Catch up, Mismatch and Learning: Some Evidence from India and Turkey

Kala Krishna
The Pennsylvania State University, NBER, CES-IFO and IGC

(with Veronica Frisancho, IDB)

July 20, 2012
Motivation

- Affirmative action (AA) policies used to advance disadvantaged
- In India, reservations for SC/ST/OBC
- In US, AA or 10% rules
- Intense debate over impact of such policies
- Shift from race-based to economic-based AA has been suggested
Plan

- Work on AA
- India: Targeting, Catchup and Mismatch in an EEI
  - Data
  - Results
- Work on Learning
- Turkey: Learning among repeat takers
  - Data
  - Results
- Implications for policy
Supreme Court to rule on this in September 2012

- Fisher vs. University of Texas
- Texas top 10 percent law response to 1996 ruling banning affirmative action in Texas’ public universities.
- In 2003, U.S. Supreme Court approved “points” for diverse student body. In response, the UT reinstated affirmative action on top of 10 percent law. Abigail Noel Fisher was denied admission and is white.
Rothstein and Yoon’s (2008, 2009): Find better school, lower graduation and bar passage rates for black students in the lowest quintile of admission credentials.

Sander (2004); Average performance gap between blacks and whites at selective law schools is large and gets larger over years, lower probability of graduation, mostly due to reduced grades.
THEOP (Texas Higher Education Opportunity Project)

- Top 10% public high schools in Texas are automatically granted admission to the University of Texas
- Do not control for the difficulty of the courses taken.
- If those admitted under the 10% rule take easy courses they may be more likely to graduate and get better GPAs, which need not imply better performance or outcomes.
Arcidiacono et al. (2011): Shows that although the GPA gap between white and black students at Duke University falls by half between the first and the last year of college, this comes primarily from smaller variance in grading during later years and a higher proportion of black students switching into easier majors.
Bertrand et al. (2010) follow a pool of applicants to Indian engineering colleges in 1996 in one state in India.

- Targeting exists as general displaced students richer than SC/ST displacing ones.
- Evidence against strong mismatch: wage increase after attending engineering college.
- But gains displacing < losses displaced.
- Problems: no grades.
Our Work on Affirmative Action in India

- We use data on class from an EEI in India
- Fertile ground to evaluate AA policies
  - Transparent admission criteria, extreme preferences, rigid course structure
  - Reservations: SC/ST take the JEE with general students but 15% and 7.5% of the seats in each major are reserved for SC and ST students, respectively.
  - Extreme: 97% (GE) vs 50% (SC/ST) and not able to fill.
  - Thus, gap between SC/ST and their non minority peers is greater in selective majors
Main Finding

- **Targeting:** AA effectively targets SC/ST students who are poorer than average displaced general student.

- **Catchup:** SC/ST students, *fall behind* their same-major peers, more so in selective majors.

- **Mismatch:** SC/ST students in more selective majors earn less than if they would have earned in less selective majors. They also experience higher stress levels and feelings of not belonging.
Main Finding

- 453 students graduating in 2008. 356 Bachelor and 97 Dual Degree/Integrated Master

- Three sources of data:
  - Institutional records: GPA and credits by semester, gender, caste, age, and major.
  - Exit survey: schooling, parents’ education, family income, first wage after graduation, among others.
  - JEE applicant data 2009: 384,977 applicants for 8295 seats

- Limitations:
  - No data on JEE scores: We use first year CGPA
  - Missing data: Multiple random imputation methods
  - No data on applicants
Displacing minority students come from poorer districts than displaced general students

SC/ST Worst prepared
Results: Caste, Income, Performance

(a) General Category

- Income, grades and caste:
  - For GE, income and grades slightly positively correlated.
  - For SC/ST: rich SC/ST look like GE.
Targeting

(a) General Category

(b) SC/ST

Creamy layer not an issue: few rich and look like GE
Catch Up?

**Catchup:** Gap in average CGPA between general and SC/ST students shrinks by 15%

**Grading differs by major?**
Cовп? 

(a) GE 

(b) SC/ST 

■ Distribution of grades for SC/ST students “improves” over time
Grades relative to the major

- Flatter curves for SC/ST, more so in selective majors
- Falling behind, not catching up
Do SC/ST gain in wages from going to selective majors?

Interval regression of wages on being in a selective major and $X$.

Let $w_i^*$ denote individual $i$’s wage at graduation:

$$w_i^* = \alpha_1 S_i + X_i \beta_1 - \epsilon_i$$

where $w_i = W_k$ if $\xi_{k-1} \leq w_i^* \leq \xi_k$

If $\alpha_1 \leq 0$, better off in less selective major: mismatch.

Better students in selective majors: need to control for selection

PSM: Control for observables in selection

But what about unobservables?
Labor Market Outcomes and Majors

- Joint estimation of selection and earnings: correlation in errors captures unobservables
- Model is represented by the system:

\[
\Pr(w_i = W_k | X_i) = \Phi(\alpha_1 S_i + X_i \beta_1 - \tilde{\zeta}_{k-1}; \sigma^2_\epsilon) \\
- \Phi(\alpha_1 S_i + X_i \beta_1 - \tilde{\zeta}_k; \sigma^2_\epsilon)
\]

\[
\Pr(S = 1 | Z_i) = \Phi(Z_i \gamma; \sigma^2_\mu)
\]

where \(X_i\) denotes observed individual characteristics and \(Z_i\) contains \(X_i\).
Run probit on being in selective major

Let $S_i^*$ be a latent continuous variable representing selectivity of $i$’s major:

$$S_i^* = Z_i \gamma - \mu_i$$

where $S_i = 1$ if $S_i^* > 0$ and $S_i = 0$ if $S_i^* < 0$.

Then run interval regression for wages

Let $w_i^*$ denote individual $i$’s wage at graduation:

$$w_i^* = \alpha_1 S_i + X_i \beta_1 - \epsilon_i$$

where $w_i = W_k$ if $\xi_{k-1} \leq w_i^* \leq \xi_k$

If $\alpha_1 \leq 0$, better off in less selective major: mismatch
**Mismatch: Results**

**Table: Effect of Attending a Selective Major at EEI on Wages ($\hat{\alpha}_1$)**

<table>
<thead>
<tr>
<th>Interval Regression</th>
<th>PS</th>
<th>Joint Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>GE</td>
<td>0.185***</td>
<td>0.176***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>SC/ST</td>
<td>0.055</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.085)</td>
<td>(0.093)</td>
</tr>
</tbody>
</table>

$\rho = 0.196$

$\rho = 0.913$

- Estimates fall as Control for observables (PSM) and unobservables (JE).
Let $y_i$ denote an aspect of individual $i$’s well being, stress, so positive coefficient is bad...:

$$y_i^* = \alpha_2 S_i + X_i \beta_2 - v_i$$

where $y_i = 1$ if $y_i^* > 0$ and $y_i = 0$ if $y_i^* < 0$.

Now, the system becomes:

$$\Pr(y_i = 1|X_i) = \Phi(\alpha_2 S_i + X_i \beta_2; \sigma_v^2)$$
$$\Pr(S = 1|Z_i) = \Phi(Z_i \gamma; \sigma_\mu^2)$$

If $\alpha_2 \geq 0$, more stress from being in selective major. Mismatch.

Negative for GE, positive for SC/ST.

$\rho > 0$, Type A personalities, stressed out people, choose more selective majors
### Table: Effect of Attending a Selective Major at EEI on Emotional and Social Well-Being ($\hat{\alpha}_2$)

<table>
<thead>
<tr>
<th></th>
<th>Probit</th>
<th>PS</th>
<th>Joint Estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stress</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>0.112*</td>
<td>0.125*</td>
<td>-0.082***</td>
</tr>
<tr>
<td></td>
<td>(0.065)</td>
<td>(0.068)</td>
<td>(0.010)</td>
</tr>
<tr>
<td>SC/ST</td>
<td>0.351***</td>
<td>0.320**</td>
<td>0.115**</td>
</tr>
<tr>
<td></td>
<td>(0.120)</td>
<td>(0.137)</td>
<td>(0.046)</td>
</tr>
<tr>
<td><strong>Hostel feels like home</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>0.005</td>
<td>0.008</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.058)</td>
<td>(0.099)</td>
</tr>
<tr>
<td>SC/ST</td>
<td>-0.159</td>
<td>-0.088</td>
<td>-0.103**</td>
</tr>
<tr>
<td></td>
<td>(0.118)</td>
<td>(0.125)</td>
<td>(0.052)</td>
</tr>
</tbody>
</table>

\(\rho\) values:
- Stress: 0.85
- SC/ST: -0.34
- Hostel feels like home: -0.09
- SC/ST: 0.78
Evidence falling behind, labor market mismatch, and social mismatch.

Suggests such extreme preferences may not even be in the interest of SC/ST
Through the Looking-Glass-By Lewis Caroll:

"Well, in our country," said Alice, still panting a little, "you’d generally get to somewhere else — if you run very fast for a long time, as we’ve been doing."

"A slow sort of country!" said the Queen. "Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!"

Wrong environment to look for catchup and learning
Nathan and Camara (1998): 55% of the juniors taking the SAT in the US improved their scores as seniors while 35% of them experienced score reductions. The higher (lower) the student’s initial SAT score as a junior, the lower (higher) is the increase in her score when retaking as a senior.

Vigdor and Clotfelter (2003): use data on evolution of SAT scores over multiple attempts. Find more learning at bottom than top.

SATs not best place to look for learning

- Less important in admissions than ÖSS
- Less difficult than ÖSS. Compromises ability to distinguish at the high end
- Perfect (294 of 1.5m) or near to perfect scores (5683 of 1.5m) in the SATs.
What is the ÖSS?

- Every year, 1.5 million students take ÖSS in Turkey.
- Placement rate is about 30%
- 2/3 of the applicants are repeat takers
- Can look for learning as same exam
- If less privileged learn more on re-tries, potential for leveling playing field
Unique dataset on about 40,000 applicants each from three high school tracks (Science, Social Studies, Turkish/Math) in 2002 to estimate learning, in terms of *score improvement*, accounting for selection.

Four data sources:

- Application form: high school track, name and type of school, repeat taker
- Survey applicant: household income, parents’ education and occupation, family size, time and money spent on *dersanes*, number of previous attempts...
- High school records: GPA at graduation
- ÖSYM records: performance in the exam
Assumptions

Key feature: data on GPA and ÖSS score

Assumptions:

i) students know their own (unobserved to the econometrician) ability

ii) learning is draw from a distribution that varies with observables/unobservables

iii) performance is partly random

iv) system is in steady state..

Develop dynamic model of retaking where costs of retaking vary over time
Three Effects

- Learning: Students who retake can learn. This raises scores and shifts the distribution of retakers to the right.
- Selection: Students choose to retake/not. If worse students tend to retake, this effect will move the distribution of scores for retakers to the left.
- Composition: If above two occur differently across groups we could have a composition effect. For example, learning may be greater for the less privileged as they are further away from their “frontier”.

Composition Effect
Distribution of Scores by GPA quartiles: Science
Distribution of Scores by GPA quartiles: Turkish-Math

Quartile I

Quartile II

Quartile III

Quartile IV

1st 2nd 3rd 4th 5th or more
Learning Gains in ÖSS

- Challenging given lack of panel data.
- Repeat takers act as “Quasi Panel” given Steady State
- Measure cumulative learning between the first and the nth attempt.
Model

- High school GPA of $i$:

  \[ g_i = X_i \alpha_0 + \theta_i + \epsilon_{i0} \]

- Scores in the $n$th try:

  \[ s_{in} = X_i \alpha_1 + \sum_{k=2}^{n} \lambda_{ik} + \theta_i + \epsilon_{in} \]

$\lambda_{ik}$: marginal learning in $k$th try. Drawn from distributions that can depend on $X$, $\theta$

$X_i$: observed characteristics

$\theta_i$: unobserved ability

$\epsilon_{ik}$: exogenous shock

All $\epsilon_{ik}, \theta_i, X$ orthogonal to each other
Estimation Strategy: The Steady State Assumption

Cross section at time $t$

- 1st
- 2nd
- 3rd
- 4th
- 5+th

Cohort $t$

- Enroll
- Exit
- Retry

Equivalent groups

$n^{th}$ time takers are students from cohort $t - (n - 1)$

(e.g., 2nd time takers are students from cohort $(t - 1)$)
**Estimation Strategy**

**1st Time Takers**

\[ g_i = X_i \alpha_0 + \theta_i + \epsilon_{i0} \]
\[ s_{i1} = X_i \alpha_1 + \theta_i + \epsilon_{i1} \]

\[ \hat{\alpha}_0, \hat{\alpha}_1 \]

**nth Time Takers**

\[ g_i = X_i \alpha_0 + \theta_i + \epsilon_{i0} \]
\[ s_{in} = X_i \alpha_1 + \sum_{k=2}^{n} \lambda_{ik} + \theta_i + \epsilon_{in} \]

\[ g_i - X_i \hat{\alpha}_0 = \theta_i + \epsilon_{i0} \]
\[ s_{in} - X_i \hat{\alpha}_1 = \sum_{k=2}^{n} \lambda_{ik} + \theta_i + \epsilon_{in} \]

\[ E(g_i - X_i \hat{\alpha}_0 | N_i = n) = E(\theta_i | N_i = n) + E(\epsilon_{i0} | N_i = n) \]
\[ \approx E(\theta_i | N_i = n) \]
\[ = \hat{\theta}_n \]

\[ E(s_{in} - X_i \hat{\alpha}_1 - \hat{\theta}_n | N_i = n) = E \left( \sum_{k=2}^{n} \lambda_{ik} + \epsilon_{in} | N_i = n \right) \]
\[ = E \left( \sum_{k=2}^{n} \lambda_{ik} | N_i = n \right) \]
Estimation Strategy: Interpretation

- Estimate the cumulative learning effect between the 1\textsuperscript{st} and the \(n\)\textsuperscript{th} attempt: \(\sum_{k=2}^{n} \lambda_{ik}\).

- Since \(\lambda_{ik}\) varies across individuals \(\forall k\), mean learning for 2\textsuperscript{nd} time takers may be different from mean learning for 3\textsuperscript{rd} time takers in their second attempt.

- Therefore, difference between 3\textsuperscript{rd} time takers’ and 2\textsuperscript{nd} time takers’ cumulative learning should not be interpreted as marginal learning.
## Results: Mean Cumulative Learning

### Cumulative Learning in Points (100 scale)

<table>
<thead>
<tr>
<th>Track</th>
<th>$2^{nd}$</th>
<th>$3^{rd}$</th>
<th>$4^{th}$</th>
<th>$5^{th}+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>9.7</td>
<td>6.8</td>
<td>6.6</td>
<td>6.2</td>
</tr>
<tr>
<td></td>
<td>(0.09)</td>
<td>(0.11)</td>
<td>(0.13)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Social Studies</td>
<td>4.8</td>
<td>7.1</td>
<td>9.0</td>
<td>11.1</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td>(0.10)</td>
<td>(0.12)</td>
<td>(0.14)</td>
</tr>
<tr>
<td>Turkish-Math</td>
<td>6.7</td>
<td>7.1</td>
<td>7.9</td>
<td>9.5</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td>(0.09)</td>
<td>(0.15)</td>
<td>(0.23)</td>
</tr>
</tbody>
</table>

Note: Standard error of the mean in parenthesis.
Results: Overall

- Learning more the less privileged in family background, school type, Coaching expenditure, ...
## Results: Decomposition of Change in Mean Scores

<table>
<thead>
<tr>
<th>Attempts</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt;</th>
<th>3&lt;sup&gt;rd&lt;/sup&gt;</th>
<th>4&lt;sup&gt;th&lt;/sup&gt;</th>
<th>5&lt;sup&gt;th&lt;/sup&gt; +</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Science</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Diff. between $s_{in}$ and $s_{i1}$</td>
<td>-0.9</td>
<td>-5.0</td>
<td>-7.1</td>
<td>-8.9</td>
</tr>
<tr>
<td>Due to $X_i$</td>
<td>-3.6</td>
<td>-6.4</td>
<td>-8.0</td>
<td>-9.0</td>
</tr>
<tr>
<td>Due to $\theta$</td>
<td>-6.9</td>
<td>-5.5</td>
<td>-5.7</td>
<td>-6.1</td>
</tr>
<tr>
<td>Due to $\sum_{k=2}^{n} \lambda_{ik}$</td>
<td>9.7</td>
<td>6.8</td>
<td>6.6</td>
<td>6.2</td>
</tr>
<tr>
<td><strong>Social Studies</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Diff. between $s_{in}$ and $s_{i1}$</td>
<td>3.7</td>
<td>6.7</td>
<td>8.7</td>
<td>9.9</td>
</tr>
<tr>
<td>Due to $X_i$</td>
<td>-0.3</td>
<td>0.3</td>
<td>0.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Due to $\theta$</td>
<td>-0.8</td>
<td>-0.7</td>
<td>-0.8</td>
<td>-1.5</td>
</tr>
<tr>
<td>Due to $\sum_{k=2}^{n} \lambda_{ik}$</td>
<td>4.8</td>
<td>7.1</td>
<td>9.0</td>
<td>11.1</td>
</tr>
<tr>
<td><strong>Turkish-Math</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Diff. between $s_{in}$ and $s_{i1}$</td>
<td>1.7</td>
<td>2.1</td>
<td>2.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Due to $X_i$</td>
<td>-1.5</td>
<td>-2.0</td>
<td>-2.6</td>
<td>-3.1</td>
</tr>
<tr>
<td>Due to $\theta$</td>
<td>-3.6</td>
<td>-3.0</td>
<td>-3.2</td>
<td>-4.1</td>
</tr>
<tr>
<td>Due to $\sum_{k=2}^{n} \lambda_{ik}$</td>
<td>6.7</td>
<td>7.1</td>
<td>7.9</td>
<td>9.5</td>
</tr>
</tbody>
</table>
Results: Cum. Learning by School Type

- **a) 2nd Attempt**
- **b) 3rd Attempt**
- **c) 4th Attempt**
- **d) 5th Attempt**

Science | Social Studies | Turkish–Math

[Graphs showing learning outcomes by school type for each attempt.]
Results: Cum. Learning by Coaching Expenditures

- **a) 2nd Attempt**
- **b) 3rd Attempt**
- **c) 4th Attempt**
- **d) 5th Attempt**

![Graphs showing learning outcomes by prep school expenditures across different attempts.](image)

Legend:
- **Science**
- **Social Studies**
- **Turkish–Math**
Results: Cum. Learning by Father’s Education

a) 2nd Attempt

b) 3rd Attempt

c) 4th Attempt

d) 5th Attempt

- Science
- Social Studies
- Turkish–Math
Conclusion and Future Agenda

- Less advantaged students exhibit larger performance improvements as they keep retaking.
  - Open competitive exams that can be taken more than once generate important learning gains for less advantaged students before they go to college.
- Though students in EEI fall back, this may be because of the moving goal post.
- Suggests two step approach
  - Identify talented
  - Provide pre-entry services

may be more fruitful in leveling playing field
First time takers perform worse (better) in Social Studies and Turkish-Math (Science).
Composition effect likely to be important

- Across all subjects, less privileged learn more than more privileged
- Science has more privileged, then Turkish Math, then Social Studies
- Suggests least movement to right in retakers in Science and most in Social studies as seen