Measuring the Price of Discrimination with Data on Poker Games

May 29, 2013

Dr. Ingo Fiedler
1. Introduction

• Economic theory suggests that discrimination is price sensitive and money an equalizer
• This means that discrimination decreases the higher the private costs for the discriminator
• Economic reasoning: Discrimination comes at a price for the discriminator
  ➢ Example: an employer discriminates against potential employees with a non-white skin tone and thus may miss the best employees who then work for a competitor.
  ➢ Imperfect competition leads to rents and rents can be used to discriminate and they can be discrimination in itself. [Jomo, 2003].
  ➢ Competition reduces discrimination
→ Is money an equalizer? And by how much?
Forms of Discrimination [Bertrand et al. 2005 and Parsons et al. 2011]

- **Taste-based discrimination**: Discriminating against a person or group for personal reasons.
- **Statistical discrimination**: Discriminating because the membership of a certain group provides information about a relevant characteristic.
- **Implicit discrimination**: Unintentional and outside of the discriminator’s awareness.
- **Positive Discrimination**: Being disproportionately likely to engage with a member of a certain group given the prevalence of members of this group.
- **Negative Discrimination**: Being less likely to engage with a member of a certain group given its prevalence.
Brief Outline of Research Idea

- Analyzing whether poker players from a certain country are
  - significantly more likely to play against people from a certain country (positive discrimination)
  - significantly less likely to play against people from a certain country (negative discrimination)
- Control if and by how much the discrimination vanishes when stakes increase
  - **Main hypothesis A**: Positive discrimination decreases with increasing stakes.
  - **Main hypothesis B**: Negative discrimination decreases with increasing stakes.

→ Determining the price elasticity of discrimination
Data: Online Poker Database of the University of Hamburg (OPD-UHH)

• Database includes (inter alia):
  - data on 4.6 million poker player identities over 6 months
  - information on origin of players
  - information on limits played (stakes)
  - Data collected 2009/2010 (before Black Friday)
  - 10 minutes between data points

• Data used: Dataset used by other studies
  - analyzing playing habits [Fiedler 2012a, 2012b, 2012c; Wilcke 2013]
Data used

- One week of No Limit Heads up tables at Full Tilt Poker for stakes NL50 to NL20,000

<table>
<thead>
<tr>
<th>Limit</th>
<th>Number of observations</th>
<th>Number of player pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL 50</td>
<td>121,028</td>
<td>60,514</td>
</tr>
<tr>
<td>NL 100</td>
<td>61,842</td>
<td>30,921</td>
</tr>
<tr>
<td>NL 200</td>
<td>35,944</td>
<td>17,972</td>
</tr>
<tr>
<td>NL 400</td>
<td>17,864</td>
<td>9,932</td>
</tr>
<tr>
<td>NL 600</td>
<td>6,602</td>
<td>3,301</td>
</tr>
<tr>
<td>NL 1,000</td>
<td>3,674</td>
<td>1,837</td>
</tr>
<tr>
<td>NL 2,000</td>
<td>1,306</td>
<td>653</td>
</tr>
<tr>
<td>NL 5,000</td>
<td>2,812</td>
<td>1,406</td>
</tr>
<tr>
<td>NL 10,000</td>
<td>568</td>
<td>284</td>
</tr>
<tr>
<td>NL 20,000</td>
<td>328</td>
<td>164</td>
</tr>
</tbody>
</table>
Steps of the Data Analysis

- Data Extraction
  - MS SQL Server 2008 Enterprise

- Data Import
  - SPSS 18

- Data Editing
  - STATA SE

- Data Querying
  - STATA SE

- Descriptive Analysis
  - Excel 2010

- Statistical Analysis
  - SPSS 18
Definition of Variables, Symmetric Differences

- **Frequencies** of player observations from different countries *relative* to all observations
- **Expected absolute** number of observations for each country pair
- **Expected frequency** of a country pair *relative* to all observations
- **Ratio** between the observed and expected observations of a country pair
- The **difference** between the observed and the expected observations of a country pair and its relative complement

→ These figures show which country pairings are observed more often or less often than expected
Definition of Variables, Non-Symmetric Differences

- *Observed frequency* of a country pair relative to a country i (and not relative to all observations)

- *Expected frequency* of a country pair relative to a country i (Please note that this figure is different for the same player pair depending on which country is country i)

- *Difference* between the observed relative frequency of a country pair and its expected relative frequency relative to a country
Controlling for Time Differences

• Time differences between countries make it more or less likely to encounter players from the other country at the tables

• Interpretation of time differences:
  - bonus on the probability of playing against players from countries which have zero or little time difference
  - discount on the probability of playing against players from countries which have a high time difference

• Capturing time differences
  - Grouping countries with identical time differences
  - Determining effects via a regression on dummies
Results #1: Positive Discrimination was found

- No significant negative discrimination was found
- Significant positive discrimination was found in the so-called hom-bias (players from country A are more likely to play against other people from country A)
- Result of Wilcoxon signed-rank test:

<table>
<thead>
<tr>
<th>Negative ranks</th>
<th>Positive ranks</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>52</td>
<td>79</td>
</tr>
</tbody>
</table>

\[
Z = -2.629 \\
\text{Asymptotic significance} = 0.009
\]
Results #2: Positive Discrimination decreases in stakes

<table>
<thead>
<tr>
<th></th>
<th>NL50</th>
<th>NL100</th>
<th>NL200</th>
<th>NL400</th>
<th>NL600</th>
<th>NL1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative ranks</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>9</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Positive ranks</td>
<td>18</td>
<td>13</td>
<td>10</td>
<td>3</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>17</td>
<td>14</td>
<td>12</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>Z</td>
<td>-3.354</td>
<td>-1.681</td>
<td>-.973</td>
<td>-1.334</td>
<td>-.533</td>
<td>-1.153</td>
</tr>
<tr>
<td>Asymptotic Significance</td>
<td>.001</td>
<td>.093</td>
<td>.331</td>
<td>.182</td>
<td>.594</td>
<td>.249</td>
</tr>
</tbody>
</table>

Test: $x_{ii} - x_{ii}^a$
Interpretation #1

- People are willing to pay a premium to play with other players from their country
- That is remarkable because the online poker environment is purely virtual
- Is it because of the possibility of chatting in the same language?
  - Rather not, because:
    - people chat very rarely in online poker and
    - all languages except English are prohibited (words in foreign languages are made unrecognizable)
  - Also: Wilcoxon signed-rank test without English speaking countries found the same result (and not a stronger one)
Interpretation #2

• Players on the higher limits are not willing to pay the (higher) price for the good „discrimination“

→ Demand for discrimination decreases with an increasing price
→ Money is an equalizer in this setting

• Further analysis revealed a price elasticity of demand for discrimination of -0.25 (very inelastic)
Limitations

• Cases of country pairs with zero observations were excluded
• Stickiness/status-quo bias: players are more likely to stick to a table they are already sitting at than to switch tables during the phase of ten minutes
• Transfer of the results from empirical analysis to the estimate of a demand curve is limited
  ➢ too few data points/limits where a significant home bias is observed
  ➢ depends on the assumption that the price is increasing proportionally with stakes
Thank you for your kind attention!

Ingo.fiedler@uni-hamburg.de