BTM301 Class Project

Effects of Performance and Non-performance Factors on AAV of Free Agent contracts in Baseball

Onejune Lee, Minsoo Kang, Minwook Kim, Sunghun Ko

KAIST

December 11, 2024

Table of Contents

Introduction

Background

- Free agent contracts in baseball are determined by various factors.
- It is clear that the salary of a player is not solely determined by performance.
- Non-performance factors that affects to salary can be thought as sort of "bubble" or mispricing of market.
- Conversely, there may be some performance factors which are seemingly affecting the salary but actually not.
- To figure out the effects of these factors, we selected several factors, built a linear model based on these factors, ran regression, and interpreted the results.

Introduction

Factors considered

Following are the factors that are not determined by performance of each player:

- WR: Win rate of the player's last team before free agency
- Atd: Average attendance at home game of the player's last team before free agency
- SU: Last season's total salary of team the player signed with
- L: Dummy variable indicating left-handedness of the player
- AGE: Age of the player at the time of signing

These can be thought as candidates of potential "bubble" factors.

Factors considered - cont'd

For the performance factors, we chose the following for hitters:

- OBP: On-base percentage of the player
- SLG: Slugging percentage of the player
- HR: Home runs of the player
- PA: Plate appearances of the player

Two of these factors are ratios, and the other two are cumulative values. We also included the league's average values for ratio factors to see if relative performance is more important than absolute performance or not.

Our initial plan was including more factors and league average of them, too, but failed to find the data for them.

Factors considered - cont'd

For the performance factors, we chose the following for pitchers:

- ERA: Earned run average of the player
- WHIP: Walks plus hits per inning pitched of the player
- SO: Strikeouts of the player
- IP: Innings pitched of the player

Again, two of these factors are ratios, and the other two are cumulative values, and we included the league's average values for ratio factors.

Model

Rationale

What we expect is that the average annual value (AAV) is proportional to the factors we mentioned. We can write this as, for hitters:

$$AAV \propto WR,\ Atd,\ SU,\ \lambda^{L}\ (\text{for some } \lambda),\ \frac{1}{AGE},\ OBP,\ SLG,\ HR,\ PA,\ OBP_{\text{avg}},\ SLG_{\text{avg}},$$

although we do not know the degree of proportionality. As you can see, taking logarithm on both sides, we can write this as a linear model. We can do the same for pitchers, too.

Model

Hitter Model

$$\begin{aligned}\log AAV_i = {} & \beta_{WR} \log WR_i + \beta_{Atd} \log Atd_i + \beta_{SU} \log SU_i + \beta_L L + \beta_{AGE} \log AGE_i \\ & + \beta_{OBP} \log OBP_i + \beta_{SLG} \log SLG_i + \beta_{HR} \log HR_i + \beta_{PA} \log PA_i \\ & + \beta_{OBP_{avg}} \log OBP_{\text{avg}} + \beta_{SLG_{avg}} \log SLG_{\text{avg}} + \beta_0 + \varepsilon_i\end{aligned}$$

Pitcher Model

$$\begin{aligned}\log AAV_i = {} & \beta_{WR} \log WR_i + \beta_{Atd} \log Atd_i + \beta_{SU} \log SU_i + \beta_L L + \beta_{AGE} \log AGE_i \\ & + \beta_{ERA} \log ERA_i + \beta_{WHIP} \log WHIP_i + \beta_{SO} \log SO_i + \beta_{IP} \log IP_i \\ & + \beta_{ERA_{avg}} \log ERA_{\text{avg}} + \beta_{WHIP_{avg}} \log WHIP_{\text{avg}} + \beta_0 + \varepsilon_i\end{aligned}$$

Data

Data Collection

- 2010-2019 free agent contracts from MLB and KBO were collected. More recent data were excluded because we were not sure about how to handle the COVID-19's effects.
- Performance data was collected from Baseball Reference and KBO's official website
- Attendance data was collected from ESPN and KBO's official website
- Salary data was collected from independent researcher's website
- Both KBO and MLB's AAV and salary data were normalized by CPI at December of the year of signing, by Korea and US, respectively.

Data

Disclaimer

- Data of those who failed to find a team were excluded.
- Players whose performance was empty for various reasons (e.g., came from other leagues, got injured) were also excluded.
- There may be a potential unknown issues with data which ocurred in the process of data collection.

Results

MLB Hitters

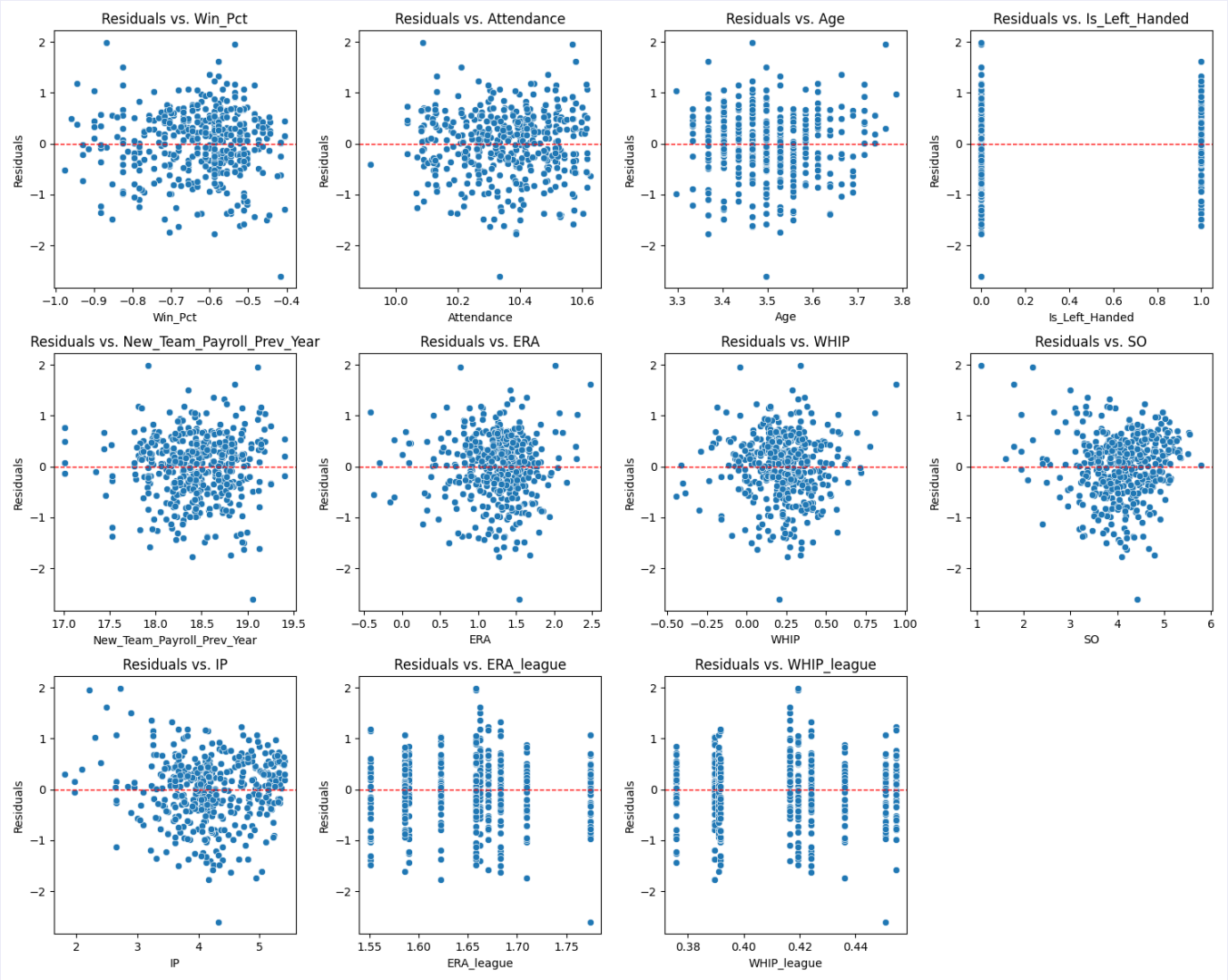
Results

MLB Pitchers: Regression Results

Variable	Coeff.	Std. Err	t-value	$P > \|t\|$
const	13.3743	2.949	4.535	0.000
Win_Pct	1.0601	0.285	3.725	0.000
Attendance	0.3321	0.225	1.479	0.140
Age	-1.7899	0.349	-5.131	0.000
Is_Left_Handed	0.0188	0.073	0.257	0.797
New_Team_Payroll_Prev_Year	0.1624	0.081	1.996	0.047
ERA	0.0660	0.116	0.566	0.572
WHIP	-1.3626	0.256	-5.330	0.000
SO	0.5457	0.141	3.882	0.000
IP	0.0896	0.140	0.642	0.521
ERA_league	0.7778	1.070	0.727	0.468
WHIP_league	-3.2791	2.445	-1.341	0.181

Results

MLB Pitchers: Residuals



Results

KBO Hitters

Results

KBO Pitchers

Discussion

Conclusion