Determinants of Baseball Player Salary and Performance

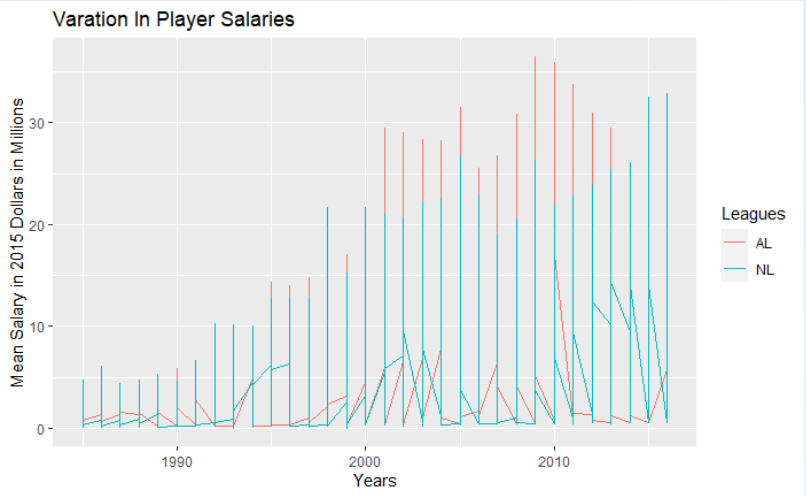
Ryan Marinelli
University of York
Master of Science Candidate in Computer Science

Abbreviated abstract: The research question is to study how a balance is struck in salary when considering the merit of an individual player and the success of the team. To study the role of salary, econometric methods are applied. An Autoregressive Distributed Lag Model is applied in this study along with clustering methods. It appears individual-level factors are weighted more heavily and have their salaries diminished by expectations based on performance.

Problem, Data, Previous Works

- The goal of the research is to understand the interplay of rewards and success between an individual and their team

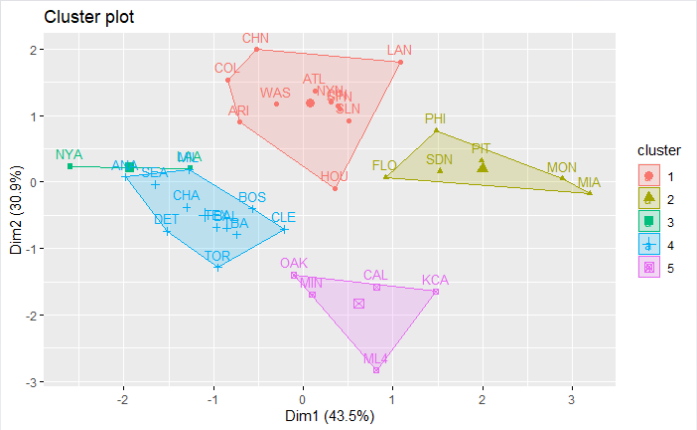
- Data is comprised from the Lahman dataset and data from Federal Reserve Bank of St.Louis(FRED)
  - Lahman Data
    - Complete batting and pitching statistics from 1871 to 2019
    - Lahman R package to access the data
  - FRED Data
    - Consumer Price Index
    - Used to normalize salary and convert salaries into 2015 dollars

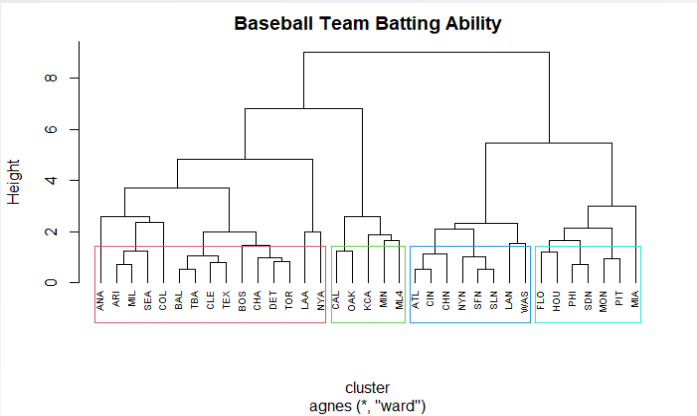


Methods

- Clustering Analysis
  - Hierarchical Clustering
  - K-Means Based
    - Silhouette Method to determine number of Clusters
    - PCA and K-Means

- Autoregressive Distributed Lag Model
  - Parametric approach with lagged terms of key variables
    - Newey-West SE
  - Dickey–Fuller Test suggested stationary
  - ARCH LM-test suggest heteroskedasticity
    - Graphed AR(1) term
      - Clustered Residuals
      - ARCH modeling
        - Not as meaningful in terms of interpretation





Results and Conclusions

- For each point in batting average, salary is associated with being 1.3 million more.
  - Amplified with home-runs
  - Negative coefficients suggest expectations in lagged terms
- Personal characteristics are more significant in salary. Team-level variables have a much smaller effect size. At least lags do not punish current salary.

Coefficient Values of Distributed Lag Model

Parameters	Estimates	Newey-West Standard Error	T-Score	P-Values
(Intercept)	-1.18	0.10	-11.4	0.00
L(adjusted_salary, 1)	0.13	0.00	33.5	0.00
L(adjusted_salary, 2)	0.07	0.00	19.6	0.00
batting_Average	1.39	0.12	12.0	0.00
L(batting_Average, 1)	-0.48	0.08	-5.8	0.00
homerun_Average	20.67	1.52	13.6	0.00
L(homerun_Average, 2)	-1.68	0.45	-3.7	0.00
rank_delta	0.59	0.02	24.1	0.00
W	0.02	0.01	1.7	0.08
L	-0.02	0.01	-2.1	0.04