

# Econometrics Topics Course: Simulation

Julian Budde

August 27, 2023

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>The Identification Problem</b>	<b>1</b>
2.1	Model Setup . . . . .	1
2.2	Extrapolation . . . . .	2
2.3	Implementation . . . . .	4
<b>3</b>	<b>The Estimation Problem</b>	<b>5</b>
3.1	Implementation . . . . .	6
<b>4</b>	<b>Simulation Setting</b>	<b>7</b>
<b>5</b>	<b>Results</b>	<b>8</b>
<b>6</b>	<b>Potential Future Analysis</b>	<b>10</b>
<b>A</b>	<b>Choice of Tolerance</b>	<b>11</b>
<b>B</b>	<b>Results for Larger Samples</b>	<b>12</b>
<b>C</b>	<b>Unstable Estimator</b>	<b>12</b>

## 1 Introduction

## 2 The Identification Problem

### 2.1 Model Setup

Mogstad, Santos, and Torgovitsky (2018) use an IV model based on a selection equation, an approach summarized for example in Heckman and Vytlačil (2007a) and Heckman and Vytlačil (2007b). The key to the model is the selection equation which determines treatment status as a function of observed covariates (including the instrument) and unobserved heterogeneity in the likelihood to select into treatment. The key selection problem usually is that this unobserved heterogeneity is correlated with treatment effects or the level of potential outcomes. The instrumental variable solves this problem by shifting people into or out of treatment in a way uncorrelated to their

unobserved heterogeneity.

The key model is formulated as follows: We consider a model with binary treatment  $D \in \{0, 1\}$ . For convenience I drop all the subscripts throughout the paper. The **outcome equation** relates potential outcomes (the outcomes observed if individuals were exogenously assigned their value  $D = d$ ) and treatment to *observed* outcomes  $Y$  by

$$Y = Y_1 D + Y_0 (1 - D). \quad (1)$$

Treatment  $D$  itself is determined by the *choice equation*, which relates treatment status to observed covariates (in particular the instrument denoted by  $Z$ ) and unobserved heterogeneity  $U$ :

$$D = I\{p(Z) - U \geq 0\}. \quad (2)$$

$U$  is modeled to follow a standard Uniform distribution, although this is not a restriction because for any continuous  $U$  we can redefine the selection equation by applying the CDF  $F_U$  on both sides of the inequality. With this normalization,  $p(Z) = P(D = 1 | Z = 1)$  and thus is the *propensity score*, the probability to take up treatment conditional on observed covariates.  $U$  can be understood as the "resistance" to treatment: conditional on the propensity score (i.e. observables), individuals with a sufficiently high realization of  $U$  will never take up treatment.

While the model can be formulated to include both exogenous covariates  $X$  and "outside" instruments  $Z_0$  (so  $Z = (Z_0, X)$ ), in what follows I focus on the case without any covariates. Thus all the following statements will not include any conditioning on  $X$ .

**IV Model:** In addition to the outcome equation and choice equation, the IV model requires three further assumptions

- I.1  $U \perp Z_0$
- I.2  $E[Y_d | Z, U] = E[Y_d | U]$  and  $E[Y_d^2] < \infty$  for  $d \in \{0, 1\}$ .
- I.3  $U$  has a uniform distribution on  $[0, 1]$  conditional on  $Z$ .

The first two assumptions guarantee exogeneity of  $Z_0$  (exogenous shift in the choice probability and no direct effect on potential outcomes). The first assumption in combination with the additive separability of the choice equation, is equivalent to the monotonicity assumption in Angrist, Imbens, and Rubin (1996) that allows identification of the LATE among instrument-compliers, a result proven by Vytlacil (2002).

For example, a binary IV  $Z \in \{0, 1\}$  with propensity score  $p(0) = \underline{u} < p(1) = \bar{u}$  allows to identify  $LATE(\underline{u}, \bar{u})$ . Intuitively, individuals with realization of  $U$  in the interval  $[\underline{u}, \bar{u}]$  are those for which the instrument realization randomly shifts them between treatment states (the compliers). Those with realizations smaller than  $\underline{u}$  always take up treatment (the always-taker), while those with realizations larger than  $\bar{u}$  never take up treatment. The next section introduces the identification or extrapolation problem.

## 2.2 Extrapolation

While the Imbens and Angrist (1994) result shows that we can identify a LATE in this model (or multiple LATE if  $Z$  takes on several values), these might not necessarily be the parameters of interest. The key insight of Mogstad, Santos, and Torgovitsky (2018) is that many target parameters of interest *as well as* identified parameters like the LATE or IV

slope coefficients are functions of the same underlying **marginal treatment response** (MTR) functions. The MTR functions are denoted  $m_0, m_1$  and defined as

$$m_d(u) = E[Y_d|U = u] \quad (3)$$

For some target parameters, which will be denoted by  $\beta^*$ , writing them in terms of MTR functions is immediate. For example,  $LATE(a, b)$  averages the difference  $m_1(u) - m_0(u)$  over the range  $u \in [a, b]$ .

More generally, target parameters can be written in the form

$$\beta^* = E \left[ \int_0^1 m_0(u, X) \omega_0^*(u, Z) d\mu^*(u) \right] + E \left[ \int_0^1 m_1(u, X) \omega_1^*(u, Z) d\mu^*(u) \right] \quad (4)$$

where  $\omega_d^*$  are identified weighting functions depending on the target parameter (e.g. 1 and  $-1$  for the ATE).

A central result in the paper (Proposition 1) is that also all **IV-like estimands** of the form  $E[s(D, Z)Y]$  are weighted averages of MTR functions:

$$\beta_s = E \left[ \int_0^1 m_0(u, X) \omega_{0s}(u, Z) du \right] + E \left[ \int_0^1 m_1(u, X) \omega_{1s}(u, Z) du \right] \quad (5)$$

where the ewights are  $\omega_{0s} \equiv s(0, z)I[u > p(z)]$ ,  $\omega_{1s} \equiv s(1, z)I[u \leq p(z)]$ .

Introduce some further notation:

- $S$ : Set of IV-like specifications implying identified parameters  $\beta_s$ .
- $\mathcal{M}$ : Space of possible MTR functions, potentially including some a priori restrictions.
- $\mathcal{M}_S \subseteq \mathcal{M}$ : Sub-space of MTR functions *consistent* with identified estimands  $\beta_s$  for all  $s \in S$ .

Then the *identified set* for  $\beta^*$  denoted by  $\mathcal{B}_S^*$  is the set of  $b \in \mathbb{R}$  that is generated by some  $m \equiv (m_0, m_1) \in \mathcal{M}_S$ .

Proposition 2 in the paper establishes that for a convex  $\mathcal{M}$  the identified is of the form  $\mathcal{B}_S^* = [\underline{\beta}^*, \overline{\beta}^*] \subseteq \mathbb{R}$ . Further, these bounds are the solution to an optimiuation problem over  $m \in \mathcal{M}_S$  that can be recast as a linear program. In this program, the objective is to make the target parameter as small (or large) as possible while satisfying the constraint that at the optimal solution the chosen MTR functions imply the identified estimands (implicit in  $m \in \mathcal{M}_S$ ).

**Sharp identified set:** Proposition 3 in the paper establihes that if we use "enough" IV-like specifications to identify  $\mathcal{B}_S^*$ , then this is the smallest set consistent with conditional means  $E[Y|Z = z, D = d]$  and the model assumptions. For example, for a binary instrument we need to use all cross moments of the form  $E[I\{Z = z\}I\{D = d\}Y]$ . Intu- itively, if we think about the numerator of the Wald estimand  $E[Y|Z = 1] - E[Y|Z = 0]$  this differences out  $E[Y_1]$  for the always-taker and  $E[Y_0]$  for the never-taker, which allows to identify the (scaled) average treatment effect for the complier subpopulation. However, these moments itself constraint the admissable MTR functions so for extrapolation we want to use estimands that contain this information.

## 2.3 Implementation

In practice we need to consider a finite-dimensional parameter space  $\mathcal{M}_{fd} \subseteq \mathcal{M}$ . For example we can model  $m_d(u, x)$  as a finite number of basis functions:

$$m_d(u) = \sum_{k=1}^{K_d} \theta_{dk} b_{dk}(u).$$

For the simulation exercise here, the setting is however a lot easier. Proposition 4 in the paper establishes that for a  $Z$  with discrete support and target weights on the MTR functions that are piecewise constant over  $u$ , a finite-dimensional space of MTR functions recovers the exact solution. In particular, we can use constant splines as the basis functions, defined over a partition of  $u$  where all relevant weights (target and identified parameters) are constant.

It is useful to define linear maps  $\Gamma$  and  $\Gamma^*$  that takes as argument some  $m \in \mathcal{M}$  and return a parameter  $\beta$ . In particular, define for identified estimands  $\beta_s$

$$\Gamma_s(m) = E \left[ \int_0^1 m_0(u, X) \omega_{0s}(u, Z) du \right] + E \left[ \int_0^1 m_1(u, X) \omega_{1s}(u, Z) du \right] \quad (6)$$

and for the target parameter  $\beta^*$

$$\Gamma^*(m) = E \left[ \int_0^1 m_0(u, X) \omega_0^*(u, Z) du \right] + E \left[ \int_0^1 m_1(u, X) \omega_1^*(u, Z) du \right]. \quad (7)$$

In both cases,  $\omega_{ds}$  are the relevant weights implied by the IV-like specification.

Then our MTR space constrained by the identified estimands becomes

$$\mathcal{M}_S \equiv \{m \in \mathcal{M} : \Gamma_s(m) = \beta_s \text{ for all } s \in S\}. \quad (8)$$

and the identified set is given by

$$\mathcal{B}_S^* \equiv \{b \in \mathcal{R} : b = \Gamma^*(m) \text{ for some } m \in \mathcal{M}_S\}. \quad (9)$$

Because these maps are linear and we use a linear combination of basis functions to approximate  $\mathcal{M}$  we can restate the problem as

$$\bar{\beta}_{fd}^* \equiv \sup_{(\theta_0, \theta_1) \in \Theta} \sum_{k=1}^{K_0} \theta_{0k} \Gamma_0^*(b_{0k}) + \sum_{k=1}^{K_1} \theta_{1k} \Gamma_1^*(b_{1k})$$

$$\text{s.t.} \quad \sum_{k=1}^{K_0} \theta_{0k} \Gamma_{0s}(b_{0k}) + \sum_{k=1}^{K_1} \theta_{1k} \Gamma_{1s}(b_{1k}) = \beta_s \quad \text{for all } s \in \mathcal{S}.$$

In the case I consider in this study with discrete  $Z$  and a target parameter (LATE) that has constant weights, Proposition 4 applies and we can get exact results using constant splines as basis functions. In particular, these have to use knots corresponding to the points at which either the target parameter or the propensity score changes. For example, with a propensity score of  $p = (0.35, 0.6, 0.7)$  and a target parameter  $LATE(0.35, 0.9)$  we use a partition of  $u$  with  $u_{part} = [0, 0.35, 0.6, 0.7, 0.9, 1]$ , implying five basis functions for each of the intervals. As argued in the paper, for

the case of constant splines (as well as Bernstein polynomials) the integrals in the linear maps  $\Gamma_d^*(b_{dj})$  and  $\Gamma_{ds}(b_{dj})$  can be solved analytically. For example, considering  $d = 1$  and some IV-like specification  $s(0, z)$ , for  $\Gamma_{ds}$  (focusing on a single constant spline) we get:

$$\begin{aligned}
\Gamma_{0s}(b_{0j}) &= E_Z \left[ \int_0^1 m_0(u) w_{0s}(u, Z) du \right] \\
&= E_Z \left[ \int_0^1 I\{u \in [\underline{u}_j, \bar{u}_j]\} \theta_{0j} s(0, Z) I\{p(Z) < u\} du \right] \\
&= \theta_{0j} E_Z \left[ s(0, Z) \int_0^1 I\{u \in [\underline{u}_j, \bar{u}_j]\} I\{p(Z) < \underline{u}_j\} du \right] \\
&= \theta_{0j} E_Z \left[ s(0, Z) I\{p(Z) < \underline{u}_j\} (\bar{u}_j - \underline{u}_j) \right] \\
&= \theta_{0j} (\bar{u}_j - \underline{u}_j) \sum_{z \in \mathcal{Z}} f_Z(z) s(0, z) I\{p(z) < \underline{u}_j\}.
\end{aligned} \tag{10}$$

The second line follows from the definition of the weights and taking a constant spline as the basis functions, where  $\theta_{0j}$  is the coefficient on the basis function  $b_{dj}$  corresponding to some element of the  $u$ -partition ranging from  $[\underline{u}_j, \bar{u}_j]$ . This  $\theta_{0j}$ , as written earlier, will be one of the choice variables in the linear program. The third line uses the fact that we use a partition such that weights (and thus also the propensity scores) are constant over a given element of the partition. Therefore,  $p(Z) < u \iff p(Z) < \underline{u}_j^1$ . The last lines then pull out constants and write out the expectation using that  $Z$  has discrete support  $\mathcal{Z}$ . An equivalent result holds for  $d = 1$ :

$$\Gamma_{1s}(b_{1j}) = \theta_{1j} (\bar{u}_j - \underline{u}_j) \sum_{z \in \mathcal{Z}} f_Z(z) s(1, z) I\{p(z) > \underline{u}_j\}. \tag{11}$$

### 3 The Estimation Problem

When observing only a random sample we cannot exactly satisfy the constraint that the optimizer *exactly* implies identified estimands  $\beta_s$ . Both the identified estimands and the weights on the constant splines will be estimated. Instead, Mogstad, Santos, and Torgovitsky (2018) propose to solve the following problem (stated for the upper bound):

$$\hat{\beta}^* = \sup_{m \in \mathcal{M}} \hat{\Gamma}^*(m) \text{ s.t. } \sum_{s \in S} |\hat{\Gamma}_s(m) - \hat{\beta}_s| \leq \inf_{m' \in \mathcal{M}} \sum_{s \in S} |\hat{\Gamma}_s(m') - \hat{\beta}_s| + \kappa_n.$$

A few things to note:

- The upper bound makes the target estimand as large as possible for some  $m \in \mathcal{M}$ , but note that the linear map  $\Gamma^*(m)$  needs to be estimated (the weights on the MTR functions are functions of the data as will be clear below).
- The constraint is reformulated:
  - All admissible  $m \in \mathcal{M}$  have to come as close to the estimated identified estimands  $\hat{\beta}_s$  as the MTR functions that are closest to satisfying it plus some tolerance  $\kappa_n$ .
  - The tolerance  $\kappa_n$  has to shrink with the sample size. If  $\kappa_n$  is too large the bounds will be too wide, while a very small  $\kappa_n$  will introduce a lot of noise (e.g. think  $\kappa_n = 0$  which leaves the minimizer on the RHS as the only solution).

---

<sup>1</sup>  $p(Z) = \underline{u}_j$  might hold with equality but then this only holds exactly for that point so the interval also evaluates to zero.

This implies that we now have to solve a first-step linear program that finds the minimizer to the problem on the RHS of the constraint. [Briefly explain how this is done using tricks for the absolute value.] I explore the choice of  $\kappa_n$  in the simulations below but generally find  $\frac{1}{N}$  or  $\frac{1}{N^2}$  to result in similar estimates with MSE considerably lower than  $\frac{1}{\sqrt{N}}$  or  $\frac{1}{N^{\frac{1}{4}}}$ .

Mogstad, Santos, and Torgovitsky (2018) propose the following plug-in estimators

$$\hat{\Gamma}_{ds}(b_{dk}) \equiv \frac{1}{n} \sum_{i=1}^n \int_0^1 b_{dk}(u, X_i) \hat{\omega}_{ds}(u, Z_i) d\mu^*(u),$$

where  $\hat{\omega}_{0s}(u, z) \equiv \hat{s}(0, z) \mathbb{I}[u > \hat{p}(z)]$   
and  $\hat{\omega}_{1s}(u, z) \equiv \hat{s}(1, z) \mathbb{I}[u \leq \hat{p}(z)]$ ,

where  $\hat{s}$  is an estimator of  $s$ , and  $\hat{p}$  is an estimator of the propensity score. An estimator of  $\hat{\Gamma}_d^*(b_{dk})$  can be constructed similarly as

$$\hat{\Gamma}^*(b_{dk}) \equiv \frac{1}{n} \sum_{i=1}^n \int_0^1 b_{dk}(u, X_i) \hat{\omega}_d^*(u, Z_i) d\mu^*(u),$$

where  $\hat{\omega}_d^*$  is an estimator of  $\omega_d^*$ , the form of which will depend on the form of the target parameter. As pointed out in the paper, these estimators simplify considerably with constant spline basis functions for some parameters because the integrals can actually be solved analytically [TODO do this in the simulation/code; add formulas for this; this could be potential bug].  $\beta_s$  can be estimated based on

$$\hat{\beta}_s \equiv \frac{1}{n} \sum_{i=1}^n \hat{s}(D_i, Z_i) Y_i.$$

Appendix Proposition S3 establishes the consistency of their procedure, in particular that  $\hat{\underline{\beta}}^* \rightarrow_p \underline{\beta}^*$  and  $\hat{\bar{\beta}}^* \rightarrow_p \bar{\beta}^*$ .

In our case, the estimators of the linear map simplify to

$$\hat{\Gamma}_{1s}(b_{1j}) = \theta_{1j}(\bar{u}_j - \underline{u}_j) \sum_{i=1}^n \hat{s}(1, Z_i) \mathbb{I}\{\hat{p}(Z_i) > \underline{u}_j\}. \quad (12)$$

and similarly for  $d = 0$ . Here, we replace the expectation by sample moments and  $s(1, z)$  and  $p(z)$  by their estimated counterparts.

### 3.1 Implementation

The estimation procedure I implement closely follows the suggestions in Mogstad, Santos, and Torgovitsky (2018). In particular I solve the estimation problem in three steps:

Before solving the two linear programs I first estimate all the required objects. These include

- The identified estimands:  $\beta_s = E[s(D, Z)Y]$  for  $s \in S$ .
- The "weights" on the  $\theta_{dj}$  choice variables implied by the target and identified estimands via the linear maps  $\Gamma^*$  and  $\Gamma_{ds}$  (see equation 12).
- The propensity score  $p(z)$ .

**Step 1 — LP for the Constraint:** The first task is to solve the minimization problem in the RHS of the constraint in equation 3. This solves for the  $\theta_{jd}$  that imply the minimal deviation from the estimates of the identified estimands  $\beta_s$  in terms of absolute loss. As suggested in the paper, to model absolute loss in the objective function it is possible to introduce dummy variables that have to satisfy two constraints corresponding to the negative and positive part of each  $|\hat{\Gamma}_{\hat{m}m_s}(m')\hat{\beta}_s|^2$ .

**Step 2 — LP for the Bounds:** With the solution from Step 1 at hand and chosen tolerance level  $\kappa_n$ , we can now solve the two linear programs corresponding to (3). These are standard linear programs with the objective of maximizing (minimizing) the implied target estimand. Now the constraint includes absolute values (the deviations from estimated target parameters in the LHS of the constraint in (3)), which we can again address by introducing dummy variables and additional constraints. The resulting optimal solutions constitute the upper and lower bounds.

I solve both linear programs using AMPL, specifically the AMPL Python API. AMPL licenses are free for researchers and graduate students. The solver I used was "HiGHS" or "gurobi". All estimation, data handling and analysis steps are performed in Python. I discuss some practical issues relating to the tolerance level and the u-partition after introducing the DGP and main simulation analysis below.

## 4 Simulation Setting

I use the main data generating process (DGP) used in the numerical example by Mogstad, Santos, and Torgovitsky, 2018. They have a discrete instrument with the following specifications:

- Support of  $Z$ :  $Z \in \{0, 1, 2\}$ ;
- Density of  $Z$ :  $f_Z(0) = 0.5, f_Z(1) = 0.4, f_Z(2) = 0.1$ ;
- Propensity score:  $P(d = 1|Z = 0) \equiv p(0) = 0.35$ ,

Note the setup has no covariates  $X$ . Following Imbens and Angrist (1994) three local average treatment effects (LATE) are point-identified: LATE(0.35, 0.6), LATE(0.6, 0.7), and LATE(0.35, 0.7). This will show up in the identification results below, which cover point-identification as a special case.

While the paper assumes a binary outcome I directly simulate (potential) outcomes using the underlying MTR functions, such that  $Y_d = m_d(u)$  which imply outcomes  $Y$ . These are of course no longer binary, but all essential moments remain the same so consistency should not be affected<sup>3</sup>.

I study a range of different targets of the form  $LATE(0.35, \bar{u})$ , where  $0.35 \leq \bar{u} \leq 1$ .

I focus on the sharp, non-parametric bounds depicted in Figure 5 of Mogstad, Santos, and Torgovitsky, 2018. These are constructed using all cross-moments of  $D, Z$  with the data  $Y$ , i.e. IV-like estimands of the form

$$\beta_s = E[I\{D = d, Z = z\}Y]$$

<sup>2</sup>For example, with a program of the form  $\min |X|$  introducing a dummy  $X'$ , and objective  $\min X'$  with two constraints  $X \leq X'$  and  $-X \leq X'$  mimicks the desired program. In our case the objective will generally be of the form  $\min \sum_{s \in S} |X_s|$  which can be modeled by introducing  $|S|$  variables and  $2|S|$  constraints. I mainly used the lpSolve manual 5.1 as a reference.

<sup>3</sup>If we want to use a binary  $Y$  consistent with the DGP one way would be to draw outcomes from  $\{0, 1\}$  with probability corresponding to  $m_{d_i}(u_i)$  for each  $i = 1, \dots, n$ . This might be preferred if we want to do simulations for an actual DGP corresponding to some observed data, because otherwise we understate the variance  $Y$ .

for  $d \in \{0, 1\}$  and  $z \in \{0, 1, 2\}$ . Proposition 3 in the paper establishes that this set of identified estimands delivers the sharpest bounds that are consistent with the conditional means of  $Y$  and the assumptions of the model.

Throughout the analysis I impose the restriction that the MTR functions  $m$  are bounded between 0, 1 (formulated as  $\theta_{jd} \in [0, 1]$ ), a restriction imposed in the paper and immediate for a binary outcome.

In the following section I report a Figure similar to Figure 8 in the paper, which reports bounds for  $LATE(0.35, \bar{u})$  for a range of values, although with a different parametric assumption on the MTRs. In addition to the identification result (which in the plot I call the "true bounds"), I also report estimation results. For a grid of  $\bar{u}$  values, I estimate the bounds and plot their means and distributions.

Throughout I use a sample size of  $N = 10000$ ,  $R = 1000$  replications and a tolerance for the identification constraint equivalent to  $\frac{1}{N}$ .

Before turning to the main results in the next section I want to discuss some practical issues I encountered during implementation. First, the choice of the **tolerance criterion**  $\kappa_n$  seems to be non-trivial. In Appendix Figure A.2 I report results for varying tolerances ( $\kappa_n \in \{\frac{1}{\sqrt{n}, \frac{3}{n}, \frac{1}{n}, \frac{1}{n^2}}\}$ ). A higher  $\kappa_n$  should result in (weakly) wider bounds because more MTR pairs satisfy the constraint. Choosing  $\kappa_n$  too lower might result in a very noisy estimator; in the extreme case  $\kappa_n = 0$  by construction we only have a single solution (assuming there is a unique minimizer which seems highly likely given the sampling). As expected I observe that bounds are significantly wider for  $\kappa_n = \frac{1}{\sqrt{n}}$  and narrower when choosing one of the smaller tolerances. Choosing a  $\kappa_n < \frac{1}{\sqrt{n}}$  also significantly reduced bias. Beyond a  $\kappa_n$  of the order  $\frac{1}{n}$  there does not seem to be much of a change, however.

Second, I found that estimation results are highly unstable when the partition of  $u$  includes values very close to each other. When I first ran the simulation below which uses a target  $LATE(0.35, \bar{u})$  I ran into this issue: Because  $p(0) = 0.35$  for a large enough  $N$ ,  $\hat{p}(1)$  will be very close to 0.35, thus resulting in a partition that includes 0.35 as well as a number very close to it. The result is a bi-modal distribution of the estimator for the lower bound, as shown in Appendix Figure C.4. Similar issues arise for the upper bound when  $\bar{u}$  comes close to  $p(1)$  or  $p(2)$ . The two modes for the lower bound exactly correspond to situations where we estimate  $\hat{p}(0) < 0.35$  (smaller mode) or  $\hat{p}(0) > 0.35$  (larger mode). A simple fix to this issue I rely on in the following results is to remove one of the close knots in the partition. This seems also the relevant usecase in practice: When a researcher estimates the (to them unknown) propensity score to be  $\hat{p}(0)$  close to 0.35 and wants to extrapolate,  $LATE(0.35, 0.9)$  is not a natural choice of the target parameter, but instead  $LATE(\hat{p}(0) \approx 0.35, 0.9)$ .

## 5 Results

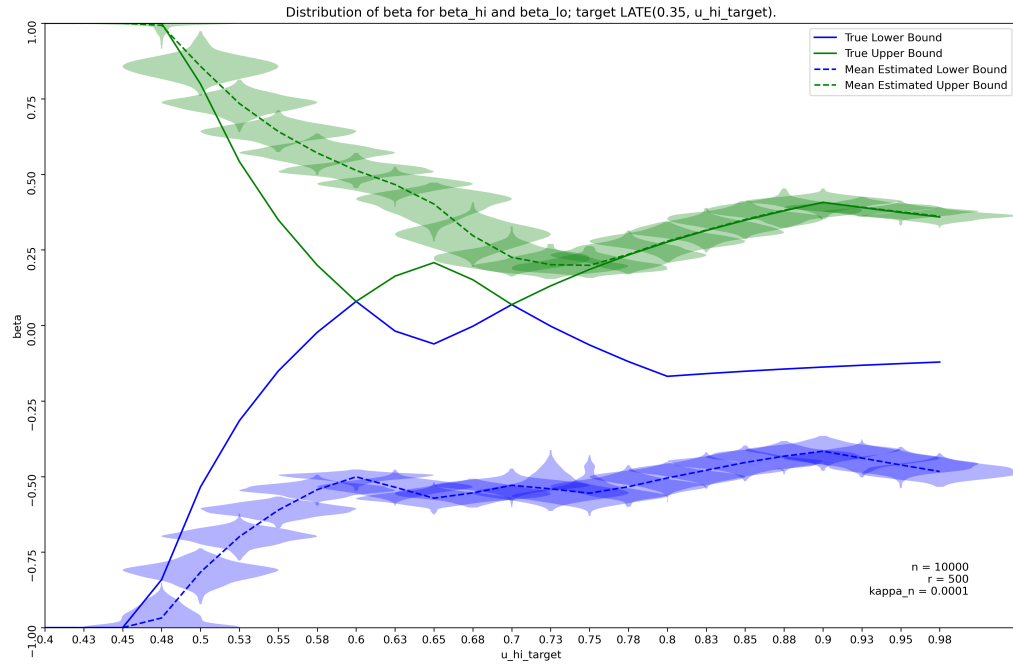
The figure below reports identification and estimation results for each  $LATE(0.35, \bar{u})$ .

### Identification:

- We achieve point identification for  $\bar{u} \in \{0.6, 0.7\}$ .
- Generally, as argued in the paper, extrapolation to parameters further away from what is point-identified results in wider bounds.
- This is particularly striking for smaller values of  $\bar{u}$ , where for  $\bar{u} < 0.45$  bounds become completely uninfor-



Figure 1: Simulation Results for Sharp Non-Parametric Bounds



mative ranging from  $-1$  to  $1$ .

- This is in stark contrast to the paper, and highlights the importance of shape restrictions (or in the case of the paper parametric assumptions) on the MTR functions.
- Bounds for larger values are a lot smaller, highlighting that in some sense the data is more informative about individuals with a high resistance to treatment, i.e. high  $u$  (Todo: Understand how so? "Informative"?).

#### Estimation:

- Estimated bounds are on average wider than the true bounds.
- Most notably, for many of the target parameters the distribution is **bimodal**.
- This is particularly true for the lower bounds which has a bimodal distribution for *all* values of  $\bar{u}$ .
- While for values up to 0.7 one of the modes seems to be close to the true bound, this is not the case for larger values of  $\bar{u}$ .
- For the upper bounds distributions are only bimodal in the range 0.6 to 0.7 (roughly or exactly?).
- For all targets  $> 0.7$  the upper bound estimator is tightly centered.

#### Questions

- Mean coverage might not be informative if for all estimators one bound covers while the other does not with roles reversing.

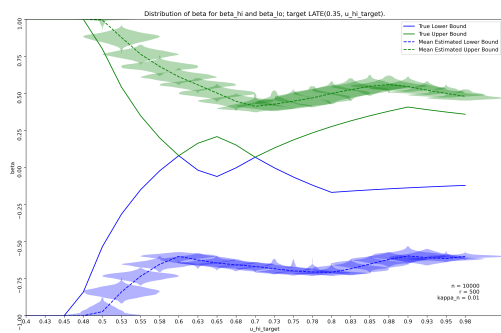
## 6 Potential Future Analysis

- Weak IV
  - Need to think about how this would be relevant; really weak IV is a failure of the regular asymptotic approximation to the finite sample distribution thereby making inference unreliable. This is not really an identification problem; i.e. for a discrete IV as long as not all propensity scores are the same we can always point identify some LATE. However, weak IV might show up in estimation of the bounds.
  - Case of weak IV/many weak IV
  - Weak IV for IV slope → Use reduced form/cross-moments instead of IV-slope coefficient; thereby avoiding weak IV problem. Is this a trade-off between estimation and inference? I.e. identification bounds become wider but inference becomes more reliable?

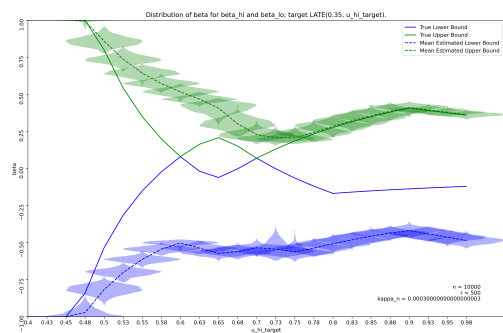
## References

- Angrist, Joshua D, Guido W Imbens, and Donald B Rubin (1996). "Identification of causal effects using instrumental variables". In: *Journal of the American statistical Association* 91.434, pp. 444–455.
- Heckman, James J and Edward Vytlacil (2007a). "Econometric evaluation of social programs, part I: Causal models, structural models and econometric policy evaluation". In: *Handbook of econometrics* 6, pp. 4779–4874.
- (2007b). "Econometric evaluation of social programs, part II: Using the marginal treatment effect to organize alternative econometric estimators to evaluate social programs, and to forecast their effects in new environments". In: *Handbook of econometrics* 6, pp. 4875–5143.
- Imbens, Guido W and Joshua D Angrist (1994). "Identification and estimation of local average treatment effects". In: *Econometrica: journal of the Econometric Society*, pp. 467–475.
- Mogstad, Magne, Andres Santos, and Alexander Torgovitsky (2018). "Using instrumental variables for inference about policy relevant treatment parameters". In: *Econometrica* 86.5, pp. 1589–1619.
- Vytlacil, Edward (2002). "Independence, monotonicity, and latent index models: An equivalence result". In: *Econometrica* 70.1, pp. 331–341.

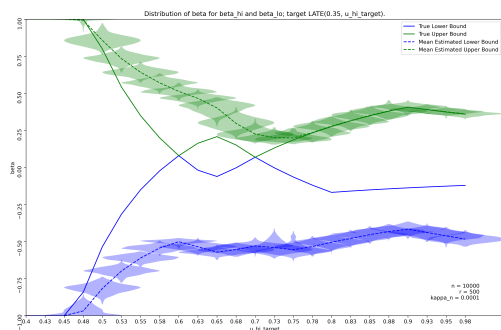
## A Choice of Tolerance

Figure A.2: Simulation Results by Choice of  $\kappa_n$ 

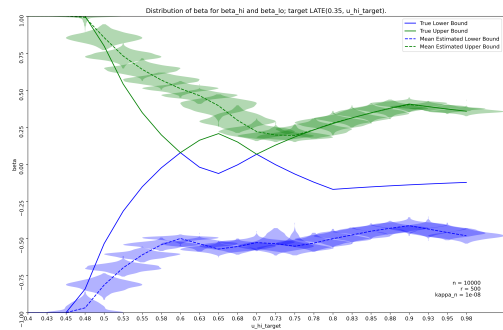
(a)  $\kappa_n = \frac{1}{\sqrt{n}}$



(b)  $\kappa_n = \frac{3}{n}$



(c)  $\kappa_n = \frac{1}{n}$

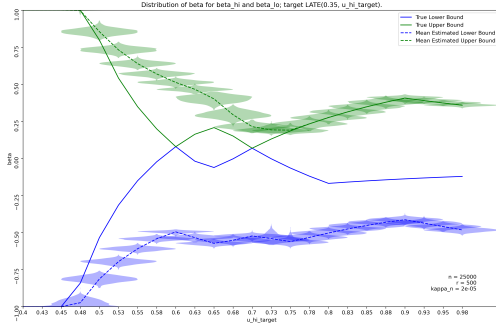


(d)  $\kappa_n = \frac{1}{n^2}$

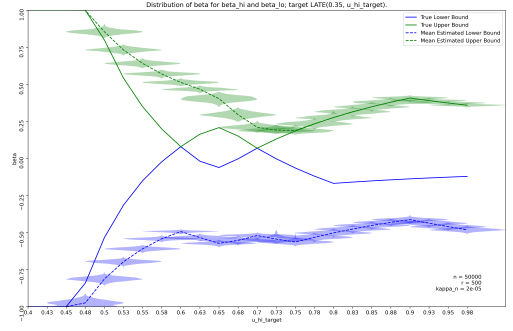
**Notes:** This figure compares simulation results for the main specification described in Section 4 for different tolerance levels  $\kappa_n$ .

## B Results for Larger Samples

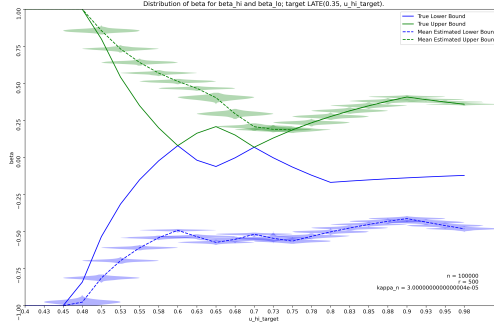
Figure B.3: Simulation Results for Larger Sample Sizes



(a)  $N = 25,000$



(b)  $N = 50,000$

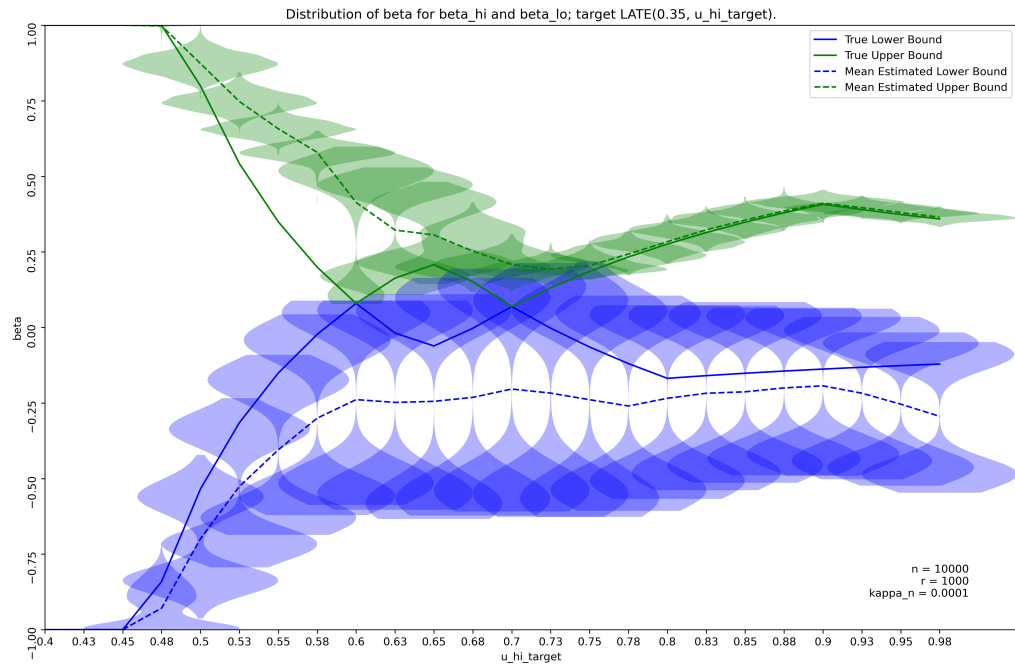


(c)  $N = 100,000$

**Notes:** This figure compares simulation results for the main specifcaiton described in Section 4 different samples sizes  $N$ .

## C Unstable Estimator

Figure C.4: Simulation Results: Bi-Modal Distribution



**Notes:** This figure shows results for the main specification when we use a partition including both estimated  $\hat{p}(0)$  close to 0.35, and 0.35 in the partition for  $u$ .