# Limited Monotonicity and the Combined Compliers LATE*


Nadja van 't Hoff
*University of Southern Denmark*
navh@sam.sdu.dk

Arthur Lewbel
*Boston College*
lewbel@bc.edu

Giovanni Mellace
*University of Southern Denmark*
giome@sam.sdu.dk


January 20, 2025


**Abstract**

We consider endogenous binary treatment with multiple binary instruments. We propose a novel limited monotonicity (LiM) assumption that is generally weaker than alternative monotonicity assumptions in the literature. We define and identify (under LiM) the combined compliers local average treatment effect (CC-LATE), which is arguably a more policy-relevant parameter than the weighted average of LATEs identified by two-stage least squares (TSLS), and is valid under more general conditions. Estimating the CC-LATE is trivial, equivalent to running TSLS with one constructed instrument on a subsample. We use our CC-LATE to empirically assess how knowledge of HIV status influences protective behaviors.



**Keywords:** Instrumental variable, Local Average Treatment Effect, monotonicity, multiple instruments.

**JEL classification: C14, C21, C26.**

*Corresponding author: Nadja van 't Hoff (navh@sam.sdu.dk). This project was made possible through generous funding by Independent Research Fund Denmark (grant 90380031B). We have benefited from discussions with Tymon Słoczyński, Phillip Heiler, Jonathan Roth, Toru Kitagawa, Michael Lechner, Volha Lazuka, and participants at the Nordic Econometric Meeting 2022, the EWMES 2022, ICEEE 2023, the SWETA workshop 2023, IAAE 2023, the Frankfurt Econometrics Workshop 2023, the $(ec)^2$ workshop 2023, and at seminars at the University of Copenhagen and the IMT Lucca.

# 1 Introduction

Instrumental variables are commonly used to address treatment endogeneity. Endogeneity arises when the treatment is not randomly assigned and individuals self-select into treatment based on observed and unobserved characteristics. In many settings, it is more realistic that treatment effects vary across individuals based on both observed and unobserved factors, rather than assuming a uniform treatment effect for all individuals.

When treatment effects are heterogeneous and multiple valid instruments are available, each instrument separately identifies the effect for the individuals whose treatment status changes in response to the instrument: the compliers. The treatment effect in the subgroup of these compliers is referred to as the local average treatment effect (LATE). The usual practice for combining instruments is to use the two-stage least squares (TSLS) estimator. For example, Mogstad et al. (2021) found that more than half of all empirical papers employing instrumental variables (IV) published in top-tier journals used TSLS, with multiple instrumental variables for a single treatment.

Imbens and Angrist (1994) show that TSLS converges to a weighted average of the instrument-pair LATEs in the case of multiple valid binary instruments. They impose a monotonicity assumption which ensures that individuals respond to a change in the instrument values in a monotone way, meaning that two-way flows in response to a change in the instrument values are ruled out. We follow Mogstad et al. (2021) in referring to this monotonicity assumption as Imbens and Angrist monotonicity (IAM).

While treatment effects are commonly allowed to be heterogeneous, choices are not: assuming IAM is equivalent to assuming choice homogeneity. This asymmetry is pointed out by Heckman et al. (2006). Mogstad et al. (2021) relax IAM to the weaker partial monotonicity (PM) assumption that allows for more choice heterogeneity. PM considers a change in a single component of the instrument while holding the values of the other instruments fixed. Goff (2024) introduces Vector Monotonicity (VM), a special case of PM, where all these changes have the same direction. For example, in a random coefficient model, PM implies random coefficients with restricted signs in the selection equation. In such a model, VM would imply the same sign on all coefficients, whereas IAM additionally restricts the magnitude of the coefficients. Mogstad et al. (2021) further show that the TSLS estimand retains the interpretation of a linear combination of LATEs with weights that add to one in the case of multiple binary instruments, with the LATEs corresponding to different response groups. They also provide testable implications to study whether these weights are convex.

Despite being common practice, TSLS has several shortcomings. First, PM may still be overly restrictive for certain applications, e.g. using twinning and same-sex siblings as exogenous variation for household size (Angrist and Evans, 1998). PM assumes that parents uniformly respond to their first two children being of the same sex when fixing the twinning instrument. While it is commonly believed that parents prefer having children of both genders, this assumption does not hold true in all contexts (De Chaisemartin, 2017; Dahl and Moretti, 2008), leading to a violation of PM. Second, even if PM holds, the weights of the TSLS estimand can be counterintuitive. These weights, which are applied to the LATEs of specific compliance types, depend on the instrument distribution and can even be negative. Moreover, these weights are not observable and cannot be estimated. When PM is violated, the interpretation of the TSLS estimand is further complicated and the weighted average of LATEs estimated by TSLS includes the LATEs of defier types.

The purpose of the present paper is to address these shortcomings of using the TSLS estimator when multiple binary instruments are available. We propose a less restrictive monotonicity assumption than PM, and we provide an estimand with a more intuitive interpretation than the weighted average of LATEs identified by TSLS. Our proposed monotonicity assumption is referred to as *limited monotonicity* (LiM). This LiM assumption only requires that the treatment status of a unit when all instruments simultaneously equal one is greater than or equal to the treatment status of that unit when all of the instruments equal zero. This means that defiers with respect to some instruments are allowed, as long as these defier types can be pushed towards compliance by other instruments.

For example, in the twinning and same-sex application studied by Angrist and Evans (1998), LiM always holds since all parents are pushed towards compliance (which in this context means having an additional child) by the twinning instrument, even if they defy the same-sex instrument.

Another example where LiM is more plausible than PM is our empirical application. Here, the treatment involves learning HIV status, with the instruments being randomly assigned cash incentives and distance to the test center. Some individuals might defy the distance instrument because of social stigma, however, a cash incentive can overcome this stigma and push those individuals towards compliance as argued in Thornton (2008). See Section 4 for more details.

LiM does not impose any restrictions on choice behavior for units that have some, but not all, of the instruments equal to one. As a result, LiM allows for rich choice heterogeneity. Put differently, LiM requires fewer choice restrictions than PM, allowing for many more response types in the population. Specifically, units can often be defiers for a subset of instruments.

Under LiM, we show that a parameter called the *combined compliers local average treatment effect* (CC-LATE) is identified, and we provide a very simple consistent estimator. The CC-LATE is defined as the average treatment effect (ATE) for all individuals who are untreated when all instruments equal zero, and who are treated when all instruments equal one. Equivalently, these are individuals who comply with at least one of the instruments, or a combination of them, when the values of all other instruments are either all zero or all one. We refer to this set of individuals as the "combined compliers". We show the combined compliers is the largest possible subset of the population for which a LATE can be constructed given the provided instruments. So the CC-LATE equals the ATE for the largest possible set of people, and in that sense is as representative of population ATE as is possible.

We claim that the CC-LATE is a more interesting and broadly applicable parameter for a policy-maker than the TSLS estimand for two reasons. Firstly, the CC-LATE isidentified in the presence of a variety of defier types. This is an attractive property of the CC-LATE, since the number of potential defier types grows rapidly with the number of available instruments. Secondly, even if PM is valid, the CC-LATE is more policy relevant than TSLS. This is because the CC-LATE equals a weighted average of LATEs among combined compliers, with weights equalling the corresponding complier shares. Thus, the weights are non-negative by construction and have an intuitive interpretation. In contrast, when PM holds (a strong restriction that CC-LATE does not require), TSLS estimates a weighted average of effects for the same compliers as for CC-LATE, but with less meaningful and sometimes negative weights.

To estimate the CC-LATE, we construct a new instrument that, for each observation, equals one if all the observed instruments equal one, and equals zero if all the observed instruments equal zero. The CC-LATE is obtained by running TSLS using this single constructed instrument on just the subset of observations where this constructed instrument is defined. This estimator generally involves discarding a large fraction of the observations in the data, however, the loss of efficiency from doing so is much less than one might expect. This is because the observations that are kept are the most informative, in the sense that this selection maximizes the size of the complier population. The result is a generally much larger first stage, which compensates for the loss of precision caused by the dropped observations. In practice, we find that standard errors are similar when using our CC-LATE estimator compared to the standard TSLS LATE estimator. Both our simulation studies and our empirical application indicate that dropping all these observations does not cause a large loss in precision. This is discussed in more detail in Section 2.3.

Another feature of the CC-LATE is that it simplifies analysis by effectively reducing to a single instrument context regardless of the number of initial instruments, essentially providing a dimensionality reduction. This also means that many results for the single instrument setting are applicable when estimating the CC-LATE. For example, this feature of the CC-LATE simplifies the inclusion of covariates, since we can immediately apply estimators that have been proposed in the literature in the context of a single instrument. See for example Tan (2006), Frölich (2007), Słoczyński et al. (2022), and Ma (2023).

We illustrate our CC-LATE by estimating the effect of learning of one's HIV status on protective behavior, such as the purchase of condoms. Thornton (2008) investigates the effect of knowing one's HIV status on the purchase of contraceptives in rural Malawi, countering selection issues by instrumenting with a financial incentive offered in the form of cash and with the distance to the recommended HIV center. Both instruments were randomly assigned. We argue that LiM is more plausible than PM in this application. We find that the CC-LATE estimates provide more evidence for protective behavior after learning of one's HIV status than the TSLS estimates. Differences between the estimates might be due to differences in the weighting schemes between the TSLS estimand and our CC-LATE and/or a violation of PM. We also show that the CC-LATE allows us to estimate the LATE on a substantially larger complier population than using each instrument individually. When using the cash instrument only (the one which generates the highest compliers' share among the instruments we consider), the relative compliers consist of 42.5% of the entire population, whereas using the distance instrument yields a share of compliers equal to 2.4% of the population. When we use both instruments and estimate the CC-LATE, the share of combined compliers increases to 44.4% of the population. Particularly compelling is that, when introducing a third instrument which indicates whether an amount above the median cash value was received (30.3% of compliers in isolation), the combined compliers make up 52.9% of the population. This is a substantial improvement over using any of the instruments alone.

Our work is most closely related to that of Mogstad et al. (2021), Frölich (2007), Goff (2024), and Sun and Wüthrich (2022). Mogstad et al. (2021) introduce PM and show that the TSLS estimand retains the interpretation of a weighted average of LATEs under this assumption. Frölich (2007) considers identification with multiple instrumental variables. One of his many estimands is equivalent to ours, but it differs in terms of interpretation as he imposes IAM. Frölich (2007) shows that this estimand gives the effect for the largest group of (pure) compliers when IAM holds, whereas we show that, under LiM, the CC-LATE refers to the combined complier population, which also includes types ruled out under IAM.

Similarly, Goff (2024) considers this estimand but under vector monotonicity (VM), which is a special form of PM and strictly stronger than our LiM. Under this assumption, Goff (2024) shows that the "all compliers" LATE (ACLATE) is identified. In the setting with two binary instruments, the combined complier population of the CC-LATE is equivalent to Goff's (2024) all compliers population, and the ACLATE and the CC-LATE coincide. Therefore, in the two instruments setting, we show that both parameters are identified under a strictly weaker assumption. When more than two instruments are available, the "all compliers" and the "combined compliers" are different, with the latter being at least as large as the former. Thus, our CC-LATE gives the ATE for at least as large a complier population, and is identified under a weaker monotonicity assumption.

Goff (2024) also introduces the set LATE (SLATE), which represents the LATE when a subset of instruments changes from zero to one. Under VM, the ACLATE is a special case of the SLATE, whereas under LiM, the CC-LATE is a special case of the SLATE. However, the identification of SLATE in the absence of VM is not addressed by Goff (2024).

"Sun and Wüthrich (2022) develop a framework for a potentially vector-valued discrete instrument $Z$, which considers pairs of instrument values $z$ and $z'$ and introduces pairwise monotonicity, meaning $D(z') \geq D(z)$ almost surely. Under pairwise monotonicity, $E(Y^1 - Y^0|D(z') > D(z))$ is identified by a Wald estimand. This approach is equivalent to assuming LiM and identifying the CC-LATE for $z = (0, 0, ..., 0)$ and $z' = (1, 1, ..., 1)$ in case of multiple binary instruments, though Sun and Wüthrich (2022) do not recognize this as an empirically relevant case. If the instruments are ordered such that each has a positive first stage when used individually, then the comparison between $z = (0, 0, ..., 0)$ and $z' = (1, 1, ..., 1)$ becomes the most empirically relevant. A positive first stage indicates that each instrument, when used separately, generates more compliers than defiers. Thus, we expect the highest compliance when comparing $z = (0, 0, ..., 0)$ to $z' = (1, 1, ..., 1)$."

Other studies either focus on relaxing the monotonicity assumption in the setting of a binary treatment and a single binary instrument (Słoczyński, 2020; Kolesár, 2013; Small et al., 2017; De Chaisemartin, 2017; Dahl et al., 2023), or on relaxing or omitting monotonicity in the case of unordered treatments (Kirkeboen et al., 2016; Hull, 2018; Salanié and Lee, 2018; Heckman and Pinto, 2018). Sigstad (2023) considers the judge IV design, introducing LATE under extreme-pair monotonicity, where the strictest judge is always harsher than the most lenient one. Sigstad (2024) studies under what monotonicity conditions MTE-based estimates continue to point-identify common parameters of interest, such as the LATE.

In the multiple instruments setting, Huntington-Klein (2020) derives identification of the

Super-Local Average Treatment Effect under a condition where monotonicity is imposed on subgroups within the data. Mogstad et al. (2020) show that each instrument has its own selection equation under PM, and they use mutual consistency of these equations to obtain information about (instrument-invariant) parameters. One strand of the literature focuses on estimating treatment effects beyond the LATE through extrapolation. For instance, Mogstad and Torgovitsky (2018) extrapolate the support of a single LATE to include observations other than compliers and provide bounds. Mogstad et al. (2018) extrapolate the LATE to a population with lower willingness to pay for treatment.

The remainder of this paper is organized as follows: Section 2 begins by introducing the LiM assumption and the CC-LATE for the setting with two binary instruments, followed by an extension to the setting with more than two binary instruments. Section D presents a comparison of LiM to other versions of the monotonicity assumption. Section 4 provides an empirical application to the impact of learning one's HIV status on contraceptive use as considered by Thornton (2008). Finally, Section 5 concludes. All the proofs, some additional results, a comparison of the CC-LATE estimand to other estimands, and some simulation studies are included in the appendix.

# 2 Limited monotonicity and the combined compliers LATE

## 2.1 Definitions and baseline assumptions

Consider the standard Imbens and Angrist (1994) LATE framework, with an outcome $Y$ and a binary treatment $D$. Assume we have $k$ binary instruments $Z_1, Z_2, ..., Z_k$[1]. Denote by $D^{z_1 z_2 \ldots z_k} \in \{0,1\}$ the potential treatment states, and by $Y^{d,z_1 z_2 \ldots z_k}$ the potential outcomes

[1]The methodology can be readily extended to contexts involving ordered instruments (see Appendix A.5).

(see, for instance, Rubin, 1974). Assume that the instruments satisfy the exclusion restriction, meaning they do not directly affect $Y^d$, and are independent of the potential treatments and outcomes. This ensures that the instruments are as good as randomly assigned. Formally, this is given by Assumption 1.[2]

**Assumption 1: Random assignment and exclusion**

$$Z_j \perp\!\!\!\perp (D^{z_1 z_2 \ldots z_k}, Y^d) \qquad \forall z_1 z_2 \ldots z_k \in \{0,1\}^k, d \in \{0,1\}, j \in \{1,2,\ldots,k\}.$$

We make the following two additional assumptions, which are standard for the LATE framework: The stable unit treatment value assumption (SUTVA) and the instrument relevance assumption. SUTVA requires that the observed outcome is equal to the potential outcome under the received treatment and ensures that the treatment assigned to any individual does not affect the potential outcomes of any other individual, that the individuals do not potentially have access to a different version of the treatment, and that there is no measurement error. The relevance assumption ensures that compliers exist.

**Assumption 2: SUTVA**

$Y = Y^d$ if $D = d$, and $D = D^{z_1 z_2 \ldots z_k}$ if $Z_1 = z_1$, $Z_2 = z_2$, ..., and $Z_k = z_k$.

**Assumption 3: Instrument overlap and relevance**

$0 < P(Z_1 \cdot Z_2 \cdot \ldots \cdot Z_k = 1) < 1$ and $0 < P((1 - Z_1) \cdot (1 - Z_2) \cdot \ldots \cdot (1 - Z_k) = 1) < 1$ and $P(D^{1 \ldots 1 \ldots 1} = 1) \neq P(D^{0 \ldots 0 \ldots 0} = 1)$.

These three assumptions alone do not guarantee identification of a meaningful causal effect. To identify the LATE with only one binary instrument, we would impose the standard mono-

[2] Assumption 1 can be replaced by mean independence when mean effects are of interest, as is the case in our setting. However, as is common in the literature, for simplicity we make the stronger (often equally plausible) assumption of independence.

tonicity assumption that rules out defiers. With multiple binary instruments, we propose a novel weaker monotonicity assumption which requires only that individuals are at least as likely to be treated if all the instruments are switched on as when all the instruments are switched off. In terms of potential treatments, this gives Assumption 4. We refer to this assumption as *limited* monotonicity, since it only imposes a constraint on $P(D^{1...1...1} \geq D^{0...0...0})$. In Section D, we compare LiM to the monotonicity assumptions proposed by Imbens and Angrist (1994) and Mogstad et al. (2021), and show that LiM is strictly weaker than the former and generally weaker than the latter.

**Assumption 4: Limited monotonicity (LiM)**

$P(D^{1...1...1} \geq D^{0...0...0}) = 1$ or $P(D^{1...1...1} \leq D^{0...0...0}) = 1$.

We assume that the instruments are defined such that positive LiM holds, i.e, $P(D^{1...1...1} \geq D^{0...0...0}) = 1$. This only requires defining all instruments such that they each have a positive first stage when used individually. [3]

## 2.2 Two binary instrument setting

We first demonstrate our results for the two binary instrument setting. Then in Section 2.3 we generalize to an arbitrary number of binary instruments.

### 2.2.1 Principal strata and types

With one binary instrument, Imbens and Angrist (1994) (see also Angrist et al., 1996) define four types of individuals: compliers, always-takers, never-takers, and defiers. These types are

[3] If the first-stage coefficient of an instrument is close to zero, it may be preferable to exclude such an instrument.

defined by the values of their potential treatments. With two binary instruments there are sixteen possible types of individuals, as listed in Table 1. Similar to the setting with one binary instrument, the never-takers ($nt$) never take up treatment and the always-takers ($at$) always take up treatment, independent of the instrument values. We follow Mogstad et al. (2021) in labeling some of the other response types: The eager compliers ($ec$), the reluctant compliers ($rc$), the first instrument compliers ($1c$), and the second instrument compliers ($2c$). These compliers respond to either one of the instruments or a combination thereof. We define combined compliers as the set $cc \equiv \{ec, rc, 1c, 2c\}$, so combined compliers are any of these four complier types.

There are different defier types with two binary instruments. Second instrument defiers ($2d$) respond more strongly to the first instrument, since $D = 1$ when $Z_1 = 1$ ($D^{11} = 1$ and $D^{10} = 1$), but they are defiers with respect to the second instrument as soon as $Z_1 = 0$ ($D^{01} = 0$ and $D^{00} = 1$). Similar reasoning can be followed for the first instrument defiers ($1d$). Eager defiers ($ed$) only take up treatment when either both instruments are switched on ($D^{11} = 1$) or when both instruments are switched off ($D^{00} = 1$), but not when a single instrument is switched on ($D^{10} = 0$ and $D^{01} = 0$). Reluctant defiers ($rd$) do not take up treatment when either both instruments are switched on ($D^{11} = 0$) or when both instruments are switched off ($D^{00} = 0$), but they do take up treatment when a single instrument is switched on ($D^{10} = 1$ and $D^{01} = 1$). Finally, there are six other defier types ($d1$, $d2$, $d3$, $d4$, $d5$, and $d6$).

Unlike the case with a single binary instrument, monotonicity with multiple instruments means there are more defier types than complier types. This is due to the existence of defiers with respect to either instrument. When only one of the instruments is observed, individuals may correspond to different types for this instrument, depending on the value that the other (possibly unobserved) instrument takes (see Table 1). For instance, consider an eager defier

($ed$). If only instrument $Z_1$ were observed, this individual would be a complier when $Z_2 = 1$, but would be a defier with respect to $Z_1$ when $Z_2 = 0$.

In the two-instrument setting, LiM reduces to the following assumption:[4]

**Limited monotonicity (LiM) in the two-instrument setting**

$P(D^{11} \geq D^{00}) = 1.$

LiM allows for 12 out of the 16 initial response types (see Table 1). It rules out four defier types, as shown in Table 1 ($d3$, $d4$, $d5$, and $d6$). These are the defier types that would take up treatment when all instruments are switched off ($D^{00} = 1$), but would not take up treatment when all instruments are switched on ($D^{11} = 0$). These response types never classify as a complier when only one of the instruments is observed. More specifically, receiving a second instrument never pushes these individuals towards compliance.

#### 2.2.2 The CC-LATE

Our parameter of interest, denoted by $\beta$, is the combined compliers local average treatment effect (CC-LATE), defined as $E(Y^1 - Y^0 | T \in cc)$, where $T$ denotes type and, for the case of two instruments, the combined compliers are the set $cc \equiv \{ec, rc, 1c, 2c\}$. The combined compliers here are individuals who become compliers when both instruments are switched on. This means that the CC-LATE is robust to the presence of all defier types except the ones that

[4]Vytlacil's equivalence result (Vytlacil, 2002) connects the LATE assumptions to selection models. Monotonicity assumptions place restrictions on choice behavior. Suppose that we have the following selection equation:

$$D_i(z_1, z_2) = \mathbb{1}[\beta_{0i} + \beta_{1i} z_1 + \beta_{2i} z_2 + \beta_{3i} z_1 z_2 \geq 0].$$

LiM only imposes that either $\beta_{1i} + \beta_{2i} + \beta_{3i} \geq 0$ or $\beta_{1i} + \beta_{2i} + \beta_{3i} \leq 0$. It neither imposes restrictions on the signs and magnitudes of the coefficients nor on direct comparisons between the coefficients. $\beta_{0i}$, $\beta_{1i}$, $\beta_{2i}$, and $\beta_{3i}$ are allowed to vary with $i$, allowing for rich choice heterogeneity.

are more likely to be treated when both instruments are turned off than when both are switched on (see Table 1). Theorem 1 gives our main result for the setting with two binary instruments.

**Theorem 1:** Let Assumptions 1, 2, 3, and 4 hold with two instruments. Then the CC-LATE is identified as

$$\beta = \frac{E\left(Y \mid Z_1 = 1, Z_2 = 1\right) - E\left(Y \mid Z_1 = 0, Z_2 = 0\right)}{E\left(D \mid Z_1 = 1, Z_2 = 1\right) - E\left(D \mid Z_1 = 0, Z_2 = 0\right)} = E(Y^1 - Y^0 | T \in cc),$$

where $T$ denotes type and the combined compliers are the set $cc \equiv \{ec, rc, 1c, 2c\}$.

**Proof** in Appendix A.1.

Table 1: Principal strata and the definition of the response types in case of two binary instruments and a binary treatment.

| Type | $D^{11}$ | $D^{10}$ | $D^{01}$ | $D^{00}$ | Type w.r.t. $Z_1$ | | Type w.r.t. $Z_2$ | | Notion | LiM | PM/VM | IAM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ($T$) | | | | | when $Z_2 = 0$ | when $Z_2 = 1$ | when $Z_1 = 0$ | when $Z_1 = 1$ | | | | |
| $at$ | 1 | 1 | 1 | 1 | Always-taker | Always-taker | Always-taker | Always-taker | Always-taker | ✓ | ✓ | ✓ |
| $ec$ | 1 | 1 | 1 | 0 | Complier | Always-taker | Complier | Always-taker | Eager complier | ✓ | ✓ | ✓ |
| $rc$ | 1 | 0 | 0 | 0 | Never-taker | Complier | Never-taker | Complier | Reluctant complier | ✓ | ✓ | ✓ |
| $1c$ | 1 | 1 | 0 | 0 | Complier | Complier | Never-taker | Always-taker | First instrument complier | ✓ | ✓ | ✓ |
| $2c$ | 1 | 0 | 1 | 0 | Never-taker | Always-taker | Complier | Complier | Second instrument complier | ✓ | ✓ | |
| $1d$ | 1 | 0 | 1 | 1 | Defier | Always-taker | Always-taker | Complier | First instrument defier | ✓ | | |
| $2d$ | 1 | 1 | 0 | 1 | Always-taker | Complier | Defier | Always-taker | Second instrument defier | ✓ | | |
| $ed$ | 1 | 0 | 0 | 1 | Defier | Complier | Defier | Complier | Eager defier | ✓ | | |
| $rd$ | 0 | 1 | 1 | 0 | Complier | Defier | Complier | Defier | Reluctant defier | ✓ | | |
| $d1$ | 0 | 1 | 0 | 0 | Complier | Never-taker | Never-taker | Defier | Defier type 1 | ✓ | | |
| $d2$ | 0 | 0 | 1 | 0 | Never-taker | Defier | Complier | Never-taker | Defier type 2 | ✓ | | |
| $d3$ | 0 | 1 | 1 | 1 | Always-taker | Defier | Always-taker | Defier | Defier type 3 | | | |
| $d4$ | 0 | 1 | 0 | 1 | Always-taker | Never-taker | Defier | Defier | Defier type 4 | | | |
| $d5$ | 0 | 0 | 1 | 1 | Defier | Defier | Always-taker | Never-taker | Defier type 5 | | | |
| $d6$ | 0 | 0 | 0 | 1 | Defier | Never-taker | Defier | Never-taker | Defier type 6 | | | |
| $nt$ | 0 | 0 | 0 | 0 | Never-taker | Never-taker | Never-taker | Never-taker | Never-taker | ✓ | ✓ | ✓ |

✓demonstrates the types allowed for under the respective forms of the monotonicity assumption. We consider a case where PM and VM are equivalent, specifically under the choice restrictions defined in Equation (4) in Appendix D. These restrictions align with those in Table 2 and Equation (7) of Mogstad et al. (2021). Response types under IAM are for the setting when all individuals prefer the incentive created by $Z_1$ over the incentive created by $Z_2$.

#### 2.2.3 Estimation and inference

To estimate the CC-LATE with two instruments $Z_1$ and $Z_2$, first drop all observations that have $z_1$ not equal $z_2$. For the remaining subsample, apply TSLS (or equivalently IV) using $\widetilde{Z} = Z_1 = Z_2$ as the sole instrument. As noted earlier, the loss in precision from dropping these observations is much less than one might expect, because the observations that are kept maximize the size of the complier population, leading to a larger first stage. This is demonstrated in our simulations and empirical application, where the precision (Wald statistic) of this CC-LATE estimator is similar to that of the standard multiple instrument LATE that applies TSLS to all of the data. See section 2.3 for details.

We can write our CC-LATE estimator as $\hat{\beta} = (D^T P_{\widetilde{Z}} D)^{-1} D^T P_{\widetilde{Z}} Y$ with $P_{\widetilde{Z}} = \widetilde{Z}(\widetilde{Z}^T \widetilde{Z})^{-1} \widetilde{Z}^T$, which reduces to $\hat{\beta} = (\widetilde{Z}' D)^{-1} \widetilde{Z}' Y$ in the just-identified case. Denote the subsample averages of $Y$ and $D$ when $z_1 = 0$ and $z_2 = 0$ by $\bar{Y}_{00}$ and $\bar{D}_{00}$, and as $\bar{Y}_{11}$, and $\bar{D}_{11}$ when $z_1 = 1$ and $z_2 = 1$. Then the CC-LATE estimator can also be written as $\hat{\beta} = \frac{\bar{Y}_{11} - \bar{Y}_{00}}{\bar{D}_{11} - \bar{D}_{00}}$, as shown in Appendix A.2. An alternative representation of this estimator using two ordinary least squares (OLS) regressions as well as method of moments (MM) estimation are provided in Appendix A.3. Based on this MM representation, standard MM estimation packages can be used to automatically generate consistent estimates and standard errors. It is also possible to estimate the CC-LATE by replacing the expectations that define the CC-LATE estimand with sample averages. If we have covariates, then after constructing $\widetilde{Z}$ we can instead apply the single instrument estimators with covariates proposed by Tan (2006), Frölich (2007), Słoczyński et al. (2022), and Ma (2023).

## 2.3 Extension to more than two instruments

Suppose we have $k \geq 2$ binary instruments. Let Assumptions 1 (Random Assignment and Exclusion), 2 (SUTVA), 3 (Instrument Overlap and Relevance), and 4 (LiM) hold. Then we show that

$$E(Y^1 - Y^0|T \in cc) = \frac{E\left(Y|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(Y|Z_1 = 0, \ldots, Z_k = 0\right)}{E\left(D|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(D|Z_1 = 0, \ldots, Z_k = 0\right)},$$

where $cc$ represents units who comply with at least one of the instruments, or a combination of them, when the values of all other instruments are either all zero or all one. A great advantage of this parameter is that it is robust to the presence of many different defier types. More specifically, it allows for all defier types for which $P(D^{11\ldots1} = D^{00\ldots0}) = 1$.

**Proof** in Appendix A.4.

The share of the population that are in the set of combined compliers is given by

$$E\left(D|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(D|Z_1 = 0, \ldots, Z_k = 0\right). \tag{1}$$

Usually, adding an additional instrument will increase this share, which in turn enlarges the denominator of the CC-LATE estimand. However, in rare cases this share could decrease. For example, in Table 2, type 2 individuals are included in the combined complier group along with type 1 when only two instruments are used, but with three instruments type 2 individuals are excluded from this group.[5] However, as long as the additional instrument has a positive first stage, it is unlikely that a large proportion of individuals in the population belong to type 2 (i.e., defiers with respect to the third instrument, when the other two both take value one), and therefore it is unlikely that the combined complier share of the population decreases when the additional instrument is included.

[5]Note that VM would rule out type 2, while PM might not.

Empirically, the above combined complier share can be easily estimated for any given set of instruments, so one can directly check if this share in fact increases with the addition of a given instrument.

Table 2: This table illustrates how adding instruments can change the types included in the combined complier population.

| | $D^{000}$ | $D^{001}$ | $D^{110}$ | $D^{111}$ |
|---|---|---|---|---|
| Type 1 | 0 | 0 | 1 | 1 |
| Type 2 | 0 | 0 | 1 | 0 |

Since adding instruments typically increases the denominator of the CC-LATE estimator, we would expect adding instruments to improve estimation precision. However, adding instruments also reduces the sample size, because only observations having instruments all equal zero and all equal one are used for estimation. This reduction in sample size decreases precision. This trade-off means that the precision of the CC-LATE estimator can either be better or worse than that of the standard LATE TSLS estimator, which uses all observations but generally has a smaller denominator. In our simulation results (see Appendix C), and in our empirical application (see Section 4.4) we find that the precision of the CC-LATE estimator remains comparable to that of TSLS.

To show this trade-off algebraically, consider the variance of the CC-LATE estimator, obtained by running TSLS using a single constructed instrument in the subsample where, for each unit, the instruments either all equal one or they all equal zero. Let $N_k$ be the number of observations in this subsample when using $k$ instruments, $\widetilde{Z}^k = \mathbb{1}\{Z_1 = Z_2 \cdots = Z_k\}$, $\pi_{cc,k} = E(D|\widetilde{Z}^k = 1) - E(D|\widetilde{Z}^k = 0)$, $\hat{\beta}_k = (\widetilde{Z}^{k\prime} D)^{-1}\widetilde{Z}^{k\prime} Y$ and $\sigma_k^2 = Var(Y - \beta_{\text{CC-LATE}} D)$.

Then, assuming homoskedasticity for simplicity, we have

$$Var(\hat{\beta}_k) = \sigma_k^2 \frac{1}{N_k} \frac{1}{\pi_{cc,k}^2 \, E(\widetilde{Z}^k)(1 - E(\widetilde{Z}^k))}. \tag{2}$$

Note that for $k = 1$, this reduces to the standard LATE variance with one instrument. This variance might not be monotonic in $k$. Adding an instrument reduces the sample size, i.e., $N_k > N_{k+1}$, but at the same time using an extra instrument generally increases the share of combined compliers: $\pi_{cc,k}^2 \leq \pi_{cc,k+1}^2$. Therefore, adding instruments can either increase or decrease the variance, despite decreasing the subsample size. In practice, one can estimate $\pi_{cc,k}$ for different values of $k$ to assess the benefit of adding instruments. A similar argument can be made for the heteroskedastic case and when comparing ordinary TSLS with our CC-LATE estimator. Both our estimator and the TSLS are not expected to perform well when the number of instruments is very large (and many instrument asymptotic theory would be more relevant in that case), so we focus on the case of a moderate number of instruments.

In conclusion, despite the potentially large decrease in sample size from estimating the CC-LATE, the CC-LATE might not entail any loss in efficiency relative to standard TSLS, particularly in applications where the instruments are strong, the number of instruments is relatively small, or the sample size is large. It is also important to emphasize that the CC-LATE has a more straightforward interpretation than the TSLS estimand. Therefore, one might prefer the CC-LATE for its greater policy relevance, even in cases where it is less precisely estimated than TSLS.

An additional consideration When choosing the number of instruments to use is that LiM validity for $k$ instruments does not guarantee LiM will hold when adding a $(k+1)^{\text{th}}$ instrument.[6]

[6]To illustrate, consider this table:

# 3 LiM compared to PM/VM and IAM

This section illustrates why LiM is generally more plausible than alternative monotonicity assumptions. Our LiM assumption, and the monotonicity assumptions by Imbens and Angrist (1994) and Mogstad et al. (2021) can be formulated as follows:

**Limited monotonicity (LiM)**

$$P(D^{1...1...1} \geq D^{0...0...0}) = 1 \text{ or } P(D^{1...1...1} \leq D^{0...0...0}) = 1.$$

**Imbens and Angrist monotonicity (IAM) (Imbens and Angrist, 1994)**

$P(D^{i...j...k} \geq D^{p...q...r}) = 1$ or $P(D^{i...j...k} \leq D^{p...q...r}) = 1$

$\forall\, i \in \{0,1\}, ..., j \in \{0,1\}, ..., k \in \{0,1\}$ and $\forall\, p \in \{0,1\}, ..., q \in \{0,1\}, ..., r \in \{0,1\}$

such that $P(D^{i...j...k}) \neq P(D^{p...q...r})$.

**Partial monotonicity (PM) (Mogstad et al., 2021)**

$P(D^{1...j...k} \geq D^{0...j...k}) = 1$ or $P(D^{1...j...k} \leq D^{0...j...k}) = 1$,

$P(D^{i...1...k} \geq D^{i...0...k}) = 1$ or $P(D^{i...1...k} \leq D^{i...0...k}) = 1$, and

$P(D^{i...j...1} \geq D^{i...j...0}) = 1$ or $P(D^{i...j...1} \leq D^{i...j...0}) = 1$

$\forall\, i \in \{0,1\}, ..., j \in \{0,1\}, ..., k \in \{0,1\}$.

Note that vector monotonicity (VM) as introduced by Goff (2024) is equivalent to PM if all inequalities have the same sign, and stronger than PM otherwise.

Obviously, all those assumptions (IAM, PM/VM, and LiM) are equivalent in the case of one

| | $D^{000}$ | $D^{001}$ | $D^{110}$ | $D^{111}$ |
|---|---|---|---|---|
| Type 3 | 1 | 0 | 1 | 0 |
| Type 4 | 1 | 0 | 0 | 1 |

Type 3 individuals violate LiM when using three instruments but not with just the first two. Conversely, type 4 individuals violate LiM with the first two instruments, but adding the third instrument resolves this violation.

binary instrument, where they reduce to either $P(D^1 \geq D^0) = 1$ or $P(D^1 \leq D^0) = 1$. When there are two or more instruments, LiM is strictly weaker than IAM. For scenarios with two or more instruments where PM and VM are equivalent, PM/VM and LiM are nested, with LiM being strictly weaker. This means that LiM is strictly weaker than PM/VM when increasing (or decreasing) instrument values consistently leads to an increase (or decrease) in treatment uptake.

In cases where this monotonic relationship does not hold, PM and LiM are non-nested, with LiM supporting considerably more heterogeneity in treatment choice by allowing for more response types. As the number of instruments increases, the difference between LiM and PM increases considerably, with LiM allowing a much larger set of response types and, consequently, far greater choice heterogeneity compared to PM. Moreover, the non-nested response types permitted by PM but not by LiM are unlikely to be empirically plausible, further complicating the justification for PM over LiM. A detailed comparison of these findings is provided in Appendix D.

# 4 Empirical application to the impact of learning of HIV status

In this section, we apply our CC-LATE methodology to estimating the effect of learning of one's HIV status on protective behaviors. Learning of a negative HIV test result could motivate individuals to further protect themselves, while learning about a positive result could motivate individuals to reduce or abstain from behaviors that could spread the disease. The effect of learning test results on the spread of HIV is very important from a policy perspective. Since learning of the test results is an individual choice, selection bias is a serious problem in this application. Thornton (2008) investigates the effect of knowing one's HIV status on the purchase of contraceptives in rural Malawi. To deal with selection issues, Thornton (2008) instruments the endogenous decision of learning one's HIV test results with two instruments: (1) a financial incentive offered in the form of cash to pick up the test result and (2) the distance to a recommended HIV center.

## 4.1 Data

For our analyses, we use the same sample as Thornton (2008). The complete-case sample contains HIV-positive and HIV-negative individuals in Balaka and Rumphi who had sex and got tested for HIV in 2004 and took part in a follow-up survey in 2005. Similar to Thornton (2008), we consider four different outcomes. The outcomes are (1) whether or not an individual bought condoms at the follow-up survey that took place two months after testing, (2) how many condoms the individual bought at the follow-up survey, (3) if the individual reported buying condoms between getting tested and the follow-up survey, and (4) whether the individual reported having sex between getting tested and the follow-up survey. The treatment is whether or not the individual obtained the HIV test results and hence is aware of their HIV status.

We consider three instruments. The first instrument equals one when an individual received *any cash* incentive and zero otherwise. The second instrument is a *distance* incentive that equals one when distance to an HIV test center is less than 1.5km and zero otherwise. We further construct a third instrument, *above median cash* incentive, that equals one if the individual received an amount of cash incentive above the median amount, and zero otherwise. The idea is that some individuals may only react to the incentive if they receive a larger amount of cash. Therefore, this instrument can potentially generate more compliers.

## 4.2 Motivation for LiM and the CC-LATE

We start by checking which version of PM could be consistent with the data. To this end, define $\bar{D}^{z_1,z_2} = \frac{1}{\sum_i z_{1,i} \cdot z_{2,i}} \sum_i z_{1,i} \cdot z_{2,i} \cdot D_i$. When we only consider the *any cash* and *distance* instruments, we have $\bar{D}^{00} = 0.388$, $\bar{D}^{10} = 0.805$, $\bar{D}^{01} = 0.392$, and $\bar{D}^{11} = 0.832$. This implies that the ordering of PM as in Equation (4) in Section D is consistent with the data, leading to the response types as listed in Table 1 in Section 2. When adding the third instrument, *above median cash*, the version of PM consistent with the data is nested with LiM and strictly stronger. It is worth noting that, if the PM condition holds, the standard TSLS estimates a weighted average of the LATEs on the types in the set of combined compliers, while our CC-LATE directly gives a single LATE for the combined complier population, which is arguably a more policy relevant causal parameter.

Moreover, we argue that assuming LiM is more plausible in this application than assuming PM. First of all, LiM is more plausible regarding the response types potentially present in the population. Living close to the recommended HIV center might encourage some individuals to learn of their HIV status due to the small effort of traveling to the center. On the other hand, it might discourage other individuals who would feel too embarrassed to visit an HIV center in their neighborhood out of fear of being recognized. These individuals are defiers with respect to the instrument for the proximity of an HIV center and defy learning of their HIV status when living close to the recommended HIV center. However, they could be willing to learn of their status if they receive a financial incentive. Thornton (2008, p. 1858–1859) emphasizes the importance of a financial incentive to push distance defiers towards compliance. She states: "[T]he evidence from this experiment in Malawi indicates that such psychological barriers, if they exist, can easily and inexpensively be overcome. Cash incentives may directly compensate for the real costs (e.g., travel expenses, missed work) or psychological costs of obtaining HIV results, or they may indirectly reduce the stigma associated with HIV testing by providing individuals with a public excuse for attending the results center." PM would be violated if, in addition to these individuals, there exist individuals who always comply with the proximity instrument. LiM, however, would still hold since it allows for the co-existence of proximity instrument compliers and proximity instrument defiers. LiM only requires that when individuals receive cash and live close to a center, they do not defy learning of their HIV test results.[7] As pointed out by Thornton (2008), social stigma can prevent individuals from learning of their HIV status. She finds that social barriers can be lifted by financial incentives, as the cash provides an excuse for visiting the HIV test center. Inclusion of our third instrument, above median cash incentive, makes LiM even more likely to hold, since it allows there to be individuals who remain distance defiers even with smaller cash incentives.

## 4.3 Instrument distribution and complier share

To estimate the CC-LATE, we only use the subsample of observations for which all instrument values are zero and those for which all instrument values equal one. In the setting with the two instruments, *any cash* and *distance*, 43% of the observations are used to estimate the CC-LATE (see Table 3). In the setting with all three instruments, 27% of the observations are used to estimate the CC-LATE (see Table 3). Including the third instrument, *above median cash*, thus leads to a loss of 16% of the total number of observations. Adding instruments always leads to the same amount of or fewer observations used for estimating the CC-LATE. However, as noted earlier in Section 2.3, the loss in estimation precision from this smaller sample size is partly or completely offset by a corresponding increase in the combined compliers' share of

[7] Stigma could also induce a violation of LiM, but only through people for whom the stigma is not sufficiently compensated for by the subsidy. If it exists, this would likely be a far smaller violation of LiM than of PM, so any resulting bias in LiM is likely to be much smaller than the bias in PM.

the total population (and hence a larger denominator in the LATE formula).

The probability of being a $Z_1$, $Z_2$ or $Z_3$ complier and the probability of being a combined complier in the two and three instrument settings are summarized in Figure 1. The share of compliers for the *distance* instrument is only 2.4%. Since adding instruments never decreases the set of combined compliers, the largest complier share of 52.9% is reached when all three instruments are used.

Table 3: Distribution of the instruments in the setting with two instruments and three instruments in the complete-case data.

| | $Z_1$ *Any cash* | $Z_2$ *Distance* | $Z_3$ *Above median cash* | No. observations | % observations |
|---|---|---|---|---|---|
| Two instruments | 0 | 0 | | 134 | 13% |
| | 1 | 0 | | 497 | 49% |
| | 0 | 1 | | 79 | 8% |
| | 1 | 1 | | 298 | 30% |
| Total no. of observations | | | | 1008 | 100% |
| Observations used by CC-LATE | | | | 432 | 43% |
| Three instruments | 0 | 0 | 0 | 134 | 13% |
| | 0 | 1 | 0 | 79 | 8% |
| | 1 | 0 | 0 | 254 | 25% |
| | 0 | 0 | 1 | 0 | 0% |
| | 1 | 1 | 0 | 154 | 15% |
| | 0 | 1 | 1 | 0 | 0% |
| | 1 | 0 | 1 | 243 | 24% |
| | 1 | 1 | 1 | 144 | 14% |
| Total no. of observations | | | | 1008 | 100% |
| Observations used by CC-LATE | | | | 278 | 28% |

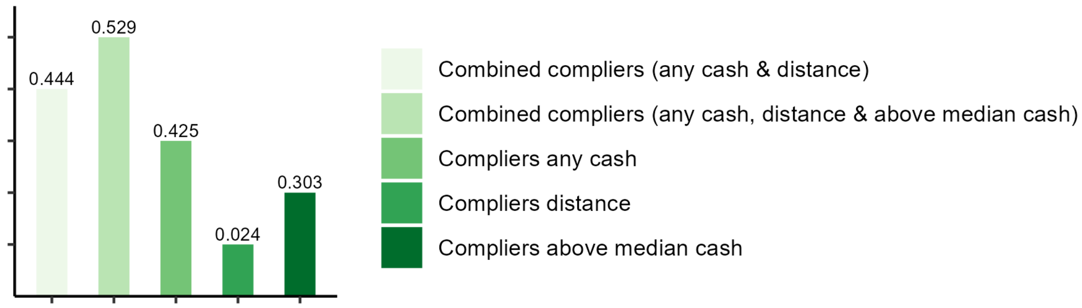



Figure 1: Shares of complier populations for different instrument configurations.

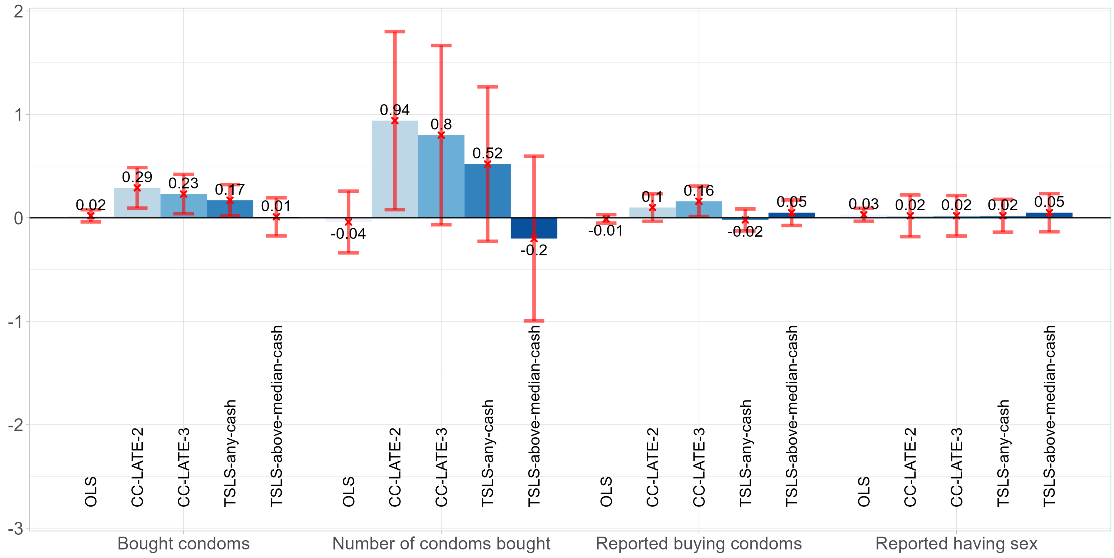
## 4.4 Results

We estimate the effect of learning of HIV status on the four aforementioned outcomes with OLS, the CC-LATE estimator, and TSLS. The first stage of the TSLS estimator is saturated in the instruments. Standard errors are robust and clustered at the village level. Controls are omitted.[8] OLS, which we expect to be downward biased, gives estimates that are rather small and never statistically significant (see Figure 2a). Possible endogeneity giving rise to downward bias could be that respondents who do not practice safe sex are more likely to choose to learn their HIV status, or that individuals who do practice safe sex are less likely to choose to learn their HIV status.

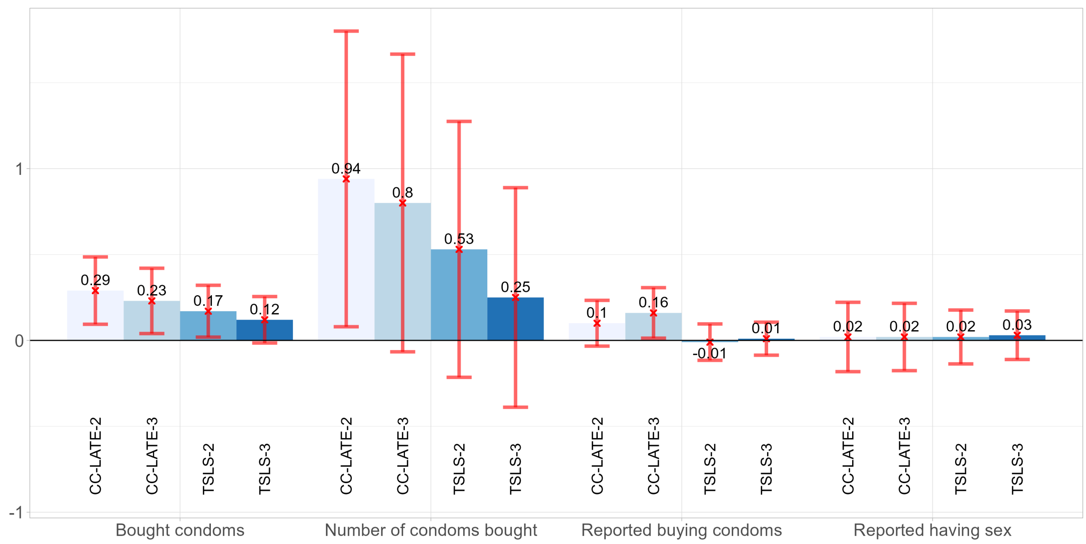
When comparing the *CC-LATE-2* and *CC-LATE-3* estimates in Figure 2a, which are the estimates when using two and three instruments, respectively, we see that adding a third instrument does not have much effect on the precision of the CC-LATE estimator in this application, as the confidence intervals are of similar lengths for all outcomes. The precision loss due to using fewer observations with three instruments is offset by the extra compliers generated by adding the instrument. The estimate decreases in magnitude when adding the *above median cash* instrument, but it should be noted that the two CC-LATEs refer to different populations. The smaller effects might be due to the fact that the additional instrument adds compliers that need extra cash to be pushed towards compliance and are thus possibly less motivated to learn of their test results.

Figure 2a also gives estimates obtained when using each instrument separately. Using the *any cash* or the *above median cash* instruments in isolation gives estimates that are always insignificant, and confidence intervals which are comparable to one or both of our CC-LATE estimators. For all four outcome variables, the estimate obtained when using the *distance*

[8] Since the instruments are randomized, omitting controls should not introduce any bias.




(a) Comparison of CC-LATE estimates to OLS estimates and the TSLS estimates resulting from using each instrument separately. Figure 3 in Appendix B.3 includes the estimate of the distance instrument, which is excluded here for ease of comparison.




(b) Comparison of CC-LATE estimates to TSLS estimates in the case of two or three instruments.

Figure 2: These figures show CC-LATE and TSLS estimates for four outcomes, with treatment defined as learning HIV status. For two instruments (*CC-LATE-2*), we use *any cash* (any financial incentive received) and *distance* (HIV center within 1.5 km). For three instruments (*CC-LATE-3*), *above median cash* (incentive above median) is added. Standard errors are clustered at the village level, with 95% confidence intervals shown in red. Estimates are also reported in Tables 6 and 7 in Appendix B.2.

instrument individually is larger in magnitude with much wider confidence intervals (see Figure 3 in Appendix B.3). The F-statistic of this instrument is rather small (approximately 3), making it a potentially "weak" instrument, which is reflected in the estimates.

We now compare CC-LATE estimates with estimates obtained using TSLS with multiple instruments, as is typically done in the literature. The estimates are depicted in Figure 2b.[9] Reassuringly, the confidence intervals of the CC-LATE estimates and TSLS estimates are comparable. Table 4 presents the ratios of standard errors between TSLS and CC-LATE estimates, indicating that the confidence intervals for TSLS estimates range between 65% and 87% of those for the CC-LATE estimates. The first outcome considered is whether an individual bought condoms at the follow-up survey. Individuals who received their test results were 23 percentage points more likely to buy condoms according to the CC-LATE estimate with three instruments (*any cash*, *distance*, *above median cash*). This is 12 percentage points for TSLS with three instruments, although it is not statistically significant at the 5% level. When using two instruments, we find a higher effect of 29 percentage points with our CC-LATE estimator compared to the 17 percentage points found using TSLS. For the second outcome, neither the CC-LATE nor the TSLS estimates are statistically significant at the 5% level when using three instruments. For the setting with two instruments, the CC-LATE estimate is not only larger in magnitude, but also significant and indicates that, among the combined compliers, individuals who learned of their HIV status bought on average 0.94 condoms more.

Interestingly, adding compliers who respond to the *above median cash* instrument leads to an increase in the CC-LATE estimate for the "reported buying condoms" outcome while, as we saw above, it leads to a decrease for the "bought condoms" outcome. While the former outcome captures whether the respondents bought condoms between getting tested and the follow-up

[9]See Figure B.3 in Appendix B.3 for Figure 2a without the *distance* instrument to allow for easier comparison of the CC-LATE estimator to the LATEs of each instrument used separately.

survey, the latter outcome captures whether the 30 cents they received at the end of the follow-up survey were subsequently used to buy subsidized condoms. The difference in estimates for two and three instruments between these two outcomes may be explained by the fact that the individuals who had to be pushed to compliance by a stronger financial incentive might be lying when responding to the question of whether they bought condoms before the follow-up survey. These individuals subsequently do not buy condoms since they would rather keep the money. The estimates for the outcome, “reported having sex“, are insignificant regardless of the estimator used.

Overall, the CC-LATE estimates provide more evidence for protective behavior after learning of one’s HIV status compared to the TSLS estimates.[10] Differences in estimates can be attributed to either differences in the estimand or to a violation of the PM assumption. The weighted average estimated by TSLS might contain either negative weights or weights that are substantially different from the relative share of the type that contributes to the weighted average. Moreover, if distance instrument defiers are present, then PM is violated, and the weighted average contains the LATE of this defier type. The CC-LATE is robust to the presence of this defier type, whereas TSLS is not. Furthermore, when we use three instruments there are 64 types in the set of combined compliers under LiM. Under PM, at most 35 response types are allowed.

[10]Note that the treatment concerns choosing to know one’s HIV status without differentiating between positive or negative test results. We find similar effects in the subsample with individuals who test negative. The subsample with individuals who test positive is too small to draw meaningful conclusions.

Table 4: Ratio of standard errors between TSLS and CC-LATE estimates.

| | Bought condoms | Number of condoms bought | Reported buying condoms | Reported having sex |
|---|---|---|---|---|
| TSLS-2 / CC-LATE-2 | 0.77 | 0.87 | 0.79 | 0.78 |
| TSLS-3 / CC-LATE-3 | 0.71 | 0.74 | 0.65 | 0.72 |

# 5 Conclusion

TSLS is often used in empirical applications to combine multiple instruments. We have noted some problems with this approach, particularly the restrictiveness of commonly invoked monotonicity assumptions like PM. We introduce a more plausible monotonicity assumption, which we refer to as LiM, and we introduce the CC-LATE, an arguably more policy-relevant causal parameter. The CC-LATE applies to a large complier population and is robust to the presence of a variety of defier types that may often exist in practice.

We apply our CC-LATE to estimate the effect of learning one's HIV status on protective behavior. In comparison to TSLS, the CC-LATE estimates provide more evidence of protective behavior. We and others have noted that the PM assumption usually invoked to justify standard TSLS LATE estimation may be violated, by the presence of distance instrument defiers. Our CC-LATE remains valid in the presence of these defiers, as long as they can be induced to comply by a high cash incentive. We find that programs encouraging learning of one's HIV status using cash and distance incentives can help prevent the spread of the disease. The statistically significant magnitudes we find for these effects are modest, but are larger than those indicated by standard TSLS LATE estimates.

# References

Angrist, J. and Evans, W. N. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*, 88:450–477.

Angrist, J., Imbens, G., and Rubin, D. (1996). Identification of Causal Effects using Instrumental Variables. *Journal of the American statistical Association*, 91(434):444–455.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bloom, H. S. (1984). Accounting for No-Shows in Experimental Evaluation Designs. *Evaluation Review*, 8(2):225–246.

Dahl, C. M., Huber, M., and Mellace, G. (2023). It Is Never Too Late: A New Look at Local Average Treatment Effects With or Without Defiers. *The Econometrics Journal*, 26(3):378–404.

Dahl, G. B. and Moretti, E. (2008). The Demand for Sons. *The Review of Economic Studies*, 75(4):1085–1120.

De Chaisemartin, C. (2017). Tolerating Defiance? Local Average Treatment Effects without Monotonicity. *Quantitative Economics*, 8(2):367–396.

Frölich, M. (2007). Nonparametric IV Estimation of Local Average Treatment Effects with Covariates. *Journal of Econometrics*, 139(1):35–75.

Goff, L. (2020). A Vector Monotonicity Assumption for Multiple Instruments. *arXiv preprint arXiv:2009.00553*.

Goff, L. (2024). A Vector Monotonicity Assumption for Multiple Instruments. *Journal of Econometrics*, 241(1):105735.

Heckman, J. J. and Pinto, R. (2018). Unordered Monotonicity. *Econometrica*, 86(1):1–35.

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

Huber, M., Lechner, M., and Wunsch, C. (2013). The Performance of Estimators based on the Propensity Score. *Journal of Econometrics*, 175(1):1–21.

Hull, P. (2018). Isolateing: Identifying Counterfactual-Specific Treatment Effects with Cross-Stratum Comparisons. *Available at SSRN 2705108*.

Huntington-Klein, N. (2020). Instruments with Heterogeneous Effects: Bias, Monotonicity, and Localness. *Journal of Causal Inference*, 8(1):182–208.

Imbens, G. and Angrist, J. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, 62(2):467–476.

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

Kolesár, M. (2013). Estimation in an Instrumental Variables Model with Treatment Effect Heterogeneity. *Unpublished*.

Lau, C. and Torgovitsky, A. (2020). mivcausal: A Stata Module for Testing the Hypothesis about the Signs of the 2SLS Weights.

Ma, Y. (2023). Identification-Robust Inference for the LATE with High-Dimensional Covariates. *arXiv preprint arXiv:2302.09756*.

Mogstad, M., Santos, A., and Torgovitsky, A. (2018). Using Instrumental Variables for Inference about Policy Relevant Treatment Parameters. *Econometrica*, 86(5):1589–1619.

Mogstad, M. and Torgovitsky, A. (2018). Identification and Extrapolation of Causal Effects with Instrumental Variables. *Annual Review of Economics*, 10:577–613.

Mogstad, M., Torgovitsky, A., and Walters, C. R. (2020). *Policy Evaluation with Multiple Instrumental Variables*. (No. w27546). National Bureau of Economic Research.

Mogstad, M., Torgovitsky, A., and Walters, C. R. (2021). The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables. *American Economic Review*, 111(11):3663–98.

Rubin, D. B. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688.

Salanié, B. and Lee, S. (2018). Identifying Effects of Multivalued Treatments. *Econometrica*, 86(6):1939–1963.

Sigstad, H. (2023). Monotonicity Among Judges: Evidence from Judicial Panels and Consequences for Judge IV Designs. *Available at SSRN 4534809*.

Sigstad, H. (2024). Marginal Treatment Effects and Monotonicity. *arXiv preprint arXiv:2404.03235*.

Słoczyński, T. (2020). When Should We (Not) Interpret Linear IV Estimands as LATE? *arXiv preprint arXiv:2011.06695*.

Słoczyński, T., Uysal, S. D., and Wooldridge, J. M. (2022). Doubly Robust Estimation of Local Average Treatment Effects Using Inverse Probability Weighted Regression Adjustment. *arXiv preprint arXiv:2208.01300*.

Small, D. S., Tan, Z., Ramsahai, R. R., Lorch, S. A., and Brookhart, M. A. (2017). Instrumental Variable Estimation with a Stochastic Monotonicity Assumption. *Statistical Science*, 32(4):561–579.

Sun, Z. and Wüthrich, K. (2022). Pairwise Valid Instruments. *arXiv preprint arXiv:2203.08050*.

Tan, Z. (2006). Regression and Weighting Methods for Causal Inference using Instrumental Variables. *Journal of the American Statistical Association*, 101(476):1607–1618.

Thornton, R. L. (2008). The Demand for, and Impact of, Learning HIV Status. *American Economic Review*, 98(5):1829–63.

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

# Appendices

## A Proofs

### A.1 Proof of Theorem 1

Assume our data consists of independent, identically distributed observations of the vector $(Y_i, D_i, Z_{1i}, Z_{2i})$ for individuals $i = 1, ..., n$. Define the following four variables:

$$R_{1i} = (1 - Z_{1i})(1 - Z_{2i}), \quad R_{2i} = Z_{1i}Z_{2i}, \quad R_{3i} = (1 - Z_{1i})Z_{2i}, \quad R_{4i} = Z_{1i}(1 - Z_{2i}).$$

Under SUTVA, the observed treatment $D_i$ assigned to an individual $i$ can be written as

$$\begin{aligned} D_i &= (1 - Z_{1i})(1 - Z_{2i})D_i^{00} + Z_{1i}Z_{2i}D_i^{11} + (1 - Z_{1i})Z_{2i}D_i^{01} + Z_{1i}(1 - Z_{2i})D_i^{10} \\ &= D_i^{00}R_{1i} + D_i^{11}R_{2i} + D_i^{01}R_{3i} + D_i^{10}R_{4i}. \end{aligned}$$

Consider the denominator of the CC-LATE estimand:

$$\begin{aligned} E\left(D|Z_1 = 1, Z_2 = 1\right) - E\left(D|Z_1 = 0, Z_2 = 0\right) &= E(D|R_2 = 1) - E(D|R_1 = 1) \\ &= E(D_i^{11}|R_2 = 1) - E(D_i^{00}|R_1 = 1) \\ &= E(D_i^{11}) - E(D_i^{00}). \end{aligned}$$

Let $\pi_t = \Pr(T \in t), t = at, rc, ec, 1c, 2c, 1d, 2d, ed, rd, d1, d2, nt$ (see Table 1). We have

$$\begin{aligned} E(D_i^{00}) =& \sum_t E(D_i^{00}|T = t)\pi_t \\ =& \pi_{at} \cdot 1 + \pi_{rc} \cdot 0 + \pi_{ec} \cdot 0 + \pi_{1c} \cdot 0 + \pi_{2c} \cdot 0 + \pi_{1d} \cdot 1 + \pi_{2d} \cdot 1 + \pi_{ed} \cdot 1 + \pi_{rd} \cdot 0 \\ &+ \pi_{d1} \cdot 0 + \pi_{d2} \cdot 0 + \pi_{nt} \cdot 0 \\ =& \pi_{at} + \pi_{1d} + \pi_{2d} + \pi_{ed} \end{aligned}$$

and

$$\begin{aligned}E(D_i^{11}) =&\sum_t E(D_i^{11}|T=t)\pi_t\\=&\pi_{at}\cdot 1+\pi_{rc}\cdot 1+\pi_{ec}\cdot 1+\pi_{1c}\cdot 1+\pi_{2c}\cdot 1+\pi_{nt}\cdot 0+\pi_{1d}\cdot 1+\pi_{2d}\cdot 1+\pi_{ed}\cdot 1\\&+\pi_{rd}\cdot 0+\pi_{d1}\cdot 0+\pi_{d2}\cdot 0\\=&\pi_{at}+\underbrace{\pi_{rc}+\pi_{ec}+\pi_{1c}+\pi_{2c}}_{\pi_{cc}}+\pi_{1d}+\pi_{2d}+\pi_{ed}\\=&\pi_{at}+\pi_{cc}+\pi_{1d}+\pi_{2d}+\pi_{ed}.\end{aligned}$$

It therefore follows that

$$E(D|Z_1=1,Z_2=1)-E(D|Z_1=0,Z_2=0)=E(D_i^{11})-E(D_i^{00})=\pi_{cc},$$

which is the probability of being any type of complier.

Let $\beta_i=Y_i^1-Y_i^0$. Note that unlike the CC-LATE $\beta$, the term $\beta_i$ is random. Under SUTVA, the observed outcome $Y$ can be written as

$$\begin{aligned}Y_i &= Y_i^1D_i+Y_i^0(1-D_i)=\beta_iD_i+Y_i^0\\&= \beta_i\left[D_i^{00}R_{1i}+D_i^{11}R_{2i}+D_i^{01}R_{3i}+D_i^{10}R_{4i}\right]+Y_i^0\\&= \beta_iD_i^{00}R_{1i}+\beta_iD_i^{11}R_{2i}+\beta_iD_i^{01}R_{3i}+\beta_iD_i^{10}R_{4i}+Y_i^0.\end{aligned}$$

Now, consider the numerator of the CC-LATE estimand,

$$\begin{aligned}&E\left(Y|Z_1=1,Z_2=1\right)-E\left(Y|Z_1=0,Z_2=0\right)\\&=E(Y|R_2=1)-E(Y|R_1=1)\\&=E(\beta_iD_i^{11}+Y_i^0|R_2=1)-E(\beta_iD_i^{00}+Y_i^0|R_1=1)\\&=E(\beta_iD_i^{11})-E(\beta_iD_i^{00}).\end{aligned}$$

We have that

$$\begin{aligned}E(\beta_i D_i^{00}) =& \sum_t E(\beta_i D_i^{00}|T=t)\cdot\pi_t\\ =& E(Y_i^1-Y_i^0|T=at)\cdot\pi_{at}+E(Y_i^1-Y_i^0|T=1d)\cdot\pi_{1d}+E(Y_i^1-Y_i^0|T=2d)\cdot\pi_{2d}\\ &+E(Y_i^1-Y_i^0|T=ed)\cdot\pi_{ed}\end{aligned}$$

and

$$\begin{aligned}E(\beta_i D_i^{11}) =& \sum_t E(\beta_i D_i^{11}|T=t)\cdot\pi_t\\ =& E(Y_i^1-Y_i^0|T=at)\cdot\pi_{at}+E(Y_i^1-Y_i^0|T\in cc)\cdot\pi_{cc}+E(Y_i^1-Y_i^0|T=1d)\cdot\pi_{1d}\\ &+E(Y_i^1-Y_i^0|T=2d)\cdot\pi_{2d}+E(Y_i^1-Y_i^0|T=ed)\cdot\pi_{ed}.\end{aligned}$$

Therefore,

$$\begin{aligned}&E(Y|Z_1=1,Z_2=1)-E(Y|Z_1=0,Z_2=0)\\ &=E(\beta_i D_i^{11})-E(\beta_i D_i^{00})\\ &=E(Y^1-Y^0|T\in cc)\cdot\pi_{cc},\end{aligned}$$

and hence

$$\begin{aligned}\beta =& \frac{E(Y\mid Z_1=1,Z_2=1)-E(Y\mid Z_1=0,Z_2=0)}{E(D\mid Z_1=1,Z_2=1)-E(D\mid Z_1=0,Z_2=0)}\\ =& \frac{E(Y\mid R_2=1)-E(Y\mid R_1=1)}{E(D\mid R_2=1)-E(D\mid R_1=1)}\\ =& E(Y^1-Y^0|T\in cc).\end{aligned}$$

## A.2 TSLS with one instrument in the subsample

Denote the subsample averages of $Y$ and $D$ when $(z_1=0,z_2=0)$ by $\bar{Y}_{00}$ and $\bar{D}_{00}$, respectively, and as $\bar{Y}_{11}$, and $\bar{D}_{11}$ when $(z_1=1,z_2=1)$. Denote the total number of observations in the subsample by $\widetilde{N}$, the number of observations for which $(z_1=0,z_2=0)$ as $N_{00}$, and the number of observations for which $(z_1=1,z_2=1)$ as $N_{11}$. Then, $N_{11}=\sum_{i=1}^{\widetilde{N}}\widetilde{Z}$ and $N_{00}=\sum_{i=1}^{\widetilde{N}}(1-\widetilde{Z})$.

$$
\begin{aligned}
\widetilde{Z}'Y &= \sum_{i=1}^{\widetilde{N}} (\widetilde{Z}_i - \bar{\widetilde{Z}})(y_i - \bar{Y}) \\
&= \sum_{i=1}^{\widetilde{N}} \widetilde{Z}_i (y_i - \bar{Y}) - \bar{\widetilde{Z}} \sum_{i=1}^{\widetilde{N}} (y_i - \bar{Y}) \\
&= \sum_{i=1}^{\widetilde{N}} \widetilde{Z}_i (y_i - \bar{Y}) \\
&= N_{11} \frac{1}{N_{11}} \sum_{i=1}^{\widetilde{N}} \widetilde{Z}_i (y_i - \bar{Y}) \\
&= N_{11} (\bar{y_1} - \bar{Y}) \\
&= N_{11} \left( \bar{y_1} - \frac{N_{00}}{\widetilde{N}} \bar{y_0} - \frac{N_{11}}{\widetilde{N}} \bar{y_1} \right) \\
&= N_{11} \left( \frac{N_{00}\bar{Y}_{11} + N_{11}\bar{Y}_{11}}{\widetilde{N}} - \frac{N_{00}\bar{Y}_{00} + N_{11}\bar{Y}_{11}}{\widetilde{N}} \right) \\
&= \frac{N_{11}N_{00}(\bar{Y}_{11} - \bar{Y}_{00})}{\widetilde{N}}
\end{aligned}
$$

In a similar fashion, one can show that $\widetilde{Z}'D = \frac{N_{11}N_{00}(\bar{D}_{11} - \bar{D}_{00})}{\widetilde{N}}$. Then:

$$
\hat{\beta} = (\widetilde{Z}'D)^{-1}\widetilde{Z}'Y = \frac{N_{11}N_{00}(\bar{Y}_{11} - \bar{Y}_{00})/\widetilde{N}}{N_{11}N_{00}(\bar{D}_{11} - \bar{D}_{00})/\widetilde{N}} = \frac{\bar{Y}_{11} - \bar{Y}_{00}}{\bar{D}_{11} - \bar{D}_{00}}.
$$

## A.3 Alternative estimation approaches

Define the following four variables:

$$
R_{1i} = (1 - Z_{1i})(1 - Z_{2i}), \quad R_{2i} = Z_{1i}Z_{2i}, \quad R_{3i} = (1 - Z_{1i})Z_{2i}, \quad R_{4i} = Z_{1i}(1 - Z_{2i}).
$$

A simple consistent estimator of the CC-LATE then consists of the following steps:[11]

[11] As they are unconditionally uncorrelated with $R_1$ and $R_4$ by construction, one could drop $R_2$ and $R_3$ from these regressions without changing the estimates. However, including them is necessary if one wants to include covariates.

1. Use OLS to estimate the coefficients $\alpha_1$ and $\alpha_2$ in

$$D_i = \alpha_1 R_{1i} + \alpha_2 R_{2i} + \alpha_3 R_{2i} + \alpha_4 R_{2i} + e_i,$$

where $e_i$ is the regression error. Denote the estimates $\widehat{\alpha}_j$.

2. Use OLS to estimate the coefficients $\gamma_1$ and $\gamma_2$ in

$$Y_i = \gamma_1 R_{1i} + \gamma_2 R_{2i} + \gamma_3 R_{3i} + \gamma_4 R_{4i} + \varepsilon_i,$$

where $\varepsilon_i$ is the regression error. Denote the estimates $\widehat{\gamma}_j$.

3. The CC-LATE estimator is then

$$\widehat{\beta} = \frac{\widehat{\gamma}_2 - \widehat{\gamma}_1}{\widehat{\alpha}_2 - \widehat{\alpha}_1}.$$

The asymptotic distributions of $\widehat{\beta}$ and $\widehat{\delta}$ can be obtained by the delta method. We can rewrite the above steps as a method of moments (MM) estimator and use a standard MM estimation package to automatically generate consistent estimates and standard errors. To do so, observe that the above regressions can be expressed as the following set of moments:

$$\begin{aligned} &E\left(\left(D_i - \alpha_1 R_{1i} - (\delta + \alpha_1) R_{2i} - \alpha_3 R_{3i} - \alpha_4 R_{4i}\right) R_{ji}\right) = 0 \quad \text{for } j = 1, 2, 3, 4, \text{ and} \\ &E\left(\left(Y_i - \gamma_1 R_{1i} - (\beta\delta + \gamma_1) R_{2i} - \gamma_3 R_{3i} - \gamma_4 R_{4i}\right) R_{ji}\right) = 0 \quad \text{for } j = 1, 2, 3, 4. \end{aligned} \tag{3}$$

Let the vector $\theta = (\beta, \delta, \alpha_1, \alpha_3, \alpha_4, \gamma_1, \gamma_3, \gamma_4)$. Then, the above eight moments can be replaced with corresponding sample moments, and the parameters $\theta$ can be directly estimated using MM estimation. The corresponding $\widehat{\delta}$ will equal $\widehat{\alpha}_2 - \widehat{\alpha}_1$, the estimated probability of an individual $i$ being a combined complier, and $\widehat{\beta}$ will equal the CC-LATE estimate $\frac{\widehat{\gamma}_2 - \widehat{\gamma}_1}{\widehat{\alpha}_2 - \widehat{\alpha}_1}$.

Alternatively, simplifications in getting the limiting distribution of $\widehat{\beta}$ with the delta method can be obtained as follows: Let $\delta = \alpha_2 - \alpha_1$, let $\zeta = \gamma_1 + \gamma_2$, and let $\widetilde{R}_i = R_{1i} + R_{2i}$. Then

$$D_i = \alpha_1 \widetilde{R}_i + \delta R_{2i} + \alpha_3 R_{3i} + \alpha_4 R_{4i} + e_i,$$

$$Y_i = \gamma_1 \widetilde{R}_i + \zeta R_{2i} + \gamma_3 R_{3i} + \gamma_4 R_{4i} + \varepsilon_i.$$

Thus, one can simply estimate the OLS regressions of $D_i$ and $Y_i$ on $\widetilde{R}_i$, $R_{2i}$, $R_{3i}$, and $R_{4i}$, and the coefficients of $R_{2i}$ will be consistent estimates of $\zeta$ and $\delta$, and $\beta = \zeta/\delta$. Note that we can also set up the MM estimator this way.

### A.4 Proof for the extension to multiple instruments

Suppose we have $k \geq 2$ binary instruments, and that Assumptions 1, 2, 3, and 4 hold. Define $D^{z_1 z_2 \ldots z_k}$ the potential treatment state, $R_1 = (1 - Z_1)(1 - Z_2)\ldots(1 - Z_k)$, and $R_2 = Z_1 Z_2 \ldots Z_k$. Under SUTVA, the observed treatment $D_i$ can be written as

$$D_i = D_i^{00\ldots0} R_{1i} + D_i^{11\ldots1} R_{2i} + \tilde{D}_i,$$

where $\tilde{D}_i$ includes all possible combinations of instrument values and the respective potential treatment states. Thus,

$$\begin{aligned} E\left(D|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(D|Z_1 = 0, \ldots, Z_k = 0\right) &= E(D|R_2 = 1) - E(D|R_1 = 1) \\ &= D_i^{11\ldots1} - D_i^{00\ldots0}. \end{aligned}$$

Let $cc$ be the set of all complier types, then

$$E\left(D|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(D|Z_1 = 0, \ldots, Z_k = 0\right) = \pi_{cc}.$$

Similarly, it is easy to show that

$$E\left(Y|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(Y|Z_1 = 0, \ldots, Z_k = 0\right) = E(Y^1 - Y^0|T \in cc)\pi_{cc}.$$

Thus,

$$\frac{E\left(Y|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(Y|Z_1 = 0, \ldots, Z_k = 0\right)}{E\left(D|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(D|Z_1 = 0, \ldots, Z_k = 0\right)} = E(Y^1 - Y^0|T \in cc).$$

An alternative way to obtain this result is as follows:

$$
\begin{aligned}
&E\left(Y_i|Z_{1i}=1,\ldots,Z_{ki}=1\right)-E\left(Y|Z_{1i}=0,\ldots,Z_{ki}=0\right)\\
&=E\left(Y_i|R_{2i}=1\right)-E\left(Y_i|R_{1i}=1\right)\\
&=E(D_i^{1\ldots1\ldots1}\cdot Y_i^1+(1-D_i^{1\ldots1\ldots1})\cdot Y_i^0|R_{2i}=1)-E(D_i^{0\ldots0\ldots0}\cdot Y_i^1+(1-D_i^{0\ldots0\ldots0})\cdot Y_i^0|R_{1i}=1)\\
&=E(D_i^{1\ldots1\ldots1}\cdot Y_i^1+(1-D_i^{1\ldots1\ldots1})\cdot Y_i^0)-E(D_i^{0\ldots0\ldots0}\cdot Y_i^1+(1-D_i^{0\ldots0\ldots0})\cdot Y_i^0)\\
&=E(D_i^{1\ldots1\ldots1}\cdot Y_i^1+(1-D_i^{1\ldots1\ldots1})\cdot Y_i^0-D_i^{0\ldots0\ldots0}\cdot Y_i^1-(1-D_i^{0\ldots0\ldots0})\cdot Y_i^0)\\
&=E(D_i^{1\ldots1\ldots1}\cdot Y_i^1+Y_i^0-D_i^{1\ldots1\ldots1}\cdot Y_i^0-D_i^{0\ldots0\ldots0}\cdot Y_i^1-Y_i^0+D_i^{0\ldots0\ldots0}\cdot Y_i^0)\\
&=E(D_i^{1\ldots1\ldots1}\cdot Y_i^1-D_i^{1\ldots1\ldots1}\cdot Y_i^0-D_i^{0\ldots0\ldots0}\cdot Y_i^1+D_i^{0\ldots0\ldots0}\cdot Y_i^0)\\
&=E((D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0})(Y_i^1-Y_i^0))\\
&=E\left(E((D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0})(Y_i^1-Y_i^0)|(D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0}))\right)\\
&=1\cdot P(D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0}=1)\cdot E\left(Y_i^1-Y_i^0\middle|\,D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0}=1\right)\\
&\quad-1\cdot P(D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0}=-1)\cdot E\left(Y_i^1-Y_i^0\middle|\,D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0}=-1\right)\\
&\quad+0\cdot P(D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0}=0)\cdot E\left(Y_i^1-Y_i^0|D_i^{1\ldots1\ldots1}-D_i^{0\ldots0\ldots0}=0\right)\\
&=E(Y_i^1-Y_i^0|D_i^{1\ldots1\ldots1}>D_i^{0\ldots0\ldots0})\cdot P(D_i^{1\ldots1\ldots1}>D_i^{0\ldots0\ldots0})\\
&\quad-E(Y_i^1-Y_i^0|D_i^{1\ldots1\ldots1}<D_i^{0\ldots0\ldots0})\cdot P(D_i^{1\ldots1\ldots1}<D_i^{0\ldots0\ldots0}).
\end{aligned}
$$

LiM rules out the second part (if LiM is violated then, similar to setting with one binary instrument, treatment effects might be positive for all individuals, but the effect of the defiers cancels out the effect of the compliers). Rewriting leads to the CC-LATE:

$$
\begin{aligned}
&E\left(Y_i|Z_{1i}=1,\ldots,Z_{ki}=1\right)-E\left(Y|Z_{1i}=0,\ldots,Z_{ki}=0\right)\\
&=E(Y_i^1-Y_i^0|D_i^{1\ldots1\ldots1}>D_i^{0\ldots0\ldots0})\cdot P(D_i^{1\ldots1\ldots1}>D_i^{0\ldots0\ldots0})
\end{aligned}
$$

Then, rewrite

$$
E(Y_i^1-Y_i^0|D_i^{1\ldots1\ldots1}>D_i^{0\ldots0\ldots0})=\frac{E\left(Y_i|Z_{1i}=1,\ldots,Z_{ki}=1\right)-E\left(Y|Z_{1i}=0,\ldots,Z_{ki}=0\right)}{P(D_i^{1\ldots1\ldots1}>D_i^{0\ldots0\ldots0})}
$$

as follows:

$$
\begin{aligned}
E(Y^1 - Y^0|T \in cc) &= \frac{E\left(Y_i|Z_{1i} = 1, \ldots, Z_{ki} = 1\right) - E\left(Y|Z_{1i} = 0, \ldots, Z_{ki} = 0\right)}{P(D_i^{1\ldots1\ldots1} - D_i^{0\ldots0\ldots0} = 1)} \\
E(Y^1 - Y^0|T \in cc) &= \frac{E\left(Y_i|Z_{1i} = 1, \ldots, Z_{ki} = 1\right) - E\left(Y|Z_{1i} = 0, \ldots, Z_{ki} = 0\right)}{P(D_i^{1\ldots1\ldots1}|Z_{1i} = 1, \ldots, Z_{ki} = 1) - P(D_i^{0\ldots0\ldots0} = 1|Z_{1i} = 0, \ldots, Z_{ki} = 0)} \\
E(Y^1 - Y^0|T \in cc) &= \frac{E\left(Y_i|Z_{1i} = 1, \ldots, Z_{ki} = 1\right) - E\left(Y|Z_{1i} = 0, \ldots, Z_{ki} = 0\right)}{E\left(D|Z_1 = 1, \ldots, Z_k = 1\right) - E\left(D|Z_1 = 0, \ldots, Z_k = 0\right)}.
\end{aligned}
$$

$0 \cdot P(D_i^{1\ldots1\ldots1} - D_i^{0\ldots0\ldots0} = 0) \cdot E\left(Y_i^1 - Y_i^0|D_i^{1\ldots1\ldots1} - D_i^{0\ldots0\ldots0} = 0\right)$ demonstrates the fact that the CC-LATE does not capture the effect for those individuals for whom a change from being exposed to none of the instruments to being exposed to all instruments simultaneously does not change the treatment status, meaning that this change is not informative for these individuals. The always-takers and never-takers belong to this group.

## A.5 Extension to unordered instruments

The concepts of LiM and the CC-LATE can be naturally extended to settings with ordered instruments. Consider, for instance, two instruments: $Z_1 \in \{0, 1\}$ and $Z_2 \in \{0, 1, 2\}$. Then, a version of LiM can be imposed where $P(D^{1,2} \geq D^{0,0}) = 1$. The combined complier population comprises all types for which $D^{1,2} > D^{0,0}$. The LATE for this subpopulation is identified as follows:

$$
E(Y^1 - Y^0|T \in cc) = \frac{E\left(Y|Z_1 = 1, Z_2 = 2\right) - E\left(Y|Z_1 = 0, Z_2 = 0\right)}{E\left(D|Z_1 = 1, Z_2 = 2\right) - E\left(D|Z_1 = 0, Z_2 = 0\right)}.
$$

Note that for estimation, observations where the instrument values equal $(z_1, z_2) = (1, 2)$ or $(z_1, z_2) = (0, 0)$ are used. It should be noted that as the number of levels that the ordered instrument can attain increases, the number of observations available for estimation is likely to decrease.[12]

[12] We are grateful to the anonymous reviewer for providing these insights.

## A.6 CC-LATE under IAM

In their appendix, Imbens and Angrist (1994) state that, under IAM,

$$E(Y|Z = z_K) = E(Y|Z = z_0) + \alpha_{z_K,z_0} \cdot (P(z_K) - P(z_0)).$$

We can rewrite this as follows:

$$\frac{E(Y|Z = z_K) - E(Y|Z = z_0)}{P(z_K) - P(z_0)} = \alpha_{z_K,z_0}$$

$$\Downarrow$$

$$\frac{E(Y|Z = z_K) - E(Y|Z = z_0)}{E(D|Z = z_K) - E(D|Z = z_0)} = E(Y(1) - Y(0)|D(z_K) \neq D(z_0))$$

$$\Downarrow$$

$$\frac{E(Y|Z = z_K) - E(Y|Z = z_0)}{E(D|Z = z_K) - E(D|Z = z_0)} = \frac{\sum_{l=1}^{K} \alpha_{z_l,z_{l-1}} \cdot (P(z_l) - P(z_{l-1}))}{P(z_K) - P(z_0)}$$

$$\Downarrow$$

$$\frac{E(Y|Z = z_K) - E(Y|Z = z_0)}{E(D|Z = z_K) - E(D|Z = z_0)} = \sum_{l=1}^{K} \frac{P(z_l) - P(z_{l-1})}{P(z_K) - P(z_0)} \cdot \alpha_{z_l,z_{l-1}}.$$

$\frac{E(Y|Z=z_K)-E(Y|Z=z_0)}{E(D|Z=z_K)-E(D|Z=z_0)} = E(Y(1) - Y(0)|D(z_K) \neq D(z_0))$ shows that this can be interpreted as the effect in the largest group of compliers. This is the same interpretation as the estimand for multiple binary instruments as proposed by Frölich (2007).

Suppose we have two binary instruments and the support $z_0 = (0, 0),\ z_1 = (0, 1),\ z_2 = (1, 0),\ z_3 = (1, 1)$, ordered such that $l < m$ implies $P_l < P_m$. Then the final line in the last expression can be re-written as:

$$
\begin{aligned}
\alpha_{30} =& \frac{(P_{z_1} - P_{z_0}) \cdot \alpha_{z_1 z_0} + (P_{z_2} - P_{z_1}) \cdot \alpha_{z_2 z_1} + (P_{z_3} - P_{z_2}) \cdot \alpha_{z_3 z_2}}{P_{z_3} - P_{z_0}} \\
=& \frac{(P_{z_1} - P_{z_0})}{P_{z_3} - P_{z_0}} \cdot \frac{E(Y|Z = z_1) - E(Y|Z = z_0)}{P_{z_1} - P_{z_0}} \\
&+ \frac{(P_{z_2} - P_{z_1})}{P_{z_3} - P_{z_0}} \cdot \frac{E(Y|Z = z_2) - E(Y|Z = z_1)}{P_{z_2} - P_{z_1}} \\
&+ \frac{(P_{z_3} - P_{z_2})}{P_{z_3} - P_{z_0}} \cdot \frac{E(Y|Z = z_3) - E(Y|Z = z_2)}{P_{z_3} - P_{z_2}} \\
=& \frac{E(Y|Z = z_1) - E(Y|Z = z_0)}{P_{z_3} - P_{z_0}} + \frac{E(Y|Z = z_2) - E(Y|Z = z_1)}{P_{z_3} - P_{z_0}} \\
&+ \frac{E(Y|Z = z_3) - E(Y|Z = z_2)}{P_{z_3} - P_{z_0}} \\
=& \frac{E(Y|Z = z_3) - E(Y|Z = z_0)}{E(D|Z = z_3) - E(D|Z = z_0)}.
\end{aligned}
$$

# B Supplementary results for HIV application

## B.1 Testing for negative weights

We use Mogstad et al.'s (2021) approach to check whether the weights remain positive under PM when IAM is violated through the presence of both $Z_1$ and $Z_2$ compliers. They are positive under a violation of this assumption if the correlation between the treatment and the instruments is positive and significant, and the partial correlation between the instruments is significant. We follow their approach and regress the treatment on each instrument separately. We also regress $Z_1$ on $Z_2$ and $Z_3$ separately, and $Z_2$ on $Z_3$. The results are presented in Table 5. The correlation between the *distance* instrument and the treatment is not significant (see Column (2) of Table 5). The partial correlation between the *above median cash* and *distance* instruments is also not positive (see Column (6) of Table 5). This indicates that TSLS might contain negative weights when the IAM assumption is replaced by the weaker PM assumption.

We perform two tests on the TSLS weights. We cannot reject the hypothesis that all weights are positive when performing TSLS with the two instruments, *any cash* instrument and *distance* instrument.[13] At the same time, we do not reject the hypothesis that one of the weights in the weighted average generated by TSLS is negative, finding a p-value of 0.207. This is concerning, since one or more of the weights being negative would complicate the interpretation of the TSLS estimates.

Table 5: Testing for negative TSLS weights when both $Z_1$ and $Z_2$ compliers exist and IAM is relaxed to PM. Each column shows the coefficient from a regression of the column on the variable in the row including a constant. Significance levels: * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Got results | Got results | Got results | Any cash | Any cash | Distance |
| Any cash | 0.425*** | | | | | |
| (Std. err.) | (0.032) | | | | | |
| Distance | | 0.024 | | 0.003 | | |
| (Std. err.) | | (0.029) | | (0.027) | | |
| Median cash | | | 0.303*** | | 0.343*** | -0.003 |
| (Std. err.) | | | (0.027) | | (0.024) | (0.031) |

[13]Using the *mivcausal* package (Lau and Torgovitsky, 2020), we obtain a p-value of 0.855 using 1000 repetitions in the bootstrap.

## B.2 Tables with the estimates of the HIV application

Table 6: Estimates corresponding to Figure 2a.

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| **Panel A: Bought condoms** | | | | | | |
| Estimates | 0.024 | 0.288 | 0.228 | 1.854 | 0.170 | 0.011 |
| (Std. err.) | (0.033) | (0.157) | (0.139) | (6.424) | (0.116) | (0.093) |
| Nr. obs. | 1008 | 432 | 278 | 1008 | 1008 | 1008 |
| **Panel B: Number of condoms bought** | | | | | | |
| Estimates | -0.035 | 0.940 | 0.799 | 5.906 | 0.521 | -0.199 |
| (Std. err.) | (0.139) | (0.489) | (0.662) | (144.096) | (0.404) | (0.4) |
| Nr. obs. | 1008 | 432 | 278 | 1008 | 1008 | 1008 |
| **Panel C: Reported buying condoms** | | | | | | |
| Estimates | -0.009 | 0.096 | 0.161 | 2.070 | -0.022 | 0.051 |
| (Std. err.) | (0.025) | (0.087) | (0.069) | (2.965) | (0.06) | (0.046) |
| Nr. obs. | 1008 | 432 | 278 | 1008 | 1008 | 1008 |
| **Panel D: Reported having sex** | | | | | | |
| Estimates | 0.032 | 0.023 | 0.022 | 0.22 | 0.019 | 0.054 |
| (Std. err.) | (0.033) | (0.146) | (0.114) | (20.002) | (0.063) | (0.116) |
| Nr. obs. | 1008 | 432 | 278 | 1008 | 1008 | 1008 |

The columns give the estimates for the different methods: (1) $\hat{\beta}_{OLS}$, (2) $\hat{\beta}_{CC-LATE-2}$, (3) $\hat{\beta}_{CC-LATE-3}$, (4) $\hat{\beta}_{TSLS-above-1.5km-distance}$, (5) $\hat{\beta}_{TSLS-any-cash}$, and (6) $\hat{\beta}_{TSLS-above-median-cash}$.

The standard errors are clustered at the village level.

Table 7: Estimates corresponding to Figure 2b.

| | $\hat{\beta}_{CC-LATE-2}$ | $\hat{\beta}_{CC-LATE-3}$ | $\hat{\beta}_{TSLS-2}$ | $\hat{\beta}_{TSLS-3}$ |
|---|---|---|---|---|
| **Panel A: Bought condoms** | | | | |
| Estimates | 0.288 | 0.228 | 0.177 | 0.118 |
| (Std. err.) | (0.157) | (0.139) | (0.135) | (0.106) |
| Nr. obs. | 432 | 278 | 1008 | 1008 |
| **Panel B: Number of condoms bought** | | | | |
| Estimates | 0.94 | 0.799 | 0.543 | 0.278 |
| (Std. err.) | (0.489) | (0.662) | (0.478) | (0.337) |
| Nr. obs. | 432 | 278 | 1008 | 1008 |
| **Panel C: Reported buying condoms** | | | | |
| Estimates | 0.096 | 0.161 | -0.013 | 0.012 |
| (Std. err.) | (0.087) | (0.069) | (0.044) | (0.052) |
| Nr. obs. | 432 | 278 | 1008 | 1008 |
| **Panel D: Reported having sex** | | | | |
| Estimates | 0.023 | 0.022 | 0.02 | 0.032 |
| (Std. err.) | (0.146) | (0.114) | (0.095) | (0.058) |
| Nr. obs. | 432 | 278 | 1008 | 1008 |

The standard errors are clustered at the village level.

### B.3 Figure 2a without the distance instrument

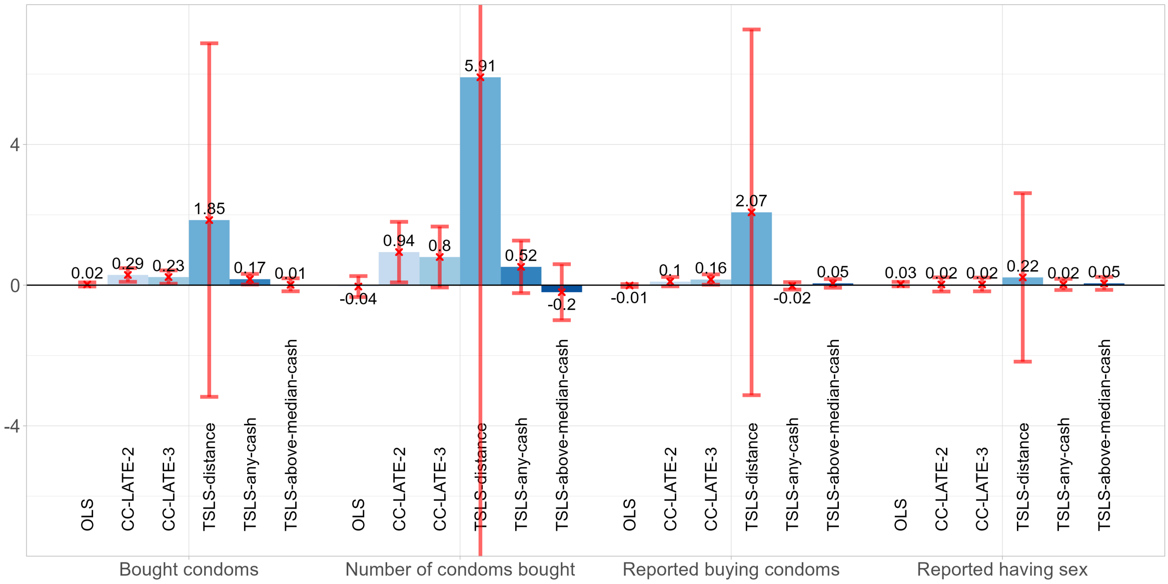



Figure 3: Comparison of CC-LATE estimates to the OLS estimates and the TSLS estimates resulting from using each instrument separately. The confidence intervals for *TSLS distance* for the outcome "number of condoms bought" is $[-12.66, 24.48]$.

## C Simulation study

In this section, we perform two different simulation studies to judge the finite sample performance of our CC-LATE estimator. First, we compare the CC-LATE estimator to the TSLS estimator in DGPs where PM is valid and others where PM is violated. Second, we compare the performance of the CC-LATE estimator when adding a weak versus strong third instrument.

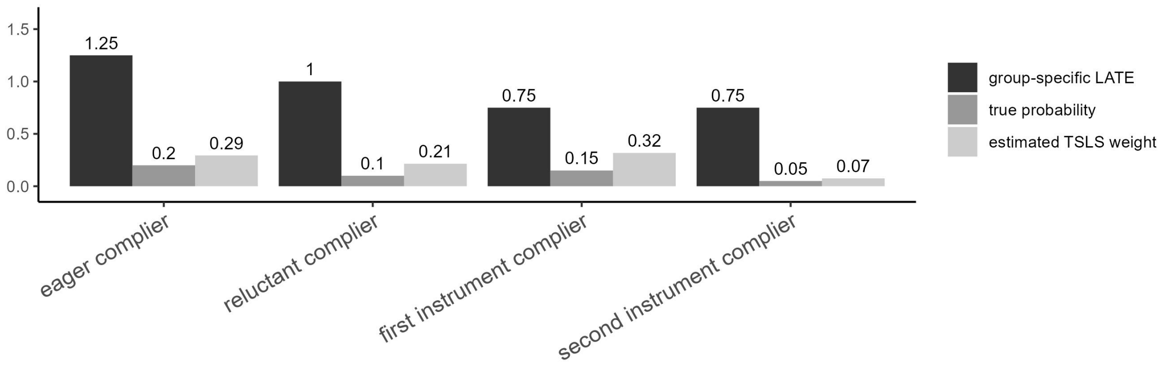
## C.1 Comparison of the CC-LATE and TSLS estimators when PM is violated

### C.1.1 Setup

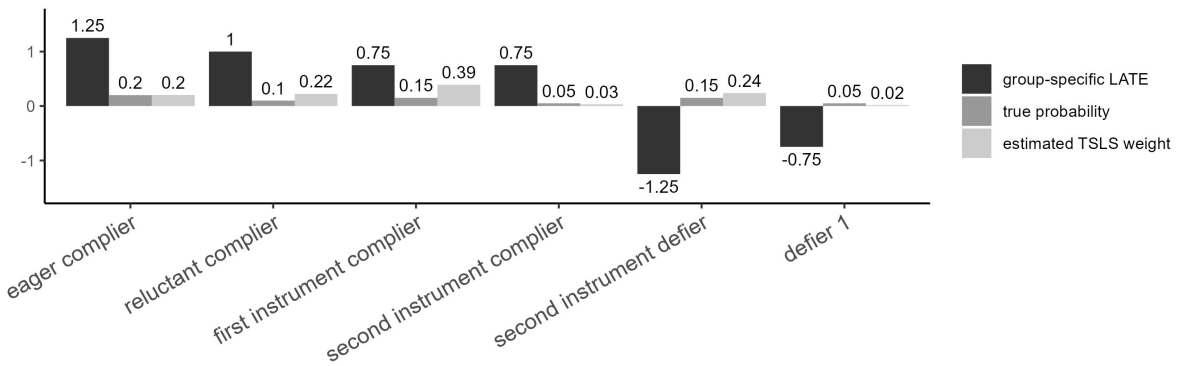
Following the idea of an empirical Monte Carlo study as in Huber et al. (2013), the DGP of the simulation largely depends on the real data of the HIV application studied in Section 4. We investigate the performance of the CC-LATE and TSLS estimator in two different settings. In the first setting, PM is valid. In the second setting, PM is violated due to the presence of defier types. Potential threats in the HIV application are the existence of second instrument defiers or defiers of type 1. This could lead to a violation of PM, while LiM would still hold.

Figure 4 depicts the true probabilities and the average effects per response type used in the simulation.[14] In Section 4, the estimated CC-LATE for the number of condoms bought when using two instruments is 0.8, and we use similar values for choosing the group-specific LATEs, $\beta_{t_i}$, of each response type. The probabilities of belonging to a certain response type are chosen based on the information that can be obtained from the HIV application. Under LiM, the response group proportions $\pi_{rd} + \pi_{d1} + \pi_{d2} + \pi_{nt}$ and $\pi_{at} + \pi_{1d} + \pi_{2d} + \pi_{ed}$ can be estimated. Under PM, the defier types are ruled out such that $\pi_{nt}$ and $\pi_{at}$ can also be estimated. We estimate these probabilities for the HIV application. We further use the estimated shares of the complier population from Figure 1 in Section 4. With these group-specific LATEs and pre-defined probabilities, the true value of the LATE for the combined compliers equals 1.

[14] Figure 4 also contains the estimated TSLS weights using equations (20) and (21) from the proof of Proposition 7 in Mogstad et al. (2021). To calculate the weights, propensity scores are predicted nonparametrically. The weights do not exactly add up to one, since they are estimated. The weights are non-negative, since our simulation study considers the setting where the instruments are monotonic in the propensity score, which is the most realistic scenario considering the HIV application.




(a) True LATEs, true weights and estimated TSLS weights when PM is not violated.




(b) True LATEs, true weights and estimated TSLS weights when PM is violated.

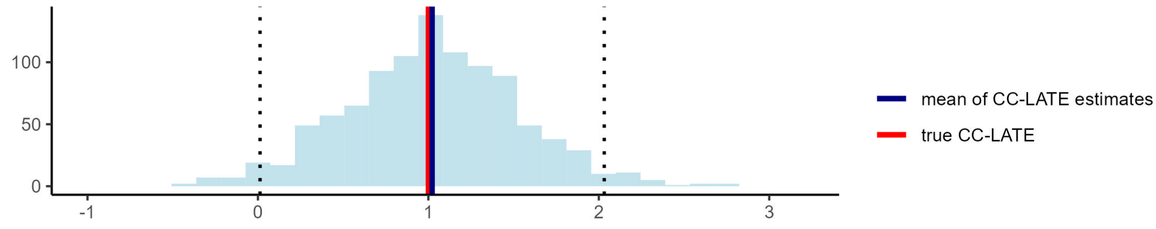
Figure 4: This figure contains the true LATEs and true weights used in the simulation study. It further shows the estimated TSLS weights when PM holds compared to when it is violated.

The sample size is $n = 1000$, which is similar to the $1,008$ observations of the HIV application. The instruments, $Z_1$ and $Z_2$, are drawn from a Bernouilli distribution with the probability set to the mean of the two binary instruments from the application, *any cash* and *distance*. Similar to the application where the instruments are randomized, the instruments are independent. The response types, $t_i$, are sampled with the pre-defined probabilities. The value of $D_i$ is then set based on the sampled response type and the instrument values. In the sample of untreated individuals, we calculate the mean, $m_y$, and the variance, $v_y$, of the outcome on the number of condoms bought. Then, $Y_i(0) = m_y$ and $Y_i = m_y + \beta_{t_i} D_i + \nu_i$, where $\nu_i \sim N(0, \upsilon_y)$. We perform 1000 simulation repetitions.

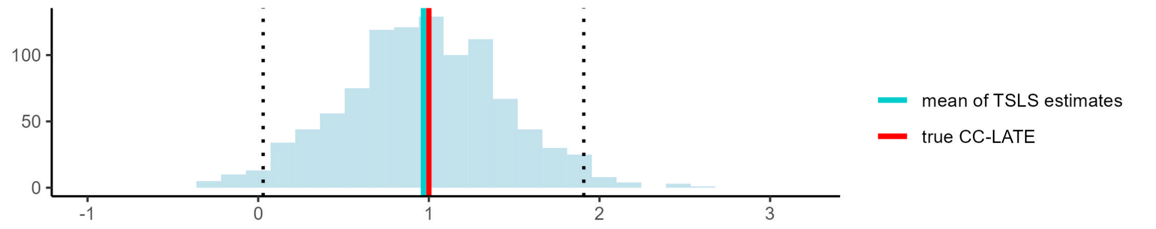
### C.1.2 Results

We compare the performance of the CC-LATE estimator and the TSLS estimator when PM is violated due to the presence of defier types. The estimates are compared to the true value of the LATE for the combined compliers, assuming that the objective of both methods is to give an estimate of the ATE for this subpopulation. Note that this objective is true for TSLS if PM is imposed such that increasing the instrument values weakly increases treatment uptake, as in Section 4.

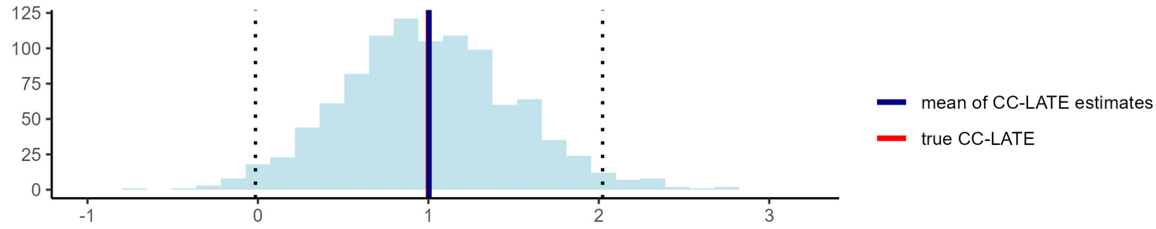
The distributions of the estimates are depicted in Figure 5, and Table 8 gives the distance to the CC-LATE, median distance, mean absolute error (MAE) with respect to the true CC-LATE, and mean squared error (MSE) with respect to the true CC-LATE. It is valuable to compare the TSLS estimand with the CC-LATE, which is also a weighted average of the LATEs for the same types but with weights accurately reflecting their relative population shares. The MSE and MAE of the CC-LATE and TSLS estimator are comparable, since the CC-LATE estimates lie closer to the true value, but are more spread out than the TSLS estimates. When PM holds, both the CC-LATE estimator and the TSLS estimator lie close to the true LATE for the combined compliers. Even though the CC-LATE estimator uses fewer observations, the standard deviation of the estimates of the two methods is comparable. Violation of PM clearly introduces downward bias in the TSLS estimates, since it now includes the LATEs of the second instrument defiers and the defiers of type 1. Interestingly, this might also explain the smaller coefficients found with TSLS in Section 4, which provides some informal evidence in favor of the existence of defier types in the HIV application. As LiM still holds in the presence



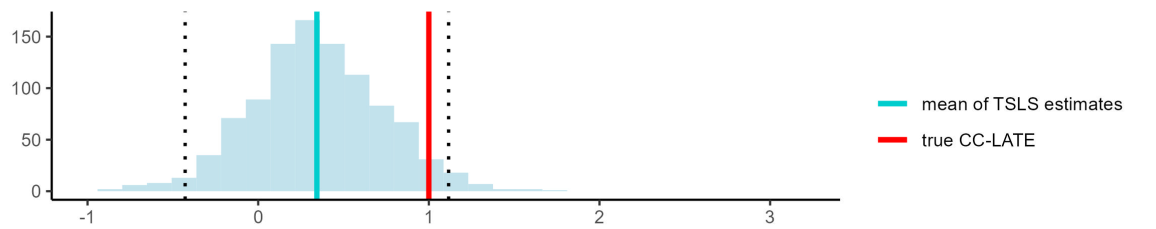

(a) Distribution of CC-LATE estimates when PM is valid.




(b) Distribution of TSLS estimates when PM is valid.




(c) Distribution of CC-LATE estimates when PM is violated.




(d) Distribution of TSLS estimates when PM is violated.

Figure 5: This figure compares the distributions of CC-LATE and TSLS estimates when PM is valid versus when PM is violated. The 95% quantiles of the Monte Carlo distribution are indicated by dashed lines.

Table 8: This table contains the estimates and measures compared to the true LATE for the combined complier population when PM is valid and when PM is violated.

| | **(1)** | | **(2)** | |
|---|---|---|---|---|
| | **PM valid** | | **PM violated** | |
| | CC-LATE estimator | TSLS estimator | CC-LATE estimator | TSLS estimator |
| Mean of estimates | 1.022 | 0.969 | 1.004 | 0.344 |
| Mean of standard errors | 0.493 | 0.465 | 0.513 | 0.377 |
| Std. dev. of estimates | 0.505 | 0.470 | 0.509 | 0.386 |
| Distance to CC-LATE | 0.022 | -0.031 | 0.004 | -0.656 |
| Median distance to CC-LATE | 0.014 | -0.035 | -0.014 | -0.667 |
| MSE (compared to CC-LATE) | 0.255 | 0.221 | 0.259 | 0.579 |
| MAE (compared to CC-LATE) | 0.397 | 0.373 | 0.402 | 0.672 |

of the introduced defier types, the bias of the CC-LATE estimator remains small when PM is violated.

## C.2 Comparison of the CC-LATE and TSLS estimators for different scenarios

In this section, we consider settings that negatively impact the performance of the CC-LATE estimator: First, unusual correlations between instruments; second, highly heterogeneous treatment effects; and third, all instruments generate few compliers.

### C.2.1 Setup

The experimental setup closely follows the framework outlined in Section C.1.1. In the first scenario, characterized by unusual correlations, we adopt the methodology from Goff (2020), implementing a transformation where $Z_{2i} = 0$ with a 95% probability when $Z_{1i} = 1$. In the second scenario, we introduce highly heterogeneous treatment effects: the local effect is 5 for eager compliers, -3 for reluctant compliers, 8 for first instrument compliers, -6 for second instrument compliers, -2 for second instrument defiers, and 3 for defiers of type 1. Consequently, the true CC-LATE is 3.2. For the scenario involving instruments that generate few compliers, we set the proportion of all complier types to 2.5%. When PM (Principal Monotonicity) holds, the proportions of always-takers and never-takers are 45% each. Conversely, when PM is violated, the proportions of always-takers and never-takers decrease to 35% each, with the proportions of second instrument defiers and defiers of type 1 increasing to 5% and 15%, respectively.

### C.2.2 Results

Table 9 presents the results of our simulation study. Panel A illustrates that introducing an unusual correlation between the instruments increases both the bias and the variance of the CC-LATE estimates compared to the results shown in Table 8, while it has a lesser impact on the TSLS estimates. This might to some extent be driven by the fact that a negative correlation between instruments also reduces the overlap in instrument values, hence substantially reducing the sample used for estimation of the CC-LATE. Moving to Panel B, we observe that the CC-LATE estimates perform reasonably well when treatment effects are highly heterogeneous, suggesting some robustness under these conditions. Panel C demonstrates that both TSLS and CC-LATE estimates are negatively impacted when all instruments are weak. However, the impact is more pronounced on the bias and variance of the CC-LATE estimator, indicating that at least one strong instrument is necessary for reasonable performance of the CC-LATE estimator.

Overall, this simulation exercise highlights the importance of exercising caution with the

CC-LATE estimator in the presence of unusual correlations between the instruments and when all instruments generate few compliers. In these scenarios, the TSLS estimator outperforms the CC-LATE estimator in terms of both bias and variance.

Table 9: This table contains the estimates and measures compared to the true LATE for the combined complier population when PM is valid and when PM is violated in three different scenarios: first, unusual correlations between instruments; second, highly heterogeneous treatment effects; and third, all instruments generate few compliers.

| | **(1)** | | **(2)** | |
|---|---|---|---|---|
| | **PM valid** | | **PM violated** | |
| Estimator | CC-LATE | TSLS | CC-LATE | TSLS |
| **Panel A: Unusual correlation between instruments** | | | | |
| (Estimated) true value | 1 | 0.958 | 1 | 0.191 |
| Mean of estimates | 1.050 | 1.017 | 1.116 | 0.081 |
| Mean of standard errors | 1.498 | 0.659 | 2.168 | 0.465 |
| Std. dev. of estimates | 1.515 | 0.671 | 2.190 | 0.455 |
| Distance to CC-LATE | 0.050 | 0.017 | 0.116 | -0.919 |
| Median distance | 0.018 | -0.008 | 0.041 | -0.915 |
| MSE (compared to CC-LATE) | 2.294 | 0.450 | 4.807 | 1.052 |
| MAE (compared to CC-LATE) | 1.167 | 0.524 | 1.240 | 0.927 |
| (Mean) nr. obs. | 147 | 1,000 | 147 | 1,000 |
| **Panel B: Highly heterogeneous treatment effects** | | | | |
| (Estimated) true value | 3.2 | 2.916 | 3.2 | 2.848 |
| Mean of estimates | 3.247 | 3.473 | 3.208 | 2.420 |

*Continued on next page*

| | | | | |
|---|---|---|---|---|
| | | | *Continued from previous page* | |
| Mean of standard errors | 0.849 | 0.753 | 0.894 | 0.602 |
| Std. dev. of estimates | 0.731 | 0.791 | 0.746 | 0.643 |
| Distance to CC-LATE | 0.047 | 0.273 | 0.008 | -0.780 |
| Median distance | 0.020 | 0.259 | -0.011 | -0.770 |
| MSE (compared to CC-LATE) | 0.537 | 0.700 | 0.556 | 1.021 |
| MAE (compared to CC-LATE) | 0.572 | 0.657 | 0.593 | 0.854 |
| (Mean) nr. obs. | 427 | 1,000 | 427 | 1,000 |
| **Panel C: All instruments generate few compliers** | | | | |
| (Estimated) true value | 0.938 | 0.533 | 0.938 | 0.140 |
| Mean of estimates | 1.289 | 0.594 | 0.062 | -0.840 |
| Mean of standard errors | 248.829 | 2.555 | 265.549 | 0.809 |
| Std. dev. of estimates | 43.908 | 2.477 | 32.410 | 0.831 |
| Distance to CC-LATE | 0.351 | -0.344 | -0.876 | -1.777 |
| Median distance | -0.204 | -0.328 | -0.069 | -1.760 |
| MSE (compared to CC-LATE) | 1,926.091 | 6.246 | 1,050.141 | 3.849 |
| MAE (compared to CC-LATE) | 6.182 | 1.769 | 5.787 | 1.789 |
| (Mean) nr. obs. | 427 | 1,000 | 427 | 1,000 |

## C.3 Impact of instrument quantity on the variance

In this section, we investigate how the variance expression of Equation (2) evolves as the number of instruments increases. To gain a deeper understanding, we conduct a simulation study.

### C.3.1 Setup

In this study, we examine a basic framework where independent instruments are sampled from a Bernoulli distribution with a probability of 0.5. The treatment effect is assumed to be homogeneous, with a true value of 0.5, and the response types are evenly distributed. To ensure computational efficiency, we limit the analysis to a maximum of 10,000 randomly selected response types, as permitted under the LiM framework. Specifically, the outcome under the control condition is set to zero ($Y_i(0) = 0$), and the observed outcome is modeled as: $Y_i = 0.5 \cdot D_i + \nu_i$, where $\nu_i \sim N(0, 1)$. We conduct 100 simulation repetitions for two distinct sample sizes: $n = 1,000$ and $n = 10,000$.

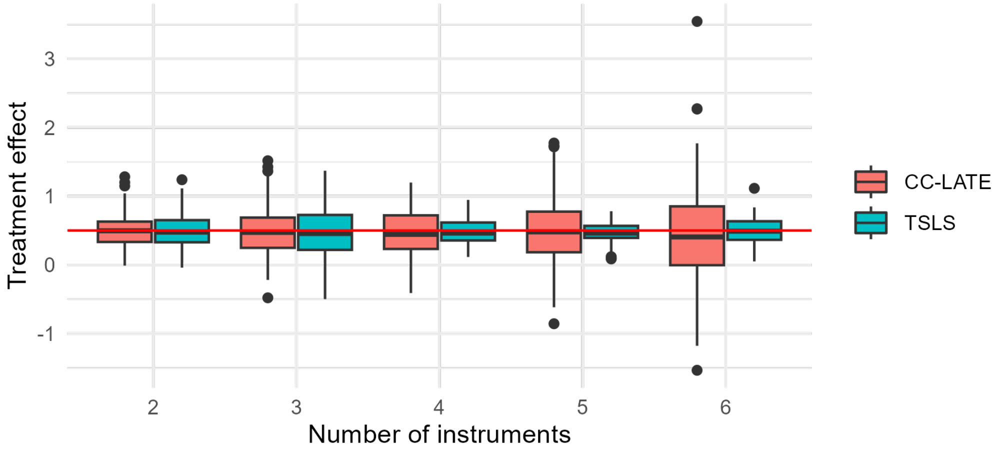
### C.3.2 Results

Figure 6 showcases the estimate distributions, while Table 10 presents the mean sample sizes used for estimating the CC-LATE. The results indicate that in the setting considered, using up to three instruments can provide a modest gain in precision, reflected by a reduced variance. However, when four or more instruments are employed, this advantage of the CC-LATE diminishes. These findings suggest that CC-LATE may offer precision gains when a few instruments are used, but this benefit becomes less pronounced as the number of instruments increases.

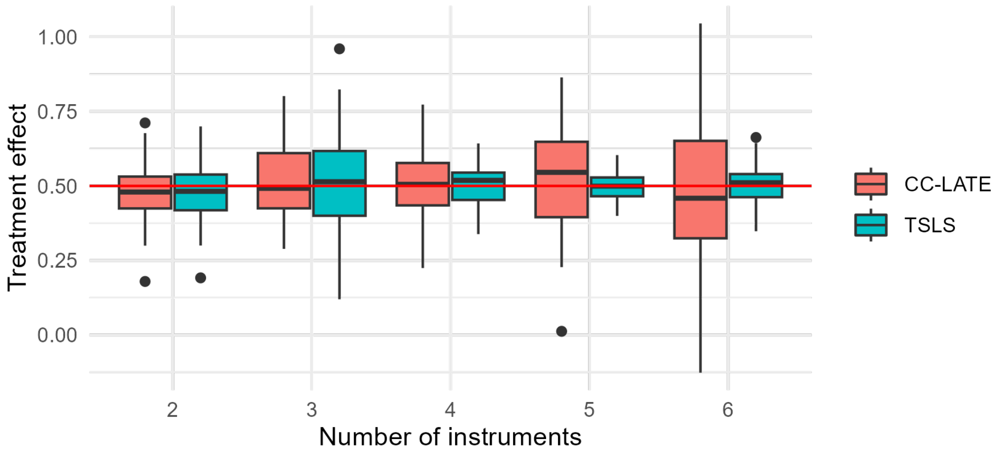
## C.4 Comparing CC-LATE estimators when adding a third (weak) instrument

### C.4.1 Setup

In this section, we study the performance of the CC-LATE estimator in two different settings where a third instrument is available. The DGPs are similar to the DGPs in Section C.1. In




(a) $n = 1,000$.




(b) $n = 10,000$.

Figure 6: This figure displays the distribution of the TSLS and CC-LATE estimates across varying numbers of instruments. The red line marks the true treatment effect of 0.5.

the first setting, the third instrument is extremely weak in that it pushes none of the individuals to compliance. The third instrument, $Z_3$, is drawn from a Bernouilli distribution with the probability equal to the mean of the *above median cash* instrument from the HIV application. The types considered in this simulation study are given in Table 11. The response types are chosen such that there are only compliers with respect to $Z_1$ and $Z_2$. Using similar notions as in the setting with two instruments, these are the eager compliers, reluctant compliers, first instrument compliers, and second instrument compliers with respect to $Z_1$ and $Z_2$. In the second setting, the third instrument is strong and adds compliers that only respond to this instrument. The third instrument complier type always takes up treatment when exposed to the third instrument, but does not influence the complier population when exposed to $Z_1$ or $Z_2$, since these response types are either always-takers or never-takers when $Z_3$ is fixed. Table 12 presents all probabilities and group-specific LATEs used in the simulation. For the second setting, the probability of being a third instrument complier equals 20%. Therefore, in this setting, the third instrument pushes many individuals towards compliance.

### C.4.2 Results

We estimate the CC-LATE using either two or three instruments where the third instrument is either weak or strong. Figure 7 depicts the estimate distributions. Table 13 contains the estimate means and the standard deviations corresponding to Figure 7. When including a third instrument that does not add any compliers, the estimated CC-LATE lies close to the true LATE of the combined compliers, which consist of the eager compliers, reluctant compliers, first instrument compliers, and second instrument compliers in this case (see Figure 7a). Since adding a third instrument reduces the number of observations used for estimation, the confidence intervals are wider (see Figure 7b).

Table 10: This table presents the average subsample size used for estimating the CC-LATE across 100 simulation repetitions.

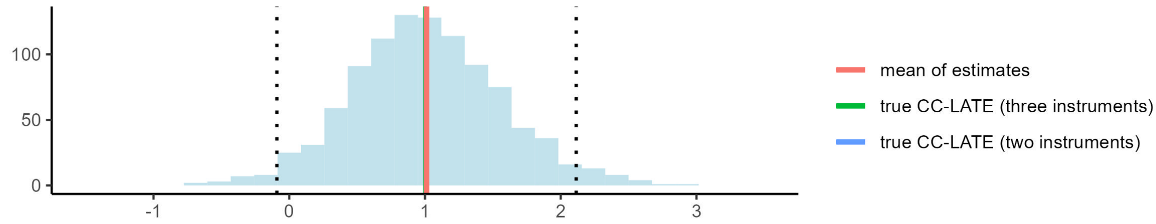
| Nr. instruments $k$ | $n_1 = 1,000$ | $n_2 = 10,000$ |
|---|---|---|
| 2 | 497.89 | 5,001.32 |
| 3 | 250.26 | 2,502.84 |
| 4 | 124.28 | 2,502.84 |
| 5 | 61.73 | 1,252.88 |
| 6 | 31.10 | 312.16 |

Table 11: Table with types considered in the simulation study.

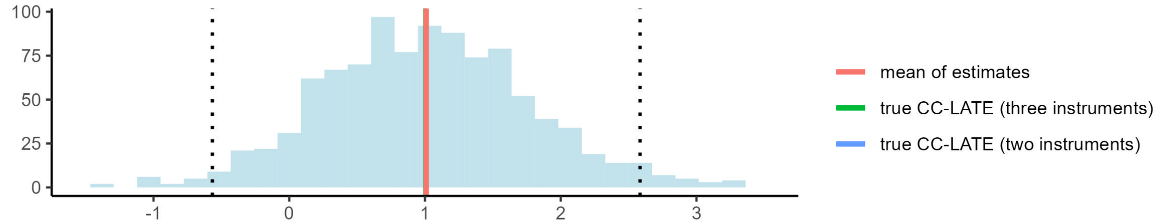
| $D^{111}$ | $D^{110}$ | $D^{101}$ | $D^{011}$ | $D^{100}$ | $D^{010}$ | $D^{001}$ | $D^{000}$ | Type when $Z_3 = 1$ | Type when $Z_3 = 0$ | Notion |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | Always-taker | Always-taker | Always-taker |
| 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | Eager complier | Eager complier | Eager complier |
| 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | Reluctant complier | Reluctant complier | Reluctant complier |
| 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | First instrument complier | First instrument complier | First instrument complier |
| 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | Second instrument complier | Second instrument complier | Second instrument complier |
| 1 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | Always-taker | Never-taker | Third instrument complier |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | Never-taker | Never-taker | Never-taker |

Table 12: Table with true average treatment effects and probabilities per response type. We compare the setting where the third instrument does not add compliers to the setting where it adds compliers.

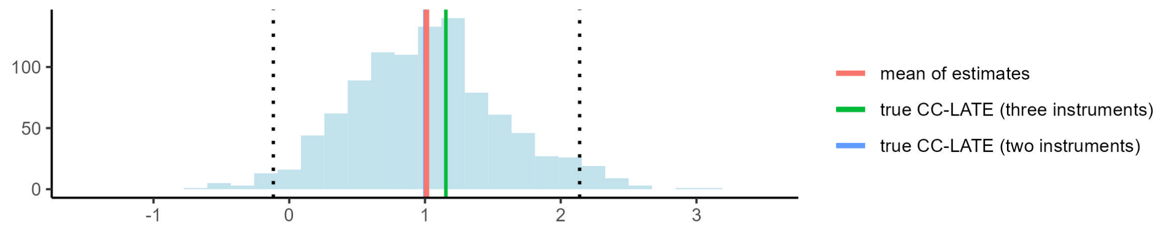
| | (1) | | (2) | |
|---|---|---|---|---|
| | **Third instrument does not add compliers** | | **Third instrument adds compliers** | |
| **Response type** | **Probability** | **True LATE** | **Probability** | **True LATE** |
| Always-taker | 0.4 | 0 | 0.3 | 0 |
| Eager complier | 0.2 | 1.25 | 0.2 | 1.25 |
| Reluctant complier | 0.05 | 0.5 | 0.05 | 0.5 |
| First instrument complier | 0.15 | 1 | 0.15 | 1 |
| Second instrument complier | 0.05 | 0.5 | 0.05 | 0.5 |
| Third instrument complier | | | 0.2 | 1.5 |
| Never-taker | 0.15 | 0 | 0.05 | 0 |
| **True CC-LATE two inst.** | 1 | | 1 | |
| **True CC-LATE three inst.** | 1 | | 1.154 | |



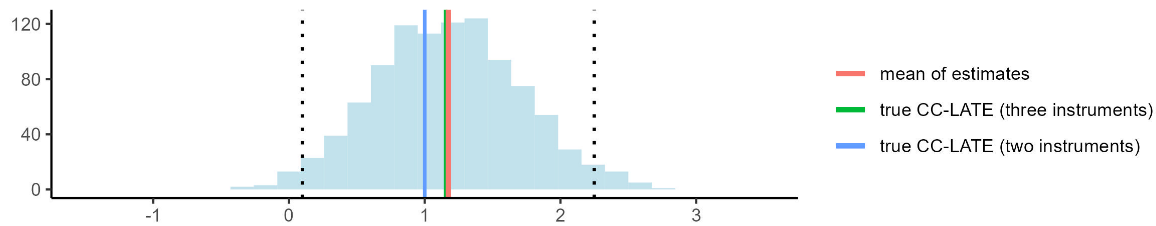

(a) Distribution of the CC-LATE estimates when using two instruments.




(b) Distribution of the CC-LATE estimates when using three instruments where the third instrument does not add any compliers.




(c) Distribution of the CC-LATE estimates when using two instruments and leaving out the third instrument when there are third instrument compliers present in the population.




(d) Distribution of the CC-LATE estimates when using three instruments where the third instrument adds third instrument compliers to the complier population.

Figure 7: This figure compares the distributions of the CC-LATE estimates for settings where two or three instruments are used and where the third instrument either adds to the complier population or does not add any compliers at all. The 95% quantiles of the Monte Carlo distribution are indicated by dashed lines.

Table 13: Table with CC-LATE estimates in case of two or three binary instruments for the setting where the third instrument does add third instrument compliers and the setting where it does not add compliers, corresponding to Figure 7.

| | **(1)** | | **(2)** | |
|---|---|---|---|---|
| | **Third instrument does not add compliers** | | **Third instrument adds compliers** | |
| | two instruments | three instruments | two instruments | three instruments |
| Mean of estimates | 1.012 | 1.009 | 1.011 | 1.174 |
| Mean of standard errors | 0.550 | 0.780 | 0.552 | 0.531 |
| Std. dev. of estimates | 0.551 | 0.788 | 0.564 | 0.537 |
| Bias | 0.012 | 0.009 | 0.011 | 0.020 |
| Median bias | -0.005 | -0.006 | -0.003 | 0.011 |
| MSE | 0.304 | 0.620 | 0.318 | 0.289 |
| MAE | 0.433 | 0.615 | 0.442 | 0.430 |

When third instrument compliers are present in the population, the mean of the CC-LATE estimates using only two instruments, $Z_1$ and $Z_2$, lies close to the true CC-LATE for the combined complier population with respect to these two instruments (see Figure 7c). Including a strong third instrument that adds third instrument compliers leads to an increase in the complier population considered. The resulting estimate gives the LATE for the eager compliers, reluctant compliers, first instrument compliers, and second instrument compliers as well as the third instrument compliers (see Figure 7d).

In conclusion, when an extremely weak instrument is added, the CC-LATE remains approximately unbiased but is less precise. When incorporating the additional instrument, the compliers that respond to this instrument are added to the complier population. While the precision remains approximately the same, the estimated LATE considers a larger subpopulation and hence might lie closer to the true ATE.

Table 14: Table with TSLS estimates in case of two or three binary instruments for the settings where the third instrument does add third instrument compliers and where it does not add compliers.

| | (1) | | (2) | |
|---|---|---|---|---|
| | **Third instrument does not add compliers** | | **Third instrument adds compliers** | |
| | two instruments | three instruments | two instruments | three instruments |
| Mean of estimates | 0.986 | 0.963 | 0.986 | 1.129 |
| Mean of standard errors | 0.526 | 0.515 | 0.528 | 0.428 |
| Std. dev. of estimates | 0.525 | 0.511 | 0.544 | 0.427 |

# D Comparison of LiM to other forms of monotonicity

This section compares LiM to other forms of monotonicity mentioned in the main text, specifically PM, VM and IAM. LiM is strictly weaker than IAM wen there are three instruments or more. To see this, consider the setting with two binary instruments: $Z_1 \in \{0, 1\}$ and $Z_2 \in \{0, 1\}$ with support $\mathcal{Z} = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. Since there are four different combinations of the instrument values, there are $\binom{4}{2} = 6$ comparisons of potential treatments, $d \in \{0, 1\}$. In other words, there are six selection probabilities $P(D^z \geq D^{z'}) = d$ with $z, z' \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$ and $z \neq z'$, that can be restricted by imposing some sort of monotonicity. IAM restricts all six comparisons. LiM always imposes only one restriction, independent of the number of instruments. To give an example, IAM imposes either $P(D^{10} \geq D^{01}) = 1$ or $P(D^{10} \leq D^{01}) = 1$. This translates to requiring that all individuals favor one instrument over the other instrument. Consequently, it is not possible to have some individuals who have a preference for $Z_1$ and other individuals who have a preference for $Z_2$. For instance, if all individuals are restricted to favor $Z_1$ over $Z_2$, then all the response types except the ones indicated in Table 1, are ruled out by IAM. In contrast, LiM allows for richer choice heterogeneity by allowing the presence of both first instrument compliers and second instrument compliers. Following the same line of reasoning, LiM is also less restrictive than IAM in settings with more than two binary instruments, as it does not impose any ordering on $P(D^{i...j...k} \geq D^{i...j...k}) \, \forall i \neq j \neq k$.

While IAM restricts all six comparisons of potential treatments for different instrument values in the case of two instruments, PM imposes four restrictions. PM requires each of the probabilities $P(D^{00} \geq D^{10})$, $P(D^{00} \geq D^{01})$, $P(D^{10} \geq D^{11})$, and $P(D^{01} \geq D^{11})$ to be either zero or one. Notice that only one of all possible PM assumptions can be consistent with the data. Estimating $E(D^{00})$, $E(D^{10})$, $E(D^{01})$, and $E(D^{11})$ reveals the version that is consistent with the considered data. This also provides a testable implication for VM, as one can check the direction of each inequalities. With two instruments, PM allows for at most seven different response types to co-exist. When increasing the values of the instruments makes participation weakly more likely, PM is equivalent to VM and imposes the following restrictions:

$$P(D^{10} \geq D^{00}) = 1, P(D^{01} \geq D^{00}) = 1, P(D^{01} \geq D^{11}) = 0, P(D^{10} \geq D^{11}) = 0. \quad (4)$$

These are equivalent to those stated in Equation (7) of Mogstad et al. (2021). The corresponding six response types consistent with the ordering in Equation (4) are given in Table 1. These choice restrictions rule out six defier types that LiM allows for.

It is straightforward to see that the restrictions in Equation (4) imply $P(D^{00} \geq D^{11}) = 0$, meaning PM and LiM are nested in this case, with LiM being strictly weaker. LiM remains strictly weaker than PM even if we reverse all inequalities in Equation (4); that is, LiM is strictly weaker than PM when increasing (or decreasing) instrument values consistently increases (or decreases) treatment uptake.

When PM do not impose any restriction on $P(D^{00} \geq D^{11})$, the two assumptions are non-nested. Note that in all these scenarios VM would also be violated. With two binary instruments, there are four possible versions of PM that are non-nested with either positive LiM, $P(D^{00} \leq D^{11}) = 1$, or negative LiM, $P(D^{00} \geq D^{11}) = 1$:

$$P(D^{10} \geq D^{00}) = 1,\ P(D^{01} \geq D^{00}) = 1,\ \text{and}\ P(D^{01} \geq D^{11}) = 1,\ P(D^{10} \geq D^{11}) = 1. \quad (5)$$

$$P(D^{10} \geq D^{00}) = 1,\ P(D^{01} \geq D^{00}) = 0,\ \text{and}\ P(D^{01} \geq D^{11}) = 0,\ P(D^{10} \geq D^{11}) = 1. \quad (6)$$

$$P(D^{10} \geq D^{00}) = 0,\ P(D^{01} \geq D^{00}) = 1,\ \text{and}\ P(D^{01} \geq D^{11}) = 1,\ P(D^{10} \geq D^{11}) = 0. \quad (7)$$

$$P(D^{10} \geq D^{00}) = 0,\ P(D^{01} \geq D^{00}) = 0,\ \text{and}\ P(D^{01} \geq D^{11}) = 0,\ P(D^{10} \geq D^{11}) = 0. \quad (8)$$

The response types that are present under these four different versions of the assumptions are listed in Table 15, together with the response types under positive and negative LiM. Clearly, in all four cases, LiM allows for substantially more choice heterogeneity than PM, allowing for a much larger number of different response types. For each of these four versions of PM, only one response type included under PM is excluded under LiM, at the cost of ruling out several other types. It is unlikely that this is a plausible scenario in empirical applications. As will be outlined below, justifying PM over LiM becomes even more difficult as the number of instruments increases.

Consider the three binary instrument setting with the three instruments $Z_1 \in \{0,1\}$, $Z_2 \in \{0,1\}$, and $Z_3 \in \{0,1\}$, and with support $\mathcal{Z} = \{(0,0,0),(0,0,1),(0,1,0),(1,0,0),(1,1,0),(0,1,1),(1,0,1),(1,1,1)\}$. Since VM is strictly stronger than PM we will only focus on PM for the rest of this section. Without imposing any restrictions, there are $2^8 = 256$ different response types, since there are eight different points of support of $Z$ for which the potential treatment status is compared pairwise. The eight different combinations of the instrument values result in $\binom{8}{2} = 28$ comparisons of potential treatments. LiM includes individuals who are compliers with respect to at least one of the instruments or a combination of instruments, but defiers for another instrument (or potentially multiple other instruments), as long as the treatment status when exposed to all instruments is at least as large as when exposed to none of the instruments. Imposing LiM ($P(D^{111} \geq D^{000}) = 1$ or $P(D^{111} \leq D^{000}) = 1$) rules out $64$ of the initial $256$ response types, allowing for a total of $192$ possible types.

The maximum number of response types under PM is only $35$, since it imposes more choice restrictions. PM imposes twelve restrictions in total that bring about $2^{12} = 4,096$ different versions of PM.[15] PM and LiM are nested in approximately 82% $(3,366/4,096 \approx 0.82)$ of these cases. In all those instances, LiM is strictly weaker than PM. PM seems rather unrealistic when it is non-nested with LiM, which entails the remaining $18\%$ of the versions of PM. These versions of PM only allow for either one, two or three additional response types excluded by LiM, at the cost of ruling out many other types that are included under LiM. In approximately $10\%$ of all cases $((730 - 324 - 12)/4,096)$, one response type is allowed for under PM that is ruled out under LiM. In approximately 8% $(324/4,096)$ of the cases, PM allows for two other response types. The maximum number of extra response types that PM allows for when non-nested with LiM is three, which occurs in 0.3% $(12/4,096)$ of the possible combinations that are consistent with the PM assumption.

The total possible number of response types is given by $2^{2^k}$. Under LiM, 75% of the response types are allowed for and 25% are ruled out, independently of the number of instruments, $k$. The combined compliers always consist of 25% of the total number of response types, meaning that $0.25 \cdot 2^{2^k}$ response types form the combined compliers. Calculating the number of response types under PM is more complicated, since the number of response types depends on the signs of the choice restrictions. Every choice restriction that is imposed eliminates at most 25% of the response types. The number of choice restrictions imposed by PM

[15] An R-script for the response types that are allowed for under the different monotonicity assumptions in case of three binary instruments is available from the authors upon request.

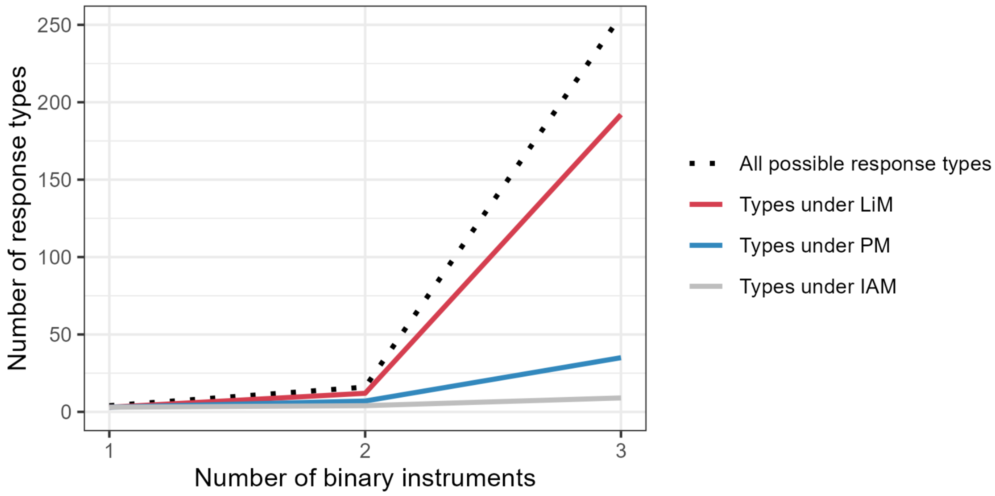



Figure 8: The maximum number of possible response types when one, two or three binary instruments are available under different versions of the monotonicity assumption is depicted. This figure shows that when more than one binary instrument is available, LiM imposes far fewer choice restrictions on the response types present in the population.

when $k$ instruments are available equals $k \cdot 2^{k-1} = \sum_{i=1}^{k} \binom{k}{i-1} \cdot (k-i-1)$.

A graphic illustration of the restrictiveness of other forms of monotonicity compared to LiM is given in Figure 8. This figure depicts the maximum number of types under each monotonicity assumption. It clearly demonstrates the advantage of imposing the LiM assumption, as the number of allowed response types increases rapidly with the available instruments. PM forces the researcher to make a choice between types. Another problem is that, depending on the types and the ordering of the propensity scores, some response types can lead to negative weights in the weighted average estimated by TSLS. This is further outlined in Appendix E.

Table 15: Principal strata and the response types in case of two binary instruments and a binary treatment when LiM and PM are non-nested.

| Type ($T$) | $D^{11}$ | $D^{10}$ | $D^{01}$ | $D^{00}$ | Notion | LiM (positive) | LiM (negative) | PM (Equation 5) | PM (Equation 6) | PM (Equation 7) | PM (Equation 8) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $at$ | 1 | 1 | 1 | 1 | Always-taker | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| $ec$ | 1 | 1 | 1 | 0 | Eager complier | ✓ | | [✓] | | | |
| $rc$ | 1 | 0 | 0 | 0 | Reluctant complier | ✓ | | | | | [✓] |
| $1c$ | 1 | 1 | 0 | 0 | First instrument complier | ✓ | | | [✓] | | |
| $2c$ | 1 | 0 | 1 | 0 | Second instrument complier | ✓ | | | | [✓] | |
| $1d$ | 1 | 0 | 1 | 1 | First instrument defier | ✓ | ✓ | | | ✓ | ✓ |
| $2d$ | 1 | 1 | 0 | 1 | Second instrument defier | ✓ | ✓ | | ✓ | | ✓ |
| $ed$ | 1 | 0 | 0 | 1 | Eager defier | ✓ | ✓ | | | | ✓ |
| $rd$ | 0 | 1 | 1 | 0 | Reluctant defier | ✓ | ✓ | ✓ | | | |
| $d1$ | 0 | 1 | 0 | 0 | Defier type 1 | ✓ | ✓ | ✓ | ✓ | | |
| $d2$ | 0 | 0 | 1 | 0 | Defier type 2 | ✓ | ✓ | ✓ | | ✓ | |
| $d3$ | 0 | 1 | 1 | 1 | Defier type 3 | | ✓ | (✓) | | | |
| $d4$ | 0 | 1 | 0 | 1 | Defier type 4 | | ✓ | | (✓) | | |
| $d5$ | 0 | 0 | 1 | 1 | Defier type 5 | | ✓ | | | (✓) | |
| $d6$ | 0 | 0 | 0 | 1 | Defier type 6 | | ✓ | | | | (✓) |
| $nt$ | 0 | 0 | 0 | 0 | Never-taker | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

✓ demonstrates the types allowed for under the respective forms of the monotonicity assumption.
(✓) denotes the one response type that is only allowed for under PM but excluded under positive LiM.
[✓] denotes the one response type that is only allowed for under PM but excluded under negative LiM.

# E Comparison of the CC-LATE to other estimands

The CC-LATE estimand is given by

$$\beta = \frac{E\left(Y|Z_1 = 1, \ldots Z_k = 1\right) - E\left(Y|Z_1 = 0, \ldots Z_k = 0\right)}{E\left(D|Z_1 = 1, \ldots Z_k = 1\right) - E\left(D|Z_1 = 0, \ldots Z_k = 0\right)}$$

when multiple binary instruments are available, and it is given by

$$\beta = \frac{E\left(Y \mid Z_1 = 1, Z_2 = 1\right) - E\left(Y \mid Z_1 = 0, Z_2 = 0\right)}{E\left(D \mid Z_1 = 1, Z_2 = 1\right) - E\left(D \mid Z_1 = 0, Z_2 = 0\right)}$$

in the case of two binary instruments.

When two binary instruments, $Z_1$ and $Z_2$, satisfy the standard assumptions including the IAM assumption, the Imbens and Angrist (1994) LATE estimands using each instrument separately are

$$\beta_1 = \frac{E\left(Y \mid Z_1 = 1\right) - E\left(Y \mid Z_1 = 0\right)}{E\left(D \mid Z_1 = 1\right) - E\left(D \mid Z_1 = 0\right)} \text{ and } \beta_2 = \frac{E\left(Y \mid Z_2 = 1\right) - E\left(Y \mid Z_2 = 0\right)}{E\left(D \mid Z_2 = 1\right) - E\left(D \mid Z_2 = 0\right)},$$

and the corresponding estimators $\widehat{\beta}_1$ and $\widehat{\beta}_2$ simply replace the above expectations with sample averages. Let $\widehat{\beta}_1$ and $\widehat{\beta}_2$ be the estimated LATEs using $Z_1$ and $Z_2$ as instruments, respectively. Under standard assumptions, $\widehat{\beta}_1$ consistently estimates $\beta_1$, the average treatment effect among all first instrument compliers, and similarly $\widehat{\beta}_2$ consistently estimates $\beta_2$, the average treatment effect among all second instrument compliers. The denominators of these expressions equal the probability of first instrument and second instrument compliers, respectively. The denominator of the CC-LATE estimand is always greater than or equal to the denominators of either $\beta_1$ or $\beta_2$, since it additionally includes eager compliers and reluctant compliers.

Imbens and Angrist (1994) show that when combining multiple instruments with TSLS under the IAM assumption, imposing choice homogeneity and using $g(Z)$ as an instrument, then TSLS gives a weighted average of the pairwise LATEs:

$$\alpha_g^{IV} = \sum_{k=1}^{K} \lambda_k \cdot E[Y_i(1) - Y_i(0)|D_i(z_k) = 1, D_i(z_{k-1}) = 0],$$

with weights

$$\lambda_k = \frac{(P(z_k) - P(z_{k-1})) \cdot \sum_{l=k}^{K} \pi_l \cdot (g(z_l) - E[g(Z)])}{\sum_{m=1}^{K}(P(z_m) - P(z_{m-1})) \cdot \sum_{l=m}^{K} \pi_l \cdot (g(z_l) - E[g(Z)])},$$

where, using Imbens and Angrist's (1994) notation, $\pi_k = Pr(Z = z_k)$, $P(z_k) = E[D_i|Z_i = z_k]$, and the support of $Z$ is ordered such that if $l < m$, then $P(z_l) \leq P(z_m)$. The weights sum to one. To guarantee positive weights, Imbens and Angrist (1994) additionally assume that $J(Z)$, the scalar instrument constructed from $Z$, depends on the propensity score $P(Z)$ in a monotone way.[16]

Mogstad et al. (2021) show that under PM, TSLS gives a weighted average of the LATEs for the response types, $g$, in the population other than the always-takers and never-takers:

$$\beta_{\text{TSLS}} = \sum_{g \in \mathcal{G}: \mathcal{C}_g \neq \emptyset} \omega_g \cdot E[Y_i(1) - Y_i(0)|G_i = g], \tag{9}$$

with weights

$$\omega_g = P(G_i = g) \sum_{k=2}^{K} (\mathbb{1}[k \in \mathcal{C}_g] - \mathbb{1}[k \in \mathcal{D}_g]) \left( \frac{\text{Cov}(D_i, \mathbb{1}[p(Z_i) \geq p(z^k)])}{\text{Var}(p(Z_i))} \right),$$

where they denote $\mathcal{C}_g$ and $\mathcal{D}_g$ to be the sets of integers $k$ for which a certain group type responds to the change from $z^{k-1}$ to $z^k$ as a complier or defier, with $\{z^1, ..., z^k\}$ the points of the instrument support ordered by the propensity scores, $p(z^1), ..., p(z^k)$. The weights sum to one. A drawback of this estimand is that its interpretation is not straightforward for two reasons: The weights are counterintuitive, and the LATEs of defier types might enter the weighted average. As is evident from the expression, negative weights can occur either if $\text{Cov}(D_i, \mathbb{1}[p(Z_i) < p(z^k)]) \leq 0$ or if $\mathbb{1}[k \in \mathcal{C}_g] - \mathbb{1}[k \in \mathcal{D}_g] = -1$. The latter expression can lead to negative weights if $\mathcal{D}_g \neq \emptyset$. When PM allows for both first instrument compliers and second instrument compliers, $\mathcal{D}_g \neq \emptyset$ always occurs for either one of these two types. Thus, a

[16] Heckman et al. (2006) show that the weights are always positive when $P(Z)$ is the instrument. Thus, the weights are always positive when the first stage of TSLS is fully saturated, since in this case $J(Z) = P(Z)$.

negative weight on the LATE for one of these complier groups is generally a cause for concern when performing TSLS under PM. Even if the resulting weight is non-negative, the magnitude of the weight will be distorted if $\mathcal{D}_g \neq \emptyset$. Interpreting the TSLS estimand becomes even more challenging when more than two instruments are available. The instruments generate a variety of different complier and defier types in this case. Consequently, there are many potential two-way flows for some change in the instrument values. Next, consider the LATEs in the weighted average. The interpretation of the TSLS estimand depends on the LATEs of the response types present in the population, which is not straightforward in the case of multiple instruments. A cause for concern is that $\mathcal{D}_g \neq \emptyset$ generally holds for defier types, causing these types to enter the weighted average in Equation (9).

An attractive property of the CC-LATE is that it always gives the effect in the population of combined compliers. The CC-LATE is robust to the many defier types that might exist under LiM. Moreover, it is not concerned with negative weights. The CC-LATE estimand can be interpreted as

$$\beta_{\text{CC-LATE}} = \sum_{g \in cc} \omega_g \cdot E[Y_i(1) - Y_i(0) | G_i = g],$$

with weights corresponding to the relative sizes of the complier groups:

$$\omega_g = P(G_i = g).$$

If PM and LiM are non-nested, as discussed in Section D, then it might not be possible to obtain an unbiased estimate of the CC-LATE if PM is true. Nevertheless, the CC-LATE parameter can still be more interesting to estimate than the TSLS parameter, because it might be close to the true LATE for the combined complier population (see Appendix F for a more detailed examination of the estimand under a violation of LiM). Particularly since the number of response types that are allowed for under PM but violating LiM are very few, as discussed previously in Section D. However, when PM is violated, one should be careful when interpreting the TSLS parameter, due to defier types entering the equation. This means that the weight can be negative, even if $\text{Cov}(D_i, \mathbb{1}[p(Z_i) < p(z^k)]) > 0$, which might even lead to the TSLS estimate having an opposite sign than the true ATE.

Frölich (2007) considers multiple instrumental variables with covariates included nonparametrically. If $D_i$ follows an index structure and under standard assumptions including the IAM assumption, which heavily restricts choice heterogeneity, Frölich (2007) derives the following LATE:

$$E[Y^1 - Y^0|\tau = c] = \frac{\int (E[Y|X = x, p(Z, X) = \bar{p}_x] - E[Y|X = x, p(Z, X) = \underline{p}_x]) f_x(x) dx}{\int (E[D|X = x, p(Z, X) = \bar{p}_x] - E[D|X = x, p(Z, X) = \underline{p}_x]) f_x(x) dx},$$

where $\bar{p}_x = max_z p(z, x)$ and $\underline{p}_x = min_z p(z, x)$. Similar to the CC-LATE, the estimation is based on the two subgroups of observations where $Z = (0.., 0..., 0)$ and $Z = (1.., 1..., 1)$. The interpretation of this estimand differs in that it considers the largest complier group, whereas the CC-LATE considers all individuals that respond to any instrument or combination thereof. From the results of Imbens and Angrist (1994), one can show that $\frac{E[Y|Z=z_K]-E[Y|Z=z_0]}{E[D|Z=z_K]-E[D|Z=z_0]} = E[Y(1) - Y(0)|D(z_K) \neq D(z_0)]$ (see Appendix A.6) which equally can be interpreted as the effect in the largest group of compliers, having the same interpretation as the estimand for multiple binary instruments as proposed by Frölich (2007).

Goff (2024) derives the "all compliers LATE" (ACLATE) under a special form of PM, which he refers to as vector monotonicity (VM). Goff (2024) shows that the ACLATE can be re-written to a weighted average over the treatment effects of the specific combined complier groups, $g \in \mathcal{G}$:

$$E[Y_i(1) - Y_i(0)|C_i = 1] = \sum_{g \in \mathcal{G}} \frac{P(G_i = g) E[c(g, Z_i)]}{E[c(G_i, Z_i)]} \cdot E[Y_i(1) - Y_i(0)|G_i = g], \tag{10}$$

where $C_i = c(G_i, Z_i) = 1$ if a unit $i$ belongs to a group of the all compliers. Identification of the ACLATE is then possible for specific choices of the function $c(g, z)$. Only in rare cases does the TSLS estimator recover the ACLATE, and Goff (2024) proposes a different estimator that is similar in construction to the TSLS estimator. He further shows that Equation (10) can be re-written to a single Wald estimand:

$$E[Y_i(1) - Y_i(0)|C_i = 1] = \frac{E[Y_i|Z_i = \bar{Z}] - E[Y_i|Z_i = \underline{\mathrm{Z}}]}{E[D_i|Z_i = \bar{Z}] - E[D_i|Z_i = \underline{\mathrm{Z}}]},$$

where $\bar{Z} = (1, 1, ..., 1)'$ and $\underline{\mathrm{Z}} = (0, 0, ..., 0)'$. Obviously, the denominator should be nonzero, and it should hold that $P(Z_i = \bar{Z}) > 0$ and $P(Z_i = \underline{\mathrm{Z}}) > 0$.

As the name suggests, "all compliers" LATE represents the LATE for units that comply at times but never defy. The ACLATE under VM is a special case of the "set" LATE (SLATE), which is also introduced by Goff (2024). The SLATE measures the effect for units that comply when a set of instruments $\mathcal{J}$ changes from all zero to all one. When all instruments are included ($\mathcal{J} = 1, .., K$), the ACLATE coincides with the SLATE.

Under VM, only all compliers, always-takers, and never-takers exist, and the ACLATE and the CC-LATE coincide. However, in a setting with more than two binary instruments, some individuals may comply with certain instrument changes while defying others. Refer to these individuals as the "all-responders". In such cases, the all compliers population is smaller than the all responders population. For instance, consider selection behavior such as:

$$D(0, 0, 0) \leq D(1, 0, 0) \geq D(1, 1, 0) \leq D(1, 1, 1).$$

Individuals with such behavior are excluded under VM but included under LiM. In this setting, where VM fails and LiM holds, the ACLATE is not necessarily identified and no longer coincides with the CC-LATE. Moreover, under LiM, the CC-LATE is a special case of the SLATE for $\mathcal{J} = 1, .., K$. In general, it holds that:

$$\text{all compliers} \subseteq \text{combined compliers} \subset \text{all-responders},$$

with the first relation being strict when $k > 2$.

# F Extensions

## F.1 Violation of LiM

In this section, we consider identification when LiM is violated. Violation of this assumption not only introduces identification issues, but also reduces the power of the instruments, which exacerbates the problem (Dahl et al., 2023). If LiM is violated, it can be shown for the setting with two binary instruments that

$$\beta = \frac{\pi_{cc}}{\pi_{cc} - \pi_{dd}} E(Y^1 - Y^0 | T \in cc) - \frac{\pi_{dd}}{\pi_{cc} - \pi_{dd}} E(Y^1 - Y^0 | T \in dd)$$

with $cc$ the set of combined compliers, $cc \equiv \{ec, rc, 1c, 2c\}$, and $dd$ the set of defiers that can never be pushed towards compliance and do not cancel out, $dd \equiv \{d3, d4, d5, d6\}$.

**Proof** in Appendix F.1.1.

This result can easily be extended to the setting with multiple binary instruments.

If the probability of being this type of defier is small, that is, $\pi_{dd}$ is small, then more weight is given to $E(Y^1 - Y^0|T \in cc)$ such that the impact of these defiers will be small. The same holds when the average treatment effect for these defiers is negligible, that is, $E(Y^1 - Y^0|T \in dd)$ is very small compared to the effect in the combined compliers group, $E(Y^1 - Y^0|T \in cc)$. The presence of these defiers is problematic when they are many and/or their treatment effect is relatively large in magnitude. In this case, they will introduce a substantial bias. There are not many settings where it is likely that these types of defiers introduce a large amount of bias, especially since LiM already allows for the existence of a rich set of defiers. The CC-LATE is identified if $E(Y^1 - Y^0|T \in cc) = E(Y^1 - Y^0|T \in dd)$.

The CC-LATE under a violation of LiM is a weighted average of the ATE for the combined compliers and the ATE for the defier types that would have been ruled out under LiM with negative weight. This is comparable to the TSLS estimand, which is a weighted average that potentially contains defier types and/or negative weights.

### F.1.1 Proof of violation of LiM

Consider the setting where limited monotonicity is violated. Let $\pi_t = \Pr(T \in t)$, $t = at, rc, ec, 1c, 2c, 1d, 2d, ed, rd, d1, d2, d3, d4, d5, d6, nt$. We have

$$
\begin{aligned}
E(D_i^{00}) =& \sum_t E(D_i^{00}|T=t)\pi_t \\
=&\pi_{at}\cdot 1 + \pi_{rc}\cdot 0 + \pi_{ec}\cdot 0 + \pi_{1c}\cdot 0 + \pi_{2c}\cdot 0 + \pi_{nt}\cdot 0 + \pi_{1d}\cdot 1 + \pi_{2d}\cdot 1 + \pi_{ed}\cdot 1 \\
&+ \pi_{d3}\cdot 1 + \pi_{d1}\cdot 0 + +\pi_{d4}\cdot 1 + \pi_{d2}\cdot 0 + \pi_{d5}\cdot 1 + \pi_{rd}\cdot 0 + \pi_{d6}\cdot 1 \\
=&\pi_{at} + \pi_{1d} + \pi_{2d} + \pi_{ed} + \pi_{d3} + \pi_{d4} + \pi_{d5} + \pi_{d6}
\end{aligned}
$$

and

$$
\begin{aligned}
E(D_i^{11}) =& \sum_t E(D_i^{11}|T=t)\pi_t \\
=&\pi_{at}\cdot 1 + \pi_{rc}\cdot 1 + \pi_{ec}\cdot 1 + \pi_{1c}\cdot 1 + \pi_{2c}\cdot 1 + \pi_{nt}\cdot 0 + \pi_{1d}\cdot 1 + \pi_{2d}\cdot 1 + \pi_{ed}\cdot 1 \\
&+ \pi_{d1}\cdot 0 + \pi_{d2}\cdot 0 + \pi_{rd}\cdot 0 + \pi_{d3}\cdot 0 + \pi_{d4}\cdot 0 + \pi_{d5}\cdot 0 + \pi_{d6}\cdot 0 \\
=&\pi_{at} + \underbrace{\pi_{rc} + \pi_{ec} + \pi_{1c} + \pi_{2c}}_{\pi_{cc}} + \pi_{1d} + \pi_{2d} + \pi_{ed} \\
=&\pi_{at} + \pi_{cc} + \pi_{1d} + \pi_{2d} + \pi_{ed}.
\end{aligned}
$$

It therefore follows that

$$E(D|Z_1=1, Z_2=1) - E(D|Z_1=0, Z_2=0) = E(D_i^{11}) - E(D_i^{00}) = \pi_{cc} - (\pi_{d3} + \pi_{d4} + \pi_{d5} + \pi_{d6}).$$

Let $\beta_i = Y_i^1 - Y_i^0$. Under SUTVA the observed outcome $Y$ can be written as

$$
\begin{aligned}
Y_i &= Y_i^1 D_i + Y_i^0(1-D_i) = \beta_i D_i + Y_i^0 \\
&= \beta_i\left[D_i^{00}R_{1i} + D_i^{11}R_{2i} + D_i^{01}R_{3i} + D_i^{10}R_{4i}\right] + Y_i^0 \\
&= \beta_i D_i^{00}R_{1i} + \beta_i D_i^{11}R_{2i} + \beta_i D_i^{01}R_{3i} + \beta_i D_i^{10}R_{4i} + Y_i^0.
\end{aligned}
$$

Now, consider the numerator of the CC-LATE estimand,

$$E\left(Y|Z_1=1,Z_2=1\right)-E\left(Y|Z_1=0,Z_2=0\right)$$

$$=E(Y|R_2=1)-E(Y|R_1=1)$$

$$=E(\beta_i D_i^{11}+Y_i^0|R_2=1)-E(\beta_i D_i^{00}+Y_i^0|R_1=1)$$

$$=E(\beta_i D_i^{11})-E(\beta_i D_i^{00}).$$

We have that

$$E(\beta_i D_i^{00})=\sum_t E(\beta_i D_i^{00}|T=t)\cdot\pi_t$$

$$=E(Y_i^1-Y_i^0|T=at)\cdot\pi_{at}+E(Y_i^1-Y_i^0|T=1d)\cdot\pi_{1d}+E(Y_i^1-Y_i^0|T=2d)\cdot\pi_{2d}$$

$$+E(Y_i^1-Y_i^0|T=ed)\cdot\pi_{ed}+E(Y_i^1-Y_i^0|T=d3)\cdot\pi_{d3}+E(Y_i^1-Y_i^0|T=d4)\cdot\pi_{d4}$$

$$+E(Y_i^1-Y_i^0|T=d5)\cdot\pi_{d5}+E(Y_i^1-Y_i^0|T=d6)\cdot\pi_{d6}$$

and

$$E(\beta_i D_i^{11})=\sum_t E(\beta_i D_i^{11}|T=t)\cdot\pi_t$$

$$=E(Y_i^1-Y_i^0|T=at)\cdot\pi_{at}+E(Y_i^1-Y_i^0|T\in cc)\cdot\pi_{cc}+E(Y_i^1-Y_i^0|T=1d)\cdot\pi_{1d}$$

$$+E(Y_i^1-Y_i^0|T=2d)\cdot\pi_{2d}+E(Y_i^1-Y_i^0|T=ed)\cdot\pi_{ed}.$$

Therefore,

$$E(Y|Z_1=1,Z_2=1)-E(Y|Z_1=0,Z_2=0)$$

$$=E(\beta_i D_i^{11})-E(\beta_i D_i^{00})$$

$$=E(Y^1-Y^0|T\in cc)\cdot\pi_{cc}-E(Y_i^1-Y_i^0|T=d3)\cdot\pi_{d3}$$

$$-E(Y_i^1-Y_i^0|T=d4)\cdot\pi_{d4}-E(Y_i^1-Y_i^0|T=d5)\cdot\pi_{d5}$$

$$-E(Y_i^1-Y_i^0|T=d6)\cdot\pi_{d6}$$

$$=E(Y^1-Y^0|T\in cc)\cdot\pi_{cc}-E(Y^1-Y^0|T\in dd)\cdot\pi_{dd}$$

with $dd$ the set of defiers that can never be pushed towards compliance and do not cancel out, $dd \equiv \{d3, d4, d5, d6\}$ and so

$$\begin{aligned}
\beta =& \frac{E\left(Y \mid Z_1 = 1, Z_2 = 1\right) - E\left(Y \mid Z_1 = 0, Z_2 = 0\right)}{E\left(D \mid Z_1 = 1, Z_2 = 1\right) - E\left(D \mid Z_1 = 0, Z_2 = 0\right)} \\
=& \frac{E(Y^1 - Y^0 | T \in cc) \cdot \pi_{cc} - E(Y^1 - Y^0 | T \in dd) \cdot \pi_{dd}}{\pi_{cc} - \pi_{dd}} \\
=& \frac{\pi_{cc}}{\pi_{cc} - \pi_{dd}} E(Y^1 - Y^0 | T \in cc) - \frac{\pi_{dd}}{\pi_{cc} - \pi_{dd}} E(Y^1 - Y^0 | T \in dd).
\end{aligned}$$

## F.2 Bloom result

In a randomized trial with one-sided noncompliance there are no never-takers. For the setting with one binary instrument, Bloom (1984) shows that IV estimates the treatment effect on the treated in randomized trials with one-sided noncompliance,

$$\frac{E(Y_i | z_i = 1) - E(Y_i | z_i = 0)}{P(D_i = 1 | z_i = 1)} = E(Y_{1i} - Y_{0i} | D_i = 1).$$

When there are two binary instruments, one-sided compliance means that

$$E(D_i | Z_1 = 0, Z_2 = 0) = P(D_i = 1 | Z_{1i} = 0, Z_{2i} = 0) = \pi_{at} + \pi_{1d} + \pi_{2d} + \pi_{ed} = 0.$$

If compliance is only possible when both instruments are offered such that $Z_{1i} = 1, Z_{2i} = 1$, then the average treatment effect on the treated (ATT) is

$$E(Y_i^1 - Y_i^0 | D_i = 1) = \frac{E(Y_i | Z_{1i} = 1, Z_{2i} = 1) - E(Y_i | Z_{1i} = 0, Z_{2i} = 0)}{P(D_i = 1 | Z_{1i} = 1, Z_{2i} = 1)}.$$

**Proof** in Appendix F.2.1.

This result can easily be extended to the setting with more than two binary instruments if it holds that compliance is only possible when an individual is exposed to all instruments.

If one-sided compliance only holds for one of the instruments, $Z_2$, and compliance is only possible when both instruments are offered, then

$$E(Y_i^1 - Y_i^0|D_i = 1) = \frac{E(Y_i|Z_{1i} = 1, Z_{2i} = 1) - E(Y_i|Z_{2i} = 0)}{P(D_i = 1|Z_{1i} = 1, Z_{2i} = 1)}.$$

### F.2.1 Proof of Bloom result

When there are two binary instruments, one-sided compliance means that

$$E(D_i|Z_1 = 0, Z_2 = 0) = P(D_i = 1|Z_{1i} = 0, Z_{2i} = 0) = 0.$$

We can re-write $E(Y_i|Z_{1i} = 1, Z_{2i} = 1)$ and $E(Y_i|Z_{1i} = 0, Z_{2i} = 0)$ as

$$E(Y_i|Z_{1i} = 1, Z_{2i} = 1) = E(Y_i^0|Z_{1i} = 1, Z_{2i} = 1) + E((Y_i^1 - Y_i^0)D_i|Z_{1i} = 1, Z_{2i} = 1) \quad (11)$$

and

$$E(Y_i|Z_{1i} = 0, Z_{2i} = 0) = E(Y_i^0|Z_{1i} = 0, Z_{2i} = 0) + E((Y_i^1 - Y_i^0)D_i|Z_{1i} = 0, Z_{2i} = 0), \quad (12)$$

where $E((Y_i^1 - Y_i^0)D_i|Z_{1i} = 0, Z_{2i} = 0) = 0$ because $D_i = 0$ if $Z_{1i} = 0, Z_{2i} = 0$. Subtracting equation (12) from equation (11) gives

$$\begin{aligned}
&E(Y_i|Z_{1i} = 1, Z_{2i} = 1) - E(Y_i|Z_{1i} = 0, Z_{2i} = 0)\\
&= E((Y_i^1 - Y_i^0)D_i|Z_{1i} = 1, Z_{2i} = 1)\\
&= E(Y_i^1 - Y_i^0|D_i = 1, Z_{1i} = 1, Z_{2i} = 1)P(D_i = 1|Z_{1i} = 1, Z_{2i} = 1)
\end{aligned}$$

where the first equality follows because $E(Y_i^0|Z_{1i} = 1, Z_{2i} = 1) = E(Y_i^0|Z_{1i} = 0, Z_{2i} = 0)$ by the independence assumption.

Note that unlike in the setting with one binary instrument where $D_i = 1$ implies $Z_i = 1$, in the setting with two binary instruments $D_i = 1$ does **not** imply $Z_{1i} = 1, Z_{2i} = 1$. So $E(Y_i^1 - Y_i^0|D_i = 1, Z_{1i} = 1, Z_{2i} = 1) \neq E(Y_i^1 - Y_i^0|D_i = 1)$. However, if compliance is only possible when both instruments are offered such that $Z_{1i} = 1, Z_{2i} = 1$, then $E(Y_i^1 - Y_i^0 | D_i = 1, Z_{1i} = 1, Z_{2i} = 1) = E(Y_i^1 - Y_i^0 | D_i = 1)$, the treatment effect on the treated is

$$E(Y_i^1 - Y_i^0 | D_i = 1) = \frac{E(Y_i | Z_{1i} = 1, Z_{2i} = 1) - E(Y_i | Z_{1i} = 0, Z_{2i} = 0)}{P(D_i = 1 | Z_{1i} = 1, Z_{2i} = 1)}.$$

This result can easily be extended to the setting with more than two binary instruments if it holds that compliance is only possible when an individual is exposed to all instruments.

## F.3 Characteristics of the complier groups

When multiple instrumental variables are available, each instrument identifies the LATE for those individuals who change their treatment status in response to a change in that specific instrument. As pointed out in Angrist and Pischke (2009), when treatment effects are heterogeneous, the LATEs might differ due to differences in complier populations. Characteristics of the different complier populations might explain some of the differences in the estimated effects. Furthermore, LATEs are criticized for their lack of external validity. Knowledge about the characteristics of the population for which the average treatment effect was estimated might be valuable when extending to other populations.

Suppose there is a binary variable, $X$, that equals one when an individual is male, and zero when an individual is female.

$$\begin{aligned}
&\frac{P(x_{1i} = 1 | D_i^{11\ldots1} > D_i^{00\ldots0})}{P(x_{1i} = 1)} \\
&= \frac{P(D_i^{11\ldots1} > D_i^{00\ldots0} | x_{1i} = 1)}{P(D_i^{11\ldots1} > D_i^{00\ldots0})} \\
&= \frac{E(D_i | Z_{1i} = 1, Z_{2i} = 1, ..., Z_{ki} = 1, x_{1i} = 1) - E(D_i | Z_{1i} = 0, Z_{2i} = 0, ..., Z_{ki} = 0, x_{1i} = 1)}{E(D_i | Z_{1i} = 1, Z_{2i} = 1, ..., Z_{ki} = 1) - E(D_i | Z_{1i} = 0, Z_{2i} = 0, ..., Z_{ki} = 0)}.
\end{aligned}$$

Thus, we can obtain the relative likelihood of a combined complier being male through the first stage and the first stage for male individuals only.