A Critique of Nardinelli and Simon’s *Quarterly Journal of Economics* paper

Artie Zillante
UNC Charlotte
Prepared for: ECON 6901, Spring 2017
Objectives of this Talk

- Provide an overview and critique of Nardinelli and Simon (1990) Quarterly Journal of Economics (QJE)
- Discuss how the paper has (and can be) extended
- Discuss how your interests can lead to a potential topic
- Throughout I will discuss dos and don’ts of presenting
Nardinelli and Simon question

Does consumer discrimination exist in the market for baseball cards?

(Always tell the audience what question you are answering very early in the talk)
Findings

Evidence supports the hypothesis of consumer discrimination

Cards of black and Hispanic players sell for less than those of white players with comparable performance measures in many of their models

(Always give findings early in the talk when people are paying attention)
Baseball Card Market

- Picture cards of current (and sometimes former) players
- Biographical and/or statistical information on the back of the card
- Card manufacturers release card “sets” each year
- At times, sets were released in series throughout the calendar year
Front and back of a baseball card
A baseball card set
Exogenous factors that affect card price regardless of the player

- Quality/popularity of sets
- Scarcity of sets/series

Those factors affect the “common card” price of the set. The common card price is the minimum price for any card in the set. The higher quality, more popular, or more scarce the set is the higher the common card price will be.
Nardinelli and Simon Data

- **Dependent variable:** Use 1989 Beckett prices for cards from the 1970 Topps set
  - Use a natural log transformation of this price relative to the common card price

- **Independent variables**
  - Hitters: H, 2B, 3B, HR, BB, SB, AB, Seasons, Postseason games, Position, Race
  - Pitchers: W, L, SV, CG, ER, K, BB, IP, H, Postseason innings, Race
Critique: Why use these data?

- **Dependent variable**
  - Beckett price guide is the industry standard
  - Wanted a long enough lag so that they were studying players whose careers were complete
  - Use natural log (typically) because of fit
  - Use $\ln(\text{price/common})$ to control for scarcity of series

- **Independent variables:**
  - Performance measures are statistics with which most baseball fans would be familiar
    - Could have used a single composite index instead of distinct performance measures
  - Position variables to control for defense
  - Race variable – variable of interest
Critique: Why use this dependent variable?

- Beckett price guide is the industry standard
- Wanted a long enough lag so that they were studying players whose careers were complete
- Use natural log (typically) because of fit
- Use ln(price/common) to control for scarcity of series
Critique: Why use these independent variables?

- Performance measures are statistics with which most baseball fans would be familiar
  - Could have used a single composite index instead of distinct performance measures
- Position variables to control for defense
- Race variable – variable of interest
Critique: Where did they get the data?

- Player statistics – baseball encyclopedia (there are now downloadable databases)
- Card prices – price guide

But what about:

- Race variables – ???
- Position variables – ???

(Readers should know where they can locate the data so they can replicate the study if they choose)
### Summary Statistics

#### TABLE I

Percent Common Players and Common Player Prices (in Dollars)

<table>
<thead>
<tr>
<th>Player ID</th>
<th>Number of players</th>
<th>Percent common players</th>
<th>Common player price</th>
<th>Maximum card price</th>
<th>Mean card price</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Hitters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–132</td>
<td>61</td>
<td>72.1</td>
<td>0.20</td>
<td>18.00</td>
<td>0.538</td>
</tr>
<tr>
<td>133–263</td>
<td>67</td>
<td>73.1</td>
<td>0.25</td>
<td>35.00</td>
<td>1.095</td>
</tr>
<tr>
<td>264–459</td>
<td>93</td>
<td>64.5</td>
<td>0.30</td>
<td>18.00</td>
<td>0.813</td>
</tr>
<tr>
<td>460–546</td>
<td>34</td>
<td>64.7</td>
<td>0.35</td>
<td>18.00</td>
<td>1.221</td>
</tr>
<tr>
<td>547–633</td>
<td>44</td>
<td>65.9</td>
<td>0.60</td>
<td>75.00</td>
<td>3.210</td>
</tr>
<tr>
<td>634–720</td>
<td>45</td>
<td>71.7</td>
<td>1.50</td>
<td>75.00</td>
<td>4.054</td>
</tr>
<tr>
<td>B. Pitchers</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–132</td>
<td>41</td>
<td>78.0</td>
<td>0.20</td>
<td>2.50</td>
<td>0.328</td>
</tr>
<tr>
<td>133–263</td>
<td>43</td>
<td>83.7</td>
<td>0.25</td>
<td>11.00</td>
<td>0.721</td>
</tr>
<tr>
<td>264–459</td>
<td>63</td>
<td>81.0</td>
<td>0.30</td>
<td>25.00</td>
<td>0.852</td>
</tr>
<tr>
<td>460–546</td>
<td>28</td>
<td>75.0</td>
<td>0.35</td>
<td>4.50</td>
<td>0.639</td>
</tr>
<tr>
<td>547–633</td>
<td>29</td>
<td>79.3</td>
<td>0.60</td>
<td>1.50</td>
<td>1.062</td>
</tr>
<tr>
<td>634–720</td>
<td>29</td>
<td>85.7</td>
<td>1.50</td>
<td>45.00</td>
<td>3.160</td>
</tr>
</tbody>
</table>

*Note. Player ID is the Topps card number. The number of observations is smaller than the total number of cards because many cards are not of individual players. Percent common players is the fraction of players who sold at the common player price.*
Theoretical Model

- Fan utility, $V$, is determined by player characteristics
  - $V$ is increasing in player performance
  - $V$ may be a function of other factors, such as the player’s race
  - $V$ is unobserved, so use card price as a proxy
Critique: Theoretical Model

- Not very rigorous
- Most of the assumptions are sensible (what you are typically critiquing in a theoretical model are the assumptions, unless the authors have done the proof incorrectly)
- Main reason for its inclusion seems to be to motivate the choice of a Tobit model for the econometric model
Econometric Models

• Separate models for hitters and pitchers due to different performance measures
• Tobit models using $\ln(\text{price}/\text{common price})$ as the dependent variable
• Probit models using a binary equal to 1 if the player was a “Star”, 0 if not
  ◦ A “Star” is a player who had a card price above the common price
Critique: Why use these models?

- **Tobit model**
  - Used because there is a lower bound on card price, the common card price
  - Want to know how each of the independent variables affects card prices

- **Probit model**
  - Used as a control on the influence of very high card prices for some star players
  - Robustness check
Important point on model selection

You will have to determine which econometric models to use when conducting your research. This is why you have taken (and/or are taking) courses in econometrics!!!
Results – Tobit models

• For hitters:
  ◦ Using a single dummy variable for minorities, find a 10% premium for White players
  ◦ Using dummies for Black and Hispanic, found 6.4% and 17% reductions compared to White players

• For pitchers:
  ◦ Using a single dummy variable, find a 13% reduction for minorities
  ◦ Using distinct dummies for Black and Hispanic players, find 16% and 12% reductions
Critique: Tobit model results

- They really only discuss the race variable
  - May be because of reviewer comments
- They make many statements about differences in prices (even for the race variables) but these differences are NOT statistically significant
  - See this shortly
- I am not certain if the chi-square tests for equal coefficients are for only Black and Hispanic coefficients or all three races
  - Based on text on page 592 I believe it is just the minority coefficients
### Table III

**Log Price Tobit Regressions and Probit Regressions for Hitters**

<table>
<thead>
<tr>
<th>Estimation method</th>
<th>Dependent variable: <em>P - P_i</em></th>
<th>STAR</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NONWHITE</strong></td>
<td>-0.3188</td>
<td>-0.8389</td>
<td>2.4</td>
<td>2.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>BLACK</strong></td>
<td>-0.2029</td>
<td>-0.3892</td>
<td>(1.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HISPANIC</strong></td>
<td>-0.5516</td>
<td>-2.0647</td>
<td>(3.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HITS</strong></td>
<td>0.0050</td>
<td>0.0084</td>
<td>(2.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DOUBLES</strong></td>
<td>-0.0008</td>
<td>-0.0044</td>
<td>(1.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TRIPLES</strong></td>
<td>0.0021</td>
<td>0.0134</td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HOME RUNS</strong></td>
<td>0.0032</td>
<td>0.0084</td>
<td>(2.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>WALKS</strong></td>
<td>0.0044</td>
<td>-0.0013</td>
<td>(1.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>STOLEN BASES</strong></td>
<td>0.0002</td>
<td>-0.0012</td>
<td>(1.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>AT BATS</strong></td>
<td>-0.0004</td>
<td>-0.0017</td>
<td>(1.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SEASONS</strong></td>
<td>-0.1417</td>
<td>-0.0872</td>
<td>(2.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>POSTSEASON GAMES</strong></td>
<td>0.0262</td>
<td>0.0515</td>
<td>(1.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First base</strong></td>
<td>0.0597</td>
<td>0.1939</td>
<td>(2.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second base</strong></td>
<td>-0.0474</td>
<td>-0.2694</td>
<td>(2.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Third base</strong></td>
<td>0.4496</td>
<td>0.7988</td>
<td>(2.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shortstop</strong></td>
<td>-0.0839</td>
<td>-0.0203</td>
<td>(1.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Catcher</strong></td>
<td>0.1402</td>
<td>0.0670</td>
<td>(2.3)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CONSTANT</strong></td>
<td>-1.2845</td>
<td>-2.0219</td>
<td>(5.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SIGMA</strong></td>
<td>0.6445</td>
<td>0.6421</td>
<td>(5.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(14.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Log likelihood</strong></td>
<td>(Slopes = 0)</td>
<td>-224.05</td>
<td>-224.05</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(14.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* 334 observations. Asymptotic t-ratios are in parentheses.

*P - P_i* is the log card price minus the log price of the common player. STAR is the proportion of players with a card price greater than the price of the common player, that is, with *P - P_i*> 0.

**SIGMA** is the estimated standard error of the regression.
Results – Probit model

- For hitters:
  - Find similar results as in their Tobit models
- For pitchers:
  - Find similar results as in their Tobit models
Critique: Probit model results

- Better at discussing statistical significance of individual coefficients
- In the pitcher model, an important variable (wins) is now NOT statistically significant
  - If you are a baseball fan this should seem strange
Important points

- Make sure that the reader knows what tests you are conducting
- Do not hide from your results
  - Your job is to take the best data you can find and use the best methods to analyze the data
  - If a coefficient estimate is not statistically different than zero, what impact does that variable really have in your model?
- Do not oversell your results
General Critique

- They do well in controlling for some factors (like the card series)
- Dismiss other factors
  - Player’s team, in particular the team the player is pictured with on the 1970 Topps card
  - Discuss year of debut, but rookie (first year) cards are/were extremely important to collectors – do not control for these
    - Beckett lists 89 (49 individual player) RCs in 1970 Topps
  - At the time 2nd year cards were also important to collectors
In hindsight, the period from which the authors drew their data was a bubble, and prices for individual cards were much more variable than prices for sets.
Extensions

- Use data from other sports with measurable performance (football, basketball, etc.)
- Examine data from other sets and/or time periods
- Examine prices of “rookie” cards – if discrimination exists then one might think it would show up when there is very little known about the player
Examining rookie card prices

- We use the first ever publicly available price for cards released from 1986-1993
- We also use a 17 year later price for the exact same cards
- Use similar models to Nardinelli and Simon
  - Find cards of black players with comparable measurable variables are more highly priced at initial release
  - Find no difference using 17 later prices for comparable players
Look to your own interests for topics

- Many times students want to tackle “big” topics
  - Big topics can be daunting
  - Many people are working on big topics
- Smaller topics may prove more fruitful
  - May want to look at a particular industry/market in which you are interested
  - Perhaps compare across different industries/markets