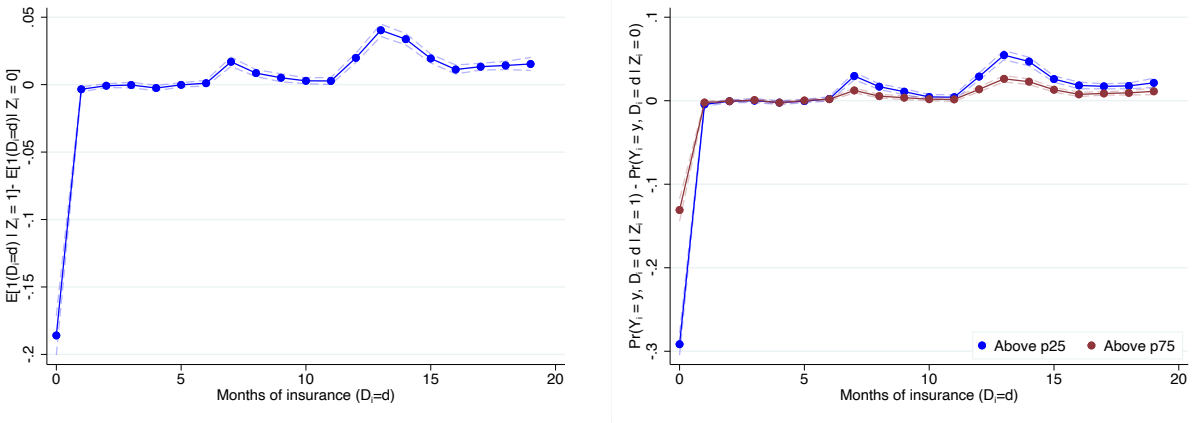
# Online Appendix

The online appendix includes the following sections:

- **Appendix A**. This section includes additional figures referenced in the main text.
- **Appendix B**. This section discusses tests of EMCO and presents Monte-Carlos simulations investigating their power.
- **Appendix C**. This section discusses how covariates can be incorporated when testing the moment inequalities described in Appendix B.
- **Appendix D**. This section presents proofs of propositions in this online appendix.

# A Additional figures

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




(a) Proposition 6 moment conditions

(b) Proposition 7 moment conditions

*Notes:* This figure shows visual tests of Propositions 6 and 7 for the OHIE replication data (Finkelstein et al., 2012). Panel (a) estimates the differences in the distribution 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 a binary indicator for whether a standardized measure of health care utilization falls above sample quantiles. Consistent with EMCO, there is a large decrease at $D_i = 0$ and positive increases elsewhere. Since administrative data on mortality and hospitalization are unavailable in the public replication 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 Testing 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 6.** *If Assumptions 1 and 2 hold, then:*
*(i)* $\Pr(D_i = 0|Z_i = 1) < \Pr(D_i = 0|Z_i = 0)$
*(ii)* $\Pr(D_i = d|Z_i = 1) \geq \Pr(D_i = d|Z_i = 0) \quad \forall d > 0$

Part (i) of Proposition 6 follows because Assumption 1 part (iii) (monotonicity) implies the instrument cannot increase treatment at its lowest level ($d = 0$). Assumption 1 part (i) (relevance) combined with Assumption 2 implies that the instruments must decrease treatment at $d = 0$, yielding the strict inequality in part (i) of Proposition 6. Part (ii) of Proposition 6 is implied by the addition of Assumption 2, but provides necessary and not sufficient conditions for it to hold.

To understand why, it is useful to consider a simple example. Table B.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 $\Delta^1_{ext}$), those shifted from $D_i(0) = 0$ to $D_i(1) = 2$ (share $\Delta^2_{ext}$), and those shifted from $D_i(0) = 1$ to $D_i(0) = 2$ (share $\Delta_{int}$). 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 $D_i$ conditional on $Z_i = 1$ stochastically dominates the distribution of $D_i$ conditional on $Z_i = 0$ 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.

Table B.1: Illustration of Proposition 6

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

Proposition 6 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.[13]

[13] Andresen and Huber (2021) consider a treatment defined by $1\{D_i \geq d^*\}$ and assume that the instrument

EMCO also places a 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 7.** *If Assumptions 1 and 2 hold, then for any measurable set $A$ in the support of $Y_i$:*
*(i)* $\Pr(Y_i \in A, D_i = 0 | Z_i = 0) \geq \Pr(Y_i \in A, D_i = 0 | Z_i = 1)$
*(ii)* $\Pr(Y_i \in A, D_i = d | Z_i = 1) \geq \Pr(Y_i \in A, D_i = d | Z_i = 0) \quad \forall d > 0$

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 with $D_i > 0$, and decreasing due to $Z_i$ for those with $D_i = 0$. Proposition 6 and 7 are closely related because Proposition 6 implies Proposition 7 if $A$ is set to $\mathcal{Y}$. Because the inequality in part (i) of Proposition 7 is weak however, it is implied by the exogeneity and monotonicity components of Assumption 1 alone. Part (ii) is implied by the addition of Assumption 2 to Assumption 1.

Proposition 7 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 6, however, Proposition 7 is necessary but not sufficient for EMCO. To see why, note that:

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

Under EMCO, the final two terms on the right-hand side disappear; the difference on the left-hand side must be non-negative, since the remaining term on the right-hand side is a density. When EMCO does not hold, however, this quantity can still be positive whenever $\Pr(Y_i(d) \in A, D_i(1) > d = D_i(0))$ is not too large. Hence there are some 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 $Y_i$.

One can test the restrictions in Propositions 6 and 7 by asking whether they appear consistent with the data. In particular, Proposition 7 implies that for any measurable $A \subseteq \mathcal{Y}$:

$$\mathbb{E}\left[1(D_i = 0) \cdot 1(Y_i \in A)|Z_i = 1\right] - \mathbb{E}\left[1(D_i = 0) \cdot 1(Y_i \in A)|Z_i = 0\right] \leq 0 \tag{B.1a}$$

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

For each $A$, Equations B.1a and B.1b imply $\bar{D} + 1$ total restrictions. These restrictions must hold for any $A$, including $A = \mathcal{Y}$. If $Y_i$ is discrete, they must hold at each point in support of $Y_i$,

only shifts individuals from below $d^*$ to above it (i.e., $D_i(0) < D_i(1) \Rightarrow D_i(0) < d^* \leq D_i(1)$. 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, for the case where $d^* = 0$. As we show in Appendix D.2, these conditions are equivalent to those in Proposition 6 part (ii). Andresen and Huber (2021) also state that these conditions should apply conditional on each value of the outcome.

yielding $(\bar{D}+1)\cdot|\mathcal{Y}|$ inequalities. Adding additional subsets of $\mathcal{Y}$ (including $A=\mathcal{Y}$) adds only redundant inequalities that can be obtained by summing B.1a and B.1b over points in the support of $Y_i$. The restrictions in Equations B.1a and B.1b must also hold for any outcome. It is also possible to incorporate covariates, as discussed below in Appendix C. Doing either can potentially yield more powerful tests.[14]

Testing the restrictions in Equations B.2a-B.2b implies testing the combined null hypothesis that *both* the LATE and EMCO assumptions are satisfied. The tests described below can thus be thought of as an omnibus test for the joint null hypothesis that Assumption 1 and Assumption 2 hold. Rejection can indicate violations of either assumption or both.

We consider two approaches to testing these restrictions. The first approach is based on the observation that for a set of choices of subsets $A$, the restrictions in Equations B.1a-B.1b define a set of $p$ differences in means comprised of the $\bar{D}+1$ differences for each $A$. We denote the sample analogs of these differences $\bar{\delta}=(\bar{\delta}_1,\bar{\delta}_2,\ldots,\bar{\delta}_p)$. When $p$ is fixed and the sample is sufficiently large, it is reasonable to invoke the asymptotic approximation that $\bar{\delta}\sim N(\delta,\Sigma)$. A simple but conservative test of the inequality restrictions in Equations B.1a-B.1b compares the maximum of the $p$ elements of $\bar{\delta}$ to the distribution of the largest element from a $N(0,\Sigma)$ random vector. We implement this approach in our simulations below, using the nonparametric bootstrap to estimate $\Sigma$.

The second approach leverages tools from the moment inequality literature. If there are many outcomes, the outcomes are continuous or take on many values, or if the researcher wishes to incorporate a large set of covariates, the resulting set of moments can be large and it may be more prudent to use tools that are better suited to testing many inequality restrictions simultaneously. To see how this can be done, define for any measurable $A\subseteq\mathcal{Y}$:

$$m_{i0}=\frac{1(D_i=0)\cdot 1(Y_i\in A)\cdot Z_i}{\Pr(Z_i=1)}-\frac{1(D_i=0)\cdot 1(Y_i\in A)\cdot(1-Z_i)}{1-\Pr(Z_i=1)} \tag{B.2a}$$

$$m_{id}=\frac{1(D_i=d)\cdot 1(Y_i\in A)\cdot(1-Z_i)}{1-\Pr(Z_i=1)}-\frac{1(D_i=d)\cdot 1(Y_i\in A)\cdot Z_i}{\Pr(Z_i=1)},\quad \text{for}\quad d>0 \tag{B.2b}$$

Proposition 7 requires that:

$$\mathbb{E}\left[m_{id}\right]\leq 0\quad\text{for}\quad d=0,\ldots,\bar{D}.$$

The tools we use to test this hypothesis require the sequence of random vectors $m_i=\{m_{i0},\ldots,m_{i\bar{D}}\}, i=1,\ldots,N$ to be independently distributed, which complicates replacing $\Pr(Z_i=1)$ with its sample analog. For this test, we therefore proceed treating $\Pr(Z_i=1)$ as known (as would be the case, for example, in an experiment where the researcher controls assignment of

[14]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: $\mathbb{E}\left[1(D_i\geq d)|Z_i=1\right]-\mathbb{E}\left[1(D_i\geq d)|Z_i=0\right]\geq 0$ for all $d$. This can be seen by taking $A=\mathcal{Y}$ and summing Equations B.2a and B.2b appropriately across $d$.

the instrument, or as may be a reasonable approximation when the sample is large). Because $E[m_i] \leq 0$ for any measurable $A \subseteq \mathcal{Y}$, one can concatenate these random vectors for multiple choices of $A$, yielding a total of $p$ restrictions.

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 given by:

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

where $j$ indexes the elements of $m_i$, $\bar{m}_j$ is the sample average of $m_{ij}$, and $S_j^2 = \frac{\sum_{i=1}^{N}(m_{ij}-\bar{m}_j)^2}{N-1}$ is an estimate of its variance. 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 are violated (Chernozhukov et al., 2018), as is the case in our setting.

To obtain critical values, we use two recently proposed procedures that are computationally tractable and work well in settings where the number of moments is potentially 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).[15] 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.[16]

## B.1 Simulation evidence

To explore the power of these tests, we construct a simple simulation based on the three treatment example in Table B.1. We assume the instrument $Z_i$ is binary and that there is a binary outcome $Y_i$.

Compliance types are as follows:

1. Non-compliers ($D_i(1) = D_i(0)$), with population share $P_{non-compliers}$.
2. Type 1 extensive margin compliers ($D_i(1) = 1 > D_i(0) = 0$)), with population share $\Delta_{ext1}$.
3. Type 2 extensive margin compliers ($D_i(1) = 2 > D_i(0) = 0$)), with population share $\Delta_{ext2}$.

[15] 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).

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

4. Intensive margin compliers ($D_i(1) = 2 > D_i(0) = 1$)), with population share $\Delta_{int}$.

The distribution of treatment under $Z_i = 1$ and $Z_i = 0$ is deterministic for each type of complier. 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:

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

We simulate from this DGP 1,000 times for several different values of $\Delta_{int}$ and $\Delta_y$ using 1,000 observations in each simulation. Each simulation holds $P_{non-compliers}$ fixed at 0.4. For simplicity, we also fix the probability of being assigned to each of the extensive margin complier types to be the same, so that $\Delta_{ext1} = \Delta_{ext2} = \Delta_{ext}$. Thus the likelihood of being an extensive margin complier of either type can be expressed as $\Delta_{ext} = (1 - \Delta_{int} - 0.4)/2$. In each iteration, we draw individual types from a multinomial distribution defined by their population shares.

To formally test EMCO's restrictions in each simulation, we use the asymptotic approximation to the distribution of the sample mean differences discussed above, and the moment inequality methods proposed by Chernozhukov et al. (2018) and Romano et al. (2014). Since $Y_i$ is discrete, we evaluate the restrictions for $A = \{0\}$ and $A = \{1\}$. Bootstraps use 1,000 iterations in each simulation. As discussed above, the moment inequality tests treat $\Pr(Z_i = 1)$ as known, although replacing it with its sample analog happens to change simulation results little in this particular case.

Figure B.1 reports the results. In Panel (a), the x-axis shows the true proportion of intensive margin compliers ($\Delta_{int}$). The y-axis shows the share of simulations in which the null hypothesis that $\Delta_{int} = 0$ is rejected. One series holds $\Delta_y$ fixed at zero, while the other considers $\Delta_y = 0.2$. Panel (b) holds fixed the proportion of intensive margin compliers at either 10 or 20 percent and examines how power increases with differences in the distribution of outcomes, $\Delta_y$.

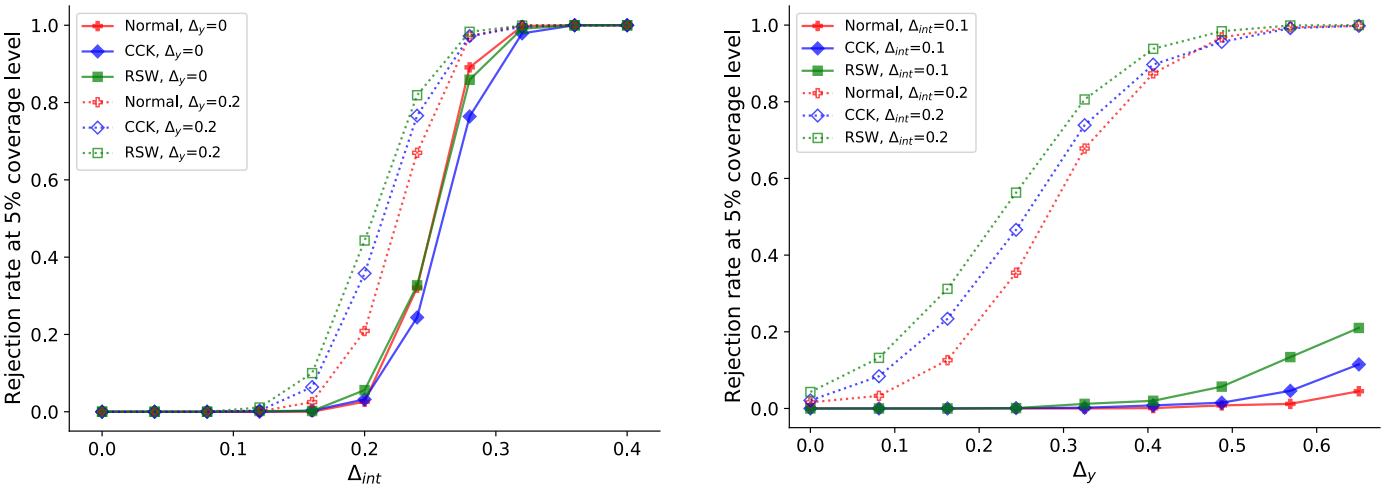
The figure shows that both tests have good finite-sample power to detect violations of EMCO large enough that the conditions in either Proposition 6 or 7 are violated. In these simulations, part (i) of Proposition 6 and part (ii) for $d = 2$ always hold in the population. Varying $\Delta_{int}$ only affects part (ii) for $d = 1$, which will be violated in this case if $\Delta_{int} > 0.2$. As result, power is effectively zero when $\Delta_{int}$ is below 0.2 and $\Delta_y = 0$, as shown in Panel (a). Rejection rates increase sharply as $\Delta_{int}$ increases beyond 0.2, however, and EMCO is almost always rejected when at least 30% of individuals are intensive margin compliers. Non-zero $\Delta_y$ yields more power because it can generate violations of part (ii) of Proposition 7 for $d = 1$ even if the conditions in Proposition 6 are satisfied.[17] When $\Delta_y = .2$, for example, the condition is violated for $A = \{1\}$ and $d = 1$

[17]Part (i) of Proposition 7 always holds in this DPG. So does part (ii) for $d = 2$.

if $\Delta_{int} > .14$. All three procedures preform similarly, although as expected the test based on asymptotic normality tends to be slightly more conservative.

Panel (b) further demonstrates how holding fixed $\Delta_{int}$, increases in $\Delta_y$ can increase rejection rates. When $\Delta_{int} = .1$, for example, the conditions in Proposition 6 are not violated, as discussed above. But if $\Delta_y$ is sufficiently large (in this case greater than about .45), the conditions in Proposition 7 kick in and EMCO can be rejected. Thus, both differences in the shares of intensive and extensive margin compliers and differences in their distribution of outcomes can be used to detect violations of EMCO. EMCO is most likely to be rejected when the share of intensive-margin compliers is large and their outcome distribution differs strongly from that of extensive-margin compliers.

Figure B.1: Simulations of tests of Assumptions 1 and 2




(a) Varying the share of intensive margin compliers ($\Delta_{int}$)

(b) Varying the difference in outcome distributions ($\Delta_y$)

*Notes:* This figure reports results from Monte-Carlo simulations based on the data generating process (DGP) described in Section B.1. The results average across 1,000 simulations for each combination of parameter values, with 1,000 observations used in each simulation. Each simulation holds the population share of compliers of all types fixed at 0.6. Panel (a) varies the proportion of intensive margin compliers ($\Delta_{int}$), with extensive margin compliers comprising the remainder of all compliers. The y-axis shows the share of simulations in which the null hypothesis that $\Delta_{int} = 0$ is rejected. The solid series plots results from simulations where all compliers have the same outcome distribution, while the dotted lines plot cases where $\Pr(Y_i = 1)$ for intensive margin compliers differs by $\Delta_y$. Panel (b) plots results from the analogous exercise varying $\Delta_y$ while holding $\Delta_{int}$ fixed at either 10 or 20 percent. "Normal" indicates the testing procedure described in Section B that compares the maximum sample moment to its estimated distribution assuming all moment inequalities bind with equality. RSW indicates results using Romano et al. (2014)'s moment inequality testing procedure. CCK indicates Chernozhukov et al. (2018)'s method. These latter two procedures treat $\Pr(Z_i = 1)$ as known.

# C Including covariates in tests of EMCO

Incorporating covariate adjustment into the moment inequality tests (and visualizations) is simple to implement. One can interact the random variables in Equations B.2a-B.2b 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 defined by the expectation of Equation B.2a must hold for every $X_i = x$:

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

Interacting the random variables in Equations B.2a-B.2b with different pre-treatment covariates will increase the number of moment inequalities. However, this is not necessarily a problem, since one of the advantages of the testing procedures proposed by Romano et al. (2014) and Chernozhukov et al. (2018) is that they can be used also in high-dimensional settings when the number of moment inequalities is potentially larger than the number of observations (Bai et al., 2019).

# D Proofs of propositions and claims in the online appendix

## D.1 Proofs of Propositions 6 and 7

To prove part (i) of Proposition 7, note that under Assumptions 1 (LATE) and 2 (EMCO):

$$\begin{aligned}
Pr(Y_i \in A, D_i = 0 | Z_i = 0) &= \Pr(D_i(0) = 0, Y_i(0) \in A | Z_i = 0) \\
&= \Pr(D_i(1) > 0, D_i(0) = 0, Y_i(0) \in A) \\
&\quad + \Pr(D_i(1) = 0, D_i(0) = 0, Y_i(0) \in A) \\
Pr(Y_i \in A, D_i = 0 | Z_i = 1) &= \Pr(D_i(1) = 0, Y_i(0) \in A | Z_i = 1) \\
&= \Pr(D_i(1) = 0, D_i(0) = 0, Y_i(0) \in A)
\end{aligned}$$

Thus $Pr(Y_i \in A, D_i = 0 | Z_i = 0) \geq Pr(Y_i \in A, D_i = 0 | Z_i = 1)$. Setting $A = \mathcal{Y}$ yields part (i) of Proposition 6 with a weak instead of strict inequality. Relevancy (Assumption 1, part (i)) requires that $\Pr(D_i(1) > 0, D_i(0) = 0, Y_i(0) \in A) > 0$ for $A = \mathcal{Y}$, which shows why the inequality in part (i) of Proposition 6 must be strict.

The proof of part (ii) is analogous, since under Assumptions 1 (LATE) and 2 (EMCO) for $d > 0$:

$$\begin{aligned}
Pr(Y_i \in A, D_i = d | Z_i = 0) &= \Pr(D_i(0) = d, Y_i(d) \in A | Z_i = 0) \\
&= \Pr(D_i(1) = d, D_i(0) = d, Y_i(d) \in A) \\
Pr(Y_i \in A, D_i = d | Z_i = 1) &= \Pr(D_i(1) = d, Y_i(d) \in A | Z_i = 1) \\
&= \Pr(D_i(1) = d, D_i(0) = d, Y_i(d) \in A) \\
&\quad + \Pr(D_i(1) = d, D_i(0) = 0, Y_i(d) \in A)
\end{aligned}$$

Thus $Pr(Y_i \in A, D_i = d | Z_i = 1) \geq Pr(Y_i \in A, D_i = d | Z_i = 0)$ for $d > 0$. Part (ii) of Proposition 6 follows when $A = \mathcal{Y}$.

## D.2 Equivalence between conditions in Proposition 6 part (ii) and Anderson and Huber (2021) conditions

Suppose that $\Pr(D_i = d | Z_i = 1) \geq \Pr(D_i = d | Z_i = 0)\ \forall d > 0$. LATE and ECMO implies that:

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

For any $d' > d > 0$:

$$\begin{aligned}
&\Pr(D_i(1) \geq d' > D_i(0)) - \Pr(D_i(1) \geq d > D_i(0)) \\
&= \sum_{k=d'}^{\bar{D}} \Pr(D_i(1) = k > D_i(0) = 0) - \sum_{l=d}^{\bar{D}} \Pr(D_i(1) = l > D_i(0) = 0) \\
&= \sum_{k=d}^{d'-1} \Pr(D_i(1) = k > D_i(0) = 0) \geq 0
\end{aligned}$$

Likewise, if $\Pr(D_i(1) \geq d' > D_i(0)) - \Pr(D_i(1) \geq d > D_i(0))$ for any $d' > d > 0$, then clearly:

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

for any $d > 0$. Thus the conditions in part (ii) of Proposition 6 are equivalent to the claim that $\Pr(D_i(1) \geq d' > D_i(0)) - \Pr(D_i(1) \geq d > D_i(0))$ for any $d' > d > 0$.