Affirmative Action in College Admissions: Theory and Estimation

Brent Hickman
University of Iowa
Affirmative Action (AA) in College Admissions

DEFINITION: preferential treatment of minorities in admissions decisions.

RATIONALE FOR AA: reduce academic race gaps.

1. The “achievement gap” — in 1996, median minority SAT score = 22\textsuperscript{nd} percentile of non-minorities

2. The “enrollment gap” — in 1996, minorities made up 17.65% of college freshmen but
   - only 11.04% in top quality quartile of colleges
   - 29.71% in bottom quartile

ADDITIONAL CONCERN: preserve incentives for academic achievement
EDUCATION DEBATE

How does AA affect incentives, behavior, and outcomes?

OPPONENTS OF AA:
Erodes incentives for students to exert effort in school; creates tradeoff between achievement and equality.

PROPONENTS OF AA:
Increases minority effort by overcoming discouragement effects; no tradeoff exists.

Evaluation of the relative merits of these arguments requires an economic model that can be empirically evaluated.
I develop a theoretical model capable of producing quantitative measurements of THREE POLICY OBJECTIVES:

(I) Narrow the achievement gap,

(II) Narrow the enrollment gap, and

(III) Promote (or at least preserve) academic achievement.

Also allowing for comparisons of alternative college admission policies. Specifically, I study THREE CANONICAL CLASSES:

- Color-Blind
- Quota (India)
- Admission Preference (US)
## Outline

- Theoretical framework
- Qualitative results
- Structural Estimation
- Estimation results/counterfactual policy experiments
College Admissions as a Bayesian Game

- **AGENTS**: $K$ heterogeneous students with privately-known study cost types: $\theta \in [\underline{\theta}, \bar{\theta}]$
  - Affluence, access to health-care, school quality, etc. $\Rightarrow \theta$

- **STRATEGIES**: achieve a score/grade $s \in \mathbb{R}_+$

- **COST OF ACHIEVEMENT**: incur a cost $C(s; \theta)$ to achieve $s$
  - $\frac{\partial C}{\partial s} > 0; \quad \frac{\partial C}{\partial \theta} > 0; \quad \frac{\partial^2 C}{\partial s^2} \geq 0; \quad \text{and} \quad \frac{\partial^2 C}{\partial s \partial \theta} \geq 0$.

- **PAYOFF OF ACHIEVEMENT**: $K$ distinct prizes (college seats), $P_K = \{p_k\}_{k=1}^K$
  - Prize rank ordering generated by $F_P(p)$
  - Payoff to type $\theta$, achieving $s$ and receiving prize $p(s)$:
    $\Pi(s; \theta) = p(s) - C(s; \theta)$. 
College Admissions as a Bayesian Game

- **DEMOGRAPHICS**: students belong to $\mathcal{M}$ or $\mathcal{N}$
  - $|\mathcal{M}| + |\mathcal{N}| = M + N = K$
  - For group $j$, $\theta \sim F_j(\theta)$, $j = \mathcal{M}, \mathcal{N}$

- **ALLOCATION MECHANISMS**:
  - **Color-Blind**: assortative matching of prizes and grades
  - **Quota**: reserve a representative set of prizes for group $\mathcal{M}$
    - Split game into two separate competitions
  - **Admission Preference**: Markup function for minority grades:
    - $\tilde{S}(s) \geq s$, $\tilde{S}'(s) > 0$ (very general)

- **SYMMETRIC EQUILIBRIUM**:
  - A set of achievement functions $\gamma_{\mathcal{M}}(\theta)$, $\gamma_{\mathcal{N}}(\theta)$
  - $\Rightarrow$ grade distributions, $G(s)$, $G_{\mathcal{M}}(s)$, $G_{\mathcal{N}}(s)$
Qualitative Results (Relative to Color-Blind)

Assume $F_M(\theta) < F_N(\theta)$

**QUOTA EQUILIBRIUM:**

- High performing (low-cost) minorities decrease effort
- Low performing (high-cost) minorities increase effort
- Opposite for non-minorities
- Zero enrollment gap (by design)
Qualitative Results (Relative to Color-Blind)

Assume $F_{\mathcal{M}}(\theta) < F_{\mathcal{N}}(\theta)$ and $C(s; \theta) = \theta s$

**Fixed Grade Markup** $\tilde{S}(s) = s + \Delta$ “Michigan rule”:

- All students decrease achievement
- Relative to $\mathcal{N}$ counterparts, $\mathcal{M}$ students decrease achievement by the amount of the markup
- No allocational effect among top colleges

The Intuition stems from how a markup alters the competitive interaction between groups $\mathcal{M}$ and $\mathcal{N}$.

**General Admission Preference** $\tilde{S}(s) = h(s)$:

- Can perform better than a Michigan rule
**Prize Data:**

- Sample of 1,314 colleges/universities
  - USN&WR institutional quality measure $\{Q_u\}_{u=1}^U$
  - Freshman enrollment by race (NCES): $\{M_u, N_u\}_{u=1}^U$

$\{Q_u, M_u, N_u\}_{u=1}^U$, characterize the sample of prizes

$$P_{K,K} = \{p_k\}_{k=1}^K = \left\{\{p_{ui}\}_{i=1}^{M_u+N_u}\right\}_{u=1}^U, \ p_{ui} = Q_u$$

$\Rightarrow F_P, F_{PM}, F_{PN}$
The Empirical Enrollment Gap

MINORITIES: Black, Hispanic, American Indian/Alaskan Native
NON-MINORITIES: White, Asian/Pacific Islander, other

Prize Allocation Distributions: Data

USN&WR QUALITY INDEX

Prize Distribution
Minority Allocation
Non-Minority Allocation

Empirical CDF
**Academic Achievement Data:**

- SAT scores & race for random sample of 92,153
- Normalize by effective zero achievement level (SAT score of 580) and drop last digit: $s \in \{0, 1, \ldots, 102\}$

$$\Rightarrow G_M(s), G_N(s)$$
Empirical Procedure

1. Measure AA practices in US college market
   - Recover grade markup $\tilde{S}(s)$ via GMM

2. Given Step 1, estimate $F_M$, $F_N$, and $C(s; \theta)$
   - Conditional on $C(s; \theta)$, recover $F_M$ and $F_N$ nonparametrically
   - Estimate cost curvature parameter via NLS

3. Given Step 2, Impose alternative admissions mechanisms on the model (i.e., Color-Blind and Quota) and evaluate changes to the stated policy objectives.
Estimating $\tilde{S}(s)$ via semiparametric GMM:

**Key Observation:** Under an admission preference, a Minority student scoring $s$ receives the same prize as a Non-Minority student scoring $\tilde{S}(s)$.

This provides a moment condition that forms the basis of a GMM estimator for $\tilde{S}$.

Intuitively, I back out the parameter values that satisfy the mapping

$$
\left(\hat{G}_M, \hat{G}_N\right) \xrightarrow{\tilde{S}(s;\Delta)} \left(\hat{F}_{PM}, \hat{F}_{PN}\right).
$$

Given a flexible specification $\tilde{S} = \Delta_0 + \Delta_1 s + \cdots + \Delta_I s^I$, I can avoid restricting the form of the markup, and the estimator reduces to a simple regression equation.
\[ \tilde{S}(s; \Delta) \] Estimation Results

Specification: \( \tilde{S} = \Delta_0 + \Delta_1 s \) (higher-order terms unimportant)

<table>
<thead>
<tr>
<th>( \hat{\Delta}_0 )</th>
<th>( \hat{\Delta}_1 )</th>
<th>( R^2 )</th>
<th>Avg. Grade Boost</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.4218</td>
<td>1.0917</td>
<td>0.99789</td>
<td>6.1611</td>
</tr>
</tbody>
</table>

(0.00277) \hspace{2cm} (0.00000199)

95% CI: 
\[ [3.3187, 3.5251] \hspace{2cm} [1.089, 1.0945] \]

Estimates corroborate and inform previous empirical work:

- Chung, Espenshade and Walling (2004, SSQ)
- Chung and Espenshade (2004, SSQ)
Type Distribution Estimation à la GPV:

In auction models, the equilibrium produces a mapping from types and type distributions into equilibrium actions; e.g.,

$$\theta, F_M, F_N \rightarrow s.$$

Unobserved $\rightarrow$ Observed.

Guerre, Perrigne and Vuong (2000, *Econometrica*) discovered that this mapping can be analytically inverted; e.g.,

$$\theta \leftarrow (s, G_M, G_N).$$

Unobserved $\leftarrow$ Observed.

**MAIN ADVANTAGE:** this method allows for recovery of $F_M$ and $F_N$ with no *a priori* distributional assumptions.
Cost Function Estimation via NLS:

**ASSUMPTION:** $C(s; \theta) = \theta e^{\alpha s}$, $\alpha > 0$.

For fixed $\alpha$, restricted GPV estimates of $\hat{F}_M(\theta; \alpha)$, $\hat{F}_N(\theta; \alpha)$ imply model-generated score distributions, $\hat{G}_M(s; \alpha)$, $\hat{G}_N(s; \alpha)$...

**NLS ESTIMATOR:**

$$\hat{\alpha} = \arg \min \left\{ \sum_{j=1}^{J} \left( \hat{G}_M(s_j; \alpha) - \hat{G}_M(s_j) \right)^2 + \left( \hat{G}_N(s_j; \alpha) - \hat{G}_N(s_j) \right)^2 \right\}$$

where $S = \{s_1, s_2, \ldots, s_J\}$ is the grid of observed grade values, and $\hat{G}_i$ is the Kaplan-Meier ECDF of SAT scores.
Model Fit:

MODEL FIT: Grade Distributions

MODEL FIT: Allocation Distributions
Type Distributions:

Private Cost Densities

Private Cost Distributions

PDF

CDF

PRIVATE COST TYPES

PRIVATE COST TYPES

BOOTSTRAP
COUNTERFACTUAL: Grade Distributions

Counterfactual Experiment: Group-Specific Grade Distributions

Grades

CDF

MINORITY (STATUS QUO)
NON-MINORITY (STATUS QUO)
MINORITY (QUOTA)
NON-MINORITY (QUOTA)
MINORITY (COLOR-BLIND)
NON-MINORITY (COLOR-BLIND)
COUNTERFACTUAL: Grade Distributions

Counterfactual Experiment:
Group-Specific Grade Distributions

CDF

MINORITY (STATUS QUO)
NON-MINORITY (STATUS QUO)
MINORITY (QUOTA)
NON-MINORITY (QUOTA)
MINORITY (COLOR-BLIND)
NON-MINORITY (COLOR-BLIND)
**Counterfactual: Enrollment Gap Changes**

**Table:** %-Changes in Enrollment, Relative to US Admission Preference

<table>
<thead>
<tr>
<th>Tier</th>
<th>Minorities Color-Blind</th>
<th>Non-Minorities Color-Blind</th>
<th>Quota</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-33.3</td>
<td>+4.3</td>
<td>+4.5</td>
</tr>
<tr>
<td></td>
<td>-24.8</td>
<td>+4.6</td>
<td>-1.2</td>
</tr>
<tr>
<td></td>
<td>+52.8</td>
<td>-6.9</td>
<td>+2.6</td>
</tr>
<tr>
<td></td>
<td>+14.3</td>
<td>+6.9</td>
<td>+3.9</td>
</tr>
<tr>
<td></td>
<td>-14.9</td>
<td>-42</td>
<td>+18.5</td>
</tr>
<tr>
<td></td>
<td>-42</td>
<td>+14.9</td>
<td>+18.5</td>
</tr>
</tbody>
</table>

**Tier:** I, II, III, IV
Admission Policy Comparisons

Table: %-Changes Relative to US Admission Preference

<table>
<thead>
<tr>
<th></th>
<th>Quantile:</th>
<th>10&lt;sup&gt;th&lt;/sup&gt;</th>
<th>25&lt;sup&gt;th&lt;/sup&gt;</th>
<th>Median</th>
<th>75&lt;sup&gt;th&lt;/sup&gt;</th>
<th>90&lt;sup&gt;th&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Achievement Gaps</strong></td>
<td>Color-Blind:</td>
<td>-2.5</td>
<td>+9.1</td>
<td>+14.1</td>
<td>+6.3</td>
<td>-11</td>
</tr>
<tr>
<td>(Objective I)</td>
<td>Quota:</td>
<td>-41.7</td>
<td>-29.8</td>
<td>-18.8</td>
<td>-1.2</td>
<td>+16.3</td>
</tr>
<tr>
<td><strong>Enrollment Gaps</strong></td>
<td>Color-Blind:</td>
<td>+56</td>
<td>+66.6</td>
<td>+80</td>
<td>+99.9</td>
<td>+106.2</td>
</tr>
<tr>
<td>(Objective II)</td>
<td>Quota:</td>
<td>-100</td>
<td>-100</td>
<td>-100</td>
<td>-100</td>
<td>-100</td>
</tr>
<tr>
<td><strong>Population Grades</strong></td>
<td>Color-Blind:</td>
<td>+4.5</td>
<td>+1.5</td>
<td>+1</td>
<td>+0.8</td>
<td>+0.9</td>
</tr>
<tr>
<td>(Objective III)</td>
<td>Quota:</td>
<td>+9.2</td>
<td>+4.2</td>
<td>+2.2</td>
<td>+1.9</td>
<td>+2</td>
</tr>
</tbody>
</table>

Key: Most preferred outcome, Least preferred outcome
Alternative Policy Proposal

Policy-Maker announces that SAT scores will be reassigned a value equal to their group percentile rank. This implies the following Admission Preference rule:

\[ \tilde{S}(s) = G_N^{-1}(G_M(s)) \]

**Properties:**

1. A grade markup function that mimics a Quota

2. Simple implementation: need only knowledge of grades, race

3. A self-adjusting grade markup rule
   - Equivalent to Color-Blind in a symmetric competition
THE END
Comparison of Alternative Admission Preference Rules

- COLOR-BLIND RULE (45° Line)
- STATUS-QUO ADMISSION PREFERENCE RULE
- QUOTA-EQUIVALENT ADMISSION PREFERENCE RULE
Competition Under a Grade Markup

With a grade markup function \( \tilde{S}(s) \), the relation between minority and non-minority achievement is summarized by

\[
C' \left[ s; \gamma_{\tilde{M}}^{-1}(s) \right] = C' \left[ \tilde{S}(s); \gamma_{\tilde{N}}^{-1}(\tilde{S}(s)) \right] \tilde{S}'(s).
\]

With linear costs, this reduces to

\[
\gamma_{\tilde{M}}^{-1}(s) = \gamma_{\tilde{N}}^{-1}(\tilde{S}(s)) \tilde{S}'(s),
\]

and with a Michigan rule, it becomes

\[
\gamma_{\tilde{M}}^{-1}(s) = \gamma_{\tilde{N}}^{-1}(s + \Delta),
\]
<table>
<thead>
<tr>
<th>Selectivity</th>
<th>Faculty Resources</th>
<th>Financial Resources</th>
</tr>
</thead>
<tbody>
<tr>
<td>acceptance rate</td>
<td>% FTIF with PhD or TD</td>
<td>ed. spend./student</td>
</tr>
<tr>
<td>yield</td>
<td>% PTIF</td>
<td>non-ed. spend./student</td>
</tr>
<tr>
<td>avg. SAT/ACT score</td>
<td>avg. faculty comp.</td>
<td></td>
</tr>
<tr>
<td>% FTF in top HS quartile</td>
<td>student/faculty ratio</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Retention</th>
<th>Alumni Satisfaction</th>
<th>Reputation</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg. graduation rate</td>
<td>alumni giving rate</td>
<td>Survey</td>
</tr>
<tr>
<td>freshman ret. rate</td>
<td></td>
<td>FTF enrollment by race</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(NCES)</td>
</tr>
</tbody>
</table>
USNWR METHODOLOGY:

1. Construct percentiles along each of 6 criteria for schools within each of 4 Carnegie classifications
   - NU, NLAC, RU, RLAC
2. Overall score is a convex combination of the various percentiles, with weights as follows:
   - Selectivity 15%
   - Faculty Resources 20%
   - Financial Resources 10%
   - Retention 25%
   - Alumni Satisfaction 5%
   - Academic Reputation 25%
Modification to USNWR method: Scoring all schools simultaneously

⇒ cannot include academic reputation score

• Spread remaining weight evenly among other criteria

• Due to high degree of correlation among the various quality measures, Estimates $\hat{F}_P$, $\hat{F}_{PM}$, and $\hat{F}_{PN}$ are remarkably robust to drastic changes among the weighting constants.
Definition of “Zero Achievement”:

Virtually impossible to score an actual zero...

ACT Score Distribution from Random Responding
(100,000 Simulations)

ACT Score Conditional on $S \leq 12$

ACT Score Conditional on $S \leq 13$

SAT Concordance study: $12 \Leftrightarrow [52, 58]$. 
Estimating $\tilde{S}(s)$ via Semiparametric GMM:

From the policy-maker’s perspective, allocations follow

$$
\pi_M(s) = F_p^{-1}\left[(1 - \mu)G_N(\tilde{S}(s)) + \mu G_M(s)\right], \quad \text{and}
$$

$$
\pi_N(s) = F_p^{-1}\left[(1 - \mu)G_N(s) + \mu G_M(\tilde{S}^{-1}(s))\right].
$$

**IMPORTANT OBSERVATION:** $\pi_M(s) = \pi_N(\tilde{S}(s)) \quad (*)$.

**DEFINITIONS:**

- Given a quantile rank $r \in (0, 1)$, define $s_N(r) \equiv G_N^{-1}(r)$
- Define $r_M(r) \equiv G_M\left(\tilde{S}^{-1}(s_N(r))\right)$ as the quantile rank of *de-subsidized* $s_N(r)$ in the minority score distribution.
Estimating $\tilde{S}(s)$ via Semiparametric GMM:

\[
(*) \implies F_{P_M}^{-1}(r_M[r]) = F_{P_N}^{-1}(r) \\
\implies G_M\left(\tilde{S}^{-1}[s_N(r)]\right) = F_{P_M}\left(F_{P_N}^{-1}[r]\right) \\
\implies G_N^{-1}(r) = \tilde{S}\left(G_M^{-1}\left[F_{P_M}\left(F_{P_N}^{-1}[r]\right)\right];\Delta\right),
\]

**GMM Estimator:**

**Step 1:** Choose $r = \{r_u\}_{u=1}^U$, where $r_u = \hat{F}_P^{-1}(p_u)$, $u = 1,\ldots,U$.

**Step 2:** $\hat{\Delta} = \arg\min \left\{ \sum_{u=1}^U \left[\hat{G}_N^{-1}(r_u) - \tilde{S}\left(\hat{G}_M^{-1}\left[\hat{F}_{P_M}\left(\hat{F}_{P_N}^{-1}[r_u]\right)\right];\Delta\right)\right]^2 \right\}$,

where $\hat{G}_M$, $\hat{G}_N$, $\hat{F}_{P_M}$, and $\hat{F}_{P_N}$ are the Kaplan-Meier ECDFs of $S_M, T_M$, $S_N, T_N$, $P_M, M$, and $P_N, N$, respectively, and inverses are computed via nearest-neighbor interpolation.
Type Distribution Estimation à la GPV:

Non-minority FOC:

\[- \frac{(1-\mu)f_N[\gamma_N^{-1}(s)](\gamma_N^{-1})'(s) + \mu f_M[\gamma_M^{-1}(\tilde{S}^{-1}(s))](\gamma_M^{-1})'(\tilde{S}^{-1}(s))\frac{d\tilde{S}^{-1}(s)}{ds}}{f_P[F_P^{-1}[1 - (1-\mu)F_N[\gamma_N^{-1}(s)] + \mu F_M[\gamma_M^{-1}(\tilde{S}^{-1}(s))]]]} = \mathcal{C}'(s; \gamma_N^{-1}(s))\]

Recall that $S_M \sim G_M(s) = 1 - F_M[\gamma_M^{-1}(s)]$.

Thus, subsidized minority test scores are distributed

$\tilde{S}(S_M) \sim \tilde{G}_M(s) = G_M(\tilde{S}^{-1}(s)) = 1 - F_M[\gamma_M^{-1}(\tilde{S}^{-1}(s))]$. 
Type Distribution Estimation à la GPV:

Following Guerre, Perrigne and Vuong (2000, *Econometrica*), the non-minority FOC can be re-written as

\[
(1 - \mu)\hat{g}_N(s) + \mu\hat{g}_M(s) \over \hat{f}_p \left[ \hat{F}_p^{-1} \left( (1 - \mu)\hat{G}_N(s) + \mu\hat{G}_M(s) \right) \right] = C' \left( s; \hat{\theta} \right),
\]

and similarly, the minority FOC can be re-written as

\[
(1 - \mu)\hat{g}_N(s) + \mu\hat{g}_M(s) \over \hat{f}_p \left[ \hat{F}_p^{-1} \left( (1 - \mu)\hat{G}_N(s) + \mu\hat{G}_M(s) \right) \right] = C' \left( s; \hat{\theta} \right),
\]

where \( \tilde{G}_N(s) = G_N\left( \tilde{S}(s) \right) \) is the distribution of *de-subsidized* non-minority test scores.
Pseudo-Private Types:
\[ \hat{\mu} = 0.17652 (0.000141); \hat{\alpha} = 0.054099 (0.001339) \]
Other Details

- **Demographic Parameter:**
  \[ \hat{\mu} = \frac{\sum_{u=1}^{U} M_u}{\sum_{u=1}^{U} (M_u + N_u)} \] using FTF enrollment

- **Distribution/Density Estimates:**
  \( \hat{f}_P, \hat{F}_P, \hat{g}_M, \hat{G}_M, \hat{g}_N, \hat{G}_N \)
  via **boundary-corrected kernel smoothing**

- **Asymptotic Properties:**
  Campo, Guerre, Perrigne and Vuong (2009) estimate an auction model with utility curvature. They prove asymptotic normality with convergence at the following rates:
  - Utility Curvature Parameter: \( N^{(R+1)/(2R+3)} \), \( R = \# \) of cts derivatives of the distribution of private information.
  - Private Type Distributions: optimal rate for non-parametric estimators
Bootstrapped 95% Confidence Bands for Type Densities (120Samples)

Bootstrapped 95% Confidence Bands for Type Distributions (120Samples)
BOUNDARY CORRECTED KERNEL SMOOTHING

GOAL: to estimate $\hat{f}_P(p)$, where $P$ is a random variable living on $[p, \bar{p}]$.

TECHNICAL CHALLENGE: kernel-smoothed density estimates are known to exhibit considerable bias near the extremes of the sample.

SOLUTION:

- Karunamuni and Zhang (2008)
- A boundary-corrected density estimator for data living on a bounded interval.
- Same bias and variance near boundaries as in the interior of the sample.
Model Fit for $\tilde{S} = \Delta_0 + \Delta_1 s$

Grade Transformation Function:
Data versus Estimation

NORMALIZED SCORES
SUBSIDIZED SCORE

NORMALIZED SCORES

0 10 20 30 40 50 60 70 80 90
0 10 20 30 40 50 60 70 80 90

AFFINE GRADE TRANSFORMATION
45° LINE
DATA