The Ecosystem of Baseball

• Major league baseball (MLB) clubs develop minor league players (younger, less experienced) in a “farm system”

• Parent club has control of a drafted player for 7 seasons once he makes the major leagues
The Case

Diamond Dollars Case Competition
presented by

What is the Perfect Cole Hamels Trade?
“People in all fields operate with beliefs and biases. To the extent you can eliminate both and replace them with data, you gain a clear advantage.”

— Michael Lewis, *Moneyball: The Art of Winning an Unfair Game*
About Us

We’re three juniors, a sophomore and a freshman from varying educational backgrounds united by our love for baseball and joined by the Stanford Sports Analytics Club.
Our Approach

Three main questions:

• How do we project the performance of Hamels and traded prospects over the coming seasons?
• How do we quantify the improvement of a team upon the addition of Hamels?
• Can we develop a quantitative metric to evaluate the value of a trade to both the Phillies and their trade partner, and how do we optimize that metric?
Our Approach

Three main questions:

• How do we project the performance of Hamels and traded prospects over the coming seasons?
• How do we quantify the improvement of a team upon the addition of Hamels?
• Can we develop a quantitative metric to evaluate the value of a trade to both the Phillies and their trade partner, and how do we optimize that metric?
What Data?

Baseball stats from Fangraphs, baseball-reference.com:
- Wins Above Replacement (WAR)
- Weighted On-base Average (wOBA)
- Fielding-Independent Pitching (FIP)

Demographic information:
- Age
- Minor/Major League Level
- Service Time
Our Database

We have:

- Minor league player-seasons data (one long Excel file...)
- List of Baseball America Top-10 rated prospects for each trade partner team
- Age, career WAR, and service time for each player on each MLB team
Projecting Minor League Prospects

How do we project the future performance of minor league prospects?

In a trade for Cole Hamels, it is likely that the Phillies will want one or multiple “top” prospects in return.

For the sake of simplicity, we only considered the top 10 prospects on each team (according to Baseball America preseason rankings).
Accumulate a database of every minor league player's seasons from 2010-2014 (all levels, all players)

For every 2015 top-10 prospect… find similar historical minor league seasons (age, level, wOBA/FIP)

Narrow down by only considering similar seasons from former top-10 minor league prospects

For each of those old players, average MLB fWAR per season to predict future of our 2015 prospect

For every player with a “similar” former season, find career MLB seasons played, career MLB fWAR

Return projected career fWARs for every 2015 prospect
Projecting Minor League Prospects

As input, the program takes the list of top-10 prospects for a trade partner team.

The program finds comparable statistical seasons to each prospect’s 2014 season adjusted for **age** (i.e., 22) and **level** (i.e., AA).

- “Comparable” season defined to be +/- 5% of comparison statistic
- Batters compared using wOBA
- Pitchers compared using FIP
An Aside: Similarity Algorithms

- Jaccard similarity
- Cosine similarity
- K-nearest neighbors
- K-means clustering

Applications: Recommender systems, classification, & more
Our Approach

Three main questions:

• How do we project the performance of Hamels and traded prospects over the coming seasons?
• How do we quantify the improvement of a team upon the addition of Hamels?
• Can we develop a quantitative metric to evaluate the value of a trade to both the Phillies and their trade partners, and how do we optimize that metric?
Correlating Wins to Playoffs

Compiled every season from 1995-2014 and noted which teams made the playoffs, aggregated into logistic regression

\[
\text{Playoff probability} = \frac{1}{1 + e^{-(A + (B \times \text{wins})}}
\]

\[A = -60.25773\]
\[B = 0.69183\]

Observation: very short “sweet spot” where probability changes dramatically with wins
The Hamels Effect

1. **Cleveland Indians**
   - Before Hamels: 85.5 W / 24.6 % playoffs
   - With Hamels: 88.4 W / 71.2 % playoffs
   - Change: +46.6%

2. **Toronto Blue Jays**
   - Before Hamels: 86.2 W / 34.6 % playoffs
   - With Hamels: 89.2 W / 80.9 % playoffs
   - Change: +46.3%

3. **Detroit Tigers**
   - Before Hamels: 86.3 W / 36.2 % playoffs
   - With Hamels: 89.2 W / 80.6 % playoffs
   - Change: +44.4%

4. **New York Yankees**
   - Before Hamels: 85.2 W / 21.1 % playoffs
   - With Hamels: 87.7 W / 60.0 % playoffs
   - Change: +38.9%

5. **Seattle Mariners**
   - Before Hamels: 84.9 W / 15.9 % playoffs
   - With Hamels: 87.3 W / 53.7 % playoffs
   - Change: +37.8%

6. **Los Angeles Angels**
   - Before Hamels: 83.7 W / 8.6 % playoffs
   - With Hamels: 86.5 W / 40.5 % playoffs
   - Change: +31.9%
Our Approach

Three main questions:

• How do we project the performance of Hamels and traded prospects over the coming seasons?
• How do we quantify the improvement of a team upon the addition of Hamels?
• Can we develop a quantitative metric to evaluate the value of a trade to both the Phillies and their trade partners, and how do we optimize that metric?
The Trade Machine

List of potential trade partners

Future projections of all tradable assets for both teams

Positional, “win now vs. win later” weights

TRADE MACHINE

Trade “scores” for every possible trade

Find highest-scoring trade!
The Trade Machine

Computes scores for Phillies and other team by simulating player exchanges between teams

For each player traded from team A to B
• Subtract that player’s value to team A from Score A
• Add that player’s value to team B from Score B

Value to each team differs based on weights

In the end, we have two scores: $S_A, S_B$
The Trade Machine

Our goal:

Maximize $S_A + S_B$
(ensures maximum utility for both teams)

subject to $|S_A - S_B| < t$, where $t$ is a threshold
(ensure that the trade is fair to both teams)
The Trade Machine: Example Input

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Player</td>
<td>Position</td>
<td>Level</td>
<td>fWAR15</td>
<td>fWAR16</td>
<td>fWAR17</td>
</tr>
<tr>
<td>2</td>
<td>Dellin Betances</td>
<td>P</td>
<td>MLB</td>
<td>1.5</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>3</td>
<td>David Carpenter</td>
<td>P</td>
<td>MLB</td>
<td>0.4</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>Nathan Eovaldi</td>
<td>P</td>
<td>MLB</td>
<td>1.7</td>
<td>0.3</td>
<td>1.8</td>
</tr>
<tr>
<td>5</td>
<td>Didi Gregorius</td>
<td>SS</td>
<td>MLB</td>
<td>1.8</td>
<td>2.0</td>
<td>2.2</td>
</tr>
<tr>
<td>6</td>
<td>Bryan Mitchell</td>
<td>P</td>
<td>MLB</td>
<td>-0.9</td>
<td>-0.8</td>
<td>-0.7</td>
</tr>
<tr>
<td>7</td>
<td>Michael Pineda</td>
<td>P</td>
<td>MLB</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>8</td>
<td>Austin Romine</td>
<td>C</td>
<td>MLB</td>
<td>0.3</td>
<td>0.3</td>
<td>0.3</td>
</tr>
<tr>
<td>9</td>
<td>Chasen Shreve</td>
<td>P</td>
<td>MLB</td>
<td>0.2</td>
<td>0.2</td>
<td>0.3</td>
</tr>
<tr>
<td>10</td>
<td>Masahiro Tanaka</td>
<td>P</td>
<td>MLB</td>
<td>3.2</td>
<td>3.2</td>
<td>3.1</td>
</tr>
<tr>
<td>11</td>
<td>Adam Warren</td>
<td>P</td>
<td>MLB</td>
<td>0.5</td>
<td>0.5</td>
<td>0.4</td>
</tr>
<tr>
<td>12</td>
<td>Chase Whitley</td>
<td>P</td>
<td>MLB</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>13</td>
<td>Justin Wilson</td>
<td>P</td>
<td>MLB</td>
<td>-0.2</td>
<td>-0.2</td>
<td>-0.2</td>
</tr>
<tr>
<td>14</td>
<td>Luis Severino</td>
<td>P</td>
<td>AA</td>
<td>2.5</td>
<td>2.7</td>
<td>2.9</td>
</tr>
<tr>
<td>15</td>
<td>Greg Bird</td>
<td>1B</td>
<td>AA</td>
<td>1.2</td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td>16</td>
<td>Gary Sanchez</td>
<td>C</td>
<td>AA</td>
<td>-2.7</td>
<td>-2.5</td>
<td>-1.3</td>
</tr>
<tr>
<td>17</td>
<td>Ian Clarkin</td>
<td>P</td>
<td>A+</td>
<td>1.2</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>18</td>
<td>Rob Refsnyder</td>
<td>2B</td>
<td>AAA</td>
<td>0.8</td>
<td>1.1</td>
<td>1.3</td>
</tr>
<tr>
<td>19</td>
<td>Jacob Lindgren</td>
<td>P</td>
<td>AA</td>
<td>1.5</td>
<td>1.7</td>
<td>2.3</td>
</tr>
<tr>
<td>20</td>
<td>Miguel Andujar</td>
<td>3B</td>
<td>A</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.1</td>
</tr>
</tbody>
</table>

List of tradable MLB assets and prospects
Projected WARs, 2015-18
Trade 1

Phillies get:
- Ketel Marte (SS, AAA)
- D.J. Peterson (3B, AA)
- Taijuan Walker (RHP, MLB)

Mariners get:
- Cole Hamels (LHP, MLB)
- Darin Ruf (1B, MLB)
- Patrick Kivlehan (3B, AA)
- Edwin Diaz (RHP, A)

Phillies Score: +2.010
Mariners Score: +2.104
Total Trade Score: +4.114
Trade 2

Phillies get:
- Greg Bird (1B, AA)
- Didi Gregorius (SS, MLB)
- Rob Refsnyder (2B, AAA)
- Jacob Lindgren (LHP, AA)

Yankees get:
- Cole Hamels (LHP, MLB)
- Ben Revere (OF, MLB)
- Gary Sanchez (C, AA)
- Greg Bird (1B, AA)

Phillies Score: +2.692
Yankees Score: +2.784
Total Trade Score: +5.476
Data + Other Sports

• Football
  – Impact of weather
  – Bayesian draft analysis
  – Fantasy Football Machine Learning (yours truly)

• Basketball
  – Player efficiency rating
  – Advanced defensive metrics

• Soccer
  – Predictive shot-taking

• Hockey
  – Elo ratings
  – Offensive line shift productivity

• Join Sports Analytics!!
Big Data/CS Classes To Take:

- CS145 – Databases
- CS124 – Natural Language Processing
- CS246 – Mining Massive Datasets
- CS221 – Artificial Intelligence
- CS229 – Machine Learning
Thank you!