)\big(T_3(1,1)-T_3(1,0)-T_3(0,1)+T_3(0,0)\big)\big]=0.
\end{aligned}$$

Rearranging these conditions with respect to $(\beta_1, \beta_2, \beta_c)$ yields the 2SLS estimand. Denote

$$\begin{aligned}
\Delta\Delta T_1 &\equiv T_1(1,1)-T_1(1,0)-T_1(0,1)+T_1(0,0)\\
\Delta\Delta T_2 &\equiv T_2(1,1)-T_2(1,0)-T_2(0,1)+T_2(0,0)\\
\Delta\Delta T_3 &\equiv T_3(1,1)-T_3(1,0)-T_3(0,1)+T_3(0,0)
\end{aligned}$$

Then the 2SLS estimand is

$$\begin{aligned}
\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_c\end{pmatrix} &= \begin{pmatrix} \mathbb{E}[T_1(1,0)-T_1(0,0)] & \mathbb{E}[T_2(1,0)-T_2(0,0)] & \mathbb{E}[T_3(1,0)-T_3(0,0)] \\ \mathbb{E}[T_1(0,1)-T_1(0,0)] & \mathbb{E}[T_2(0,1)-T_2(0,0)] & \mathbb{E}[T_3(0,1)-T_3(0,0)] \\ \mathbb{E}[\Delta\Delta T_1] & \mathbb{E}[\Delta\Delta T_2] & \mathbb{E}[\Delta\Delta T_3] \end{pmatrix}^{-1}\\
&\quad\times \begin{pmatrix} \mathbb{E}\big[\Delta_1(T_1(1,0)-T_1(0,0))+\Delta_2(T_2(1,0)-T_2(0,0))+\Delta_c(T_3(1,0)-T_3(0,0))\big] \\ \mathbb{E}\big[\Delta_1(T_1(0,1)-T_1(0,0))+\Delta_2(T_2(0,1)-T_2(0,0))+\Delta_c(T_3(0,1)-T_3(0,0))\big] \\ \mathbb{E}\big[\Delta_1(\Delta\Delta T_1)+\Delta_2(\Delta\Delta T_2)+\Delta_c(\Delta\Delta T_3)\big] \end{pmatrix}.
\end{aligned}$$

Therefore, in general, the weights attached to Δ_2 and Δ_c in the 2SLS estimand of β_1 are nonzero, because they are determined by the inverse matrix multiplied to the second matrix. Invoking Assumption 4 (Monotonicity) does not remove these terms; it only ensures nonnegativity of $T_1(1,0)-T_1(0,0)$, $T_2(0,1)-T_2(0,0)$, and $T_3(1,1)-T_3(0,0)$. Hence, under Assumptions 1–4, the 2SLS estimand is generally a linear combination of Δ_1, Δ_2, and

Δ_c.

A.1.1 Proof of weight averaging to 1 or 0 in Proposition 1

Starting from the expression obtained in Appendix A.1, the 2SLS estimand satisfies

$$\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_c\end{pmatrix} = \begin{pmatrix}\mathbb{E}[a_1] & \mathbb{E}[a_2] & \mathbb{E}[a_3]\\ \mathbb{E}[a_4] & \mathbb{E}[a_5] & \mathbb{E}[a_6]\\ \mathbb{E}[a_7] & \mathbb{E}[a_8] & \mathbb{E}[a_9]\end{pmatrix}^{-1} \times \begin{pmatrix}\mathbb{E}\big[\Delta_1[a_1] + \Delta_2[a_2] + \Delta_c[a_3]\big]\\ \mathbb{E}\big[\Delta_1[a_4] + \Delta_2[a_5] + \Delta_c[a_6]\big]\\ \mathbb{E}\big[\Delta_1[a_7] + \Delta_2[a_8] + \Delta_c[a_9]\big]\end{pmatrix},$$

Then, denoting $A = \begin{pmatrix}a_1 & a_2 & a_3\\ a_4 & a_5 & a_6\\ a_7 & a_8 & a_9\end{pmatrix}$,

$$\beta_1 = \begin{pmatrix}1 & 0 & 0\end{pmatrix}(\mathbb{E}A)^{-1}\big[\mathbb{E}[\Delta_1 A\begin{pmatrix}1\\0\\0\end{pmatrix}] + \mathbb{E}[\Delta_2 A\begin{pmatrix}0\\1\\0\end{pmatrix}] + \mathbb{E}[\Delta_c A\begin{pmatrix}0\\0\\1\end{pmatrix}]\big]$$

Therefore, weight attached to Δ_1 for expression of β_1 is equal to

$$w_1^1 = \begin{pmatrix}1 & 0 & 0\end{pmatrix}(\mathbb{E}A)^{-1}A\begin{pmatrix}1\\0\\0\end{pmatrix}$$

Then, $\mathbb{E}w_1^1 = \begin{pmatrix}1 & 0 & 0\end{pmatrix}(\mathbb{E}A)^{-1}\mathbb{E}A\begin{pmatrix}1\\0\\0\end{pmatrix} = 1.$

It can be shown similarly for the other weights as well.

A.2 Proof of Proposition 2

Starting from the expression obtained in Appendix A.1, the 2SLS estimand satisfies

$$\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_c\end{pmatrix} = \begin{pmatrix}\mathbb{E}[T_1(1,0)-T_1(0,0)] & \mathbb{E}[T_2(1,0)-T_2(0,0)] & \mathbb{E}[T_3(1,0)-T_3(0,0)]\\ \mathbb{E}[T_1(0,1)-T_1(0,0)] & \mathbb{E}[T_2(0,1)-T_2(0,0)] & \mathbb{E}[T_3(0,1)-T_3(0,0)]\\ \mathbb{E}[\Delta\Delta T_1] & \mathbb{E}[\Delta\Delta T_2] & \mathbb{E}[\Delta\Delta T_3]\end{pmatrix}^{-1}$$
$$\times \begin{pmatrix}\mathbb{E}\big[\Delta_1\{T_1(1,0)-T_1(0,0)\}+\Delta_2\{T_2(1,0)-T_2(0,0)\}+\Delta_c\{T_3(1,0)-T_3(0,0)\}\big]\\ \mathbb{E}\big[\Delta_1\{T_1(0,1)-T_1(0,0)\}+\Delta_2\{T_2(0,1)-T_2(0,0)\}+\Delta_c\{T_3(0,1)-T_3(0,0)\}\big]\\ \mathbb{E}\big[\Delta_1(\Delta\Delta T_1)+\Delta_2(\Delta\Delta T_2)+\Delta_c(\Delta\Delta T_3)\big]\end{pmatrix},$$

where

$$\Delta\Delta T_k \equiv T_k(1,1)-T_k(1,0)-T_k(0,1)+T_k(0,0), \quad k\in\{1,2,3\}, \quad T_3 = T_1\times T_2.$$

Under Assumption 5 (No cross effects),

$$T_1(Z_1,1)=T_1(Z_1,0), \qquad T_2(1,Z_2)=T_2(0,Z_2),$$

and

$$T_3(1,0)=T_3(0,0), \qquad T_3(0,1)=T_3(0,0).$$

Hence,

$$T_1(0,1)-T_1(0,0)=0, \quad T_2(1,0)-T_2(0,0)=0, \quad \Delta\Delta T_1=0, \quad \Delta\Delta T_2=0.$$

Moreover,

$$T_1(1,0)-T_1(0,0)\geq 0 \quad T_2(0,1)-T_2(0,0)\geq 0$$
$$\Delta\Delta T_3 = T_3(1,1)-T_3(1,0)-T_3(0,1)+T_3(0,0) = T_3(1,1)-T_3(0,0)\geq 0$$

by Assumption 4 (Monotonicity) and Assumption 5 (No cross effects). Therefore the coefficient matrix becomes diagonal:

$$\begin{pmatrix}\beta_1\\ \beta_2\\ \beta_c\end{pmatrix} = \begin{pmatrix}\mathbb{E}[T_1(1,0)-T_1(0,0)] & 0 & 0\\ 0 & \mathbb{E}[T_2(0,1)-T_2(0,0)] & 0\\ 0 & 0 & \mathbb{E}[\Delta\Delta T_3]\end{pmatrix}^{-1} \begin{pmatrix}\mathbb{E}\big[\Delta_1\{T_1(1,0)-T_1(0,0)\}\big]\\ \mathbb{E}\big[\Delta_2\{T_2(0,1)-T_2(0,0)\}\big]\\ \mathbb{E}\big[\Delta_c\,\Delta\Delta T_3\big]\end{pmatrix}.$$

Applying Assumption 4 and 5 to all possible 256 potential–treatment patterns restricts admissible patterns to the seven types in Proposition 2. Among these,

$$T_1(1,0) - T_1(0,0) = 1 \text{ iff type} \in \{\text{T1-complier},\ \text{Dutiful complier}\},$$

$$T_2(0,1) - T_2(0,0) = 1 \text{ iff type} \in \{\text{T2-complier},\ \text{Dutiful complier}\},$$

$$\Delta\Delta T_3 = 1 \text{ iff type} = \text{Dutiful complier}.$$

Thus,

$$\begin{aligned}
\beta_1 &= \mathbb{E}[\Delta_1 \,|\, T_1 \text{ complier},\ \text{Dutiful complier}] \\
\beta_2 &= \mathbb{E}[\Delta_2 \,|\, T_2 \text{ complier},\ \text{Dutiful complier}] \\
\beta_c &= \mathbb{E}[\Delta_c \,|\, \text{Dutiful complier}]
\end{aligned}$$

A.3 Proof of Proposition 3

Rewrite equation (6) as

$$Y = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_c Z_3 + \epsilon,$$

where $Z_3 = Z_1 \times Z_2$. Also denote $T_3 = T_1 \times T_2$ as in A.1.

Then,

$$Y = Y(0,0) + \Delta_1 T_1 + \Delta_2 T_2 + \Delta_c T_3$$

where each T_1, T_2, T_3 can be expressed using potential choice introduced in 3.1:

$$\begin{aligned}
T_1 &= T_1(0,0) + (T_1(1,0) - T_1(0,0))Z_1 + (T_1(0,1) - T_1(0,0))Z_2 \\
&\quad + (T_1(1,1) - T_1(1,0) - T_1(0,1) + T_1(0,0))Z_3 \\
T_2 &= T_2(0,0) + (T_2(1,0) - T_2(0,0))Z_1 + (T_2(0,1) - T_2(0,0))Z_2 \\
&\quad + (T_2(1,1) - T_2(1,0) - T_2(0,1) + T_2(0,0))Z_3 \\
T_3 &= T_3(0,0) + (T_3(1,0) - T_3(0,0))Z_1 + (T_3(0,1) - T_3(0,0))Z_2 \\
&\quad + (T_3(1,1) - T_3(1,0) - T_3(0,1) + T_3(0,0))Z_3
\end{aligned}$$

Combining the four expressions above, we can connect the reduced-form coefficients to potential outcome and individual treatment effects.

$$
\begin{aligned}
&E[Y|Z_1, Z_2, Z_3] = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_c Z_1 Z_2 \\
&= E[Y(0,0) + \Delta_1 T_1(0,0) + \Delta_2 T_2(0,0) + \Delta_c T_3(0,0)] \\
&+ Z_1 E[\Delta_1(T_1(1,0) - T_1(0,0)) + \Delta_2(T_2(1,0) - T_2(0,0)) + \Delta_c(T_3(1,0) - T_3(0,0))] \\
&+ Z_2 E[\Delta_1(T_1(0,1) - T_1(0,0)) + \Delta_2(T_2(0,1) - T_2(0,0)) + \Delta_c(T_3(0,1) - T_3(0,0))] \\
&+ Z_3 E[\Delta_1(T_1(1,1) - T_1(1,0) - T_1(0,1) + T_1(0,0)) \\
&\quad + \Delta_2(T_2(1,1) - T_2(1,0) - T_2(0,1) + T_2(0,0)) \\
&\quad + \Delta_c(T_3(1,1) - T_3(1,0) - T_3(0,1) + T_3(0,0))]
\end{aligned}
$$

Thus

$$
\begin{aligned}
\gamma_1 &= \mathbb{E}\big[\Delta_1(T_1(1,0) - T_1(0,0)) + \Delta_2(T_2(1,0) - T_2(0,0)) + \Delta_c(T_3(1,0) - T_3(0,0))\big], \\
\gamma_2 &= \mathbb{E}\big[\Delta_1(T_1(0,1) - T_1(0,0)) + \Delta_2(T_2(0,1) - T_2(0,0)) + \Delta_c(T_3(0,1) - T_3(0,0))\big], \\
\gamma_c &= \mathbb{E}\Big[\Delta_1\, \Delta\Delta T_1 + \Delta_2\, \Delta\Delta T_2 + \Delta_c\, \Delta\Delta T_3\Big],
\end{aligned}
$$

where for $k \in \{1, 2, 3\}$,

$$\Delta\Delta T_k \equiv T_k(1,1) - T_k(1,0) - T_k(0,1) + T_k(0,0).$$

These expressions show that each reduced-form coefficient is a linear combination of the three individual-level effects $(\Delta_1, \Delta_2, \Delta_c)$ with weights determined by potential treatment choices. Assumption 4 (Monotonicity) implies nonnegativity of certain differences ($T_1(1,0) - T_1(0,0) \geq 0$, $T_2(0,1) - T_2(0,0) \geq 0$, $T_3(1,1) - T_3(0,0) \geq 0$), but it does *not* set the cross terms to zero. Hence, under Assumptions 1–4, $\gamma_1, \gamma_2, \gamma_c$ generally mix $\Delta_1, \Delta_2, \Delta_c$. Therefore, reduced-form estimates may not be informative if researchers' interests are in Δ. Additional restrictions on potential choices—such as Assumption 5 (No cross effects)—are required to purge these cross terms.

A.4 Proof of Proposition 4

Impose Assumptions 4 (Monotonicity) and 5 (No cross effects) on the expressions derived in Proposition 3 and in Appendix A.3. Under these restrictions, all cross-effects vanish and only own-instrument effects remain:

$$\begin{aligned}\gamma_1 &= \mathbb{E}\big[\Delta_1\{T_1(1,0) - T_1(0,0)\}\big],\\ \gamma_2 &= \mathbb{E}\big[\Delta_2\{T_2(0,1) - T_2(0,0)\}\big],\\ \gamma_c &= \mathbb{E}\big[\Delta_c\,\Delta\Delta T_3\big],\end{aligned}$$

where $\Delta\Delta T_3 \equiv T_3(1,1) - T_3(1,0) - T_3(0,1) + T_3(0,0)$. Applying these assumptions to the full set of $4^4 = 256$ potential-treatment patterns restricts admissible types to the seven listed in Table 2. Among these types,

$$T_1(1,0) - T_1(0,0) = 1 \iff \text{type} \in \{\text{T1-complier},\ \text{Dutiful complier}\},$$

$$T_2(0,1) - T_2(0,0) = 1 \iff \text{type} \in \{\text{T2-complier},\ \text{Dutiful complier}\},$$

$$\Delta\Delta T_3 = 1 \iff \text{type} = \text{Dutiful complier}.$$

Because each difference indicator equals one only for the corresponding complier group,

$$\begin{aligned}\gamma_1 &= \mathbb{E}[\Delta_1 \cdot \mathbf{1}\{T_1\text{-complier or Dutiful complier}\}]\\ &= \Pr(T_1\text{-complier or Dutiful complier}) \cdot \mathbb{E}[\Delta_1 \mid T_1\text{-complier or Dutiful complier}]\,,\\ \gamma_2 &= \mathbb{E}[\Delta_2 \cdot \mathbf{1}\{T_2\text{-complier or Dutiful complier}\}]\\ &= \Pr(T_2\text{-complier or Dutiful complier}) \cdot \mathbb{E}[\Delta_2 \mid T_2\text{-complier or Dutiful complier}]\,,\\ \gamma_c &= \mathbb{E}[\Delta_c \cdot \mathbf{1}\{\text{Dutiful complier}\}]\\ &= \Pr(\text{Dutiful complier}) \cdot \mathbb{E}[\Delta_c \mid \text{Dutiful complier}]\,.\end{aligned}$$

This establishes Proposition 4.

A.5 Proof of Proposition 5 and 6

The first stage for 2SLS estimation is:

$$\begin{aligned}T_1 &= \alpha_1^0 + \alpha_1^1 Z_1 + \alpha_1^2 Z_2 + \alpha_1^c (Z_1 \times Z_2) + \eta_1\\ T_2 &= \alpha_2^0 + \alpha_2^1 Z_1 + \alpha_2^2 Z_2 + \alpha_2^c (Z_1 \times Z_2) + \eta_2\\ (T_1 \times T_2) &= \alpha_c^0 + \alpha_c^1 Z_1 + \alpha_c^2 Z_2 + \alpha_c^c (Z_1 \times Z_2) + \eta_c\end{aligned}$$

Let $Z_3 \equiv Z_1 Z_2$ and $T_3 \equiv T_1 T_2$. Taking expectation of each treatment take-up which can be corresponded to the potential choice and also invoking Assumption 2:

$$\begin{aligned}
\mathbb{E}[T_1 \mid Z_1, Z_2, Z_3] &= \mathbb{E}[T_1(0,0)] + \mathbb{E}[(T_1(1,0) - T_1(0,0))]Z_1 + \mathbb{E}[(T_1(0,1) - T_1(0,0))]Z_2 \\
&\quad + \mathbb{E}[(T_1(1,1) - T_1(1,0) - T_1(0,1) + T_1(0,0))]Z_3 \\
\mathbb{E}[T_2 \mid Z_1, Z_2, Z_3] &= \mathbb{E}[T_2(0,0)] + \mathbb{E}[(T_2(1,0) - T_2(0,0))]Z_1 + \mathbb{E}[(T_2(0,1) - T_2(0,0))]Z_2 \\
&\quad + \mathbb{E}[(T_2(1,1) - T_2(1,0) - T_2(0,1) + T_2(0,0))]Z_3 \\
\mathbb{E}[T_3 \mid Z_1, Z_2, Z_3] &= \mathbb{E}[T_3(0,0)] + \mathbb{E}[(T_3(1,0) - T_3(0,0))]Z_1 + \mathbb{E}[(T_3(0,1) - T_3(0,0))]Z_2 \\
&\quad + \mathbb{E}[(T_3(1,1) - T_3(1,0) - T_3(0,1) + T_3(0,0))]Z_3
\end{aligned}$$

Therefore,

$$\begin{aligned}
\alpha_1^1 &= \mathbb{E}\big[T_1(1,0) - T_1(0,0)\big], & \alpha_1^2 &= \mathbb{E}\big[T_1(0,1) - T_1(0,0)\big], & \alpha_1^c &= \mathbb{E}\big[\Delta\Delta T_1\big], \\
\alpha_2^1 &= \mathbb{E}\big[T_2(1,0) - T_2(0,0)\big], & \alpha_2^2 &= \mathbb{E}\big[T_2(0,1) - T_2(0,0)\big], & \alpha_2^c &= \mathbb{E}\big[\Delta\Delta T_2\big], \\
\alpha_c^1 &= \mathbb{E}\big[T_3(1,0) - T_3(0,0)\big], & \alpha_c^2 &= \mathbb{E}\big[T_3(0,1) - T_3(0,0)\big], & \alpha_c^c &= \mathbb{E}\big[\Delta\Delta T_3\big],
\end{aligned}$$

where $\Delta\Delta T_k \equiv T_k(1,1) - T_k(1,0) - T_k(0,1) + T_k(0,0)$.

(i) Proposition 5. Impose Assumption 5 (No cross effects):

$$T_1(Z_1,1) = T_1(Z_1,0), \quad T_2(1,Z_2) = T_2(0,Z_2), \quad T_3(1,0) = T_3(0,0), \quad T_3(0,1) = T_3(0,0).$$

Then

$$\alpha_1^2 = \alpha_1^c = \alpha_2^1 = \alpha_2^c = \alpha_c^1 = \alpha_c^2 = 0.$$

Moreover, by Assumption 4 (Monotonicity),

$$T_1(1,0) - T_1(0,0), \;\; T_2(0,1) - T_2(0,0), \;\; \Delta\Delta T_3 = T_3(1,1) - T_3(0,0) \;\; \text{are all in } \{0,1\}.$$

Under Assumptions 4–5, the admissible potential–choice patterns reduce to the seven types in Table 2. In that list,

$$T_1(1,0) - T_1(0,0) = \mathbf{1}\{\text{T1-complier or Dutiful complier}\},$$

$$T_2(0,1) - T_2(0,0) = \mathbf{1}\{\text{T2-complier or Dutiful complier}\},$$

$$\Delta\Delta T_3 = \mathbf{1}\{\text{Dutiful complier}\}.$$

Taking expectations yields

$$\begin{aligned}\alpha_1^1 &= \Pr(\text{T1-complier or Dutiful complier}),\\ \alpha_2^2 &= \Pr(\text{T2-complier or Dutiful complier}),\\ \alpha_c^c &= \Pr(\text{Dutiful complier}),\end{aligned}$$

with all cross coefficients equal to zero, as claimed in Proposition 5.

(ii) Proposition 6. Now add the further restriction that there are *no* T1-only or T2-only compliers (i.e., rule out types "T1-complier" and "T2-complier" in Table 2). Under this restriction,

$$T_1(1,0) - T_1(0,0) = \mathbf{1}\{\text{Dutiful complier}\}, \qquad T_2(0,1) - T_2(0,0) = \mathbf{1}\{\text{Dutiful complier}\},$$

and still $\Delta\Delta T_3 = \mathbf{1}\{\text{Dutiful complier}\}$ from above. Therefore,

$$\alpha_1^1 = \alpha_2^2 = \alpha_c^c = \Pr(\text{Dutiful complier}),$$

and the remaining coefficients are zero. Hence the instrument's effect on T_1, T_2, and T_1T_2 is driven by the *same* complier group (Dutiful compliers), establishing Proposition 6.