Evaluating Risks and Rewards of Prospects in Major League Baseball

Johnathan Hsu, Department of Sociology
Mentor: Professor Emilio Zagheni, Department of Sociology

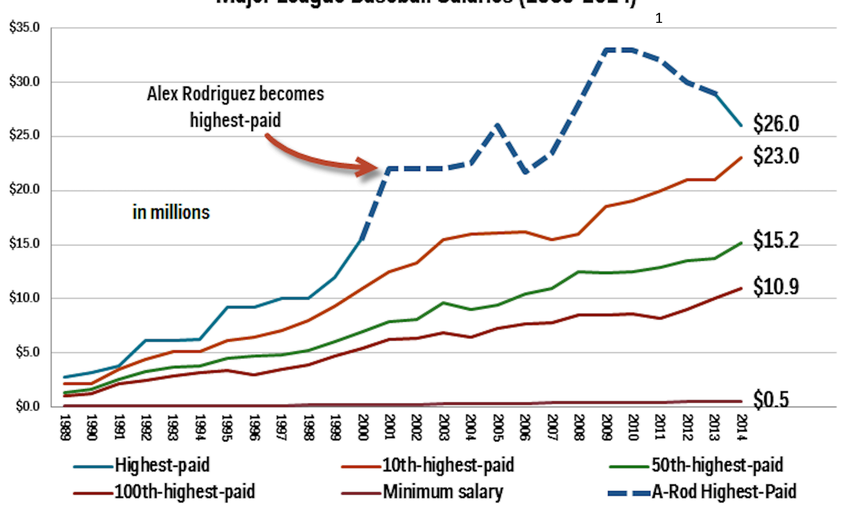
Research Question

Do age, draft order, position, or physicality serve as good indicators of the career value of prospective players?

Why Study Major League Baseball Drafting?

- Major League Baseball has 40 rounds, 30 teams, which equals approximately 1200 players drafted each year.
- Unlike other sports such as NBA and NFL, drafted players are not MLB ready immediately after drafting.
- Players can be drafted upon high school graduation, or after 21 years old.
- Current player salary is at an all time high, even after adjusting for inflation.

Major League Baseball Salaries (1989-2014)



Methods

Subjects

- Players who have played at least 1 game in the Major Leagues

Data

- Two subjective calculations[2] for Wins Above Replacement (W.A.R.) as measurement of player value.
- BeautifulSoup 4 library in Python for webscrape[3,4].
- RStudio for linear model data analysis
  - W.A.R. per year played as the dependent variable.
  - Round, height, weight, and age drafted as the independent variables.

Data Selection Process

All draft picks n = 54,308 → Signed n = 36,111 → Games ≠ 0 n = 5,855 → Exclude incomplete data n = 5,703 → Position Players n = 2,792 / Pitchers n = 3,009

Pitcher Linear Model

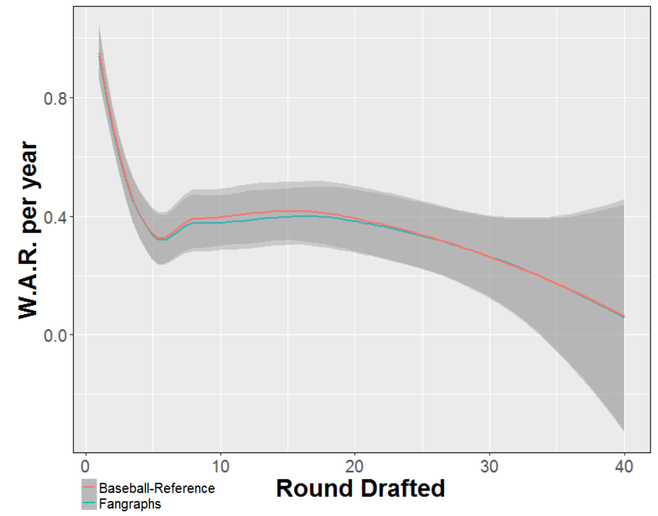
Baseball-Reference	B	SE	t	p
(Intercept)	0.065738	0.627839	0.105	0.916616
Round	-0.042906	0.009477	-4.527	6.21e-06 ***
Height	-0.077248	0.375028	-0.206	0.836822
Weight	0.005606	0.002212	2.535	0.011302 *
Age (Ref=18)	0.133981	0.036230	3.698	0.000221 ***
Fangraphs Pitcher	**β**	**SE**	**t**	**p**
(Intercept)	-1.100653	0.619108	-1.778	0.0755
Round	-0.062358	0.009345	-6.673	2.99e-11 ***
Height	0.663106	0.369813	1.793	0.0731
Weight	0.004218	0.002181	1.934	0.0532
Age (Ref=18)	0.146835	0.035726	4.110	4.06e-05 ***

Position Player Linear Model

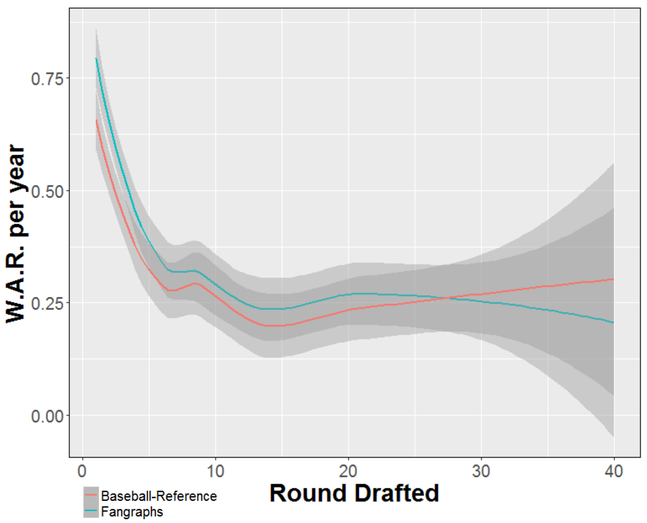
BR Position	β	SE	t	p
(Intercept)	0.086336	0.881728	0.098	0.92201
Round	-0.071938	0.014442	-4.981	6.71e-07***
Height	-0.510426	0.547106	-0.933	0.35093
Weight	0.016748	0.003129	5.353	9.36e-08 ***
Age (Ref=18)	0.144492	0.048844	2.958	0.00312 **
Fangraphs Position	**β**	**SE**	**t**	**p**
(Intercept)	-0.477047	0.853675	-0.559	0.57633
Round	-0.069536	0.013982	-4.973	7.00e-07 ***
Height	-0.254979	0.529699	-0.481	0.63030
Weight	0.017600	0.003029	5.810	6.96e-09 ***
Age (Ref=18)	0.142752	0.047290	3.019	0.00256 **

p < 0.001: «***» p < 0.01: «**» p < 0.05: «*»

Pitcher Loess at 95% CI



Position Player Loess at 95% CI



Discussion

- **Round drafted and weight** are significant factors for position players, but not for pitchers.
- **The round and age drafted** has the greatest impact on player W.A.R.
- Although **weight** in **Fangraphs Pitcher** is not significant, it is far from being trivial.
- The value of a drafted player is less consistent after **Round 10**.

Conclusion

- **Drafting younger players** is riskier in general.
- **Early round draftees** are more valuable than later round draftees.
- As shown when selecting data, **only 10% of drafted players compete in the MLB** so more qualitative and quantitative research is recommended on scouting.
- International players can be signed at a much younger age, what impact does this have on the development of these players and international baseball?

References

[1] http://www.businessinsider.com/baseballs-salaries-2014-8
[2] http://www.fangraphs.com/library/war/differences-fwar-rwar/
[3] http://www.fangraphs.com
[4] http://www.baseball-reference.com/

Acknowledgements

I would like to thank Professor Emilio Zagheni, Professor Stewart Tolnay, Kivan Polimis, Frank Edwards, and Susanna Hansson from the Department of Sociology for their continuous support.