On recoding ordered treatments as binary indicators*

Evan K. Rose Yotam Shem-Tov

November, 2021

Abstract

Researchers using instrumental variables to investigate the effects of ordered treatments (e.g., years of education, months of healthcare coverage) often recode treatment into a binary indicator for *any* exposure (e.g. any college, any healthcare coverage). The resulting estimand is difficult to interpret unless the instruments only shift compliers from no treatment to some positive quantity and not from some treatment to more—i.e., there are extensive margin compliers only (EMCO). When EMCO holds, recoded endogenous variables capture a weighted average of treatment effects across complier groups that can be partially unbundled into each group's treated and untreated means. Invoking EMCO along with the standard Local Average Treatment Effect assumptions is equivalent to assuming choices are determined by a simple two-factor selection model in which agents first decide whether to participate in treatment at all and then decide how much. The instruments must only impact relative utility in the first step. Although EMCO constrains unobserved counterfactual choices, it places testable restrictions on the joint distribution of outcomes, treatments, and instruments.

*Evan K. Rose: Saieh Family Research Fellow, University of Chicago, ekrose@uchicago.edu. Yotam Shem-Tov, Assistant Professor, Department of Economics, University of California, Los Angeles; shemtov@econ.ucla.edu. We thank Denis Chetverikov, Avi Feller, Jacob Goldin, Peter Hull, John Loeser, Juliana Londoño-Vélez, Sam Norris, Patrick Kline, Rodrigo Pinto, Jonathan Roth, Andres Santos, and Lucas Zhang for helpful comments and discussions.

1 Introduction

A large literature uses instrumental variables to estimate the effects of ordered treatments such as years of education or duration of health insurance coverage (e.g., Angrist and Krueger, 1991; Goldin et al., 2020). In these settings, it is common practice to recode the endogenous variable into a binary indicator for *any* treatment, e.g., any college or any insurance (Card, 1995; Finkelstein et al., 2012). Angrist and Imbens (1995) argued that doing so is a mistake. In the Local Average Treatment Effect (LATE) framework, the estimand recovered by two-stage least squares (2SLS) is the causal effect of any treatment exposure plus bias generated by intensive margin increases in treatment. This linear combination of effects is difficult to map to potential policies and may fall outside the range of treatment effects possible given the support of the outcome.

Instead, Angrist and Imbens (1995) advocated for leaving an ordered endogenous variable unchanged. 2SLS then recovers the "Average Causal Response" (ACR), a weighted average of effects of different doses of treatment across compliers. The ACR, however, is also difficult to interpret (Heckman et al., 2006). The compliers groups that generate it are not mutually exclusive. When treatment effects are potentially non-linear or heterogeneous—as in, for example, health responses to pharmaceuticals or the influence of unemployment duration on reemployment wages—the ACR may also differ in sign and magnitude from the effects of interest to the researcher or policymaker.

This paper investigates an assumption that significantly simplifies 2SLS analysis of ordered treatments. This restriction requires that the instruments induce units to shift from no treatment to some positive quantity but not from some treatment to more. In the health insurance example, the instruments must decrease the likelihood of being uninsured but leave the duration of coverage unchanged for individuals who would have obtained insurance regardless. In other words, the restriction requires that there are "extensive margin compliers only" (EMCO). In settings with one-sided non-compliance—that is, where individuals with one value of the instrument all receive no treatment—EMCO holds automatically.

Under EMCO, recoding an ordered treatment into an indicator is no longer a mistake. 2SLS estimates of the effect of "any treatment" recover a weighted average of treatment effects for

mutually exclusive groups of compliers. Each group is shifted to a different positive quantity of treatment from no treatment. The estimand averages the effects for each group of receiving this quantity vs. no treatment. Weights on each group are easy to recover. Moreover, under EMCO treated means for each complier group are identified, as well as the average of untreated means across all groups, allowing for a partial unbundling of the estimand. While average treatment effects for each complier group are not identified, they can be bounded. This makes it possible to test whether the data is consistent with certain hypotheses, such as that all complier groups or doses have positive average treatment effects. Bounds can be tightened using shape restrictions motivated by the setting or economic theory, as in recent work on partial identification (Chetverikov et al., 2018).

The power of EMCO comes from the restrictions it places on choice behavior. In fact, in the spirit of Vytlacil (2002), we show that taken together the standard LATE assumptions and EMCO imply that choices are rationalized by a two-step selection model where units first decide whether to participate in treatment at all and then pick treatment levels. EMCO requires the instrument affect relative utility in the first step and not the second. Importantly, the two-step selection process implies that at least two distinct latent factors govern treatment choices. While the presence of two factors makes marginal treatment effect analysis more complex, two represent a substantial dimension reduction relative to what is generated by LATE alone, which implies selection is governed by possibly as many latent factors as levels of treatment.[1] This dimension reduction explains why some quantities such as complier means are identified under EMCO, but not otherwise.

Although EMCO restricts counterfactual choices, it generates restrictions on the joint distribution of the outcome, treatment, and instruments that provide multiple testable implications. Building on the growing literature testing instrument validity (e.g., Kitagawa, 2015; Huber and Mellace, 2015; Mourifié and Wan, 2017; Frandsen et al., 2019; Norris, 2019), we provide visual and formal tests of these restrictions that have the power to detect violations in real-world settings. These tests extend results for binary treatments from Balke and Pearl (1997) and Heckman and

[1]For example, if treatment D falls in $\{0, 1, 2, \ldots, \bar{D}\}$, then LATE alone is equivalent to assuming there are $\bar{D}$ separate selection equations with distinct latent factors (Vytlacil, 2006).

Vytlacil (2005) and are based on the observation that outcome densities must be non-negative for all complier groups. They also require that the instrument does not *decrease* the mass of individuals at any positive level of treatment. To test both sets of restrictions, we utilize tools from the moment inequality literature for inference that is well-suited to settings where the number of moment restrictions may exceed the number of observations (Romano et al., 2014; Chernozhukov et al., 2018; Bai et al., 2019).

Andresen and Huber (2021) show similar restrictions on treatment distributions must hold when "binarizing" an ordered treatment at a given threshold. Our results build on theirs by deriving novel identification and choice modeling implications of EMCO, which binarzies treatment around zero, and showing how incorporating restrictions on outcome densities can increase power to detect violations. Our results thus connect the assumptions behind binarizing approaches to the literature on identification of complier means (Imbens and Rubin, 1997; Abadie, 2003), latent factor representations of LATE models (Vytlacil, 2002; Heckman and Pinto, 2018), and tests of instrument validity (Kitagawa, 2015). While we focus on the extensive margin, analogous results would apply when the instrument shifts individuals exclusively from any given level of treatment to more.

We conclude by examining the implications and plausibility of EMCO in Finkelstein et al. (2012)'s analysis of the Oregon Health Insurance Experiment (OHIE), which randomized low-income individuals' access to Medicaid. Details of the experiment make EMCO highly likely to hold in the OHIE. To be eligible to enroll, for example, participants randomized into treatment had to be uninsured for at least six months. Our formal tests confirm that EMCO cannot be rejected in Finkelstein et al. (2012)'s data; we use its identifying power to unpack treatment effects across complier groups. The results reveal clear patterns of intensive-margin adverse selection in the experimental data. Participants induced to remain on Medicaid the longest have the highest levels of healthcare utilization and worst self-reported health.

2 Setting and notation

Consider a setting with a single binary instrument $Z_i \in \{0, 1\}$ and a discrete, ordered treatment $D_i \in \{0, 1, ..., \bar{D}\}$. Let $D_i(z)$ denote the treatment status of individual i when $Z_i = z$. Observed treatment is $D_i = D_i(1)Z_i + D_i(0)(1 - Z_i)$. Let $Y_i(d)$ denote the potential outcome of interest under treatment status d. Observed outcomes are $Y_i = \sum_{d=0}^{\bar{D}} 1(D_i = d)Y_i(d)$. Assume that Z_i satisfies the standard assumptions of the LATE framework (Imbens and Angrist, 1994; Angrist et al., 1996) and its extension to ordered treatments (Angrist and Imbens, 1995):

Assumption 1. ***(LATE framework)***

(i) $\mathbb{E}[D_i|Z_i = 1] > \mathbb{E}[D_i|Z_i = 0]$ *(relevance)*

(ii) $(Y_i(0), Y_i(1), \ldots Y_i(\bar{D}), D_i(1), D_i(0)) \perp\!\!\!\perp Z_i$ *(exogeneity and exclusion)*

(iii) $D_i(1) \geq D_i(0) \quad \forall i$ *(monotonicity)*

Angrist and Imbens (1995) show that under these assumptions the Wald estimand recovers the average causal response (ACR):

$$\beta_{ACR} \equiv \frac{\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]}{\mathbb{E}[D_i|Z_i = 1] - \mathbb{E}[D_i|Z_i = 0]} = \sum_{d=1}^{\bar{D}} \omega_d \mathbb{E}[Y_i(d) - Y_i(d-1)|D_i(1) \geq d > D_i(0)] \quad (1)$$

where $\omega_d = \frac{\Pr(D_i(1) \geq d > D_i(0))}{\sum_{k=1}^{\bar{D}} \Pr(D_i(1) \geq k > D_i(0))}$.

The ACR captures a weighted average of effects of exposure to different "doses" of treatment (i.e., $\mathbb{E}[Y_i(d) - Y_i(d-1)]$) for potentially overlapping sets of compliers. While the ACR captures a well defined causal parameter, it does not correspond to a clear treatment manipulation or policy counterfactual (Heckman et al., 2006). When treatment effects are non-linear or heterogeneous, the ACR may in fact differ in sign and magnitude from the causal effects of interest.

Given the difficulty of interpreting the ACR, a common practice in applied research is to simply ignore the ordered nature of the treatment, recode it as binary, and interpret estimates as capturing the effects of "any" treatment. Research on the effects of incarceration, for example, commonly uses an indicator for any prison sentence as the endogenous variable of interest (e.g., Aizer and

Doyle, 2015; Bhuller et al., 2020; Norris et al., Forthcoming). Angrist and Imbens (1995) showed that doing so may produce a biased estimator of the ACR. Specifically, Proposition 1 shows that the recoded endogenous variable model recovers a linear combination of effects for those shifted from zero to some treatment and those shifted from some treatment to more:

Proposition 1. *Let $\beta_{recoded}$ be the Wald estimand when the endogenous variable is $1(D_i > 0)$. Then under Assumption 1:*

$$\beta_{recoded} = \underbrace{\mathbb{E}\left[Y_i(D_i(1)) - Y_i(D_i(0))|D_i(1) > D_i(0) = 0\right]}_{Extensive\ margin} +$$
$$\underbrace{\mathbb{E}\left[Y_i(D_i(1)) - Y_i(D_i(0))|D_i(1) > D_i(0) > 0\right]}_{Intensive\ margin}\frac{\Pr(D_i(1) > D_i(0) > 0)}{\Pr(D_i(1) > D_i(0) = 0)}$$

The estimator $\beta_{recoded}$ captures an unintuitive mixture of effects and thus suffers from similar interpretation issues as the ACR.[2] Moreover, because the estimand is a linear combination and not an average, $\beta_{recoded}$ may fall outside the range of treatment effects physically possible given the support of Y_i. Intuitively, the primary issue is that the instrument is no longer excludable after recoding. Some individuals' outcomes may change even though their treatment status ($1(D_i > 0)$) does not.[3]

These issues can be avoided if one is willing to rule out the existence of such individuals, an assumption we call "extensive margin compliers only":

Assumption 2. ***Extensive Margin Compliers Only (EMCO)***

$$D_i(1) > D_i(0) \Rightarrow D_i(0) = 0 \quad \forall i$$

Assumption 2 requires that the instrument only causes some individuals to switch from no treatment into some, and not from some to more. This assumption will be automatically satisfied in any setting with one-sided non-compliance—i.e., where individuals with $Z_i = 0$ must have

[2] Angrist and Imbens (1995)'s result is that $\beta_{\text{recoded}} = \beta_{\text{ACR}} \cdot (1 + \kappa)$ where $\kappa \equiv \frac{\sum_{l=2}^{\bar{D}} \Pr(D_i(1) \geq l > D_i(0))}{\Pr(D_i(1) \geq 1 > D_i(0))}$. To our knowledge, the formulation in Proposition 1 is new (see Appendix B.1 for a proof).

[3] Note that when defining the treatment as $1(D_i = d)$, compliers with $D_i(1) > d = D_i(0)$ can be viewed as defiers, since they are shifted by the instrument from $1(D_i(0) = d) = 1$ to $1(D_i(1) = d) = 0$.

$D_i = 0$—but can also hold in more general settings. EMCO implies the second term in Proposition 1 disappears and 2SLS using $1(D_i > 0)$ as the endogenous variable recovers the average effect of the treatment on individuals shifted from no exposure to some positive amount.

$$\begin{aligned}\beta_{recoded} &\equiv \frac{\mathbb{E}\left[Y_i|Z_i = 1\right] - \mathbb{E}\left[Y_i|Z_i = 0\right]}{\mathbb{E}\left[1(D_i > 0)|Z_i = 1\right] - \mathbb{E}\left[1(D_i > 0)|Z_i = 0\right]} \\ &= \sum_{d=1}^{\bar{D}} \omega_d^r \mathbb{E}\left[Y_i(D_i(1)) - Y_i(0)|D_i(1) = d > D_i(0) = 0\right]\end{aligned}$$

where $\omega_d^r = \frac{\Pr(D_i(1)=d, D_i(0)=0)}{\sum_{l=1}^{\bar{D}} \Pr(D_i(1)=l, D_i(0)=0)}$.[4]

EMCO is attractive because when it holds, β_{recoded} has a clear interpretation as a weighted average of causal effects for mutually exclusive complier populations with weights proportional to their population size.[5]

3 Advantages of EMCO: Complier means and bounds

The EMCO assumption places strong restrictions on the DGP; however, in settings in which it is satisfied, it also provides meaningful advantages. In addition to giving $\beta_{recoded}$ a clear causal interpretation, EMCO identifies other interesting quantities. Imbens and Rubin (1997) and Abadie (2002) show that when the treatment is binary (i.e., $D_i \in \{0, 1\}$), compliers' treated and untreated mean potential outcomes are identified by 2SLS regressions using $Y_i D_i$ or $Y_i(1 - D_i)$, respectively, as the outcome and D_i or $(1 - D_i)$ as the endogenous variable. When the treatment has multiple levels, it is tempting to attempt to estimate complier means using $Y_i 1(D_i = d)$ as the outcome. However, doing so without assuming EMCO yields a mixture of outcome means for multiple groups.

[4] The importance of ruling out intensive-margin compliers for interpreting $\beta_{recoded}$ has been previously mentioned in the literature. See, for example, Marshall (2016); Andresen and Huber (2021); Goldin et al. (2020); Rose and Shem-Tov (2019); Bhuller et al. (2020); Norris et al. (Forthcoming).

[5] Of course, β_{ACR} also retains a valid causal interpretation if EMCO holds, albeit one less intuitive than $\beta_{recoded}$. Under EMCO, the complier groups summed over in β_{ACR} are simply defined by $D_i(1) \geq d, D_i(0) = 0$ for each $d \geq 1$.

To see why, note that the reduced form effect of Z_i on this outcome is:

$$\begin{aligned}
&\mathbb{E}\left[Y_i 1(D_i = d)|Z_i = 1\right] - \mathbb{E}\left[Y_i 1(D_i = d)|Z_i = 0\right] \qquad (2)\\
&= \underbrace{\mathbb{E}\left[Y_i(d)|D_i(1) = d > D_i(0) = 0\right]\Pr(D_i(1) = d > D_i(0) = 0)}_{\text{Shifted } \underline{\text{into}} \ d \text{ from zero (extensive margin)}}\\
&+ \underbrace{\mathbb{E}\left[Y_i(d)|D_i(1) = d > D_i(0) > 0\right]\Pr(D_i(1) = d > D_i(0) > 0)}_{\text{Shifted } \underline{\text{into}} \ d \text{ from } D_i(0) > 0 \text{ (intensive margin)}}\\
&- \underbrace{\mathbb{E}\left[Y_i(d)|D_i(1) > D_i(0) = d\right]\Pr(D_i(1) > D_i(0) = d)}_{\text{Shifted } \underline{\text{out}} \text{ of } d \text{ to } D_i(1) > d \text{ (intensive margin)}}
\end{aligned}$$

Hence changes in $Y_i 1(D_i = d)$ due to Z_i reflect individuals both moving into $D_i = d$ from multiple sources (extensive- and intensive-margin shifts) and moving away into higher levels of treatment (intensive-margin shifts). Clearly Equation 2 when rescaled by the first stage would not yield a meaningful potential outcome mean.

However, EMCO implies that $\Pr(D_i(1) = d > D_i(0) > 0)$ and $\Pr(D_i(1) > D_i(0) = d)$ are both zero. Hence treated potential outcome means of each group of "d-type" compliers (i.e., individuals with $D_i(1) = d > 0 = D_i(0)$) are identified, as well as an average of potential outcomes under no treatment. The following proposition formalizes this claim:

Proposition 2. *If Assumptions 1 and 2 hold, then:*

$$\begin{aligned}
(i) \quad \frac{\mathbb{E}\left[Y_i 1(D_i = 0)|Z_i = 1\right] - \mathbb{E}\left[Y_i 1(D_i = 0)|Z_i = 0\right]}{\mathbb{E}\left[1(D_i = 0)|Z_i = 1\right] - \mathbb{E}\left[1(D_i = 0)|Z_i = 0\right]} &= \mathbb{E}\left[Y_i(0)|D_i(1) > D_i(0) = 0\right]\\
&= \sum_{d=1}^{\bar{D}} \omega_d^r \mathbb{E}\left[Y_i(0)|D_i(1) = d > D_i(0) = 0\right]
\end{aligned}$$

and for any $d > 0$ such that $\mathbb{E}\left[1(D_i = d)|Z_i = 1\right] - \mathbb{E}\left[1(D_i = d)|Z_i = 0\right] > 0$:

$$(ii) \quad \frac{\mathbb{E}\left[Y_i 1(D_i = d)|Z_i = 1\right] - \mathbb{E}\left[Y_i 1(D_i = d)|Z_i = 0\right]}{\mathbb{E}\left[1(D_i = d)|Z_i = 1\right] - \mathbb{E}\left[1(D_i = d)|Z_i = 0\right]} = \mathbb{E}\left[Y_i(d)|D_i(1) = d > D_i(0) = 0\right]$$

We illustrate how these results can generate additional insights below using data from the Oregon Health Insurance Experiment. Proposition 2 can be thought of as an extension of the results in Imbens and Rubin (1997), Abadie (2002), and Abadie (2003) for the binary case to

multi-valued treatments. In Appendix B.2, we present a more general version of Proposition 2 that is analogous to Theorem 3.1 in Abadie (2003) and implies that covariates X_i of each group of d-type compliers can also be identified:

$$\frac{\mathbb{E}\left[X_i 1(D_i = d)|Z_i = 1\right] - \mathbb{E}\left[X_i 1(D_i = d)|Z_i = 0\right]}{\mathbb{E}\left[1(D_i = d)|Z_i = 1\right] - \mathbb{E}\left[1(D_i = d)|Z_i = 0\right]} = \mathbb{E}\left[X_i|D_i(1) = d > D_i(0) = 0\right] \ \forall d > 0$$

Unfortunately, we cannot identify $\mathbb{E}\left[Y_i(0)|D_i(1) = d > D_i(0) = 0\right]$ without additional assumptions. Since Z_i shifts individuals from $D_i(0) = 0$ to multiple positive levels of treatment, we cannot separately identify the untreated counterfactual for each group of d-type compliers. As noted above, only a weighted average of untreated means across all complier groups is identified. Thus, while treated d-type complier means are identified, d-type treatment *effects* are not:

$$\underbrace{\mathbb{E}\left[Y_i(d) - Y_i(0)|D_i(1) = d > D_i(0) = 0\right]}_{\text{Not identified}} =$$

$$\underbrace{\mathbb{E}\left[Y_i(d)|D_i(1) = d > D_i(0) = 0\right]}_{\text{Identified}} - \underbrace{\mathbb{E}\left[Y_i(0)|D_i(1) = d > D_i(0) = 0\right]}_{\text{Not identified}}$$

Consequently, $\beta_{recoded}$ cannot be fully decomposed into its constituent causal components. The researcher can, however, construct bounds on d-type treatment effects.

Specifically, let Y_d^0 be the unknown quantity $\mathbb{E}\left[Y_i(0)|D_i(1) = d > D_i(0) = 0\right]$. By the above, $Y_d^d = \mathbb{E}\left[Y_i(d)|D_i(1) = d > D_i(0) = 0\right]$ is point identified for all $d > 0$. ω_d^r is likewise identified by the ratio of $\Pr(D_i = d|Z_i = 1) - \Pr(D_i = d|Z_i = 0)$ to $\Pr(D_i > 0|Z_i = 1) - \Pr(D_i > 0|Z_i = 0)$. Bounds on d-type treatment effects are given by the solution to the linear program:

$$\begin{aligned} & min/max_{\{Y_d^0\}_{d=1}^{\bar{D}}} \ Y_d^d - Y_d^0 \\ & \quad s.t. \quad Y_d^0 \in \mathcal{Y} \quad \forall d > 0 \\ & \mathbb{E}\left[Y_i(0)|D_i(1) > D_i(0)\right] = \sum_{d=1}^{\bar{D}} \omega_d^r Y_d^0 \end{aligned} \tag{3}$$

where $\mathcal{Y}$ is the support of Y_i.

Given that the unknown quantities $\{Y_d^0\}_{d=1}^{\bar{D}}$ are disciplined by only two sets of restrictions, these

bounds are likely to be wide without further assumptions. If the support of Y_i is not bounded, implying the first set of restrictions does not bind, then bounds for specific complier group's treatment effects will be uninformative. Imposing other shape restrictions, such as that treatment effects are decreasing in d, can help tighten bounds in this case. Inference can be conducted using methods that are suited to situations where the standard bootstrap fails, such as Fang and Santos (2019) or Hong et al. (2020).

Rather than bounding individual complier groups' treatment effects, it may also be interesting to test whether the data are consistent with certain joint hypotheses, such as that treatment effects for all complier groups are weakly positive. Positive treatment effects requires that $Y_d^0 \leq Y_d^d$ for all $d > 0$. Hence testing this hypothesis is equivalent to asking whether there exists a set of Y_d^0 such that:

$$\begin{aligned} & Y_d^d - Y_d^0 > 0 \ \forall d > 0 \\ & Y_d^0 \in \mathcal{Y} \quad \forall d > 0 \\ & \mathbb{E}\left[Y_i(0) | D_i(1) > D_i(0)\right] = \sum_{d=1}^{\bar{D}} \omega_d^r Y_d^0 \end{aligned} \tag{4}$$

As with individual bounds, such tests are likely to be trivially satisfied when the support of Y_i is unbounded. Inference can be conducted by viewing the problem as a shape constrained generalized method of moments problem and applying methods developed by Chernozhukov et al. (2020).

4 Implications for choice behavior

Vytlacil (2002, 2006) showed that the LATE framework (Assumption 1) alone is equivalent to assuming choices are governed by $\bar{D} - 1$ selection equations:

$$1(D_i \geq d) = 1(C^d(Z_i) - V_i^d \geq 0), \text{ for } d \in \{1, \dots \bar{D}\} \tag{5}$$

where V_i^d are random variables and C^d are unknown functions of the instruments satisfying $C^{d-1}(Z_i) - V_i^{d-1} \geq C^d(Z_i) - V_i^d \quad \forall i, d$, and $(V_i^1, \dots, V_i^{\bar{D}}) \perp\!\!\!\perp Z_i$. As in the rest of the paper,

for expositional convenience we suppress implicit conditioning on observables X_i.

While equivalent to the LATE assumptions, the selection model in Equation 5 is difficult to work with due to the $\bar{D}-1$ dimensions of the unobserved heterogeneity. EMCO restricts this unobserved heterogeneity sharply. In fact, along with the LATE framework assumptions, assuming EMCO is equivalent to assuming a two-step decision making process. In the first step, the individual chooses whether to participate or not. In the second step, the individual chooses the level of treatment. A classic example of such behavior is two-stage budgeting (e.g., see Deaton and Muellbauer, 1980). Proposition 3 formalizes this equivalence by showing that the LATE assumptions and EMCO jointly impose the same restrictions on choice behavior as a simple two-factor "hurdle" selection model:

$$1(D_i = d) = \begin{cases} 1(\pi_0(Z_i) - U_i^{Ext} \geq 0) \;\; \text{if} \;\; d = 0 \\ 1(\pi_0(Z_i) - U_i^{Ext} < 0) 1(\pi_{d+1} \leq U_i^{Int} < \pi_d) \;\; \text{if} \;\; d > 0 \end{cases} \tag{6}$$

where $(U_i^{Ext}, U_i^{Int}) \perp\!\!\!\perp Z_i$, $(U_i^{Ext}, U_i^{Int}) \sim F(u^{Ext}, u^{Int})$, and $\pi_d \geq \pi_{d+1}$.[6,7] This model includes only *two* latent factors: U_i^{Ext}, which governs the decision of whether or not to participate, and U_i^{Int}, which relates to the level of participation/treatment. Although all threshold functions π_d can depend on covariates X_i, only $\pi_0(Z_i)$ is a function of Z_i.

Proposition 3. *Assumptions 1 and 2 are equivalent to the model in Equation 6.*

The equivalence in Proposition 3 is in the sense of Vytlacil (2002): the model in Equation 6 satisfies Assumptions 1 and 2. But not only that, the model can also be derived using only these assumptions. Thus, the model in Equation 6 imposes the same restrictions on choice behavior as those imposed by the combination of the LATE framework assumptions and EMCO.

A direct implication of Proposition 3 is that so-called "single-index" models commonly used to reduce the dimensionality of unobserved heterogeneity in Equation 5 (e.g., see Meghir and Palme,

[6] We use πs as notation for thresholds because $\pi_0(Z_i) = \Pr(D_i = 0|Z_i)$ when F is the bivariate uniform distribution, which as Appendix B shows can be assumed without loss of generality. If U_i^{Ext} and U_i^{Int} are independent, $\pi_d(Z_i) - \pi_{d+1}(Z_i) = \Pr(D_i = d|Z_i, D_i > 0)\ \forall d > 0$.

[7] These assumptions pin down the model's implications for observed and counterfactual treatment choices. To complete the model, we also require assumptions on outcomes that guarantee exogeneity and exclusion: $(Y_i(0), \ldots, Y_i(\bar{D}), U_i^{Ext}, U_i^{Int}) \perp\!\!\!\perp Z_i$.

1999; Dahl, 2002; Heckman et al., 2006; Rose and Shem-Tov, 2021) are likely to be inconsistent with EMCO. To see why, consider the model in Equation 5 and assume there is only one dimension of unobserved heterogeneity, i.e., that $V_i^d = V_i \ \ \forall d \in \{1, \ldots, \bar{D}\}$. Under this single index assumption, choices can be expressed as:

$$1(D_i = d) = 1(\pi_{d+1}(Z_i) \leq U_i < \pi_d(Z_i)), \quad U_i \sim Uniform[0, 1] \tag{7}$$

Whenever $\pi_d(0) < \pi_{d-1}(0) < 1$ and $\pi_d(0) < \pi_d(1)$, there will be intensive margin compliers with $D_i(1) > d = D_i(0)$. In fact, compatibility with EMCO requires that $\pi_d(1) = \pi_d(0) \ \forall d > 1$ whenever $\Pr(D_i(0) > 0) > 0$. In other words, compliers can only be shifted from no treatment to $D_i = 1$ unless no individuals obtain positive levels of treatment under $Z_i = 0$.[8]

Nevertheless, instruments that satisfy EMCO-like behavioral restrictions are commonly used in economics. For example, the need for such instruments arises naturally in labor economics when researchers studying wages seek to correct for the choice to work at all in the style of Gronau (1974) and Heckman (1974). Mulligan and Rubinstein (2008) use these techniques to estimate the influence of the changing composition of women in the labor force on the gender wage gap. Their instrument—a mother's number of children aged zero to six interacted with marital status—must impact the decision to work but not labor supply among mothers already working.

4.1 Implications of EMCO for marginal treatment effect analysis

The inconsistency of EMCO with single-index models makes marginal treatment effect (MTE) analysis (Heckman and Vytlacil, 1999) more complex. While two factors represents a substantial dimension reduction relative to the $\bar{D} - 1$ implied by LATE alone, modeling treatment effect heterogeneity in two dimensions is significantly more challenging than the single dimension considered in the MTE literature (e.g., Heckman, 2010).

One simplifying assumption that would restore the validity of standard MTE tools is that potential outcomes are not affected by latent factors that govern selection along the intensive

[8] It is simple to empirically test the special case in which the data is consistent with both EMCO and the single index assumption since the distributions of $D_i(0)$ and $D_i(1)$ conditional on $D_i(1) > D_i(0)$ are both identified under EMCO.

margin:

$$\mathbb{E}\left[Y_i(d)|U_i^{Ext}, U_i^{Int}\right] = \mathbb{E}\left[Y_i(d)|U_i^{Ext}\right] \tag{8}$$

However, this assumption precludes selection into treatment based on gains and levels along the intensive margin except through correlation between U_i^{Ext} and U_i^{Int}.

An alternative approach is to model potential outcomes as functions of both U_i^{Ext} and U_i^{Int}. The researcher can then estimate or bound other treatment effects of interest (such as an average treatment effect) consistent with the moments identified under EMCO. Specifically, let $m_d(u_1, u_2) = \mathbb{E}\left[Y_i(d)|U_i^{Ext} = u_1, U_i^{Int} = u_2\right]$ represent treatment response functions. d-type compliers' potential outcome means are given by:

$$\begin{aligned} &\mathbb{E}\left[Y_i(d)|D_i(1) = d > 0 = D_i(0)\right] = \\ &= \int_{\pi_{d+1}}^{\pi_d}\int_{\pi_0(0)}^{\pi_0(1)} m_d(u_1, u_2)dF(u_1, u_2) / \int_{\pi_{d+1}}^{\pi_d}\int_{\pi_0(0)}^{\pi_0(1)} dF(u_1, u_2) \end{aligned} \tag{9}$$

The researcher can then pick explicit functional forms for treatment response functions or flexibly approximate them in the style of Mogstad et al. (2018) and Marx (2020). Each identified complier mean serves to discipline these functions. Since there are two unobserved factors, access to additional instruments or shape constraints would likely be critical to obtaining informative estimates.

5 Testable implications of EMCO

Combined with the LATE assumptions, EMCO places two sets of restrictions on the DGP. First, it restricts the distribution of treatments under $Z_i = 1$ vs. $Z_i = 0$. Specifically, EMCO implies that individuals are only shifted from no treatment into positive levels of treatment, ruling out any movement of individuals already at a positive level of treatment. Thus, there cannot be a *decrease* in observations at positive treatment levels due to Z_i:

Proposition 4. *If Assumptions 1 and 2 hold, then:*

(i) $\Pr(D_i = 0|Z_i = 1) < \Pr(D_i = 0|Z_i = 0)$

(ii) $\Pr(D_i = d|Z_i = 1) \geq \Pr(D_i = d|Z_i = 0) \quad \forall d > 0$

Proposition 4 provides necessary but not sufficient conditions for EMCO. To understand why, it is useful to consider a simple example. Table 1 presents a DGP for an ordered treatment with three levels. There are three potential types of compliers: Those shifted from $D_i(0) = 0$ to $D_i(1) = 1$ (with population share Δ^1_{ext}), those shifted from $D_i(0) = 0$ to $D_i(1) = 2$ (share Δ^2_{ext}), and those shifted from $D_i(0) = 1$ to $D_i(0) = 2$ (share Δ_{int}). The existence of a positive quantity of this final complier type would violate EMCO. When $\Delta_{int} = 0$ and $\Delta^1_{ext}, \Delta^2_{ext} \geq 0$, the distribution of $D_i|Z_i = 1$ stochastically dominates the distribution of $D_i|Z_i = 0$ for all $d > 0$. If $\Delta_{int} > \Delta^1_{ext}$, condition (ii) will be violated and EMCO will be rejected. But if $\Delta^1_{ext} \geq \Delta_{int} > 0$, condition (ii) will be satisfied even though EMCO is violated.

Proposition 4 also suggests a very simple visual test of EMCO. When plotting the distribution of treatments under $Z_i = 1$ and $Z_i = 0$, the former should stochastically dominate for $d > 0$. In research designs where the instrument is only conditionally randomly assigned, the researcher can plot estimates of the effect of Z_i on $1(D_i \geq d)$ conditional on a specific set of controls and examine whether the coefficients are uniformly non-negative.[9]

EMCO also places a second set of restrictions on the joint distribution of (Y_i, D_i, Z_i). Specifically, the same arguments made in Balke and Pearl (1997) and Heckman and Vytlacil (2005) for the binary treatment case imply that:

Proposition 5. *If Assumptions 1 and 2 hold, then for any measurable set A in the support of Y_i:*

(i) $\Pr(Y_i \in A, D_i = d|Z_i = 1) \geq \Pr(Y_i \in A, D_i = d|Z_i = 0) \quad \forall d > 0$

(ii) $\Pr(Y_i \in A, D_i = 0|Z_i = 0) \geq \Pr(Y_i \in A, D_i = 0|Z_i = 1)$

Intuitively, because EMCO requires that the instrument induces shifts from $D_i = 0$ to $D_i > 0$ only, the density of Y_i across its support must be weakly increasing due to Z_i for all individuals

[9] Andresen and Huber (2021) consider a treatment defined by $1\{D_i \geq d^*\}$ and assume that the instrument only shifts individuals from below d^* to above it (i.e., $D_i(1) > D_i(0) \Rightarrow D_i(0) < d^* \leq D_i(1)$. They show this assumption implies: $\Pr(D_i(1) \geq d' > D_i(0)) \leq \Pr(D_i(1) \geq d > D_i(0))$ for every $d' > d > d^*$. The same result was also independently derived in Rose and Shem-Tov (2019), Footnote 8. As we show in Appendix B, these moment inequalities are equivalent to the ones in Proposition 4. Andresen and Huber (2021) also state that these conditions should apply conditional on each value of the outcome.

with $D_i > 0$, and decreasing due to Z_i for those with $D_i = 0$.[10] Proposition 5 can also be tested visually by comparing densities of Y_i conditional on $D_i = d$ for $Z_i = 1$ and $Z_i = 0$.

Just as with Proposition 4, however, Proposition 5 is necessary but not sufficient for EMCO. To see why, note that:

$$\begin{aligned}\mathbb{E}\left[1(Y_i \in A)|D_i = d, Z_i = 1\right] - \mathbb{E}\left[1(Y_i \in A)|D_i = d, Z_i = 0\right] = \\ \Pr(Y_i \in A|D_i(1) = d > D_i(0) = 0) + \\ \Pr(Y_i \in A|D_i(1) = d > D_i(0) > 0) - \Pr(Y_i \in A|D_i(1) > d = D_i(0))\end{aligned}$$

Under EMCO, the final two terms disappear; the difference must be non-negative, since the right-hand-side is a density. When EMCO does not hold, however, this quantity can still be positive whenever $\Pr(Y_i \in A|D_i(1) > d = D_i(0))$ is not too large. Hence there are some potential DGPs under which EMCO fails but both propositions are satisfied. The ability to detect violations of EMCO depends on both the share of compliers with $D_i(0) > 0$ and the density of such individuals in regions of Y_i. We show below, however, that in Monte Carlo simulations, both propositions have a meaningful bite.

To illustrate, consider the DGP in Table 1 and a binary outcome $Y_i \in \{0, 1\}$. Let Δ_y be the difference in the distribution of Y_i between compliers who are shifted to $D_i(1) = 1$ and intensive margin compliers who are shifted from $D_i(0) = 1$ to $D_i(1) = 2$:

$$\Delta_y \equiv \Pr(Y_i = 1|D_i(1) = 1 > D_i(0)) - \Pr(Y_i = 1|D_i(1) > D_i(0) = 1) \tag{10}$$

Since the outcome is binary, differences in the probability that $Y_i = 1$ characterize differences in the full distribution of Y_i. When the outcome has broader support, there are additional restrictions

[10] In the binary treatment case, Kitagawa (2015) and Mourifié and Wan (2017) proposed using these moment conditions to construct a test for the validity of Assumption 1. It is simple to adapt their tests to the ordered treatment case without requiring that EMCO holds. This can be used to check if Assumption 1 alone is satisfied. In the ordered treatment case it is possible to non-parametrically identify the density of $Y(0)$ for compliers (i.e., $\Pr(Y_i(0)|D_i(1) > D_i(0) = 0)$) and the density of $Y_i(\bar{D})$ for compliers (i.e., $\Pr(Y_i(\bar{D})|D_i(1) = \bar{D} > D_i(0))$), where $\bar{D}$ is the highest/top dose of the treatment. The rest follows almost immediately from Kitagawa (2015), but instead of testing for positive densities of compliers' $Y_i(1)$ and $Y_i(0)$, one tests for positive densities of compliers' $Y_i(\bar{D})$ and $Y_i(0)$.

and hence more opportunities to detect the existence of intensive margin compliers.

Figure 1 reports the results of Monte-Carlo simulations based on the above DGP.[11] We use 1,000 simulations for combinations of values of Δ^1_{ext}, Δ^2_{ext}, Δ_{int}, and Δ_y. To formally test EMCO's restrictions we use the methods proposed by Chernozhukov et al. (2018) and Romano et al. (2014), which can be used also when the number of moment inequalities exceeds the number of observations. Appendix D provides implementation details. In Panel (a), the x-axis shows the true proportion of intensive margin compliers (Δ_{int}). The y-axis shows the share of simulations in which the null hypothesis that $\Delta_{int} = 0$ is rejected holding Δ_y fixed. The rejection rate when $\Delta_{int} = 0$ represents the size of the test, i.e., the likelihood of Type-I errors. As can be seen clearly from in Panel (a), there are no finite sample issues with either of the two tests. Panel (b) holds fixed the proportion of intensive margin compliers and examines how power increases with differences in the distribution of outcomes (Δ_y).

The figure shows that both tests have good finite-sample power to detect relatively small violations of EMCO. Panel (a), for example, shows that EMCO is almost always rejected when at least 30% of compliers respond along the intensive margin. For any given share of intensive margin compliers, differences in outcome distributions generate more power, since these differences create violations of Proposition 5. Both testing procedures perform similarly, although Chernozhukov et al. (2018)'s method is slightly more conservative.

6 Empirical application: The effects of health insurance

In 2008, a group of low-income adults in Oregon were randomly given the opportunity to apply for Medicaid. Finkelstein et al. (2012) use this experiment—dubbed the Oregon Health Insurance Experiment (OHIE)—to study the effects of access to Medicaid on health care utilization and financial and physical well-being. They find that insurance increases both primary and emergency care utilization, lowers some health care expenditures, and increases self-reported physical and mental health.

To analyze the experiment, the researchers primarily use 2SLS specifications with "ever on

[11] Appendix E describes the full DGP in detail.

Medicaid" as the endogenous variable. However, they note that the treatment in their setting—duration of Medicaid coverage—is continuous, and argue that coding the treatment as the "'number of months on Medicaid' may be more appropriate than 'ever on Medicaid' where the effect of insurance on the outcome is linear in the number of months insured." While a continuous endogenous variable would be appropriate regardless of linearity in the effects of insurance, a binary endogenous variable for "any Medicaid" is also appropriate in this setting because EMCO is highly likely to hold for institutional reasons. The OHIE randomized admission into the Oregon Health Standard Plan, which had been closed to enrollment since a 2004. Subjects lotteried into treatment were able to enroll and remain on the plan so long as they were eligible, which required being uninsured for at least six months prior to enrolling and ineligible for other public health insurance programs.[12] Individuals who would have obtained some Medicaid in the control group therefore could not increase their duration of coverage if lotteried into treatment, since they would be ineligible.[13]

Our tests of EMCO's restrictions support its applicability to the OHIE. Appendix Figure A.1 shows that individuals lotteried into Medicaid ($Z_i = 1$) were much less likely than controls to have zero months of insurance over the follow-up period. They are more likely, however, to have coverage for all positive durations, with particularly large spikes around 6-7 months and 12-13 months. Consistent with EMCO, there are no decreases in density at any positive level of coverage. A formal test of Propositions 4 and 5 using Romano et al. (2014)'s method confirms the impression left by Appendix Figure A.1. We cannot reject at the 5% significance level the null hypothesis that the data are consistent with EMCO ($T_n = 3.46$; critical value 3.96).[14]

EMCO gives the Finkelstein et al. (2012) results a simple causal interpretation. The two percentage point increase in hospital admissions due to Medicaid presented in their Table IV, for example, reflects increases relative to how often subjects would have been admitted without any insurance. EMCO also allows the reseracher to further decompose treatment effects across

[12]Candidates also had to be 19–64 years of age, US citizens or legal immigrants, have income under 100 percent of the federal poverty level, and possess assets of less than $2,000.

[13]It is possible some subjects would have obtained Medicaid through other means if not lotteried into treatment. As we show below, the data in this case are consistent with EMCO holding nevertheless.

[14]We test Proposition 5 using indicators for deciles of standardized measures of health care utilization and self-reported health as the outcomes.

complier populations. Table 2 illustrates this using survey data from Finkelstein et al. (2012) on self-reported health and health care utilization. Since there are many survey questions relevant for these outcomes, we create standardized indices by averaging normalized answers to seven health questions and four utilization questions (as in Table V and IX in Finkelstein et al. (2012)). Column 1 shows 2SLS estimates of the effect of any Medicaid on these indices. Consistent with the original results, Medicaid increases both utilization and health significantly. The former rises by 0.1 standard deviations, and the latter by 0.2.

Under EMCO, all compliers in the OHIE are shifted from zero months of Medicaid to some positive amount. Column 2 shows shows the average untreated outcomes for all compliers. Compliers appear to be significantly negatively-selected on health—untreated means are 0.15 standard deviations below the sample average. Utilization is slightly higher than the sample average, but not significantly so. Columns 3-6 then show treated mean outcomes for compliers shifted to 7-12 months of Medicaid, 13-16 months, etc. No individuals were shifted to 1-6 months of coverage. The final row of the table reports the share of all compliers each group comprises. Roughly 33% of compliers, for example, are shifted from zero months to 7-12 months.

These means reveal several interesting patterns. For example, all complier groups with positive density have higher treated health than the average under no insurance. But compliers induced to remain on Medicaid for the longest have the worst health. This is consistent with the adverse selection patterns noted in Finkelstein et al. (2012)—only the sickest remain on Medicaid for long.[15] Compliers who remain on Medicaid for 7-12 months, by contrast, have better than average health. While the data are consistent with positive effects for all groups, it is possible that treatment effects are negative for compliers induced to stay on Medicaid the longest. Any negative effects for this group, however, would have to be outweighed by positive effects for others.[16]

Utilization follows a similar pattern. Individuals induced to remain on Medicaid for the longest also have the highest levels of utilization. Thus the sickest compliers are also the most expensive.

[15]For example, comparing OLS to 2SLS estimates, the authors note that "differences suggest that at least within a low-income population, individuals who select health insurance coverage are in poorer health (and therefore demand more medical care) than those who are uninsured, just as standard adverse selection theory would predict."

[16]The 2SLS estimate in Column 1 is the population share-weighted average of means in Columns 3-5 minus the average in Column 2. As noted above, means under no insurance for each d-type complier group are not identified. The data are consistent with any pattern of treatment effects where the weighted average of complier means under no insurance equals the means in Column 2.

Utilization results also suggest that increases in utilization due to Medicaid are not due to one-shot "pent-up" demand, since utilization is increasing in duration on Medicaid. As with health outcomes, the data are consistent with positive treatment effects on utilization for each complier group. It is possible, however, that the 7-12 month group decreased utilization.

Taken together, the results show the power of EMCO in a setting where it is institutionally plausible. Not only do the simple 2SLS regressions presented in Finkelstein et al. (2012) have a coherent causal interpretation, but identification of complier means reveal important insights into selection patterns.

7 Conclusion

2SLS estimates of the effects of ordered treatments (e.g., years of education, months of prison or Medicaid coverage) can be difficult to interpret. The estimand is a weighted average of causal effects for overlapping sets of compliers and may differ from the estimand of interest to the researcher or relevant for policy. In some settings, an auxiliary assumption that restricts complier types—extensive margin compliers only (EMCO)—may be reasonable. EMCO asserts that the instruments only induce units to switch from zero units of treatment to some positive amount. Under EMCO, 2SLS estimates of the effect of treatment recoded into a binary indicator for *any* treatment capture the average of effects of treatment for mutually exclusive groups of compliers shifted into varying units of treatment due to the instruments. Treated means for each complier type can be recovered using standard techniques, as well as the average untreated mean, allowing for a partial decomposition of the estimand.

When should researchers invoke EMCO? While they have meaningful power to detect violations, our formal tests are based on necessary but not sufficient conditions. In practice, the validity of EMCO is likely to hinge on the institutional details of the experiment at hand. Along with the standard LATE assumptions, invoking EMCO is equivalent to assuming the data is generated by a two-stage selection process where individuals first decide whether to participate in treatment at all and then pick treatment levels. EMCO requires that the instruments only affect the utility of any participation, but not the relative utility of positive treatment levels. In many settings,

such as those with one-sided non-compliance only, it may be clear *a-priori* whether or not this is reasonable. Due to institutional factors, for example, EMCO seems highly likely to hold in the OHIE. When EMCO does hold, we argue researchers should invoke it explicitly to justify their choice of models and make use of the additional identifying power it provides.

References

Abadie, Alberto, "Bootstrap Tests for Distributional Treatment Effects in Instrumental Variable Models," *Journal of the American Statistical Association*, 2002, pp. 284–293.

_ , "Semiparametric instrumental variable estimation of treatment response models," *Journal of econometrics*, 2003, *113* (2), 231–263.

Aizer, Anna and Joseph J. Doyle, "Juvenile Incarceration, Human Capital, and Future Crime: Evidence from Randomly Assigned Judges," *The Quarterly Journal of Economics*, 2015, *130* (2), 759–803.

Allen, Roy, "Testing moment inequalities: Selection versus recentering," *Economics Letters*, 2018, *162*, 124–126.

Andresen, Martin E and Martin Huber, "Instrument-based estimation with binarised treatments: issues and tests for the exclusion restriction," *The Econometrics Journal*, 2021.

Andrews, Donald W. and Panle Jia Barwick, "Inference for parameters defined by moment inequalities: A recommended moment selection procedure," *Econometrica*, 2012, *80* (6), 2805–2826.

_ **and Xiaoxia Shi**, "Inference based on conditional moment inequalities," *Econometrica*, 2013, *81* (2), 609–666.

Andrews, Donald WK and Gustavo Soares, "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 2010, *78* (1), 119–157.

Andrews, Isaiah, Jonathan Roth, and Ariel Pakes, "Inference for linear conditional moment inequalities," Technical Report, National Bureau of Economic Research 2019.

Angrist, Joshua D. and Alan B. Krueger, "Does compulsory school attendance affect schooling and earnings?," *The Quarterly Journal of Economics*, 1991, *106* (4), 979–1014.

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

_ **,** _ **, and Donald B. Rubin**, "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 1996, *91* (434), 444–455.

Armstrong, Timothy B., "On the choice of test statistic for conditional moment inequalities," *Journal of Econometrics*, 2018, *203* (2), 241–255.

Bai, Yuehao, Andres Santos, and Azeem Shaikh, "A practical method for testing many moment inequalities," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2019, (2019-116).

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

Bhuller, Manudeep, Gordon B. Dahl, Katrine V. Løken, and Magne Mogstad, "Incarceration, Recidivism, and Employment," *Journal of Political Economy*, 2020, *128* (4), 1269–1324.

Canay, Ivan A. and Azeem M. Shaikh, "Practical and theoretical advances in inference for partially identified models," in "Advances in Economics and Econometrics: Eleventh World Congress," Vol. 2 Cambridge University Press Cambridge 2017, pp. 271–306.

Card, David, "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in "Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp" University of Toronto Press, Toronto 1995, pp. 201–222.

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato, "Inference on Causal and Structural Parameters using Many Moment Inequalities," *The Review of Economic Studies*, 11 2018, *86* (5), 1867–1900.

_ , Whitney K Newey, and Andres Santos, "Constrained conditional moment restriction models," 2020.

Chetverikov, Denis, Andres Santos, and Azeem M. Shaikh, "The Econometrics of Shape Restrictions," *Annual Review of Economics*, 2018, *10* (1), 31–63.

Cox, Gregory and Xiaoxia Shi, "Simple Adaptive Size-Exact Testing for Full-Vector and Subvector Inference in Moment Inequality Models," *arXiv preprint arXiv:1907.06317*, 2020.

Dahl, Gordon B., "Mobility and the Return to Education: Testing a Roy Model with Multiple Markets," *Econometrica*, 2002, *70* (6), 2367–2420.

Deaton, Angus and John Muellbauer, *Economics and consumer behavior*, Cambridge university press, 1980.

Fang, Zheng and Andres Santos, "Inference on directionally differentiable functions," *The Review of Economic Studies*, 2019, *86* (1), 377–412.

Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P. Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group, "The Oregon Health Insurance Experiment: Evidence from the First Year," *The Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106.

Frandsen, Brigham R, Lars J. Lefgren, and Emily C. Leslie, "Judging judge fixed effects," Technical Report, National Bureau of Economic Research 2019.

Goldin, Jacob, Ithai Z Lurie, and Janet McCubbin, "Health Insurance and Mortality: Experimental Evidence from Taxpayer Outreach," *The Quarterly Journal of Economics*, 09 2020, *136* (1), 1–49.

Gronau, Reuben, "Wage comparisons–A selectivity bias," *Journal of political Economy*, 1974, *82* (6), 1119–1143.

Heckman, James, "Shadow prices, market wages, and labor supply," *Econometrica: journal of the econometric society*, 1974, pp. 679–694.

Heckman, James J., "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy," *Journal of Economic Literature*, June 2010, *48* (2), 356–98.

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

_ **and** _ , "Structural Equations, Treatment Effects, and Econometric Policy Evaluation," *Econometrica*, 2005, *73* (3), 669–738.

Heckman, James J and Rodrigo Pinto, "Unordered monotonicity," *Econometrica*, 2018, *86* (1), 1–35.

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

Hong, Han, Jessie Li et al., "The numerical bootstrap," *The Annals of Statistics*, 2020, *48* (1), 397–412.

Huber, Martin and Giovanni Mellace, "Testing Instrument Validity for LATE Identification Based on Inequality Moment Constraints," *The Review of Economics and Statistics*, 2015, *97* (2), 398–411.

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

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

Kitagawa, Toru, "A test for instrument validity," *Econometrica*, 2015, *83* (5), 2043–2063.

Marshall, John, "Coarsening bias: How coarse treatment measurement upwardly biases instrumental variable estimates," *Political Analysis*, 2016, *24* (2), 157–171.

Marx, Philip, "Sharp Bounds in the Latent Index Selection Model," *arXiv preprint arXiv:2012.02390*, 2020.

Meghir, Costas and Marten Palme, "Assessing the effect of schooling on earnings using a social experiment," *Working paper*, 1999.

Mogstad, Magne, Andres Santos, and Alexander Torgovitsky, "Using instrumental variables for inference about policy relevant treatment parameters," *Econometrica*, 2018, *86* (5), 1589–1619.

Mourifié, Ismael and Yuanyuan Wan, "Testing local average treatment effect assumptions," *Review of Economics and Statistics*, 2017, *99* (2), 305–313.

Mulligan, Casey B and Yona Rubinstein, "Selection, investment, and women's relative wages over time," *The Quarterly Journal of Economics*, 2008, *123* (3), 1061–1110.

Norris, Samuel, "Examiner inconsistency: Evidence from refugee appeals," *University of Chicago, Becker Friedman Institute for Economics Working Paper*, 2019, (2018-75).

_ , Matthew Pecenco, and Jeffrey Weaver, "The Intergenerational and Sibling Effects of Incarceration: Evidence from Ohio," *American Economic Review*, Forthcoming.

Romano, Joseph P, Azeem M Shaikh, and Michael Wolf, "A practical two-step method for testing moment inequalities," *Econometrica*, 2014, *82* (5), 1979–2002.

Rose, Evan K. and Yotam Shem-Tov, "Does Incarceration Increase Crime?," *Working Paper*, 2019.

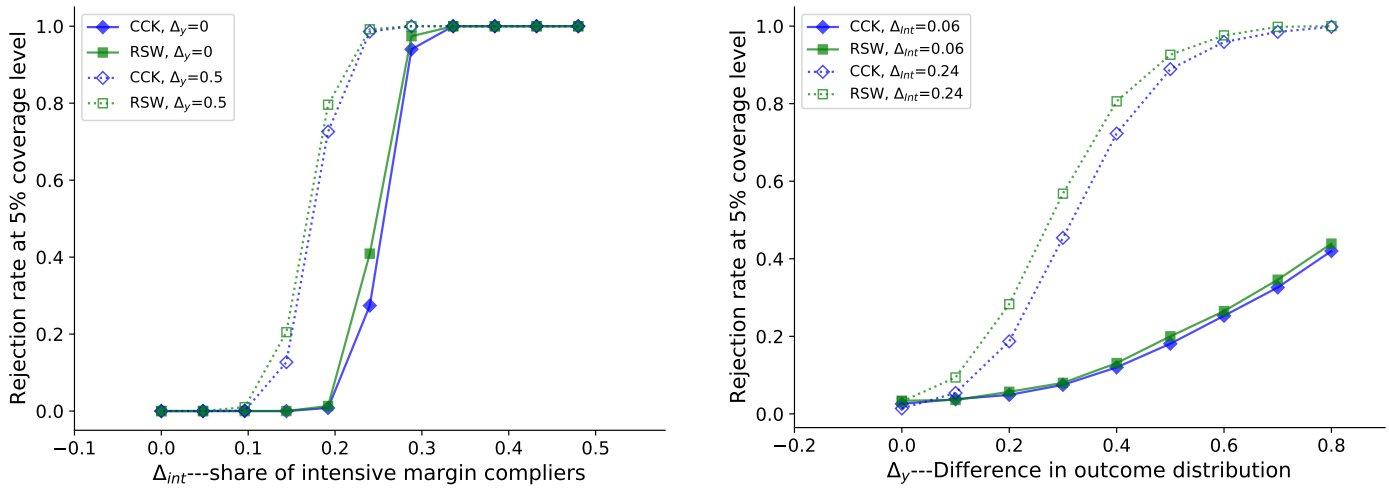
_ and _ , "How does incarceration affect crime? Estimating the dose-response function," *Journal of Political Economy*, 2021.

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

_ , "Ordered Discrete-Choice Selection Models and Local Average Treatment Effect Assumptions: Equivalence, Nonequivalence, and Representation Results," *The Review of Economics and Statistics*, 2006, *88* (3), 578–581.

Figures and Tables

Figure 1: Simulations of tests of Assumptions 1 and 2 in DGP in Table 1



(a) Varying the share of intensive margin compliers (Δ_{Int})

(b) Varying the difference in outcome distributions (Δ_y)

Notes: This figure reports results from Monte-Carlo simulations based on the data generating process (DGP) in Table 1. We used 1,000 simulations for each combination of parameter values (i.e., each point in the figure). In Panel (a), the x-axis shows the true proportion of intensive margin compliers (Δ_{int}). The y-axis shows the share of simulations in which the null hypothesis that $\Delta_{int} = 0$ is rejected holding Δ_y fixed. The rejection rate when $\Delta_{int} = 0$ represents the size of the test, i.e., the likelihood of Type-I errors. Panel (b) examines power when intensive margin compliers have a different distribution of outcomes than extensive margin compliers. We use a binary outcome and model differences in the distribution of Y_i uhttps://www.overleaf.com/project/5fa84d8bf67ba20e8474126dsing the parameter: $\Delta_y \equiv \Pr(Y_i = 1|D_i(1) = 1 > D_i(0)) - \Pr(Y_i = 1|D_i(1) > D_i(0) = 1)$. Panel (b) holds fixed the proportion of intensive margin compliers and shows that detection power increases when varying Δ_y. RSW indicates results using Romano et al. (2014)'s moment inequality testing procedure. CCK indicates Chernozhukov et al. (2018)'s method.

Table 1: Illustration of Proposition 4

Treatment $(D_i = d)$	$\Pr(D_i(0) = d)$	$\Pr(D_i(1) = d)$
0	a	$a - \Delta^1_{ext} - \Delta^2_{ext}$
1	b	$b + \Delta^1_{ext} - \Delta_{int}$
2	$1 - a - b$	$1 - a - b + \Delta^2_{ext} + \Delta_{int}$

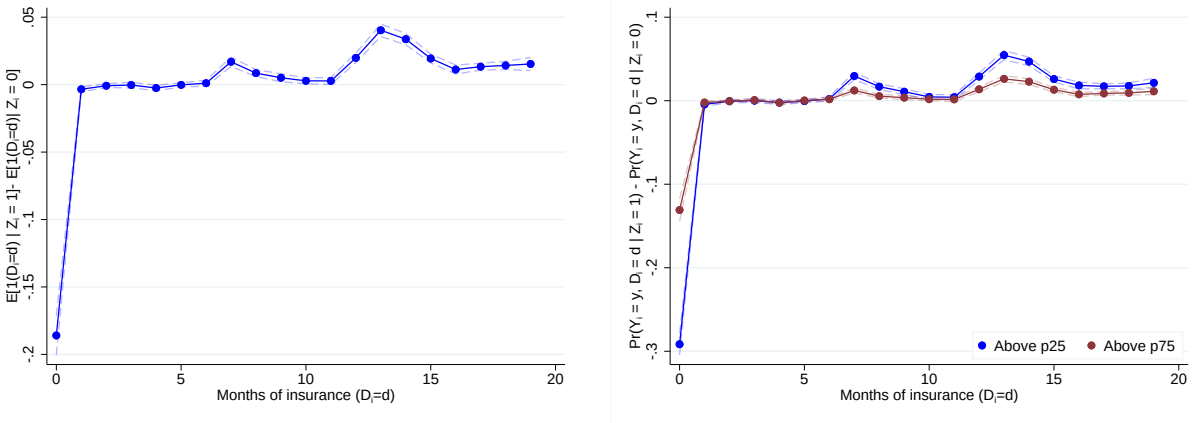
Table 2: Decomposition of Treatment Effects in the OHIE

	(1) 2SLS	(2) 0m	(3) 7-12m	(4) 13-18m	(5) 19-24m
			Health		
Effect of any Medicaid	0.180***				
	(0.0436)				
Complier mean Y		-0.152***	0.0785*	-0.0197	-0.114
		(0.0366)	(0.0360)	(0.0242)	(0.116)
			Utilization		
Effect of any Medicaid	0.108**				
	(0.0416)				
Complier mean Y		-0.0429	-0.0671	0.150***	0.171
		(0.0327)	(0.0415)	(0.0261)	(0.134)
N	17896	17896	17896	17896	17896
Share of compliers		1	0.325	0.620	0.0735

Notes: This table shows 2SLS effects of any Medicaid and complier mean outcome levels for health and utilization indices. The health index is the average of seven survey questions on health, each normalized to be mean zero and standard deviation 1, and likewise with four questions on utilization. Positive values of each reflect improved mental and physical health and increased utilization, respectively. Column 1 shows the 2SLS estimate of the effect of any Medicaid on each index. As in Finkelstein et al. (2012), the specification includes lottery fixed effects and clusters standard errors at the household level. Column 2 shows complier mean outcomes under no insurance. They are estimated from an identical 2SLS specification with $Y1(D_i = 0)$ as the outcome and $1(D_i = 0)$ as the endogenous variable. Columns 3-6 show mean outcomes for compliers shifted from zero to the quantity of insurance listed in the header. They are estimated from 2SLS regressions with $Y1(D_i = d)$ as the outcome and $1(D_i = d)$ as the endogenous variable. The final row of the table reports the share of each d-type compliers. For example, column 2 identifies the untreated mean for 100% of compliers, since it averages across all d-types. Column 3 shows treated means for compliers shifted to 7-12 months, which comprises 33% of all compliers.

A Additional Figures

Figure A.1: Visual Tests of EMCO in Finkelstein et al. (2012)'s Analysis of the OHIE



(a) Proposition 4 moment conditions

(b) Proposition 5 moment conditions

Notes: This figure shows visual tests of Propositions 4 and 5. Panel (a) estimates the difference in density of months of insurance coverage between individuals lotteried into Medicaid ($Z_i = 1$) and those not ($Z_i = 0$). Consistent with EMCO, there is a large decrease in mass at $D_i = 0$ and positive increases for all $D_i > 0$. Panel (b) estimates differences in the intersection of outcome values and treatment levels. The outcome is binary indicator for whether a standardized measure of health care utilization falls above sample quantiles. Consistent with EMCO, there is a large decreases at $D_i = 0$ and positive increase elsewhere. The figure was produced using the replication data from Finkelstein et al. (2012). Since administrative data on mortality and hospitalization are unavailable in the public data, we use the survey-sample and estimate differences by regressing indicators for $D_i = d$ or $(D_i = d)(Y_i = y)$ on the lottery dummy controlling for lottery strata, as in the original analysis. Standard errors are clustered at the household level.

B Proofs

B.1 Proof of Proposition 1

The IV estimator $\beta_{recoded}$ is equal to:

$$\begin{aligned}
\beta_{recoded} &\equiv \frac{\mathbb{E}\left[Y_i|Z_i=1\right]-\mathbb{E}\left[Y_i|Z_i=0\right]}{\mathbb{E}\left[1(D_i>0)|Z_i=1\right]-\mathbb{E}\left[1(D_i>0)|Z_i=0\right]} \\
&= \frac{\mathbb{E}\left[Y_i(D_i(1))|Z_i=1\right]-\mathbb{E}\left[Y_i(D_i(0))|Z_i=0\right]}{\Pr(D_i(1)>D_i(0)=0)} \\
&= \frac{\mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|D_i(1)>D_i(0)=0\right]\Pr(D_i(1)>D_i(0)=0)}{\Pr(D_i(1)>D_i(0)=0)} \\
&\quad + \frac{\mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|D_i(1)>D_i(0)>0\right]\Pr(D_i(1)>D_i(0)>0)}{\Pr(D_i(1)>D_i(0)=0)} \\
&= \mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|D_i(1)>D_i(0)=0\right] \\
&\quad + \mathbb{E}\left[Y_i(D_i(1))-Y_i(D_i(0))|D_i(1)>D_i(0)>0\right]\frac{\Pr(D_i(1)>D_i(0)>0)}{\Pr(D_i(1)>D_i(0)=0)}
\end{aligned} \tag{B.1}$$

where the first equality follows from Assumption 1, specifically the random assignment of Z_i (i.e., independent of potential outcomes) and monotonicity (i.e., $D_i(1) \geq D_i(0)$ for all i).

B.2 Proof of Proposition 2

We begin by presenting a more general version of Proposition 2. Proposition 6 is analogous to Theorem 3.1 in Abadie (2003), generalizing this result for the binary treatment case to multi-valued treatments. A direct implication of Proposition 6 is that compliers' observable pre-treatment characteristics can be identified as is described in Equation 3.

Proposition 6. *Let $g(\cdot)$ be a measurable function on (Y_i, D_i, X_i) such that $\mathbb{E}\left[g(Y_i, D_i, X_i)\right] < \infty$ and assume that $\Pr(D_i(1) = d > D_i(0)) > 0 \ \forall d > 0$. Then, under Assumptions 1 and 2:*

$$\begin{aligned}
(i)\ & \frac{1}{\Pr(D_i(1)>D_i(0))}\mathbb{E}\left[\kappa g(Y_i, 1(D_i>0), X_i)\right] = \mathbb{E}\left[g(Y_i, 1(D_i>0), X_i)|D_i(1)>D_i(0)=0\right], \\
(ii)\ & \frac{1}{\Pr(D_i(1)=d>D_i(0))}\mathbb{E}\left[\kappa_{(d)} g(Y_i, X_i)\right] = \mathbb{E}\left[g(Y_i(d), X_i)|D_i(1)=d>D_i(0)=0\right] \quad for \quad d>0, \\
(iii)\ & \frac{1}{\Pr(D_i(1)>D_i(0))}\mathbb{E}\left[\kappa_{(0)} g(Y_i, X_i)\right] = \mathbb{E}\left[g(Y_i(0), X_i)|D_i(1)>D_i(0)=0\right]
\end{aligned}$$

where:

$$\kappa = 1 - \frac{[1 - 1(D_i > 0)]\, Z_i}{\Pr(Z_i = 1)} - \frac{1(D_i > 0)\,(1 - Z_i)}{\Pr(Z_i = 0)}$$
$$\kappa_{(d)} = 1(D_i = d)\frac{Z_i - \Pr(Z_i = 1)}{\Pr(Z_i = 0)\Pr(Z_i = 1)},\ d > 0$$
$$\kappa_{(0)} = [1 - 1(D_i > 0)]\,\frac{(1 - Z_i) - \Pr(Z_i = 0)}{\Pr(Z_i = 0)\Pr(Z_i = 1)}$$

Proof of part (i)

Recall that both the overall share of compliers and the share of each d-type complier group can be identified:

$$\Pr(D_i(1) > D_i(0)) = \mathbb{E}\left[1(D_i > 0)|Z_i = 1\right] - \mathbb{E}\left[1(D_i > 0)|Z_i = 0\right] \tag{B.2}$$
$$\Pr(D_i(1) = d > D_i(0)) = \mathbb{E}\left[1(D_i = d)|Z_i = 1\right] - \mathbb{E}\left[1(D_i = d)|Z_i = 0\right] \quad \forall d > 0$$

The intuition behind the proof of part (i) is simple. The κ term "extracts" the compliers by taking the overall mean and "subtracting" from it the "never-takers" and "always-takers." More formally, note that:

$$\begin{aligned}\mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i(1) > D_i(0)\right]\Pr(D_i(1) > D_i(0)) &= \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)\right] \\ - \underbrace{\mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i(1) = D_i(0) > 0\right]\Pr(D_i(1) = D_i(0) > 0)}_{\text{"Always-takers"}} \\ - \underbrace{\mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i(1) = D_i(0) = 0\right]\Pr(D_i(1) = D_i(0) = 0)}_{\text{"Never-takers"}}\end{aligned} \tag{B.3}$$

Next we examine each of the expressions ins $\mathbb{E}\left[g(\cdot)\kappa\right]$. The first is $\mathbb{E}\left[g(\cdot)\cdot 1\right]$ and the second is:

$$\begin{aligned}&\mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)(1 - 1(D_i > 0))Z_i\right] = \\ &\quad \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i = 0, Z_i = 1\right]\Pr(D_i = 0, Z_i = 1) \\ &= \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i(1) = 0\right]\Pr(D_i(1) = 0|Z_i = 1)\Pr(Z_i = 1) \\ &= \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i(1) = D_i(0) = 0\right]\Pr(D_i(1) = D_i(0) = 0)\Pr(Z_i = 1)\end{aligned} \tag{B.4}$$

and the third expression is:

$$\begin{aligned}&\mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)1(D_i > 0)(1 - Z_i)\right] = \\ &\quad \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i > 0, Z_i = 0\right]\Pr(D_i > 0, Z_i = 0) \\ &= \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i(0) > 0\right]\Pr(D_i(0) > 0|Z_i = 0)\Pr(Z_i = 0) \\ &= \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i)|D_i(1) = D_i(0) > 0\right]\Pr(D_i(1) = D_i(0) > 0)\Pr(Z_i = 0)\end{aligned} \tag{B.5}$$

Finally, substituting Equations B.3, B.4 and B.5 into $\mathbb{E}\left[\kappa \cdot g(Y_i, 1(D_i > 0), X_i)\right]$ completes the proof that:

$$\frac{1}{\Pr(D_i(1) > D_i(0))}\mathbb{E}\left[\kappa \cdot g(Y_i, 1(D_i > 0), X_i)\right] = \mathbb{E}\left[g(Y_i, 1(D_i > 0), X_i) | D_i(1) > D_i(0)\right]$$

Proof of part (ii)

First note that:

$$\begin{aligned} &\mathbb{E}\left[g(Y_i, X_i)\kappa_{(d)}\right] = \\ &= \mathbb{E}\left[\frac{g(Y_i, X_i)1(D_i = d)Z_i}{\Pr(Z_i = 1)\Pr(Z_i = 0)}\right] - \mathbb{E}\left[\frac{g(Y_i, X_i)1(D_i = d)}{\Pr(Z_i = 0)}\right] \end{aligned} \tag{B.6}$$

Next we examine each of the two expressions in Equation B.6:

$$\begin{aligned} &\mathbb{E}\left[g(Y_i, X_i)1(D_i = d)Z_i\right] = \\ &= \mathbb{E}\left[g(Y_i, X_i)|D_i = d, Z_i = 1\right]\Pr(D_i = d, Z_i = 1) \\ &= \mathbb{E}\left[g(Y_i(d), X_i)|D_i(1) = d\right]\Pr(D_i = d|Z_i = 1)\Pr(Z_i = 1) \\ &= \mathbb{E}\left[g(Y_i(d), X_i)|D_i(1) = D_i(0) = d\right]\Pr(D_i(1) = D_i(0) = d)\Pr(Z_i = 1) \\ &\quad + \mathbb{E}\left[g(Y_i(d), X_i)|D_i(1) = d > D_i(0)\right]\Pr(D_i(1) = d > D_i(0))\Pr(Z_i = 1) \end{aligned} \tag{B.7}$$

and

$$\begin{aligned} &\mathbb{E}\left[g(Y_i, X_i)1(D_i = d)\right] = \\ &= \mathbb{E}\left[g(Y_i, X_i)|D_i = d\right]\Pr(D_i = d) \\ &= \mathbb{E}\left[g(Y_i(d), X_i)|D_i = d, Z_i = 1\right]\Pr(D_i = d, Z_i = 1) \\ &\quad + \mathbb{E}\left[g(Y_i(d), X_i)|D_i = d, Z_i = 0\right]\Pr(D_i = d, Z_i = 0) \\ &= \mathbb{E}\left[g(Y_i(d), X_i)|D_i(1) = d\right]\Pr(D_i = d|Z_i = 1)\Pr(Z_i = 1) \\ &\quad + \mathbb{E}\left[g(Y_i(d), X_i)|D_i(0) = d\right]\Pr(D_i = d|Z_i = 0)\Pr(Z_i = 0) \\ &= \mathbb{E}\left[g(Y_i(d), X_i)|D_i(1) = d > D_i(0)\right]\Pr(D_i(1) = d > D_i(0))\Pr(Z_i = 1) \\ &\quad + \mathbb{E}\left[g(Y_i(d), X_i)|D_i(1) = D_i(0) = d\right]\Pr(D_i(1) = D_i(0) = d) \end{aligned} \tag{B.8}$$

Substituting the above Equations B.7 and B.8 into Equation B.6 and rearranging completes the proof.

Proof of part (iii)

The proof of part (iii) is similar to that of part (ii) but we include it for completeness. We begin, same as before, by noting that:

$$\begin{aligned}&\mathbb{E}\left[g(Y_i,X_i)\kappa_{(0)}\right]= \\ &=\mathbb{E}\left[\frac{g(Y_i,X_i)(1-1(D_i>d))(1-Z_i)}{\Pr(Z_i=1)\Pr(Z_i=0)}\right]-\mathbb{E}\left[\frac{g(Y_i,X_i)(1-1(D_i>d))}{\Pr(Z_i=1)}\right]\end{aligned} \tag{B.9}$$

Next we examine each of the two expressions in Equation B.9:

$$\begin{aligned}&\mathbb{E}\left[g(Y_i,X_i)(1-1(D_i>d))(1-Z_i)\right]= \\ &=\mathbb{E}\left[g(Y_i,X_i)|D_i=0,Z_i=0\right]\Pr(D_i=0,Z_i=0) \\ &=\mathbb{E}\left[g(Y_i(0),X_i)|D_i(0)=0,Z_i=0\right]\Pr(D_i=0|Z_i=0)\Pr(Z_i=0) \\ &=\mathbb{E}\left[g(Y_i(0),X_i)|D_i(0)=0\right]\Pr(D_i(0)=0)\Pr(Z_i=0) \\ &=\mathbb{E}\left[g(Y_i(0),X_i)|D_i(1)=D_i(0)=0\right]\Pr(D_i(1)=D_i(0)=0)\Pr(Z_i=0) \\ &\quad+\mathbb{E}\left[g(Y_i(0),X_i)|D_i(1)>D_i(0)=0\right]\Pr(D_i(1)>D_i(0)=0)\Pr(Z_i=0)\end{aligned} \tag{B.10}$$

and

$$\begin{aligned}&\mathbb{E}\left[g(Y_i,X_i)(1-1(D_i>d))\right]= \\ &=\mathbb{E}\left[g(Y_i,X_i)|D_i=0\right]\Pr(D_i=0) \\ &=\mathbb{E}\left[g(Y_i(0),X_i)|D_i(0)=0,Z_i=0\right]\Pr(D_i(0)=0|Z_i=0)\Pr(Z_i=0) \\ &\quad+\mathbb{E}\left[g(Y_i(0),X_i)|D_i(0)=0,Z_i=1\right]\Pr(D_i(1)=0|Z_i=1)\Pr(Z_i=1) \\ &=\mathbb{E}\left[g(Y_i(0),X_i)|D_i(1)>D_i(0)=0\right]\Pr(D_i(1)>D_i(0)=0)\Pr(Z_i=0) \\ &\quad+\mathbb{E}\left[g(Y_i(0),X_i)|D_i(1)=D_i(0)=0\right]\Pr(D_i(1)=D_i(0)=0)\end{aligned} \tag{B.11}$$

Substituting the above Equations B.10 and B.11 into Equation B.9 and rearranging completes the proof.

B.3 Proof of Proposition 3

The proof that the model in Equation 6 satisfies the restrictions LATE and EMCO place on treatment choices is trivial. To show that LATE and EMCO imply the selection model representation in Equation 6, first note that distributional assumptions on U_i^{Ext} and U_i^{Int} are inconsequential. For any strictly increasing distribution function G, an equivalent model can be written with $\tilde{U_i}^{Ext}=G^{-1}(F_{U^{Ext}}(U_i^{Ext}),\tilde{U_i}^{Int}=G^{-1}(F_{U^{Int}}(U_i^{Int})),\tilde{\lambda}(Z_i)=G^{-1}(F_{U^{Ext}}(\lambda(Z_i)))$, and $\tilde{\pi_d}=G^{-1}(F_{U^{Ext}}(\pi_d))$, where $F_{U^{Ext}}$ and $F_{U^{Int}}$ are the marginal CDFs of U_i^{Ext} and U_i^{Int}, respectively.

Hence it is without loss of generality to assume that U_i^{Ext} and U_i^{Int} are jointly uniform random variables.

EMCO restricts choice behavior to three categories of individuals: never-takers, who have $D_i(1) = D_i(0) = 0$; $\bar{D} - 1$ types of always-takers, each with $D_i(1) = D_i(0) = d$; and $\bar{D} - 1$ types of compliers, each with compliers $D_i(1) = d > 0 = D_i(0)$. These populations map into the distribution of U_i^{Ext} and U_i^{Int} as follows:

1. Never-takers: $U_i^{Ext} \in [0, \pi_0(1)]$.
2. Type-d always-takers: $U_i^{Ext} \in (\pi_0(0), 1]$, $\pi_{d+1} \leq U_i^{int} < \pi_d$
3. Type-d compliers: $U_i^{Ext} \in (\pi_0(1), \pi_0(0)]$, $\pi_{d+1} \leq U_i^{int} < \pi_d$

The model in Equation 6 therefore implies the same set of observed and counterfactual treatment choices as under LATE and EMCO. The LATE framework assumptions can therefore be represented by a two latent factor hurdle model without imposing any additional restrictions on the DGP.

B.4 Equivalence between $\Pr(D_i = d|Z_i = 1) - \Pr(D_i = d|Z_i = 0) \geq 0$ and $\Pr(D_i \geq d|Z_i = 1) - \Pr(D_i \geq d|Z_i = 0) \geq \Pr(D_i \geq d+1|Z_i = 1) - \Pr(D_i \geq d+1|Z_i = 0)$

The LATE assumptions (Assumption 1) and EMCO implies that $D_i(1) > D_i(0) \Rightarrow D_i(0) = 0$, and thus:

$$\Pr(D_i = d|Z_i = 1) - \Pr(D_i = d|Z_i = 0) = \Pr(D_i(1) = d > D_i(0) = 0) \tag{B.12}$$

and

$$\begin{aligned} \Pr(D_i \geq d|Z_i = 1) - \Pr(D_i \geq d|Z_i = 0) &= \Pr(D_i(1) \geq d > D_i(0) = 0) \\ &= \sum_{k=d}^{\bar{D}} \Pr(D_i(1) = k > D_i(0) = 0) \end{aligned} \tag{B.13}$$

and hence:

$$\begin{aligned} &\Pr(D_i \geq d|Z_i = 1) - \Pr(D_i \geq d|Z_i = 0) - [\Pr(D_i \geq d+1|Z_i = 1) - \Pr(D_i \geq d+1|Z_i = 0)] \\ &= \sum_{k=d}^{\bar{D}} \Pr(D_i(1) = k > D_i(0) = 0) - \sum_{k=d+1}^{\bar{D}} \Pr(D_i(1) = k > D_i(0) = 0) \\ &= \Pr(D_i(1) = d > D_i(0) = 0) \geq 0 \end{aligned} \tag{B.14}$$

This completes the proof.

B.5 Proofs of Propositions 4 and 5

The arguments for Propositions 4 and 5 are similar to the ones from Balke and Pearl (1997) for the binary treatment (D_i) setting. Assumptions 1 (LATE) and 2 (EMCO) imply that: $D_i(1) > D_i(0) \Rightarrow D_i(0) = 0$. Thus, the instrument shifts individuals from zero to different values of the treatment $D_i \in \{1, \ldots, \bar{D}\}$. Hence, the mass of every value of $d > 1$ (and every value of $y \in \mathcal{Y}$) should be lower under $Z_i = 0$ than $Z_i = 1$.

B.5.1 Proof of Proposition 4

We begin with the derivations of Proposition 4 (i):

$$\begin{aligned}\mathbb{E}\left[1(D_i = 0)|Z_i = 1\right] &= \Pr(D_i = 0|Z_i = 1) = \Pr(D_i(1) = 0) \\ &= \underbrace{\Pr(D_i(1) = 0, D_i(0) > 0)}_{= 0 \text{ by monotonicity (Assumption 1, (iii))}} + \Pr(D_i(1) = 0, D_i(0) = 0)\end{aligned} \tag{B.15}$$

where the second equality follows from the independence between Z_i and $D_i(z)$ (Assumption 1, (ii)), and

$$\begin{aligned}\mathbb{E}\left[1(D_i = 0)|Z_i = 0\right] &= \Pr(D_i = 0|Z_i = 0) = \Pr(D_i(0) = 0) \\ &= \underbrace{\Pr(D_i(1) > 0, D_i(0) = 0)}_{> 0 \text{ by relevancy (Assumption 1, (i))}} + \Pr(D_i(1) = 0, D_i(0) = 0) \\ &\Rightarrow \Pr(D_i = 0|Z_i = 1) - \Pr(D_i = 0|Z_i = 0) \leq 0\end{aligned} \tag{B.16}$$

Next we turn to the derivations of Proposition 4 (ii):

$$\begin{aligned}\mathbb{E}\left[1(D_i = d)|Z_i = 1\right] &= \Pr(D_i = d|Z_i = 1) = \Pr(D_i(1) = d) \\ &= \underbrace{\Pr(D_i(1) = d, D_i(0) = 0)}_{\geq 0 \text{ by relevancy (Assumption 1, (i))}} + \underbrace{\Pr(D_i(1) = d > D_i(0) > 0)}_{= 0 \text{ by EMCO (Assumption 2)}} \\ &\quad + \underbrace{\Pr(D_i(1) = D_i(0))}_{\text{never-takers+always-takers}}\end{aligned} \tag{B.17}$$

and

$$\begin{aligned}\mathbb{E}\left[1(D_i = d)|Z_i = 0\right] &= \Pr(D_i = d|Z_i = 0) = \Pr(D_i(0) = d) \\ &= \underbrace{\Pr(D_i(0) = d > D_i(1))}_{=0 \text{ by monotonicity (Assumption 1)}} \\ &\quad + \underbrace{\Pr(D_i(1) > d = D_i(0))}_{=0 \text{ by EMCO (Assumption 2)}} + \underbrace{\Pr(D_i(1) = D_i(0))}_{\text{never-takers+always-takers}} \\ &\Rightarrow \Pr(D_i = d|Z_i = 1) - \Pr(D_i = d|Z_i = 0) \geq 0\end{aligned} \tag{B.18}$$

This completes the proof of Proposition 4.

B.5.2 Proof of Proposition 5

We begin with the derivations of Proposition 5 (i). For any measurable set A:

$$\begin{aligned}\mathbb{E}\left[1(D_i = 0)1(Y_i \in A)|Z_i = 1\right] &= \Pr(D_i = 0, Y_i \in A|Z_i = 1) \\ &= \Pr(D_i(1) = 0, Y_i(D_i(1)) \in A) \\ &= \Pr(D_i(1) = 0, Y_i(0) \in A) \\ &= \Pr(D_i(1) = 0, D_i(0) = 0, Y_i(0) \in A)\end{aligned} \tag{B.19}$$

and using similar derivations it follows that:

$$\begin{aligned}\mathbb{E}\left[1(D_i = 0)1(Y_i \in A)|Z_i = 0\right] &= \Pr(D_i(0) = 0, Y_i(D_i(0)) \in A|Z_i = 0) \\ &= \underbrace{\Pr(D_i(1) > 0, D_i(0) = 0, Y_i(0) \in A)}_{>0 \text{ by relevancy (Assumption 1, (i))}} \\ &\quad + \Pr(D_i(1) = 0, D_i(0) = 0, Y_i(0) \in A) \\ &\Rightarrow \Pr(D_i = 0, Y_i \in A|Z_i = 1) - \Pr(D_i = 0, Y_i \in A|Z_i = 0) \leq 0\end{aligned} \tag{B.20}$$

Finally, we turn next to the derivations of Proposition 5 (ii). For any measurable set A and

$d > 0$:

$$\begin{aligned}
\mathbb{E}\left[1(D_i = d)1(Y_i \in A)|Z_i = 1\right] &= \Pr(D_i = d, Y_i \in A|Z_i = 1) \\
&= \Pr(D_i(1) = d, Y_i(d) \in A) \\
&= \underbrace{\Pr(D_i(1) = d, D_i(0) = 0, Y_i(d) \in A)}_{\geq 0 \text{ by relevancy (Assumption 1, (i))}} \\
&\quad + \underbrace{\Pr(D_i(1) = d > D_i(0) > 0, Y_i(d) \in A)}_{= 0 \text{ by EMCO (Assumption 2)}} \\
&\quad + \underbrace{\Pr(D_i(1) = D_i(0), Y_i(d) \in A)}_{\text{never-takers+always-takers}}
\end{aligned} \tag{B.21}$$

and

$$\begin{aligned}
\mathbb{E}\left[1(D_i = d)1(Y_i(d) \in A)|Z_i = 0\right] &= \Pr(D_i = d, Y_i(d) \in A|Z_i = 0) \\
&= \Pr(D_i(0) = d, Y_i(d) \in A) \\
&= \underbrace{\Pr(D_i(1) = D_i(0), Y_i(d) \in A)}_{\text{never-takers+always-takers}} \\
&\quad + \underbrace{\Pr(D_i(1) > D_i(0) = d, Y_i(d) \in A)}_{= 0 \text{ by EMCO (Assumption 2)}} \\
&\Rightarrow \Pr(D_i = d, Y_i(d) \in A|Z_i = 1) - \Pr(D_i = d, Y_i(d) \in A|Z_i = 0) \geq 0
\end{aligned} \tag{B.22}$$

This completes the proof of Proposition 5.

C Including covariate adjustment in the moment inequality tests

Incorporating covariate adjustment into the moment inequality tests (and visualizations) is simple to implement. We present (and implement in our code) two options. First, we can interact the moment conditions in Equations D.1a-D.1d with specific values of the covariates $X_i = x$. This can be used to construct additional and even more powerful tests of EMCO. For example, the moment condition in Equation D.1a must hold for every $X_i = x$:

$$\mathbb{E}\left[\frac{1}{\Pr(Z_i=1)}1(D_i=0)1(X_i=x)Z_i - \frac{1}{1-\Pr(Z_i=1)}1(D_i=0)1(X_i=x)(1-Z_i)\right] \leq 0$$

Interacting the moment conditions in Equations D.1a-D.1d with different pre-treatment covariates (as is demonstrated above) will increase the number of moment inequalities exponentially. However, this is not necessarily a problem, since one of the advantages of the testing procedures we discuss in Section **??** (Romano et al., 2014; Chernozhukov et al., 2018) is that they can both be used also in high-dimensional settings when the number of moment inequalities is larger than the number of observations (Bai et al., 2019).

The second option for incorporating covariates is by including them as linear controls in an OLS specification. Consider for example the following specification:

$$1(D_i = d) = \alpha Z_i + X_i'\gamma + e_i \tag{C.1}$$

when X_i only includes an intercept then:

$$\hat{\alpha} \to \mathbb{E}\left[1(D_i=d)|Z_i=1\right] - \mathbb{E}\left[1(D_i=d)|Z_i=0\right] = \Pr(D_i(1)=d>0=D_i(0))$$

similarly we can express all the moment conditions in Propositions 4 and 5 as simple OLS specifications without covariates except an intercept.

Another observation is that by Frisch–Waugh–Lovell α is the same as $\tilde{\alpha}$:

$$1(D_i = d) = \tilde{\alpha}\tilde{Z}_i + \tilde{e}_i \tag{C.2}$$

where $\tilde{Z}_i$ is the residuals from projecting Z_i on X_i, i.e., $Z_i = X_i'\gamma + \tilde{Z}_i$. The equivalence between equations C.1 and C.2 implies that any OLS coefficient can be reformulated as an unconditional moment condition and thus the moment inequality tests of Romano et al. (2014) and Chernozhukov et al. (2018) can be then be used one-sided hypothesis.

D Formal tests of EMCO

Propositions 4 and 5 imply a set of inequality restrictions must hold in the population distribution of (Y_i, D_i, Z_i). To jointly test the restrictions imposed by LATE and EMCO, tools from the moment inequality literature can be applied to the sample analogues of these restrictions. Specifically, LATE and EMCO imply that the following inequalities are satisfied:

$$\mathbb{E}\left[\frac{1}{\Pr(Z_i=1)}1(D_i=0)Z_i - \frac{1}{1-\Pr(Z_i=1)}1(D_i=0)(1-Z_i)\right] \leq 0 \tag{D.1a}$$

$$\mathbb{E}\left[\frac{1}{1-\Pr(Z_i=1)}1(D_i=d)(1-Z_i) - \frac{1}{\Pr(Z_i=1)}1(D_i=d)Z_i\right] \leq 0 \quad \forall\, d>0 \tag{D.1b}$$

$$\mathbb{E}\left[\frac{1}{\Pr(Z_i=1)}\cdot 1(D_i=0)\cdot 1(Y_i\in A)\cdot Z_i - \frac{1}{1-\Pr(Z_i=1)}\cdot 1(D_i=0)\cdot 1(Y_i\subset A)\cdot(1-Z_i)\right] \leq 0 \quad \forall A\in\mathcal{Y} \tag{D.1c}$$

$$\mathbb{E}\left[\frac{1}{1-\Pr(Z_i=1)}\cdot 1(D_i=d)\cdot 1(Y_i\in A)\cdot(1-Z_i) - \frac{1}{\Pr(Z_i=1)}\cdot 1(D_i=d)\cdot 1(Y_i\in A)\cdot Z_i\right] \leq 0 \quad \forall\, d>0 \tag{D.1d}$$

where D.1a and D.1b ($\bar{D}+1$ total restrictions) follow from Proposition 4, and D.1c and D.1d are the result of Proposition 5. The latter must hold for any measureable set A in the support of Y_i. If Y_i is discrete, for example, these restrictions add $(\bar{D}+1)\cdot|\mathcal{Y}|$ total inequalities. Note that if $A=\mathcal{Y}$, the second two sets of restrictions nest the first two sets. These restrictions must also hold for any outcome, as well as conditional on any set of covariates. Including additional moments for other outcomes or interacting the moments with covariates can potentially yield sharper tests.[17]

Let $\bar{m}=(\bar{m}_1,\bar{m}_2,\ldots,\bar{m}_p)$ collect the sample analogs of the moments in Equations D.1a-D.1d, with $\bar{m}_p=\frac{1}{N}\sum_{i=1}^{N}m_{ip}$ and with variance estimates of each $S_j^2=\frac{\sum_{i=1}^{N}(m_{ij}-\bar{m}_j)^2}{N-1}$.

We are interested in testing the null hypothesis that:

$$H_0:\mathbb{E}[m_j]\leq 0 \quad \text{for all} \quad j=1,\ldots,p$$

against the alternative:

$$H_1:\mathbb{E}[m_j]>0 \quad \text{for some} \quad j=1,\ldots,p$$

The econometrics literature has developed a variety of methods for testing moment inequalities of this form (see Canay and Shaikh, 2017, for a recent review). We use a Kolmogorov-Smirnov (KS) variance adjusted test statistic:

$$T_n=\max\left\{\max_{1\leq j\leq p}\left\{\frac{\sqrt{N}\bar{m}_j}{S_j}\right\},0\right\}$$

The KS test statistic is commonly used in the literature on testing moment inequalities (e.g., Bai et al., 2019). It is powerful when the objective is to detect whether any of the inequalities is violated

[17] These restrictions also capture a direct implication of Assumption 1 shown by Angrist and Imbens (1995) to hold regardless of whether or not EMCO is true: $\mathbb{E}[1(D_i\geq d)|Z_i=1]-\mathbb{E}[1(D_i\geq d)|Z_i=0]\geq 0$ for all d.

(Chernozhukov et al., 2018), as is the case in our setting.[18]

To obtain critical values, we use two recently proposed procedures that are computationally tractable and work well in high-dimensional settings (e.g., when the number of moments is larger than the number of observations). The first is Romano et al. (2014)'s two-step moment recentering approach. The second is Chernozhukov et al. (2018), who proposed a two-step bootstrap procedure based on moment selection (Andrews and Soares, 2010; Andrews and Barwick, 2012; Andrews and Shi, 2013).[19] Both of the tests require two steps. In the first step, a moment recentering or selection procedure takes place. The second step conducts inference on the test statistic using the bootstrap over the recentered or selected moments.[20]

Figure 1 presents results from Monte-Carolo simulations (described more in the detail in Section 5 and Appendix E) using both the procedures advocated by Chernozhukov et al. (2018) and Romano et al. (2014). Recently, Allen (2018) and Bai et al. (2019) compared the detection power of these two procedures and concluded that the recentering approach proposed by Romano et al. (2014) is at least as powerful as the moment selection procedure proposed by Chernozhukov et al. (2018). These findings are consistent with our simulation results in Figure 1.

[18] Armstrong (2018) shows that KS test statistics can have power advantages over Cramer–vonMises-style test statistics.

[19] Chernozhukov et al. (2018) also propose methods based on self-normalized sums that allow one to analytically calculate critical values and are faster computationally. However, these approaches are generally less powerful than the bootstrap procedure. We use the bootstrap procedure (as in Bai et al. (2019)) when comparing the power of the test proposed by Chernozhukov et al. (2018) to that proposed by Romano et al. (2014).

[20] Other methods proposed recently for inference and testing on moment inequality models includes Andrews et al. (2019)'s for linear conditional moment inequalities, Chernozhukov et al. (2020), and Cox and Shi (2020).

E Monte-Carlo simulations details

In this appendix, we describe in detail how the Monte-Carlo simulations in Figure 1 are conducted. The DGP is based on the example in Table 1. The outcome Y_i is binary, the instrument Z_i is binary, and the treatment $D_i \in \{0, 1, 2\}$.

There are 4 types of individuals:

1. Non-compliers ($D_i(1) = D_i(0)$), with sample share $P_{non-compliers}$.
2. Type 1 extensive margin compliers ($D_i(1) = 1 > D_i(0) = 0$)), with sample share Δ_{ext1}.
3. Type 2 extensive margin compliers ($D_i(1) = 2 > D_i(0) = 0$)), with sample share Δ_{ext2}.
4. Intensive margin compliers ($D_i(1) = 2 > D_i(0) = 1$)), with sample share Δ_{int}.

In each simulation, we assign units to each type to match the sample shares.

The distribution of treatment under $Z_i = 1$ and $Z_i = 0$ for each complier type is given by the above. Non-compliers, whose treatment does not depend on Z_i, are assigned to each level of D_i with equal probability. The remainder of the DGP is:

- $Z_i \sim Bernulli(0.5)$
- $Y_i \sim Bernulli(0.3)$ for all individuals who are not intensive margin compliers, and for intensive margin compliers $Y_i \sim Bernulli(0.3 + \Delta_y)$

We used 1,000 simulations, each with 1,000 observations, for each combination of parameter values $(\Delta_{ext1}, \Delta_{ext2}, \Delta_{int}, \Delta_y)$, i.e., each point in Figure 1. For simplicity, we fix share of extensive margin compliers of each type to be the same ($\Delta_{ext1} = \Delta_{ext2}$). The other parameters are fixed across the different simulations. To conduct inference we used 1,000 bootstrap draws in each simulations.