Online Appendix: An Economic Approach to Generalizing Findings from Regression-Discontinuity Designs

Nirav Mehta

July 11, 2018

Online Appendix 1 Allow the Probability of Enrollment to Vary by x

If the ρ_ω are no longer assumed to be uniform within intended treatment group, we adapt Manuscript Equation 10 to obtain the administrator's problem:

$$\max_{\tilde{\kappa}} \left(\int_0^{\tilde{\kappa}} \rho_0(x)(\Delta(x) - \chi)dx \right) + \left(\int_{\tilde{\kappa}}^1 \rho_1(x)(\Delta(x) - \chi)dx \right).$$

Note that a necessary condition for κ^* being optimal is $(\rho_1(\kappa^*) - \rho_0(\kappa^*))\Delta(\kappa^*) = (\rho_1(\kappa^*) - \rho_0(\kappa^*))\chi \Rightarrow \Delta(\kappa^*) = \chi$, which is identical to Condition 1(i) for the sharp design. Optimality of κ^* further implies:

$$\text{(A.1)} \quad \int_{\kappa^*}^1 \rho_1(x)\Delta(x)dx \geq \int_{\kappa^*}^1 \rho_1(x)\chi dx = \overline{\rho_1}\chi(1-\kappa^*) \Rightarrow \underbrace{\frac{\int_{\kappa^*}^1 \rho_1(x)\Delta(x)dx}{1-\kappa^*}}_{\text{ITT}} \geq \overline{\rho}_1\chi = \overline{\rho}_1\Delta(\kappa^*),$$

where $\overline{\rho}_1 \equiv \frac{\int_{\kappa^*}^1 \rho_1(x)dx}{1-\kappa^*}$ is the average attendance probability among the intended treated and the second equality follows because $\chi = \Delta(\kappa^*)$. Equation A.1 shows that if the average attendance probability among the intended treated ($\overline{\rho}_1$) were known then the RD estimate of the treatment effect could be used to provide a lower bound for the mean effect of intending to treat among the (intended) treated (ITT).

We can also bound the mean effect of intending to not treat the untreated (ITUT) by considering the second part of the sufficient conditions for this case, that is, adapting Manuscript Equation 4:

$$\int_0^{\kappa^*} \rho_0(x)\Delta(x)dx < \int_0^{\kappa^*} \rho_0(x)\chi dx = \overline{\rho}_0\chi\kappa^* \Rightarrow \underbrace{\frac{\int_0^{\kappa^*} \rho_0(x)\Delta(x)dx}{\kappa^*}}_{\text{ITUT}} < \overline{\rho}_0\chi = \overline{\rho}_0\Delta(\kappa^*), \tag{A.2}$$

where $\overline{\rho}_0 \equiv \frac{\int_0^{\kappa^*} \rho_0(x)dx}{\kappa^*}$ is the average attendance probability among the intended untreated and, as with the ITT above, the second equality follows because $\chi = \Delta(\kappa^*)$. Analogous to the case for the ITT, Equation A.2 shows that if the average attendance probability among the intended untreated is known then the RD estimate of the treatment effect could be used to provide an upper bound for the mean effect of intending to not treat among the (intended) untreated (ITUT).

Online Appendix 2 Treatment Effect Uncertainty

Suppose the administrator is uncertain about the average treatment effect $\Delta(x)$ but has observed $\check{\Delta}(x)$, an unbiased signal of $\Delta(x)$. Let $\check{\Delta}(x) = \Delta(x) + \epsilon_i$, where ϵ is distributed independently of x, denote the administrator's noisy signal of the average treatment effect for students with index x. Because the administrator has unbiased beliefs about $\Delta(x)$ at every point x, it must be the case that $\mathrm{E}[\epsilon_i] = 0$. The administrator chooses a cutoff to maximize her expected objective:

$$\begin{aligned}
\max_{\widetilde{\kappa}} \mathrm{E}\left[\beta\left(\int_{\widetilde{\kappa}}^{1} \check{\Delta}(x)dx\right) - c\,(1-\widetilde{\kappa})\right] &\Leftrightarrow \max_{\widetilde{\kappa}} \beta\, \mathrm{E}\left[\left(\int_{\widetilde{\kappa}}^{1} (\Delta(x)+\epsilon)dx\right)\right] - c\,(1-\widetilde{\kappa}) \\
&\Leftrightarrow \max_{\widetilde{\kappa}} \beta\left(\int_{\widetilde{\kappa}}^{1} \Delta(x)dx\right) + \beta\,\mathrm{E}[\epsilon] - c\,(1-\widetilde{\kappa}) \\
&\Leftrightarrow \max_{\widetilde{\kappa}} \beta\left(\int_{\widetilde{\kappa}}^{1} \Delta(x)dx\right) - c\,(1-\widetilde{\kappa})\,. \qquad \text{(A.3)}
\end{aligned}$$

The first equivalence follows from the fact that the measure of students treated $(1-\widetilde{\kappa})$ is known because it is chosen by the administrator. The second follows from the independence assumption and the third from unbiasedness. The last expression is the administrator's original problem, Manuscript Equation 2. Therefore, the analysis of this case is identical.

Intuitively, uncertainty does not affect the administrator's problem because it is linear in the amount gained.

We can also use this setup to examine what would happen if the administrator instead only had access to a biased measure of $\Delta(\cdot)$. Define $\delta(x) \equiv \mathrm{E}[\epsilon_i|x]$, that is, the conditional expectation of ϵ given x. In the case of an unbiased $\check{\Delta}(x)$ we have $\delta(x) = 0$ for all x. I consider two types of biased beliefs.

Constant Bias First suppose $\delta(x) = \delta \neq 0$, that is, ϵ is biased, but mean independent of x. In this case, Condition 1 would not be affected, as the optimal cutoff κ^* would not change from the unbiased case. Intuitively, if $\delta(\cdot)$ does not depend on x the bias does not affect the administrator's objective at the intensive margin.

Condition 2, however would be affected. Consider first the augmented participation condition:

$$\int_{\kappa^*}^{1} \Delta(x)dx \geq (\chi - \delta)(1 - \kappa^*),$$

which implies the ATT lower bound would be shifted downwards by the constant amount δ. Similarly, the non-extension condition would become

$$\int_{\hat{\kappa}}^{\kappa^*} \Delta(x)dx < \chi(\kappa^* - \hat{\kappa}) = (\chi - \delta)(1 - \kappa^*),$$

that is, the upper bound on the ATUT would also be shifted down by the constant δ. These changes to Condition 2 would propagate to the other bounds results.

Differential Bias in x Now let $\delta(x)$ be variable in x. The augmented participation condition becomes

$$\int_{\kappa^*}^{1} \Delta(x)dx \geq \chi(1 - \kappa^*) - \int_{\kappa^*}^{1} \delta(x)dx$$

and the augmented non-extension condition becomes

$$\int_{\hat{\kappa}}^{\kappa^*} \Delta(x)dx < \chi(\kappa^* - \hat{\kappa}) - \int_{\hat{\kappa}}^{\kappa^*} \delta(x)dx.$$

Consider the following two cases: (i) $\mathrm{E}[\delta(x)|x \in [\hat{\kappa}, \kappa^*)] < 0 < \mathrm{E}[\delta(x)|x \geq \kappa^*]$, $\forall \hat{\kappa} \in [0, \kappa^*)$ and (ii) $\mathrm{E}[\delta(x)|x \geq \kappa^*] < 0 < \mathrm{E}[\delta(x)|x \in [\hat{\kappa}, \kappa^*)]$, $\forall \hat{\kappa} \in [0, \kappa^*)$. In case (i) the augmented participation and non-extension conditions would reduce the lower bound on the

ATT and increase the upper bound on the ATUT. That is, all bounds would be looser. In case (ii) the opposite would happen, that is, bounds would tighten.

Online Appendix 3 Omitted Proofs from Section III.B.

Corollary $\overline{3a}$. *Suppose $\overline{\kappa}^* = 1 - \overline{\mu}$ solves Equation $\bar{2}$. The following ATUT bounds can be obtained:*

(i) There is an informative upper bound, of $ATUT^{UB} = [\Delta(\overline{\kappa}^*)\kappa^*_{\overline{a}} + \overline{\Delta}(\overline{\kappa}^* - \kappa^*_{\overline{a}})]/\overline{\kappa}^*$, if and only if $\kappa^*_{\overline{a}}$ is known.

(ii) There is an informative lower bound, of $ATUT^{LB} = \kappa^*_{\overline{a}}\underline{\Delta}/\overline{\kappa}^*$, if and only if $\kappa^*_{\overline{a}}$ is known and $\underline{\Delta} < 0$.

Proof. (i) Upper bound: By Proposition 2 the upper bound for the total treatment effect over $[0, \kappa^*_{\overline{a}}]$ is $\chi\kappa^*_{\overline{a}}$; if χ is unknown then apply Condition $\overline{1a}$ to form the upper bound $\Delta(\overline{\kappa}^*)\kappa^*_{\overline{a}}$. The administrator would have chosen to treat students $x \in [\kappa^*_{\overline{a}}, \overline{\kappa}^*]$, meaning there would be no informative upper bound, resulting in an upper bound of $\overline{\Delta}(\overline{\kappa}^* - \kappa^*_{\overline{a}})$ for the total treatment effect. Sum these and divide by the measure of untreated students to obtain $\text{ATUT}^{UB} = [\Delta(\overline{\kappa}^*)\kappa^*_{\overline{a}} + \overline{\Delta}(\overline{\kappa}^* - \kappa^*_{\overline{a}})]/\overline{\kappa}^*$.

(ii) Lower bound: As in the unconstrained problem, the interval below $\kappa^*_{\overline{a}}$ has an uninformative lower bound. Adapting the first part of Condition 2 for the interval $[\kappa^*_{\overline{a}}, \overline{\kappa}^*]$, we obtain $\chi(\overline{\kappa}^* - \kappa^*_{\overline{a}}) < \int_{\kappa^*_{\overline{a}}}^{\overline{\kappa}^*} \Delta(x)dx$; if χ is unknown then by using the reasoning in Corollary $\overline{2a}$, $0 < \int_{\kappa^*_{\overline{a}}}^{\overline{\kappa}^*} \Delta(x)dx$. Sum over these intervals and divide by the measure of untreated students to obtain $\text{ATUT}^{LB} = \kappa^*_{\overline{a}}\underline{\Delta}/\overline{\kappa}^*$.

For either upper or lower bound, if $\kappa^*_{\overline{a}}$ was unknown then one would have to adopt the worst-case scenario corresponding to the loosest bounds, setting $\kappa^*_{\overline{a}} = 0$ for the upper bound and $\kappa^*_{\overline{a}} = \overline{\kappa}^*$ for the lower bound, resulting in uninformative bounds. □

Corollary $\overline{3b}$. *Suppose $\overline{\kappa}^* = \kappa^*_{\overline{b}}$ solves Equation $\bar{2}$. The following ATUT bounds can be obtained:*

(i) There is an informative upper bound, which is tighter if $\kappa^*_{\overline{a}}$ is known. If $\kappa^*_{\overline{a}}$ is unknown the upper bound is $ATUT^{UB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1-\overline{\mu})) + \overline{\Delta}(1-\overline{\mu})]/\overline{\kappa}^*$; if $\kappa^*_{\overline{a}}$ is known the upper bound is $ATUT^{UB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - ((1-\overline{\mu}) - \kappa^*_{\overline{a}})) + \overline{\Delta}((1-\overline{\mu}) - \kappa^*_{\overline{a}})]/\overline{\kappa}^*$.

(ii) There is an informative lower bound, which is tighter if $\kappa^*_{\overline{a}}$ is known. If $\kappa^*_{\overline{a}}$ is unknown the lower bound is $ATUT^{LB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1-\overline{\mu})) + \underline{\Delta}(1-\overline{\mu})]/\overline{\kappa}^*$; if $\kappa^*_{\overline{a}}$ is known the lower bound is $ATUT^{LB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - \kappa^*_{\overline{a}}) + \underline{\Delta}\kappa^*_{\overline{a}}]/\overline{\kappa}^*$.

Proof. (i) Upper bound: Note that Equation $\overline{4b}$ implies that $\int_{1-\overline{\mu}}^{\overline{\kappa}^*} \Delta(x)dx < \Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1 - \overline{\mu}))$. By Assumption 1(iii), the average treatment effect for any treated unit cannot exceed $\overline{\Delta}$, which if $\kappa^*_{\overline{a}}$ is not known is the upper bound for the average treatment effect for $x \in [0, 1-\overline{\mu}]$. This results in an upper bound of $\text{ATUT}^{UB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1-\overline{\mu})) + \overline{\Delta}(1-\overline{\mu})]/\overline{\kappa}^*$. If $\kappa^*_{\overline{a}}$ is known, then $\int_0^{\kappa^*_{\overline{a}}} \Delta(x)dx < \Delta(\overline{\kappa}^*)\kappa^*_{\overline{a}}$ by Manuscript Equation 4, which results in a tighter upper bound of $\text{ATUT}^{UB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - ((1-\overline{\mu}) - \kappa^*_{\overline{a}})) + \overline{\Delta}((1-\overline{\mu}) - \kappa^*_{\overline{a}})]/\overline{\kappa}^*$.

(ii) Lower bound: Suppose that $\kappa^*_{\overline{a}}$ is known. Use the same argument as for the lower bound in Corollary $\overline{3a}$, but also include the interval $[1-\overline{\mu}, \overline{\kappa}^*]$, which has a total lower bound of $\underline{\Delta}(\overline{\kappa}^* - (1-\overline{\mu}))$, and note that in Case $\overline{b}$ we have $\Delta(\overline{\kappa}^*) = \chi$. This would result in a ATUT lower bound $[\Delta(\overline{\kappa}^*)((1-\overline{\mu}) - \kappa^*_{\overline{a}}) + \underline{\Delta}(\kappa^*_{\overline{a}} + (\overline{\kappa}^* - (1-\overline{\mu})))]/\overline{\kappa}^*$. However, this expression does not exploit the fact that we are in a Case-$\overline{b}$-binding constrained scenario. By adapting Condition 2 to take as the upper limit $\overline{\kappa}^*$ (instead of 1), this latter piece of information implies that $\int_{\kappa^*_{\overline{a}}}^{\overline{\kappa}^*} \Delta(x)dx \geq \chi(\overline{\kappa}^* - \kappa^*_{\overline{a}}) = \Delta(\overline{\kappa}^*)(\overline{\kappa}^* - \kappa^*_{\overline{a}})$; intuitively, the unconstrained administrator would have treated $x \in [\kappa^*_{\overline{a}}, \overline{\kappa}^*]$. Using (only) this information, the ATUT lower bound would be $[\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - \kappa^*_{\overline{a}}) + \underline{\Delta}\kappa^*_{\overline{a}}]/\overline{\kappa}^*$. Taking the maximum (that is, tightest lower bound) results in $\text{ATUT}^{LB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - \kappa^*_{\overline{a}}) + \underline{\Delta}\kappa^*_{\overline{a}}]/\overline{\kappa}^*$. If $\kappa^*_{\overline{a}}$ is not known, then we must choose $\kappa^*_{\overline{a}}$ to minimize the lower bound, that is, $\kappa^*_{\overline{a}} = (1-\overline{\mu})$, which would result in a looser bound of $\text{ATUT}^{LB} = [\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1-\overline{\mu})) + \underline{\Delta}(1-\overline{\mu})]/\overline{\kappa}^*$. □

Corollary $\overline{4a}$. *Suppose $\overline{\kappa}^* = 1 - \overline{\mu}$ solves Equation $\bar{2}$. The following ATE bounds can be obtained.*

(i) If $\underline{\Delta} < 0$ and $\kappa^*_{\overline{a}}$ is unknown then there is an informative lower bound, of $ATE^{LB} = \underline{\Delta}\overline{\kappa}^*$, which is tightened to $ATE^{LB} = \underline{\Delta}\kappa^*_{\overline{a}}$ if $\kappa^*_{\overline{a}}$ is known.

(ii) There is an informative upper bound, of $ATE^{UB} = \Delta(\overline{\kappa}^*)\kappa^*_{\overline{a}} + \overline{\Delta}(1 - \kappa^*_{\overline{a}})$, if and only if $\kappa^*_{\overline{a}}$ is known.

Proof. (i) Lower bound: If $\kappa^*_{\overline{a}}$ is unknown, by Corollary $\overline{2a}$, the lower bound on the ATE for units with $x \geq \overline{\kappa}^*$ is 0, tightening the ATE lower bound from $\underline{\Delta}$ to $\underline{\Delta}\overline{\kappa}^*$. If $\kappa^*_{\overline{a}}$ is known, then by Corollary $\overline{3a}$(ii) the lower bound on for units $x \in [\kappa^*_{\overline{a}}, \overline{\kappa}^*]$ is 0, further tightening the ATE lower bound to $\underline{\Delta}\kappa^*_{\overline{a}}$.

(ii) Upper bound: If $\kappa^*_{\overline{a}}$ is known, then by Corollary $\overline{3a}$(i) the upper bound on the total gain for the untreated is $\Delta(\overline{\kappa}^*)\kappa^*_{\overline{a}} + \overline{\Delta}(\overline{\kappa}^* - \kappa^*_{\overline{a}})$, reducing the ATE upper bound from $\overline{\Delta}$ to $\Delta(\overline{\kappa}^*)\kappa^*_{\overline{a}} + \overline{\Delta}(1 - \kappa^*_{\overline{a}})$. □

Corollary $\overline{4b}$. *Suppose $\overline{\kappa}^* = \kappa^*_{\overline{b}}$ solves Equation $\bar{2}$. The following ATE bounds can be obtained.*

(i) *If $\kappa^*_{\overline{a}}$ is unknown there is an informative lower bound, of* $ATE^{LB} = \Delta(\overline{\kappa}^*)(1-(1-\overline{\mu})) + \underline{\Delta}(1-\overline{\mu})$, *which is tightened to* $ATE^{LB} = \Delta(\overline{\kappa}^*)(1-\kappa^*_{\overline{a}}) + \underline{\Delta}\kappa^*_{\overline{a}}$, *if $\kappa^*_{\overline{a}}$ is known.*

(ii) *If $\kappa^*_{\overline{a}}$ is unknown there is an informative upper bound, of* $ATE^{UB} = \Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1-\overline{\mu})) + \overline{\Delta}(1-(\overline{\kappa}^* - (1-\overline{\mu})))$, *which is tightened to* $ATE^{UB} = \Delta(\overline{\kappa}^*)(\overline{\kappa}^* - ((1-\overline{\mu}) - \kappa^*_{\overline{a}})) + \overline{\Delta}(1-(\overline{\kappa}^* - ((1-\overline{\mu}) - \kappa^*_{\overline{a}})))$ *if $\kappa^*_{\overline{a}}$ is known.*

Proof. For each bound, first assume $\kappa^*_{\overline{a}}$ is not known.

(i) Lower bound: Measure $\overline{\kappa}^*$ units are untreated, and Corollary $\overline{3b}$(ii) shows that the ATUT is no smaller than untreated of $[\Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1-\overline{\mu})) + \underline{\Delta}(1-\overline{\mu})]/\overline{\kappa}^*$. Analogously, $1-\overline{\kappa}^*$ units are treated, and Equation $\overline{3b}$ implies that the ATT is no smaller than $\Delta(\overline{\kappa}^*)$. Integrate and sum the two parts to form $\text{ATE}^{\text{LB}} = \Delta(\overline{\kappa}^*)(1-(1-\overline{\mu})) + \underline{\Delta}(1-\overline{\mu})$. If $\kappa^*_{\overline{a}}$ is known, we can use the tighter lower bound on the ATUT from Corollary $\overline{3b}$(ii), resulting in a tighter lower bound of $\text{ATE}^{\text{LB}} = \Delta(\overline{\kappa}^*)(1-\kappa^*_{\overline{a}}) + \underline{\Delta}\kappa^*_{\overline{a}}$.

(ii) Upper bound: Use the expression for the upper bound of ATUT from Corollary $\overline{3b}$(i) and the fact that the upper bound on the treated students, with $x \in [\overline{\kappa}^*, 1]$, is $\overline{\Delta}$ (by Assumption 1(iii)), to integrate and sum to form $\text{ATE}^{\text{UB}} = \Delta(\overline{\kappa}^*)(\overline{\kappa}^* - (1-\overline{\mu})) + \overline{\Delta}(1-(\overline{\kappa}^* - (1-\overline{\mu})))$. If $\kappa^*_{\overline{a}}$ is known, then, similar to Corollary $\overline{3b}$(i), we can shift the measure $\kappa^*_{\overline{a}}$ from having an upper bound of $\overline{\Delta}$ to $\Delta(\overline{\kappa}^*)$, resulting in a tighter upper bound of $\text{ATE}^{\text{UB}} = \Delta(\overline{\kappa}^*)(\overline{\kappa}^* - ((1-\overline{\mu}) - \kappa^*_{\overline{a}})) + \overline{\Delta}(1-(\overline{\kappa}^* - ((1-\overline{\mu}) - \kappa^*_{\overline{a}})))$. □

Online Appendix 4 Weighted Objective

The administrator's original problem 2 was utilitarian, that is, it weighed gains for all students equally. The most natural alternative to the unweighted objective would be a redistributive policy, which assigned people with lower running variable indices larger weights. For example, if x measured quantiles of incoming human capital, then putting more weight on gains for students with lower indices allows the administrator to place additional value on students' becoming proficient. In this case, we can adapt Manuscript Equation 2 to allow the administrator to weigh gains for students depending on their index x by using weights $\phi(x)$, where $\phi' \leq 0$:

$$\max_{\widetilde{\kappa}} \left(\int_{\widetilde{\kappa}}^{1} \phi(x)\Delta(x)dx \right) - \chi\left(1-\widetilde{\kappa}\right), \tag{$\hat{2}$}$$

and proceed with the analysis.

Condition $\hat{1}$ (Necessity). *For problem $\hat{2}$, the following necessary conditions must hold for κ^*:*

(i) Marginal Benefit=Marginal Cost: $\phi(\kappa^*)\Delta(\kappa^*) = \chi$

(ii) Increasing Marginal Benefit: $\Delta'(\kappa^*) \geq 0$.

Proof. Differentiate the administrator's problem $\hat{2}$ with respect to $\tilde{\kappa}$ to obtain (i). Note that if the derivative is negative at a candidate solution satisfying (i), the administrator would gain by not treating students just above κ^*, thereby obtaining (ii). The inequality is strict if $\phi' < 0$. □

Condition $\hat{2}$ (Sufficiency). *The fact the program was implemented implies that the total gain from treating those units was at least as large as the total costs, that is:*

$$\text{Participation: } \int_{\kappa^*}^{1} \phi(x)\Delta(x)dx \geq \chi(1-\kappa^*). \tag{$\hat{3}$}$$

The fact the program was not extended to $\hat{\kappa} < \kappa^*$ implies that treating these units would be sub-optimal, that is:

$$\int_{\hat{\kappa}}^{\kappa^*} \phi(x)\Delta(x)dx < \chi(\kappa^* - \hat{\kappa}). \tag{$\hat{4}$}$$

Proposition 1 remains true when $\phi' \leq 0$. To see this, divide Equation $\hat{3}$ by the measure of treated students and combine with Condition $\hat{1}$(i) to obtain $\left(\int_{\kappa^*}^{1} \phi(x)\Delta(x)dx\right) / (1-\kappa^*) \geq \phi(\kappa^*)\Delta(\kappa^*)$. Because $\phi' \leq 0$, this implies that $\left(\int_{\kappa^*}^{1} \Delta(x)dx\right) / (1-\kappa^*) \geq \Delta(\kappa^*)$, where the inequality is strict if $\phi' < 0$. Intuitively, the gains for treating the treated must be even larger than the LATE if the administrator values such gains less. Analogous reasoning applied to Corollary 3 shows that the ATUT is bounded above by the LATE when $\phi' \leq 0$, and that this bound is strict when $\phi' < 0$. Therefore, the corollaries, in particular Corollary 4 bounding the ATE, also still obtain with the weighted problem Equation $\hat{2}$. In summary, all of the bounds from the unweighted problem, including Corollary 4, which bounds the ATE, are also obtained for the weighted problem $\hat{2}$.

Online Appendix 5 Variable Marginal Cost of Treatment

Begin by relaxing Assumption 1(i), replacing it with

Assumption 1′. *(i) The cost function $c(\cdot)$ is known and is non-negative, strictly increasing, and differentiable. The marginal cost function $c'(\cdot)$ is monotonic.*

Note that Assumption 1′(i) still implies that the marginal cost of providing treatment is strictly positive. The second part of Assumption 1′(i) relaxes the constant marginal cost assumption. Note that the cost can be variable in μ, but does not vary stochastically or *directly* with respect to x. However, it is possible to *indirectly* to pick up variation in costs with respect to x by using a reduced-form cost function $c_{\text{rf}}(\mu)$ in place of $c(\mu)$. Suppose the marginal cost was composed of two components and also took as an argument x: $c'_{\text{both}}(\mu, x) = c'(\mu) + c_x(x)$, where $c'(\mu)$ represented the marginal cost of the cost function in Assumption 1′ and $c_x(\cdot)$ was monotonic in x. For example, suppose the first component was constant, that is, $c'(\mu) = \chi$. Then, if $c_x(\cdot)$ is a constant χ_x the reduced-form marginal cost function $c'_{\text{rf}}(\mu) = \chi + \chi_x$ would also be constant. However, if $c_x(\cdot)$ is strictly increasing (decreasing) in x then the reduced-form total cost function would be $c_{\text{rf}}(\mu) = \int_{1-\mu}^{1} (c'(x) + c_x(x))\, dx$, which depends on the order in which students are treated. Then, the reduced form, $c'_{\text{rf}}(\mu)$, would be strictly decreasing (increasing), because students are added by extending the cutoff downward from 1. Indeed, an increasing $c_x(\cdot)$ could potentially transform an increasing marginal cost to a constant or even decreasing reduced-form marginal cost function, which would then be the one used in the analysis.

I first adapt the conditions characterizing κ^*, in terms of $\Delta(\cdot)$ and qualitative features of the (potentially reduced-form) cost function $c(\cdot)$. Specifically, I consider three cases for Assumption 1′(i): where the marginal cost is constant, decreasing, and increasing; these correspond to linear, concave, and convex cost functions, respectively. I then provide results bounding treatment effect parameters of interest.

Condition 1′ (Necessity)**.** *The following necessary conditions must hold for κ^*:*

(i) Marginal Benefit=Marginal Cost: $\Delta(\kappa^*) = c'(1-\kappa^*)$ for any cost function $c(\cdot)$ satisfying Assumption 1′

(ii) Increasing Marginal Benefit: $\Delta'(\kappa^*) \geq 0$ if the marginal cost is constant or decreasing; this inequality is strict if the marginal cost is decreasing.

Proof. Differentiate the administrator's problem 2 with respect to $\tilde{\kappa}$ to obtain (i). Note that if the derivative is negative at a candidate solution satisfying (i) but the marginal cost is nonincreasing, the administrator would gain by not treating students just above κ^*, thereby obtaining (ii). □

Condition 1′ is similar to Condition 1, except that Condition 1′(ii) has a strict inequality if the marginal cost of treatment is decreasing. As before, to guarantee uniqueness, inspection of Manuscript Equation 2 implies two additional conditions sufficient for characterizing κ^*. These conditions are identical to those in Condition 2, the only difference being that χ no longer enters either expression.

Condition 2′ (Sufficiency)**.** *The fact the program was implemented implies that the total gain from treating those units was at least as large as the total costs, that is:*

$$\text{(3}'\text{)} \qquad \textit{Participation:} \quad \int_{\kappa^*}^{1} \Delta(x)dx \geq c(1-\kappa^*).$$

The fact the program was not extended to $\hat{\kappa} < \kappa^*$ implies that treating these units would be sub-optimal, that is:

$$\text{(4}'\text{)} \qquad \int_{\hat{\kappa}}^{\kappa^*} \Delta(x)dx < c(1-\hat{\kappa}) - c(1-\kappa^*).$$

As before, a corollary immediately follows.

Corollary 1′. *The following are globally true about $\Delta(\cdot)$ if the marginal cost of treatment is nonincreasing:*

(i) $\Delta(\cdot)$ *cannot be constant.*

(ii) $\Delta(\cdot)$ *is not globally monotonically decreasing in x.*

Proof. Identical to proof of Corollary 1. □

As before, I next examine what can be deduced about averages of treatment effects for subsets of students.

Corollary 2′. *The ATT is positive for any cost function $c(\cdot)$ satisfying Assumption 1′.*

Proof. The left side of Equation $3'$ in Condition $2'$ is the total effect of treatment on the treated, that is, $(\int_{\kappa^*}^{1} \frac{\Delta(x)}{(1-\kappa^*)} dx)\,(1-\kappa^*)$. Because the marginal cost of treatment is positive (Assumption $1'$(i)), Equation $3'$ implies that

$$\int_{\kappa^*}^{1} \Delta(x)dx \geq c(1-\kappa^*) > 0.$$

Divide through by $(1-\kappa^*)$ to obtain the result:

$$\underbrace{\int_{\kappa^*}^{1} \frac{\Delta(x)}{(1-\kappa^*)}dx}_{\text{ATT}} \geq \underbrace{\frac{c(1-\kappa^*)}{(1-\kappa^*)}}_{\text{avg. cost of treating treated}} > 0. \qquad \square$$

Although Corollary $2'$ provides a lower bound for the average effect of treatment on the treated, there is no informative (that is, lower than $\overline{\Delta}$) upper bound. Corollary $2'$ makes no further assumptions about the shape of the cost function. However, if the marginal cost of treating students is nonincreasing, the lower bound on the average effect of treatment on the treated increases.

Proposition $1'$. *If the marginal cost of treatment is nonincreasing, the ATT is bounded below by the LATE at the treatment cutoff.*

Proof. If the marginal cost of treatment is nonincreasing then $c'(1-\kappa^*) \leq \frac{c(1-\kappa^*)}{1-\kappa^*}$, that is, the marginal cost of treatment for $1-\kappa^*$ is no greater than the average cost of providing treatment for treated students. Insert this inequality into Equation $3'$ and combine with this with Condition $1'$(i) to obtain

$$\underbrace{\frac{\int_{\kappa^*}^{1} \Delta(x)dx}{1-\kappa^*}}_{\text{ATT}} \geq \frac{c(1-\kappa^*)}{1-\kappa^*} \geq c'(1-\kappa^*) = \underbrace{\Delta(\kappa^*)}_{\text{LATE at } \kappa^*}. \qquad \square$$

As with Proposition 1, Proposition $1'$ shows that, if the marginal cost of treatment is nonincreasing, the discontinuity-based estimate provides a lower bound for the average effect of treatment on the treated. One should note that only qualitative information about the shape, not the level, of the marginal cost of treatment is all that is required for this result.

Although Corollary $1'$(ii) rules out an average treatment effect that is decreasing everywhere (if the marginal cost of treatment is nonincreasing), it could be the case that $\Delta(\cdot)$ is decreasing for some $x < \kappa^*$.[1] Therefore, as before, it is useful to bound averages of $\Delta(\cdot)$

[1] Note that Corollary $1'$(ii) would also obtain when the marginal cost of treatment was strictly increasing,

itself for strict subsets of untreated students.

Proposition 2′. *There exists an informative upper bound for* $\int_a^b \Delta(x)dx$ *for* $0 \leq a < b \leq \kappa^*$ *if* $\underline{\Delta} > \frac{c(1-a)-c(1-\kappa^*)-\overline{\Delta}(b-a)}{\kappa^*-b}$.

Proof. Suppose we would like to characterize $\Delta(\cdot)$ for values less than $\hat{x} < \kappa^*$. Let $\hat{\mu}$ be the measure of students under consideration and split Equation 4′ into two parts at $\hat{x}$ and rearrange terms:

(5′)

$$\int_{\hat{x}-\hat{\mu}}^{\hat{x}} \Delta(x)dx < c(1-(\hat{x}-\hat{\mu}))-c(1-\kappa^*)-\int_{\hat{x}}^{\kappa^*} \Delta(x)dx \Rightarrow \int_{\hat{x}-\hat{\mu}}^{\hat{x}} \Delta(x)dx < c(1-(\hat{x}-\hat{\mu}))-c(1-\kappa^*)-\underline{\Delta}\,(\kappa^* - \hat{x})\,,$$

where the implication follows from Assumption 1(iii).[2] □

Setting the measure of students to whom the treatment is extended equal to κ^* provides the following result about the ATUT.

Corollary 3′. *The ATUT has an informative upper bound. If the marginal cost of treatment is nonincreasing, this upper bound is the LATE at the treatment cutoff.*

Proof. Let $\hat{x} = \hat{\mu} = \kappa^*$ in Equation 5′ and divide through by κ^* to obtain the first result:

$$\underbrace{\int_0^{\kappa^*} \frac{\Delta(x)}{\kappa^*}dx}_{\text{ATUT}} < \underbrace{\frac{c(1) - c(1-\kappa^*)}{\kappa^*}}_{>0,<\infty}, \tag{8′}$$

where the right hand side is positive from Assumption 1′(i). For the second result, note that a nonincreasing marginal cost implies

$$\frac{c(1) - c(1-\kappa^*)}{\kappa^*} < c'(1-\kappa^*) = \underbrace{\Delta(\kappa^*)}_{\text{LATE at }\kappa^*}\ ,$$

where the equality follows from Proposition 1′(i). □

Analogously to the upper bound for the ATT, although Corollary 3′ bounds the average of treatment effects for all untreated students, there is no informative (that is, greater than $\underline{\Delta}$) lower bound.

if the administrator could choose which side of the cutoff to treat.

[2] As with Proposition 2, this bound will be informative for all but very low values of $\underline{\Delta}$. This condition is satisfied when using Proposition 2′ to bound the ATUT or ATE if the average cost of treating the untreated is less than $\overline{\Delta}$.

To summarize, optimality of κ^* implies a lower bound on the ATT and an upper bound on the ATUT. If the marginal cost of treatment is constant or decreasing then it must be the case that ATUT $< \Delta(\kappa^*) \leq$ ATT. Though the ATT and ATUT are respectively bounded below and above by the cutoff LATE when marginal costs are nonincreasing, the LATE does not bound these moments when the marginal cost of treatment is strictly increasing.

5.1 Bounding the ATE

This section studies the interplay between qualitative features of the cost of treatment and inferences about treatment effects, by comparing three cases: constant, decreasing, and increasing marginal cost of treatment, where each marginal cost curve passes through the point $(\kappa^*, \Delta(\kappa^*))$. A decreasing marginal cost ($c'' < 0$) might result from economies of scale, while an increasing marginal cost ($c'' > 0$) might result from congestion effects, say if it becomes increasingly difficult to find a good fit for the program.

To begin, suppose the cost function is $c(\mu) = \mu\chi$. Then, as was shown in Section III.A., the ATE lower bound is $\text{ATE}^{\text{LB}} = \underline{\Delta}\kappa^* + \chi(1-\kappa^*)$ and the ATE upper bound is $\text{ATE}^{\text{UB}} = \chi\kappa^* + \overline{\Delta}(1-\kappa^*)$, because $\chi = \Delta(\kappa^*)$ by Condition 1(i). Figure A.1 builds on the example in Manuscript Figure 1 to provide intuition for how the marginal cost of treatment bounds the ATE.

Start with the solid horizontal red line representing a constant marginal cost of treatment, and rotate the cost function counterclockwise about the point $(\kappa^*, \Delta(\kappa^*))$ to represent a decreasing marginal cost of treatment (long-dashed red line).[3] This rotation implies the ATT must be higher than the case corresponding to the constant marginal cost in order to satisfy Equation 3′. Analogously, the maximum ATUT must be lower when marginal costs are decreasing; were they the same as with constant marginal costs, the administrator might gain from extending treatment to untreated units given that they now have a lower cost of being treated, violating Equation 4′. The opposite holds true for when we rotate the cost curve clockwise about the point $(\kappa^*, \Delta(\kappa^*))$, to reflect an increasing marginal cost of treatment (dot-dashed red line). Table A.1 summarizes these results, showing that when the marginal cost of treatment is decreasing, bounds on the population ATE are tighter than they would be with a constant marginal cost, while when marginal cost is increasing, bounds on the population ATE are looser.

[3] Recall that this line is decreasing in x because the treatment is being extended from $x = 1$ downwards.

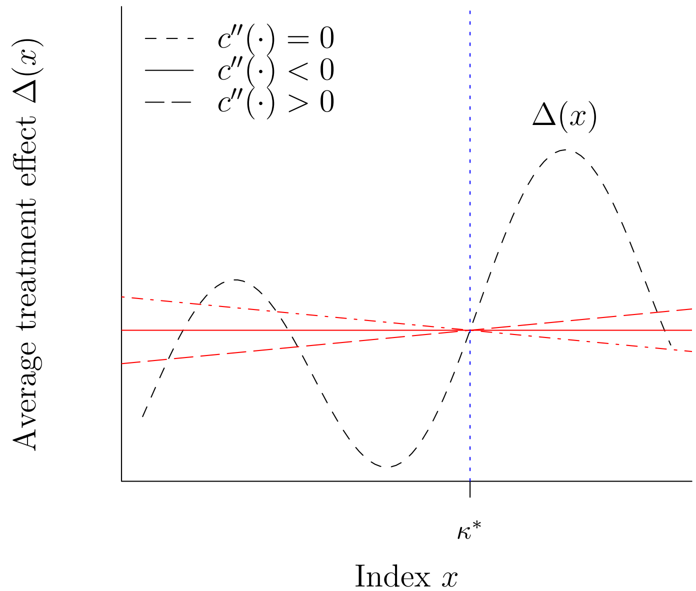

Figure A.1: Example with different cost functions

Table A.1: Summary of bounds on population ATE

	ATE bounds	
Marginal cost	Lower	Upper
Const. ($c'' = 0$)	$=\text{ATE}^{\text{LB}}$	$= \text{ATE}^{\text{UB}}$
Dec. ($c'' < 0$)	$> \text{ATE}^{\text{LB}}$	$< \text{ATE}^{\text{UB}}$
Inc. ($c'' > 0$)	$< \text{ATE}^{\text{LB}}$	$> \text{ATE}^{\text{UB}}$