Online Supplement to "Reducing the Filtering Effect in Public School Admissions: A Bias-aware Analysis for Targeted Interventions"

Authors: Yuri Faenza, Swati Gupta, Aapeli Vuorinen, and Xuan Zhang.

Because of the page limit, some of the proofs are shortened, and the simplest ones omitted, from this Electronic Companion. Full details of the proofs, as well as extended discussions, appear in the ArXiv version of the paper (Faenza et al. 2020).

Appendix EC.1: Discussion on discrete versus continuous models

Traditionally, matching markets are assumed to be discrete (Gale and Shapley 1962, Roth and Sotomayor 1992). In recent years, however, there has been an interest in models where one or both sides of the markets are continuous (Arnosti 2019, Azevedo and Leshno 2016). This is justified by the fact that in many applications, markets are large, hence predictions in continuous markets often translate with a good degree of accuracy to discrete ones. Moreover, continuous markets are often analytically more tractable than discrete ones (see, again, Arnosti (2019), Azevedo and Leshno (2016)). Our case is no exception: the continuous model allows us to deduce precise mathematical formulae, while we show through experiments that those formulae are a good approximation to the discrete case. We remark that the goal of this study is not to provide a mechanism to admits students to schools, for which the assumption of all rankings of schools as well as of students being the same would be too simplistic. On the contrary, as we want to understand the impact of bias at a macroscopic level, we believe our approximation to be meaningful and useful, since as in our model, any reasonable mechanism would output the same assignment.

Appendix EC.2: Impact on Schools

We define the *utility* $u_\gamma(s)$ of a school s under matching $\gamma \in \{\mu, \widehat{\mu}\}$, as

$$u_\gamma(s) := \int_{\theta \in \gamma^{-1}(s) \cap G_1} Z(\theta) dF_1(\widehat{Z}(\theta)) + \int_{\theta \in \gamma^{-1}(s) \cap G_2} Z(\theta) dF_2(\widehat{Z}(\theta)). \tag{EC.1}$$

That is, the utility of a school is the average true potential of admitted students. Continuing from Example 1, let s_M (resp. s_L) be the school Maya (resp. Lisa) is assigned to in the biased setting. The utilities of s_M, s_L in the unbiased setting are $u_\mu(s_M) = 1.283$ and $u_\mu(s_L) = 1.324$, while in the biased setting, they are $u_{\widehat{\mu}}(s_M) =$

1.280 and $u_{\widehat{\mu}}(s_L) = 1.320$. Hence, the change in the utilities of the two school between the two settings is negligible. We next validate these observations formally.

Let $s \in [0,1]$ denote the school that is ranked in the $s \times 100\%$ position among the continuous range of schools. As the next proposition shows, the impact on the utilities of schools is negligible for all schools other than the lowest ranked schools: for each school, although the average potential of assigned G_1 students is lower than it should be, its assigned G_2 students have much higher true potentials. Thus, the toll on the utility due to unqualified G_1 students is partially canceled out by the overqualified G_2 students and the net effect is minimal. On the other hand, some lower ranked schools that only admit G_2 students fare better in the biased setting (since they admit over-qualified G_2 candidates).

PROPOSITION EC.1. *For school s, its utility under the unbiased (resp. biased) models are respectively*

$$u_{\mu}(s) = s^{-\frac{1}{\alpha}} \quad \text{and} \quad u_{\widehat{\mu}}(s) = \begin{cases} \dfrac{1-p+p\beta^{\alpha}}{1-p+p\beta^{\alpha+1}} \left(\dfrac{s}{1-p+p\beta^{\alpha}} \right)^{-\frac{1}{\alpha}} & \text{if } s \leq 1-p+p\beta^{\alpha}, \\ \left(\dfrac{s-(1-p)}{p} \right)^{-\frac{1}{\alpha}} & \text{if } s > 1-p+p\beta^{\alpha}. \end{cases}$$

The key idea in the proof is to first compute the *cutoffs* at each school for each of the two groups, that is, the minimum *perceived* potential needed for a student to be matched to a given school. Once these are known, using Bayes' rule, we deduce the minimum *real* potential needed by students of each group to attend the school. From the latter, we can immediately compute the average utility of each school. The simple calculations leading to the formula from Proposition EC.1 are omitted: we refer to the ArXiv version (Faenza et al. 2020) for details.

As one readily observes from Proposition EC.1, the negative impact of bias on schools' utility is negligible. Hence, from an operational perspective, it may be hard to convince schools to autonomously put in place mechanisms to alleviate the effect of bias given the limited impact on them.

Appendix EC.3: Proof of Theorem 1

The analysis we give leads to a more general statement than Theorem 1, and has the goal of investigating conditions under which giving vouchers can improve over the status quo. More formally, for bounded functions $f, g : G_2 \to \mathbb{R}$, we write $f \succ g$ if we can partition G_2 in two sets S, S' (with possibly $S' = \emptyset$) so that $f(\theta) = g(\theta)$ for $\theta \in S'$ and $\sup_{\theta \in S} f(\theta) > \sup_{\theta \in S} g(\theta)$.

Note that $\succ$ is transitive and antisymmetric, and can be interpreted as a continuous equivalent of the classical lexicographic ordering for discrete vectors. In particular, if we let $f=\gamma-\mu$ and $g=\gamma'-\mu$ for matchings γ,γ', then $\sup_{\theta\in G_2}(\gamma-\mu)(\theta)>\sup_{\theta\in G_2}(\gamma'-\mu)(\theta)$ implies $f\succ g$ (taking $S=G_2$). Now suppose we debias student in $T=[Z_1,Z_2]$ for some $T\in\mathcal{T}(\widehat{c})$, and let $f:=\widehat{\mu}-\mu$, $g:=\mu_T-\mu$. Table EC.1 provides conditions under which $f\succ g$ (i.e., intervention reduces the maximum mistreated experienced by G_2 students). In particular it shows that for certain combinations of the data and the choice of Z_1 and Z_2, giving vouchers may actually lead to worse (according to $\succ$) matchings. One can check that if $p<1-\beta^\alpha$, all conditions given in Table EC.1 for different cases are satisfied.

CASE	subcase	condition for $\widehat{\mu}-\mu\succ\mu_T-\mu$
I. $\beta Z_2\geq Z_1$	1. $1\leq\beta Z_1$	$p<1-\left(\frac{Z_1}{Z_2}\right)^\alpha$
	2. $\beta Z_1\leq 1\leq\beta Z_2$	$p<1-\left(\frac{1}{\beta Z_2}\right)^\alpha$
II. $\beta Z_2\leq Z_1$	1. $1\leq\beta Z_1$	$p<1-\beta^\alpha$
	2. $\beta Z_1\leq 1\leq\beta Z_2$	$p\left(\left(\frac{1}{Z_1}\right)^\alpha-\left(\frac{1}{Z_2}\right)^\alpha\right)<(1-p)\left(\frac{1}{\beta^\alpha}-1\right)\left(\beta^\alpha-\left(\frac{1}{Z_2}\right)^\alpha\right)$
	3. $\beta Z_2\leq 1$	Not possible: $g\succ f$ in this case.

Table EC.1 Sufficient conditions for $\widehat{\mu}-\mu\succ\mu_T-\mu$ by cases, where $T=[Z_1,Z_2]$. Each strict inequality, when replaced with its non-strict counterpart, gives instead a necessary condition.

As it turns out, the optimal choice of Z_1,Z_2 is either case I.2 or case II.2, and exactly which one the optimal solution is depends on the amount of resources, i.e., the value of $\widehat{c}$. We refer the interested reader to the ArXiv version (Faenza et al. 2020) for details.

Appendix EC.4: Proof of Theorem 2

Assume $T=[Z_1,Z_2]$ is the range of true potentials of G_2 students we want to debias. For simplicity, as in previous sections, let $\widetilde{\mu}$ denote μ_T. In order to obtain the minimizer of $\sigma(\widetilde{\mu}-\mu)$, first, we want to compute $\sigma(\widetilde{\mu}-\mu)$ for each of the cases in Table EC.1.

For $1\leq t_1\leq t_2\in\mathbb{R}\cup\{+\infty\}$, let $\sigma_{t_1}^{t_2}(f):=\int_{t_1}^{t_2}\max(f(t),0)dF_1(t)$ for any function $f:[1,\infty]\to[0,1]$. When $t_1=1$ and $t_2=\infty$, we simply write $\sigma(f)$, which is consistent with previous notations. Note that with $\sigma(\widehat{\mu}-\mu)$ as a reference, it actually suffices to compute only $\sigma_{Z_1}^{Z_2/\beta}(\widetilde{\mu}-\mu)$, because minimizing $\sigma(\widetilde{\mu}-\mu)$ is equivalent to maximizing $\sigma_{Z_1}^{Z_2/\beta}(\widehat{\mu}-\mu)-\sigma_{Z_1}^{Z_2/\beta}(\widetilde{\mu}-\mu)$ since $(\widehat{\mu}-\mu)(\theta)=(\widetilde{\mu}-\mu)(\theta)$ for all $\theta\in G_2$ with $Z(\theta)\notin[Z_1,Z_2/\beta]$.

For each case, we give an explicit formula for $\sigma_{Z_1}^{Z_2/\beta}(\widehat{\mu}-\mu)-\sigma_{Z_1}^{Z_2/\beta}(\widetilde{\mu}-\mu)$. These formulae can be computed via simply integration, and are thus omitted. In addition, we analyze how this value changes (increase or decrease) with respect to Z_1 and Z_2. Details can be found in the ArXiv version of the paper (Faenza et al. 2020).

Appendix EC.5: Proof of Lemma 1

Assume T is not of the form $\{\theta\in\Theta : Z(\theta)\geq\delta\}$ or $\{\theta\in\Theta : Z(\theta)>\delta\}$ for some δ and let U be a connected and inclusionwise maximal subset of T that is bounded. Take the smallest number $\delta_1\in[1,\infty)$ so that $Z(\theta)\leq\delta_1$ for all $\theta\in U$. Let $\overline{\theta}\in\theta$ such that $Z(\overline{\theta})=\delta_1$.

Assume first that $\overline{\theta}\in U$. Then, for each $\varepsilon_1>0$, there exists $\varepsilon\in t(0,\varepsilon_1]$ such that $Z^{-1}(\delta_1+\varepsilon)\notin T$. Since $\beta<1$ and by continuity of $Z(\cdot)$, there exists $\varepsilon_2>0$ such that $\beta(\delta_1+\varepsilon)<\delta_1$ for all $\varepsilon<\varepsilon_2$. We can then take an appropriate $x\in[\delta_1,\delta_1+\varepsilon_2]\setminus T$ and $x'=\delta_1$ to show that T is not incentive compatible. Next assume $\overline{\theta}\notin U$. In particular, we have $\overline{\theta}\notin T$. Similarly to the case above, we can find $\varepsilon>0$ such that $x'=\delta_1-\varepsilon$ satisfies $Z^{-1}(x')\in U\subseteq T$ and $\beta\delta_1<x'$. Setting $x=\delta_1$, we deduce that T is not incentive compatible.

Appendix EC.6: Proofs from Section 5.1

EC.6.1. Auxiliary results for Section 5.1

Recall that, in this section, we consider the generalization of the model from Section 2 where students' true potential follow a generic continuous, integrable cdf F. Moreover, we write $[x]^+:=\max(0,x)$. Similarly to Section 5.1, we abuse notation and identify a student θ with their potential $Z(\theta)$.

Lemma EC.1. *Let ρ be an RVP. Under any continuous distribution of potentials F, we have*

$$\mu_\rho(\theta)=\rho(\theta)\left((1-p)\int_\theta^\infty dF+p\left[\int_\theta^{\theta/\beta}\rho\,dF+\int_{\theta/\beta}^\infty dF\right]\right)+(1-\rho(\theta))\cdot \left((1-p)\int_{\beta\theta}^\infty dF+p\left[\int_{\beta\theta}^{\theta}\rho\,dF+\int_{\theta}^\infty dF\right]\right), \tag{EC.2}$$

$$m_\rho(\theta)=[(1-\rho(\theta))(1-p)\int_{\beta\theta}^{\theta}dF+p\left[(1-\rho(\theta))\int_{\beta\theta}^{\theta}\rho\,dF-\rho(\theta)\int_\theta^{\theta/\beta}(1-\rho)\,dF\right]]^+. \tag{EC.3}$$

We next report more useful expressions for μ_ρ and μ'_ρ.

PROPOSITION EC.2. *Let ρ be an RVP. For all $\theta \in \Theta$, we have*

$$\mu_\rho(\theta) = -\rho(\theta)\left((1-p)\int_{\beta\theta}^{\theta} dF + p\left[\int_{\theta}^{\theta/\beta}(1-\rho)\,dF + \int_{\beta\theta}^{\theta}\rho\,dF\right]\right) + \left((1-p)\int_{\beta\theta}^{\infty} dF + p\left[\int_{\beta\theta}^{\theta}\rho\,dF + \int_{\theta}^{\infty} dF\right]\right). \tag{EC.4}$$

Moreover, if μ_ρ is differentiable at θ, we have

$$\begin{aligned}\mu'_\rho(\theta) = &-f(\theta) - \rho'(\theta)\left[p\int_{\theta}^{\theta/\beta}(1-\rho)\,dF + (1-p)\int_{\beta\theta}^{\theta} dF + p\int_{\beta\theta}^{\theta}\rho\,dF\right] \\ &- p\rho(\theta)\left[\frac{1}{\beta}(1-\rho(\theta/\beta))f(\theta/\beta) - (1-\rho(\theta))f(\theta)\right] \\ &+ (1-\rho(\theta))\left[(1-p)(f(\theta)-\beta f(\beta\theta)) + p(f(\theta)\rho(\theta) - \beta f(\beta\theta)\rho(\beta\theta))\right].\end{aligned} \tag{EC.5}$$

DEFINITION EC.1. The RVP that assigns no vouchers, denoted ρ_0, is defined by $\rho_0(\theta) := 0$ for all $\theta \in [1,\infty)$. Note that $\mu_{\rho_0}(\theta) = (1-p)\int_{\beta\theta}^{\infty} dF + p\int_{\theta}^{\infty} dF$ and $m_{\rho_0}(\theta) = m_{\widehat{\mu}}(\theta) = (1-p)\int_{\beta\theta}^{\theta} dF$.

EC.6.2. Necessary and sufficient conditions for incentive compatibility

In this section we develop necessary and sufficient conditions for incentive compatibility through the concept of well-behavedness and prove an important technical lemma.

DEFINITION EC.2 (WELL-BEHAVED RVP). We call an RVP ρ *well-behaved* if it is everywhere continuously differentiable except for a set of isolated points where it has non-negative, right-continuous jump discontinuities.

LEMMA EC.2 **(Necessary and sufficient conditions for incentive compatibility)**. *Let ρ be a well-behaved RVP and F be an arbitrary continuous distribution of potentials. ρ is incentive compatible with respect to F if and only if, for all θ such that ρ is continuously differentiable at θ, we have $\rho'(\theta) \geq 0$ or $\mu'_\rho(\theta) \leq 0$.*

Using the previous fact, one easily shows the following.

LEMMA EC.3. *Suppose ρ is a well-behaved RVP such that for all θ where ρ is continuously differentiable, we have $\rho'(\theta) \geq -\phi(\theta)$, with $\phi(\theta) := \frac{\alpha(1-p)}{\theta[p(1-\beta^\alpha)+(1-p)(\beta^{-\alpha}-1)]}$. Then, ρ is incentive compatible.*

EC.6.3. Proof of Theorem 3: properties of PropMs

We next prove Lemmas EC.4, EC.5, and EC.8, which together constitute Theorem 3.

LEMMA EC.4. *The proportional-to-mistreatment RVP ρ_m is $\frac{2\widehat{c}\alpha}{1-\beta^\alpha}$-individually fair.*

Proof. ρ_m is everywhere continuous and continuously differentiable on Θ, except at $\theta = 1/\beta$. ρ_m is therefore Lipschitz for a constant given by the supremum of the absolute value of the derivative, which occurs at $\theta = 1$ where $\rho_m'(\theta) = \frac{2\widehat{c}\alpha}{1-\beta^\alpha}$. □

LEMMA EC.5. *The proportional-to-mistreatment RVP ρ_m is incentive compatible for $\widehat{c} \le \frac{1-p}{2[p(1-\beta^\alpha)+(1-p)(\beta^{-\alpha}-1)]}$.*

Proof. Applying Lemma EC.3, it suffices to show

$$-\rho_m'(\theta) = 2\alpha\widehat{c}\beta^{-\alpha}\theta^{-\alpha-1} \le \phi(\theta) = \frac{\alpha(1-p)}{\theta\left[p(1-\beta^\alpha)+(1-p)(\beta^{-\alpha}-1)\right]}. \tag{EC.6}$$

for $\theta \ge 1/\beta$ (since $\rho_m'(\theta) \ge 0$ for $\theta < 1/\beta$). Note that this is tightest when $\theta = 1/\beta$, which gives the condition for $\widehat{c}$. □

LEMMA EC.6. *For the proportional-to-mistreatment RVP ρ_m, we have*

$$\sup_{\theta\in\Theta}(1-\rho_m(\theta))\int_{\beta\theta}^{\theta} dF = (1-\beta^\alpha)\xi(\widehat{c}), \qquad \textit{where} \quad \xi(\widehat{c}) := \begin{cases} 1-2\widehat{c}, \widehat{c} \le 1/4, \\ \frac{1}{8\widehat{c}}, \widehat{c} > 1/4. \end{cases}$$

LEMMA EC.7. *For the proportional-to-mistreament RVP, ρ_m, the maximum mistreament mm_{ρ_m} satisfies $mm_{\rho_m} \le (1-p(1-2\widehat{c}))(1-\beta^\alpha)\xi(\widehat{c})$, where $\xi(\cdot)$ is defined as in Lemma EC.6.*

Proof. Abbreviate $\rho = \rho_m$. Apply $\rho(\theta) \le \rho(1/\beta)$ to (EC.3) and simplify to get

$$\begin{aligned} m_{\rho_m}(\theta) &\le \left[(1-p)(1-\rho(\theta))\int_{\beta\theta}^{\theta} dF + p(1-\rho(\theta))\rho(1/\beta)\int_{\beta\theta}^{\theta} dF\right]^+ \\ &= \left[(1+p(\rho(1/\beta)-1))(1-\rho(\theta))\int_{\beta\theta}^{\theta} dF\right]^+. \end{aligned}$$

The thesis follows by taking the supremum over $\theta \in \Theta$, applying Lemma EC.6, and substituting $\rho(1/\beta) = 2\widehat{c}$. □

Recall that we let $mm^*(\widehat{c})$ be the maximum mistreatment achieved by the optimal policy from Theorem 1 with the amount of resources being $\widehat{c}$. We have

$$mm^*(\widehat{c}) = \begin{cases} (1-p-\widehat{c})(1-\beta^\alpha)+\widehat{c}p & \text{if } \widehat{c} \le \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}, \\ (1-p)(1-\beta^\alpha)\frac{1-\widehat{c}}{1-p\beta^\alpha} & \text{otherwise.} \end{cases} \tag{EC.7}$$

LEMMA EC.8. *Let* $p < \min\{1-\beta^\alpha, 1/2\}$ *and* $1 - \frac{p+1-\beta^\alpha}{4p(1-\beta^\alpha)} \le \widehat{c} \le \frac{(1-p)(1-\beta^\alpha)}{1-p+1-\beta^\alpha}$. *Then* $mm_{\rho_m} \le mm^*(\widehat{c})$.

Proof. Let $Q := mm^*(\widehat{c}) - mm_{\rho_m}$. We need to show $Q \ge 0$. Using Theorem 1 and Lemma EC.7, compute

$$Q = mm^*(\widehat{c}) - mm_{\rho_m} \ge (1-p)(1-\beta^\alpha) - \widehat{c}(1-\beta^\alpha - p) - (1+p(2\widehat{c}-1))(1-\beta^\alpha)\xi(\widehat{c}).$$

Simple algebraic manipulations lead to the thesis. □

Appendix EC.7: Proofs from Section 5.2

Proof of Lemma 2. Define $m_{\rho,\kappa}(\theta)$ as the expected mistreatment experienced by a disadvantaged student $\theta \in \Theta \cap G_2$. Recall from Lemma EC.1 that $m_\rho(\theta) = [M_\rho]^+(\theta)$, where

$$M_\rho(\theta) = (1-\rho(\theta))(1-p)\int_{\beta\theta}^{\theta} dF + p\left[(1-\rho(\theta))\int_{\beta\theta}^{\theta}\rho\, dF - \rho(\theta)\int_{\theta}^{\theta/\beta}(1-\rho)\, dF\right].$$

Similarly, we have $m_{\rho,\kappa}(\theta) = [M_{\rho,\kappa}(\theta)]^+$, where

$$M_{\rho,\kappa}(\theta) = (1-\kappa\rho(\theta))(1-p)\int_{\beta\theta}^{\theta} dF + p\left[(\kappa-\kappa^2\rho(\theta))\int_{\beta\theta}^{\theta}\rho\, dF - \kappa\rho(\theta)\int_{\theta}^{\theta/\beta}(1-\kappa\rho)\, dF\right].$$

We can then write,

$$\begin{aligned}
m_{\rho,\kappa}(\theta) &= [M_{\rho,\kappa}(\theta)]^+ \\
&= \Bigg[M_\rho(\theta) + (1-\kappa)\rho(\theta)(1-p)\int_{\beta\theta}^{\theta} dF - p(1-\kappa)\int_{\beta\theta}^{\theta}\rho\, dF + p(1-\kappa^2)\rho(\theta)\int_{\beta\theta}^{\theta}\rho\, dF \\
&\quad + p(1-\kappa)\rho(\theta)\int_{\theta}^{\theta/\beta} dF - p(1-\kappa^2)\rho(\theta)\int_{\theta}^{\theta/\beta}\rho\, dF\Bigg]^+ \\
&\le \left[M_\rho(\theta) + (1-\kappa)\rho(\theta)(1-p)\int_{\beta\theta}^{\theta} dF + p(1-\kappa^2)\rho(\theta)\int_{\beta\theta}^{\theta}\rho\, dF + p(1-\kappa)\rho(\theta)\int_{\theta}^{\theta/\beta} dF\right]^+ \\
&= \left[M_\rho(\theta) + (1-\kappa)\rho(\theta)\left((1-p)\int_{\beta\theta}^{\theta} dF + p(1+\kappa)\int_{\beta\theta}^{\theta}\rho\, dF + p\int_{\theta}^{\theta/\beta} dF\right)\right]^+ \\
&\le \left[M_\rho(\theta) + (1-\kappa)\|\rho\|_\infty\left((1-p)\int_{\beta\theta}^{\theta} dF + p(1+\kappa)\|\rho\|_\infty\int_{\beta\theta}^{\theta} dF + p\int_{\theta}^{\theta/\beta} dF\right)\right]^+ \\
&\le [M_\rho(\theta)]^+ + (1-\kappa)(1-\beta^\alpha)\|\rho\|_\infty\left(1 + p(1+\kappa)\|\rho\|_\infty\right) \\
&= m_\rho(\theta) + (1-\kappa)(1-\beta^\alpha)\|\rho\|_\infty\left(1 + p(1+\kappa)\|\rho\|_\infty\right),
\end{aligned}$$

where the second relation is obtained by the using the formula for $M_\rho(\theta)$, the third by dropping nonnegative terms, the fourth by rearranging terms, the fifth by replacing $\rho(\theta)$ by its maximum, the sixth by upper bounding both $\int_{\beta\theta}^{\theta} dF$ and $\int_{\theta}^{\theta/\beta} dF$ by $(1-\beta^\alpha)$, and moving nonnegative terms out of the $[\cdot]^+$ function, and the last by definition of $m_\rho(\theta)$. By taking the supremum of $m_{\rho,\kappa}(\theta)$ for $\theta \in \Theta \cap G_2$, the upper bound follows. □

Proof of Lemma 3. From the proof of Lemma 2, we have that, for $\theta \in \Theta$,

$$M_\rho(\theta) = (1-r)(1-p)\int_{\beta\theta}^{\theta} dF + p\left[(1-r)r\int_{\beta\theta}^{\theta} dF - r(1-r)\int_{\theta}^{\theta/\beta} dF\right]$$
$$= (1-r)\left[(1-p+pr)\int_{\beta\theta}^{\theta} dF - pr\int_{\theta}^{\theta/\beta} dF\right],$$

and that

$$M_{\rho,\kappa}(\theta) = (1-\kappa r)(1-p)\int_{\beta\theta}^{\theta} dF + p\left[(\kappa-\kappa^2 r)r\int_{\beta\theta}^{\theta} dF - \kappa r(1-\kappa r)\int_{\theta}^{\theta/\beta} dF\right]$$
$$= (1-\kappa r)\left[(1-p+pkr)\int_{\beta\theta}^{\theta} dF - pkr\int_{\theta}^{\theta/\beta} dF\right] = M_{\kappa\rho}.$$

□

Proof of Lemma 4. Assume that the potential of students is described by a Pareto distribution with $\alpha = 1$. Let $p = 0.8, \beta = 0.2, \kappa = 0.4; \rho(\theta) = 0.8$ and $\widehat{\rho}(\theta) = 0.5$ for $\theta \in \Theta$. Let $\rho^*(\theta) = r$ for $\theta \in \Theta$. Note that, for every $\widehat{\kappa}$, $mm_{\rho^*,\widehat{\kappa}}$, is achieved at the value of θ that maximizes the mistreatment under $\widehat{\mu}$, that is at $\widehat{\theta} = 1/\beta = 5$. Compute $\int_{\beta\widehat{\theta}}^{\widehat{\theta}} dF = 0.8$, $\int_{\widehat{\theta}}^{\widehat{\theta}/\beta} dF = 0.16$ and substitute them into the formula for $M_{\rho^*}(5)$ computed in the proof of Lemma 3:

$$M_{\rho^*}(\widehat{\theta}) = (1-r)[(0.2+0.8r)(0.8) - 0.8r(0.16)] = (1-r)(0.16+0.512r).$$

Thus we obtain $mm_\rho = 0.11392$, $mm_{\widehat{\rho}} = 0.208$, $mm_{\rho,\kappa} = mm_{0.4\rho} > 0.22$, $mm_{\widehat{\rho},\kappa} = mm_{0.4\widehat{\rho}} < 0.21$, which verifies the thesis. □

Appendix EC.8: Increasing-with-Potential RVPs

Proof of Lemma 5. Directly from Lemma EC.2. □

Proof of Theorem 4. We claim that there is $\delta > 0$ with $\delta < \theta(1-\beta)$ such that the following properties hold for all $t \in I := (\theta-\delta, \theta+\delta)$: ρ is continuous and differentiable at t; ρ is monotonically decreasing at t; and $0 < \rho(t) \le (1+\rho(\theta))/2$. An interval that satisfies the first two properties exists since ρ is continuously differentiable in some neighborhood of θ and has strictly negative derivative. The third property holds since ρ has a strictly negative derivative at θ, so it must be strictly bounded away from 0 and 1 itself. Then, one can restrict δ to guarantee the same for t close to θ. Note also that $I \subset (\beta\theta, \theta/\beta)$.

Next, fix $\varepsilon > 0$, then one can construct a distribution f that satisfies the following conditions: f is continuous and differentiable everywhere; $f(\theta) = \varepsilon$; $f(t) = 0$ for $t \notin I$; and $\int_{\theta}^{\theta+\delta} f(t)\,dt \geq \frac{1}{2}$. This can be done for instance by constructing a piece-wise constant function that satisfies all but the first condition, then smoothing it out with an appropriate bump function via standard techniques.

From (EC.5) and $p(1-\rho(\theta))(1-2\rho(\theta))+\rho(\theta) \in [0,1]$, we compute $\mu'_\rho(\theta) \geq \frac{1}{4}(-\rho'(\theta))p(1-\rho(\theta)) - \varepsilon$ (see the ArXiv version for detailed calculations). The first term in the right-hand side is strictly positive, and we can freely choose ε strictly smaller in magnitude to get $\mu'_\rho(\theta) > 0$. Note that although ρ might not be well-behaved everywhere, it is well behaved on I, and we can apply Lemma EC.2 to this point to get that ρ is not incentive compatible for θ, completing the proof. □

Appendix EC.9: Alternate Models of Bias

We now investigate deviations from the model studied so far, which assumes that disadvantaged students suffer from a uniform multiplicative bias.

Budget	Multiplicative			Additive			Difference	
$\widehat{c}$	PAUC	MM	Difference	PAUC	MM	Difference	PAUC	MM
0.1	44.4%	45.2%	0.9%	45.4%	46.1%	0.7%	1.1%	0.9%
0.2	37.6%	39.3%	1.7%	39.7%	41.1%	1.4%	2.1%	1.8%
0.3	30.8%	33.3%	2.5%	34.1%	35.9%	1.8%	3.3%	2.6%
0.4	26.4%	28.5%	2.0%	30.2%	31.8%	1.6%	3.7%	3.3%
0.5	22.0%	23.7%	1.7%	26.0%	27.4%	1.4%	4.0%	3.7%
0.6	17.6%	19.0%	1.4%	21.6%	22.8%	1.2%	4.0%	3.9%
0.7	13.2%	14.2%	1.0%	17.0%	18.0%	0.9%	3.8%	3.7%
0.8	8.8%	9.5%	0.7%	12.1%	12.7%	0.7%	3.3%	3.3%

Table EC.2 Proportion of disadvantaged students above theoretically optimal debiasing ranges under multiplicative/additive models and PAUC/maximum mistreatment aggregate mistreatment measures for $\alpha = 3$ and $p = 1/4$, with $\beta = 0.8$ for multiplicative models and $\gamma = 0.252$ for additive. For instance, the first value, 44.4% indicates that under budget $\widehat{c} = 0.1$ and the multiplicative model, the top-44.4% to 54.4% of students were debiased.

EC.9.1. An additive bias model

We first extend some of the theoretical results in Section 4 to the case of additive (in place of multiplicative) bias. We consider the same setting as Section 2 where the multiplicative bias model was introduced, but now the perceived potential of a student is given by $\widehat{Z}(\theta) = Z(\theta) - \gamma$ if $\theta \in G_2$ (and θ does not receive a voucher), with $\widehat{Z}(\theta) = Z(\theta)$ otherwise. This

then gives $F_1(t) = 1 - t^{-\alpha}$ and $F_2(t) = 1 - (t+\gamma)^{-\alpha}$, with domains $[1, \infty)$ and $[1-\gamma, \infty)$, respectively. The expression given in (1) for the biased matching $\widehat{\mu}(\theta)$ continues to hold under the additive definition of $\widehat{Z}$, and this fact yields the following result, which is an analogue of Proposition 1.

PROPOSITION EC.3. *Under additive bias of $\gamma \geq 0$, for any student $\theta \in G_2$, the displacement under $\widehat{\mu}$ is given by:*

$$\text{disp}_{\widehat{\mu}}(\theta) = \begin{cases} (1-p)\left((Z(\theta)-\gamma)^{-\alpha} - (Z(\theta))^{-\alpha}\right), & \text{if } Z(\theta) \geq 1+\gamma, \\ (1-p)\left(1 - (Z(\theta))^{-\alpha}\right), & \text{if } Z(\theta) < 1+\gamma. \end{cases} \tag{EC.8}$$

For any student $\theta \in G_1$, we have $\text{disp}_{\widehat{\mu}}(\theta) = -p((Z(\theta))^{-\alpha} - (Z(\theta)+\gamma)^{-\alpha})$. Thus, the maximum displacement of $(1-p)(1-(1+\gamma)^{-\alpha})$ is experienced by a G_2 student with potential $1+\gamma$; and the most significant negative displacement of $-p(1-(1+\gamma)^{-\alpha})$ is experienced by a G_1 student with potential 1.

We next extend Theorem 1 which establishes the optimal debiasing interval under maximum mistreatment to the additive model. Let $\mathcal{S}^c(\widehat{c})$ be the set of closed and connected subsets of $[1, \infty)$ such that for $S \in \mathcal{S}^c(\widehat{c})$, $\int_S dF_1 \leq \widehat{c}$. Similarly to the multiplicative case, we define μ_S to be the matching under the additive model when students in S receive vouchers. Let then $\mathcal{S}^c_{mm}(\widehat{c}) = \arg\min_{S \in \mathcal{S}^c(\widehat{c})} mm(\mu_S)$ be the collection of sets in $\mathcal{S}^c(\widehat{c})$ that minimize maximum mistreatment. The following result is analogous to Theorem 1.

THEOREM EC.1. *The set $\mathcal{S}^c_{mm}(\widehat{c})$ consists of a unique set $S = [Y_1^*, Y_2^*]$ where Y_1^* and Y_2^* are computed as follow: $Y_1^* = (\widehat{c} + (Y_2^*)^{-\alpha})^{-1/\alpha}$ and $Y_2^* = \min\{U_1, U_2\}$, where U_1 is the positive solution to $(1-p)(1-\widehat{c}) = \mathbb{P}(F \geq U_1 - \gamma) - p\mathbb{P}(F \geq U_1)$ and U_2 is the positive solution to $(1-p)(1-\widehat{c}) = (1-p)\mathbb{P}(F \geq U_2 - \gamma) + p\widehat{c}$.*

Proof. Recall $F \sim \text{Pareto}(1, \alpha)$ is the distribution of true potentials of all students, and let ϕ be a *bias map* that gives the perceived potential of a disadvantaged student given their true potential. ϕ therefore encodes information both about the nature of bias, as well as any debiasing steps taken. In this case we consider an additive bias model where we debias an interval $S = [Y_1, Y_2]$, so that $\phi(x) = x$ if $x \in S$ and $\phi(x) = x - \gamma$ otherwise. Let $H \sim \phi(F)$ be the distribution of perceived potentials of disadvantaged students, and then note that $(1-p)F + pH$ is the distribution of perceived potentials of all students in aggregate in the presence of bias and any debiasing. In the fair matching (where ϕ is the identity map), a

disadvantaged student is matched to school with rank $\mu(x) = \mathbb{P}(F \geq x)$, and in the ranking with bias and debiasing of S, they are matched to $\mu_\phi(x) = \mathbb{P}((1-p)F + pH \geq \phi(x))$. We therefore have under S the displacement

$$\mathrm{disp}_S(x) = \mu_\phi(x) - \mu(x) = \begin{cases} p\left[\mathbb{P}(F \leq x) - \mathbb{P}(\phi(F) \leq x)\right], & x \in S, \\ \mathbb{P}(F \leq x) - \left[p\mathbb{P}(\phi(F) \leq x - \gamma) + (1-p)\mathbb{P}(F \leq x - \gamma)\right], & x \notin S. \end{cases}$$

By carefully dividing into the cases where $Y_2 - Y_1 < \gamma$ and $Y_2 - Y_1 \geq \gamma$, one can show

$$\mathrm{disp}_S(x) = (1-p)\mathbb{P}(F \in [x-\gamma, x])\,\mathbb{1}_{\{x \notin S\}} + \begin{cases} -p\mathbb{P}(F \in [Y_2, \max\{x+\gamma, Y_2\}]), & x \in (Y_1, Y_2), \\ p\mathbb{P}(F \in [\max\{x-\gamma, Y_1\}, Y_2]), & x \in [Y_2, Y_2+\gamma). \end{cases} \tag{EC.9}$$

Let $S = [Y_1, Y_2]$ be some interval to debias, then (EC.9) implies that $m_\emptyset(x) = m_S(x)$ for $x \notin (Y_1, Y_2+\gamma)$, $m_S(x) = 0$ for $x \in (Y_1, Y_2)$, $m_S(x) \geq m_\emptyset(x)$ for $x \in [Y_2, Y_2+\gamma)$, and that $m_S(x)$ is decreasing for $x \geq \max\{1+\gamma, Y_2\}$.

Further, if $1+\gamma \notin S$ then $\max_x m_S(x) \geq \max_x m_\emptyset(x)$.We therefore know that the optimal $S \in \mathcal{S}^c(\widehat{c})$ contains $1+\gamma$ and that it minimizes $\max\{m_S(Y_1), m_S(Y_2)\}$. Now write

$$\begin{aligned} m_S(Y_1) &= (1-p)\mathbb{P}(F \in [Y_1-\gamma, Y_1]) = (1-p)\mathbb{P}(F \leq Y_1), \\ m_S(Y_2) &= (1-p)\mathbb{P}(F \in [Y_2-\gamma, Y_2]) + p\mathbb{P}(F \in [\max\{Y_2-\gamma, Y_1\}, Y_2]). \end{aligned}$$

Recall that $\mathbb{P}(F \in [Y_1, Y_2]) = \widehat{c}$. We need $m_S(Y_1) = m_S(Y_2)$, so write

$$\begin{aligned} &(1-p)\mathbb{P}(F \leq Y_1) = (1-p)\mathbb{P}(F \in [Y_2-\gamma, Y_2]) + p\mathbb{P}(F \in [\max\{Y_2-\gamma, Y_1\}, Y_2]) \\ \iff &(1-p)(1-\widehat{c}) = \min\left\{\mathbb{P}(F \geq Y_2-\gamma) - p\mathbb{P}(F \geq Y_2), (1-p)\mathbb{P}(F \geq Y_2-\gamma) + p\widehat{c}\right\}. \end{aligned}$$

Since both terms inside the minimum are decreasing in Y_2, the Y_2 that solves this equation is given by $\min\{U_1, U_2\}$ where U_1 solves $(1-p)(1-\widehat{c}) = \mathbb{P}(F \geq U_1 - \gamma) - p\mathbb{P}(F \geq U_1)$, and U_2 solves $(1-p)(1-\widehat{c}) = (1-p)\mathbb{P}(F \geq U_2 - \gamma) + p\widehat{c}$. These are exactly the expressions sought for in the theorem. □

Model	**Metric**	**Value of Metric**	**Beta fit β**
Additive	KL-divergence	0.0321	-36.9
Multiplicative	KL-divergence	**0.0293**	0.902
Additive	Wasserstein distance	9.52	-49.0
Multiplicative	Wasserstein distance	**5.36**	0.882

Table EC.3 Comparison of Additive and Multiplicative Models using KL-divergence and Wasserstein Distance Metrics

We remark that it is possible to similarly compute the optimal DDS under the PAUC measure. In the proof of Theorem EC.1 we gave an expression for the mistreatment of a given student, and one can then integrate this against F_1 to compute PAUC. One easily argues that this expression has a global minimum, but the resultant expressions do not seem to be amenable to analytical computations with a clean final statement.

We close by noting that we fitted both the multiplicative as well as the additive model to our real data, and as measured by both Wasserstein distance and KL-divergence, the multiplicative model finds a slightly better fit, leading to us focusing on this model (see Table EC.3).

EC.9.2. Impact of model misspecification

We next study the robustness of our framework under model misspecification. That is, we investigate the impact of applying our simple model of constant multiplicative bias when the true process by which bias arises is more complicated. In particular, we study additive models and models where idiosyncratic randomness exists within the bias factor or potentials of disadvantaged students. Based on computational experiments we show that our main takeaway holds, and that applying our results would lead to little efficiency loss except in the case of very high randomness.

Setup: We generate simulated data with parameters chosen to match those we fit to our real data. In the language of Section 2, all students $\theta \in \Theta$ have a true potential $Z(\theta)$ sampled i.i.d. from a Pareto$(1, \alpha)$ distribution with $\alpha = 9$, and we identify students with their potential, so we write θ for their true potential. A proportion $p = 0.3$ of students is disadvantaged and they appear at a perceived potential $\widehat{Z}(\theta)$, where $\widehat{Z}$ is some random variable with $\widehat{Z}(\theta) \leq \theta$. We study various models for $\widehat{Z}(\cdot)$.

The central planner has a budget $\widehat{c}$ and applies our model with $\widehat{Z}(\theta) = \beta\theta$ to choose an interval[19] of disadvantaged students to debias as instructed by Theorem 2. We call this the *theoretical* debias interval. Because of randomness, it does not make sense to measure the maximum mistreatment, so we concentrate solely on the positive area under the mistreatment curve (PAUC). We then compare the theoretical interval to the optimal empirical interval if the full bias process were known to the central planner a priori. We compute such an interval using grid search, which we call the *empirical* debias interval.

[19] We assume the central planner cannot observe the true potentials and must therefore debias disadvantaged students chosen based on perceived potentials. For the case of uniform multiplicative bias, this makes no difference, but when the bias process has randomness, this is an important detail.

Models: For each model, we let η be some fixed parameter. We report results for the cases where the true bias process takes each of the following forms:

1. $\widehat{Z}(\theta) = \theta - \eta$, a deterministic additive model;
2. $\widehat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Normal}(0, .02)$, minor Gaussian noise in bias factor;
3. $\widehat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Uniform}(-.05, .05)$, minor uniform noise in bias factor;
4. $\widehat{Z}(\theta) = \theta - \eta + \varepsilon$ for $\varepsilon \sim \text{Normal}(0, .1)$, additive, medium Gaussian noise in bias factor;
5. $\widehat{Z}(\theta) = \eta\theta + \varepsilon$ for $\varepsilon \sim \text{Normal}(0, .1)$, medium Gaussian noise in potential;
6. $\widehat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Uniform}(-.15, .15)$, medium uniform noise in bias factor;
7. $\widehat{Z}(\theta) = \eta\theta + \varepsilon$ for $\varepsilon \sim \text{Uniform}(-.3, .3)$, large uniform noise in potential; and
8. $\widehat{Z}(\theta) = (\eta + \varepsilon)\theta$ for $\varepsilon \sim \text{Uniform}(-.3, .3)$, large uniform noise in bias factor.

The first model in particular is the additive model studied in EC EC.9, and the fourth model is exactly the statistical discrimination model of Phelps (1972). All models except the first can be interpreted as adaptations of a statistical discrimination model, as they contain the key feature of increased variance of the disadvantaged group. We choose for each model the fixed parameter η in such a way that if the central planner would apply our theoretical model of constant multiplicative bias, they would fit exactly $\beta = 0.88$ as the Wasserstein metric minimizing parameter. This yields a set of experiments that can be readily compared. The best fit for the η parameter for each model is shown in Table EC.4.

Model	1	2	3	4	5	6	7	8
Parameter	0.130	0.878	0.876	0.151	0.860	0.857	0.843	0.848
Error	0.011	0.002	0.004	0.046	0.037	0.035	0.092	0.106

Table EC.4 Best-fits for the model parameter η. The error row shows the minimum achieved Wasserstein distance between the data under the assumed true bias distribution, and the simplified theoretical model (with Pareto parameters $\alpha = 9$ and $\beta = 0.88$).

Simulations: We perform experiments with two budgets, $\widehat{c} = 0.1$ and $\widehat{c} = 0.4$, and run simulations with 1 million students in order to adequately approximate the continuous market. Many of these models increase the variance of the distribution of disadvantaged student scores, so the theoretical interval would often debias a significantly smaller proportion of students than allowed by the budget, naturally leading to a lower PAUC reduction and making comparison difficult. Because of this phenomenon, we fix the upper endpoint of the theoretical interval, but choose the lower endpoint such that it fills up the budget.

Table EC.5 and Table EC.6 summarize the results for $\widehat{c} = 0.1$ and $\widehat{c} = 0.4$ respectively. For each model, we report the aggregate mistreatment as measured by the PAUC metric

(introduced in Section 4) for three cases: no debiasing, under the empirically optimal debiasing, and under the theoretically optimal debiasing. We report the reduction in PAUC given by both debiasing methods as well as their difference. Based on our computations, we conclude that our results are highly robust to model misspecification.

Model	PAUC			PAUC Reduction		Difference
	No debiasing	Empirical	Theoretical	Empirical	Theoretical	
1	0.2261	0.1870	0.1870	17.28%	17.28%	0.00%
2	0.2398	0.2009	0.2009	16.21%	16.20%	0.00%
3	0.2406	0.2029	0.2029	15.67%	15.65%	0.01%
4	0.2276	0.1978	0.1980	13.12%	13.00%	0.12%
5	0.2406	0.2105	0.2108	12.50%	12.39%	0.11%
6	0.2395	0.2138	0.2149	10.72%	10.27%	0.45%
7	0.2360	0.2048	0.2057	13.23%	12.82%	0.41%
8	0.2337	0.2033	0.2042	13.03%	12.64%	0.38%

Table EC.5 Comparison of PAUC reductions between theoretically and empirically optimal intervals for $\widehat{c}=0.1$.

Low Budget: In the small budget case ($\widehat{c}=0.1$, see Table EC.5), the difference in using the empirically optimal and the theoretically optimal debiasing intervals is minuscule in every case. In the case of largest difference in PAUC reduction, the empirical interval is able to reduce PAUC by 10.72%, whereas the theoretically optimal interval would have reduced it by 10.27%, a difference of only 0.45%.

Model	PAUC			PAUC Reduction		Difference
	No debiasing	Empirical	Theoretical	Empirical	Theoretical	
1	0.2261	0.0850	0.0850	62.39%	62.38%	0.00%
2	0.2398	0.0994	0.0994	58.56%	58.55%	0.01%
3	0.2406	0.1062	0.1064	55.86%	55.80%	0.06%
4	0.2276	0.1146	0.1194	49.66%	47.54%	2.12%
5	0.2406	0.1254	0.1311	47.87%	45.51%	2.36%
6	0.2395	0.1284	0.1402	46.37%	41.47%	4.90%
7	0.2360	0.1015	0.1202	56.99%	49.08%	7.92%
8	0.2337	0.0991	0.1187	57.62%	49.22%	8.40%

Table EC.6 Comparison of PAUC reductions between theoretically and empirically optimal intervals for $\widehat{c}=0.4$.

High Budget: For the large budget case ($\widehat{c}=0.4$, see Table EC.6), the difference is more pronounced, yet still limited. The difference is negligible for models 1–3. Medium Gaussian noise in either potential or bias factor causes a small difference in PAUC reduction, in the order of 2–2.5%. In the case of medium uniform noise in bias factor, the difference starts

being more noticeable at 4.9%, and with the cases of large uniform noise in potential or bias, the difference goes to 7.92% and 8.4% respectively.

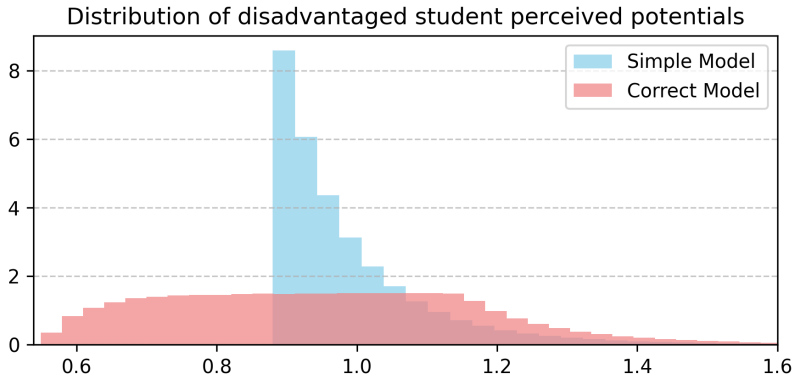

Figure EC.1 **Difference in disadvantaged student perceived potentials for model 8, comparing the naive model and the true model.**

Most Misspecified Case: In the worst case, the absolute difference in PAUC is 0.0196, meaning that on average, disadvantaged students in the correctly specified model achieve a ranking of approximately 2 percentage points higher than if the central planner were to apply the misspecified simpler model. We note that to achieve such a difference, the model has to be highly misspecified: for example, our model would predict the perceived potentials of disadvantaged students to have mean 0.99 and standard deviation 0.125, whereas the correct model in this case has mean 0.95 and standard deviation 0.231. One can see this difference in Figure EC.1, one would clearly observe that the model is not appropriate.

Additive Case: Our simple model has excellent performance under the case that the true model is any of models 1–3. In particular, in the case of the true model being additive, or the bias factor being slightly idiosyncratic, our model predictions apply virtually unchanged.

Incentive Compatibility: We assume in the section that the participants in the process, including the central planner, observe only perceived potentials and make decisions based on them. If the student knows only their perceived potential[20] and is able to manipulate it, then this case is no different to the treatment of incentive-compatibility in the main body. Voucher allocations that use a strict cutoff are always non-incentive-compatible.

[20] One could also consider the question of whether disadvantaged students would benefit from being able to manipulate their true potentials before the randomized bias process. This is certainly an interesting mathematical question and could easily be computed, but it is nonsensical in the present setup to assume that students are able to manipulate their true potential before the bias process takes place. One could imagine that the bias occurs in the testing process, but then randomness should be present for both disadvantages as well as non-disadvantaged students.

Fit to real data: To complement the synthetic comparison above, we also fit each model family directly to the real data from the case study in Section 6. Models 3, 6 and 8 above share the functional form $(\eta+\varepsilon)\theta$ with $\varepsilon \sim \text{Uniform}(-w, w)$, differing only in the value of w. We therefore group the eight models into six families, and report them alongside the deterministic multiplicative model $\widehat{Z}(\theta) = \eta\theta$ for reference. For each family we fit both the location parameter η and, where applicable, the noise width. As in the main body, we perform this fitting by assuming that the distribution of the true potentials of non-disadvantaged and disadvantaged students ought to match. We then apply the bias function on the scores of the non-disadvantaged student scores, and find the parameters that most closely match this distribution with that of the unmodified disadvantaged student scores[21]. Table EC.7 summarizes the best fits on both the full sample and on the truncated sample used in the main body[22]. Across all multiplicative families, the best-fit location parameter is essentially indistinguishable from the value of 0.88 assumed throughout the paper, with Wasserstein error within 0.001 of the best overall fit. Moreover, in the families where the noise term is added to the output, the best-fit noise width collapses to zero, so the data is not well explained by homoskedastic noise. This lends further empirical support to our choice of the simple multiplicative bias model in the main analysis.

Family	Model	Full sample			Truncated sample		
		η	w	Wass.	η	w	Wass.
$\widehat{Z}(\theta) = \eta\theta$	—	0.882	—	0.010	0.895	—	0.013
$\widehat{Z}(\theta) = \theta - \eta$	1	0.103	—	0.020	0.116	—	0.005
$\widehat{Z}(\theta) = (\eta+\varepsilon)\theta,\ \varepsilon \sim \text{Normal}(0, w)$	2	0.882	0.053	0.009	0.896	0.046	0.009
$\widehat{Z}(\theta) = (\eta+\varepsilon)\theta,\ \varepsilon \sim \text{Uniform}(-w, w)$	3,6,8	0.882	0.092	0.009	0.897	0.075	0.009
$\widehat{Z}(\theta) = \theta - \eta + \varepsilon,\ \varepsilon \sim \text{Normal}(0, w)$	4	0.104	0.001	0.020	0.116	0.009	0.005
$\widehat{Z}(\theta) = \eta\theta + \varepsilon,\ \varepsilon \sim \text{Normal}(0, w)$	5	0.882	0.001	0.010	0.895	0.046	0.010
$\widehat{Z}(\theta) = \eta\theta + \varepsilon,\ \varepsilon \sim \text{Uniform}(-w, w)$	7	0.882	0.001	0.010	0.896	0.071	0.010

Table EC.7 Best-fits of each family's location parameter (η) and noise width (σ or w) against the real data, for the full sample and for the truncated sample used in the main body, together with the corresponding Wasserstein error.

[21] This is therefore the reverse to how we fit in the main body, but is equivalent under the assumption. We have to perform it this way as the bias functions in this section include a noise component, which cannot be "undone".

[22] In the main body, we truncate the distributions to only students whose true potential under the $\beta = 0.88$ case would exceed the cutoff of 475, in order to produce a balanced market.