The Role of Batteries in Pitcher Success

POLI 521 Final Project

By: Christopher Yurris

Introduction

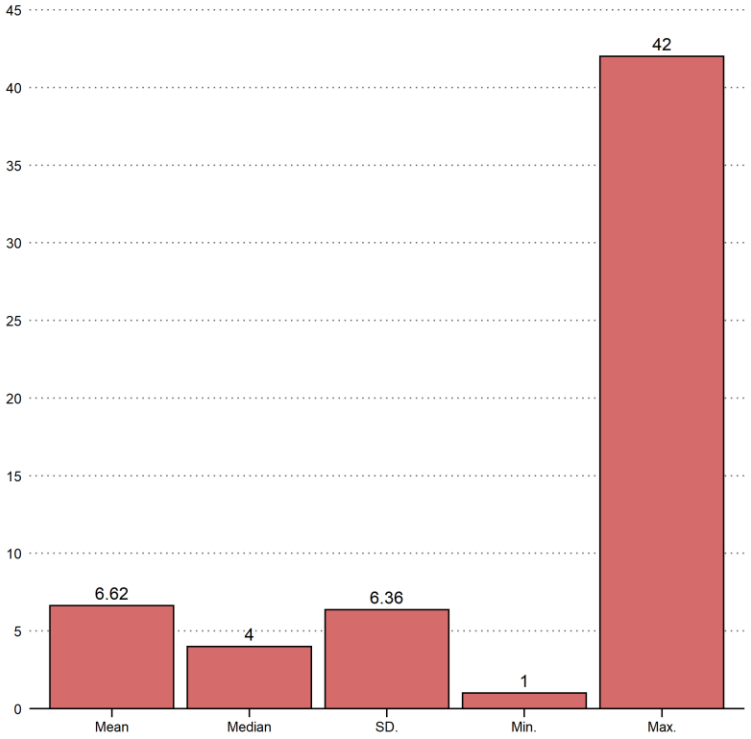
- Focusing on the role of the pitcher-catcher relationship in baseball
- Known as the “battery”
- Inspired by Adam Wainwright and Yadier Molina recently breaking the MLB record for starts as a battery (325)

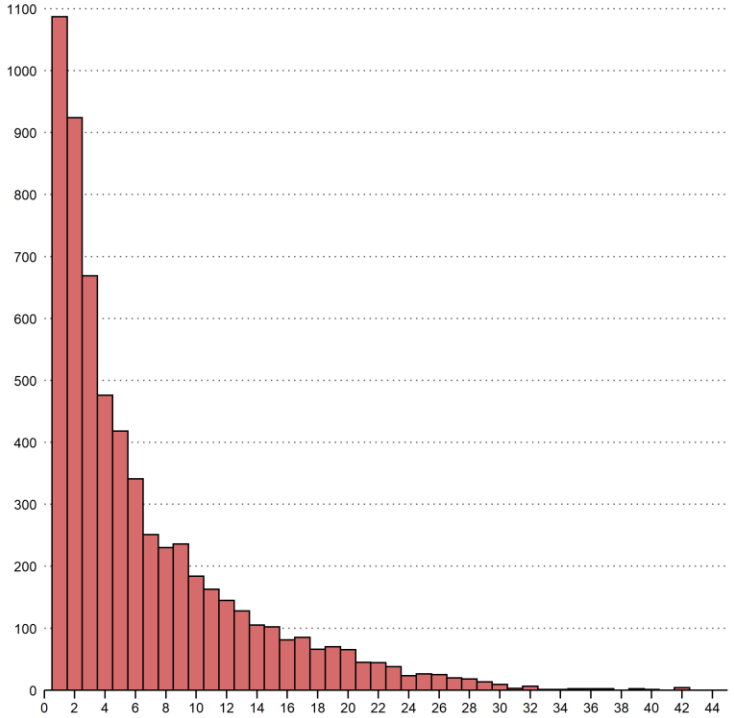
Data Sources

- Game log data from every MLB game since 1900 (Retrosheet)
- Player career statistics ("Lahman" R Package and Baseball-Reference Datafiles)

Bill Freehan (catcher) and Mickey Lolich (pitcher) – 324 Starts as a battery (1963-75)

Descriptive Statistics – Degree Centrality





Research Questions

- Do long-term battery connections have a positive impact on pitcher performance?
- What measurement most effectively conceptualizes the pitcher-catcher relationship?

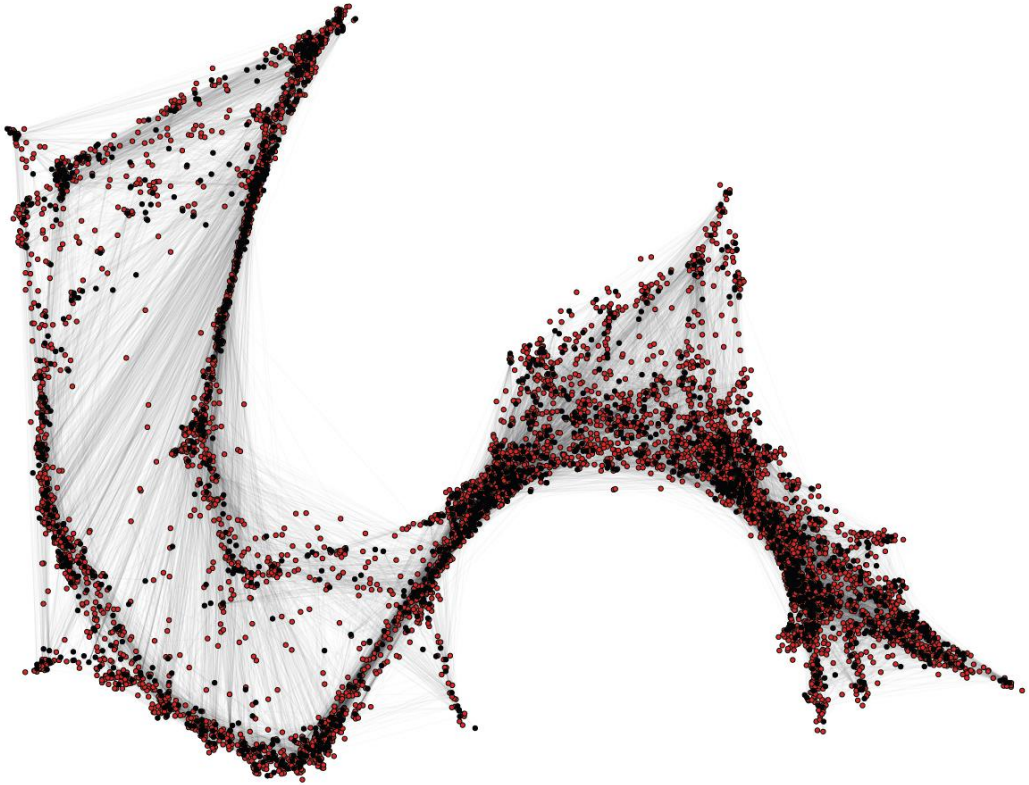
Bartolo Colon started with 42 different catchers in his career (1997-2018), tied for the MLB record.

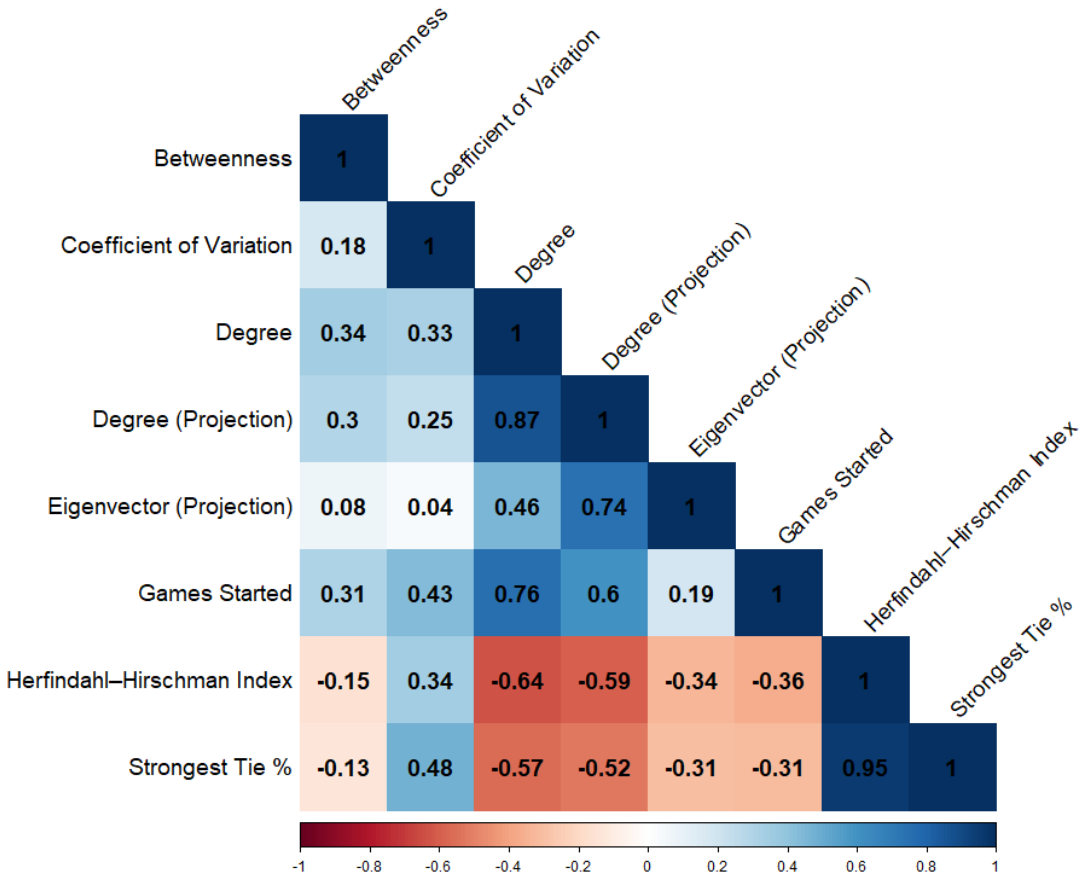
Hypotheses

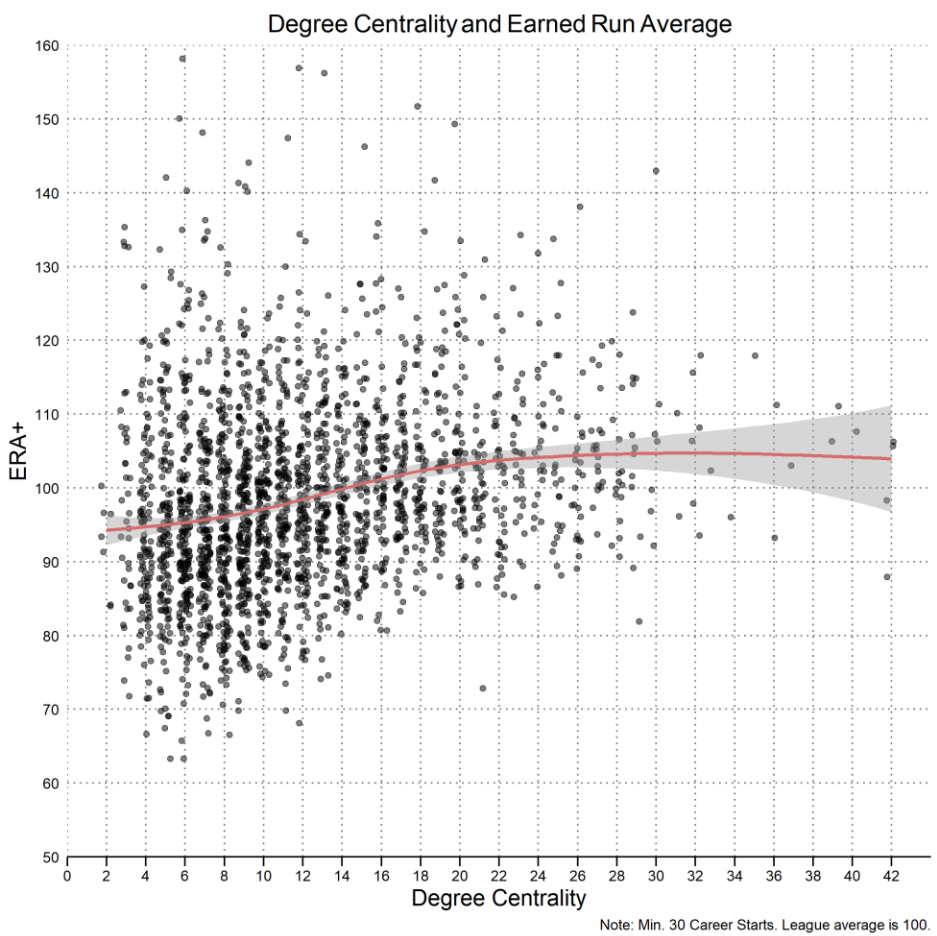
- H1: There is a negative relationship between a pitcher's degree centrality and on-field performance
- H2: Coefficient of Variation of Edge Weights will have a positive relationship with a pitcher's performance
- H3: Other standard measures of centrality (betweenness, eigenvector, projections) will have little relationship with performance

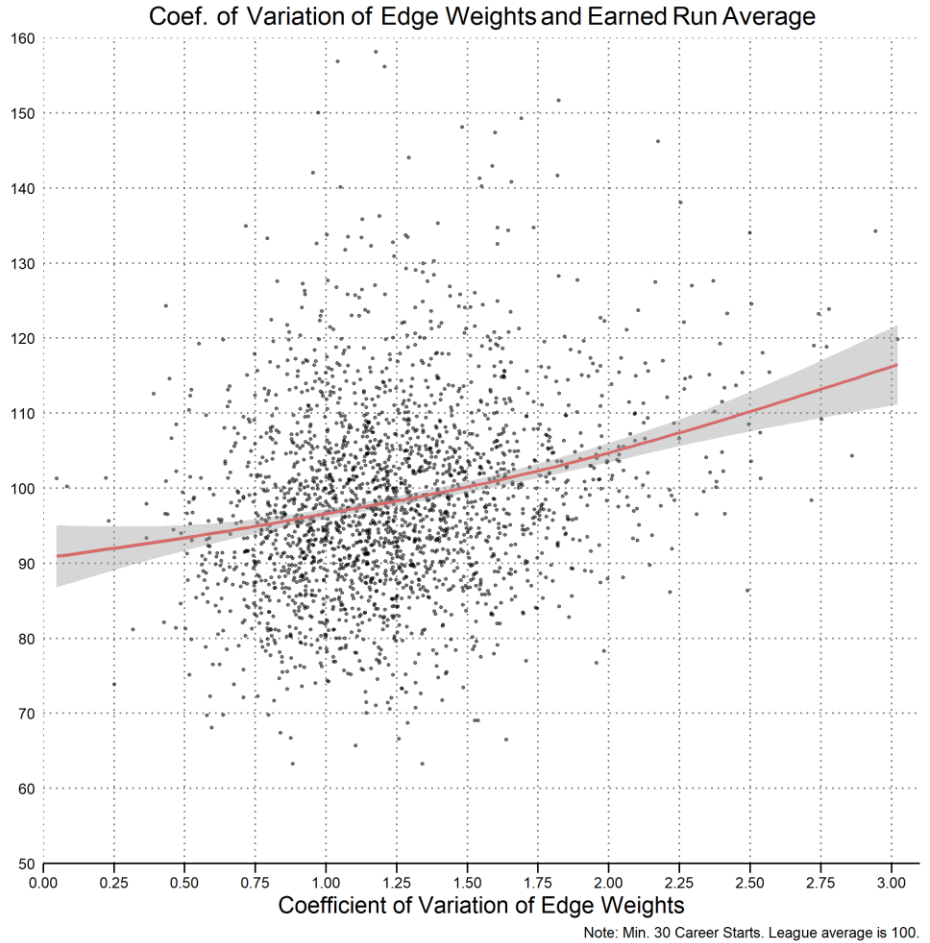
Network Characteristics

- Weighted Bipartite Network
- 7745 Total Nodes:
  - 6,118 Starting Pitchers
  - 1,627 Catchers
- 40483 unique edges




Betweenness
Coefficient of Variation
Degree
Degree (Projection)
Eigenvector (Projection)
Games Started
Herfindahl–Hirschman Index
Strongest Tie %
1
0.18 1
0.34 0.33 1
0.3 0.25 0.87 1
0.08 0.04 0.46 0.74 1
0.31 0.43 0.76 0.6 0.19 1
-0.15 0.34 -0.64 -0.59 -0.34 -0.36 1
-0.13 0.48 -0.57 -0.52 -0.31 -0.31 0.95 1
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1


Degree Centrality and Earned Run Average
ERA+
Degree Centrality
Note: Min. 30 Career Starts. League average is 100.


Coef. of Variation of Edge Weights and Earned Run Average
Coefficient of Variation of Edge Weights
Note: Min. 30 Career Starts. League average is 100.

Regression Model

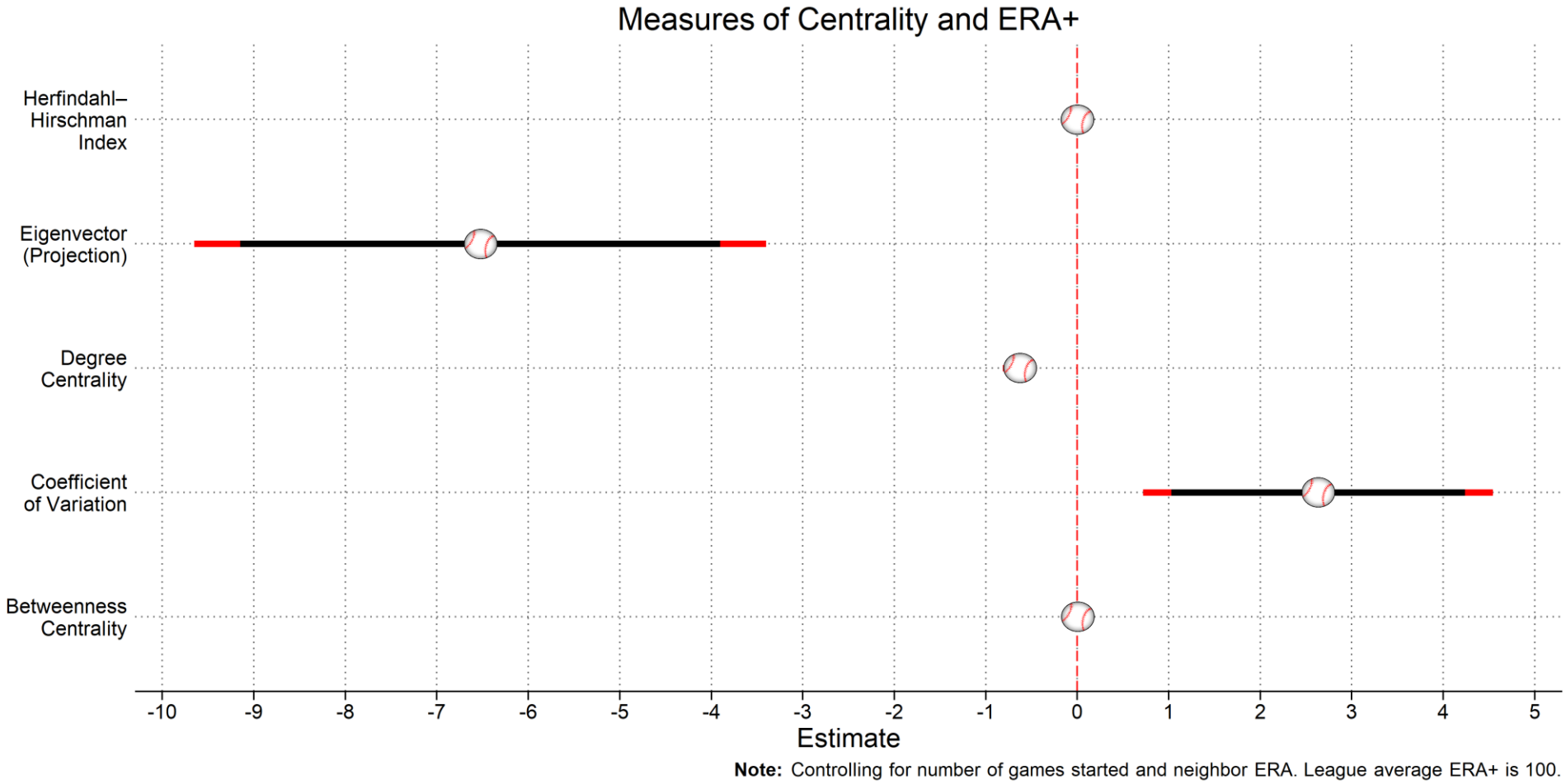
$$Y_i = \beta_0 + \beta_1 Degree + \beta_2 CV + \beta_3 HHI + \beta_4 Eigen + \beta_5 Between + \beta_6 N + \beta_7 GS$$

- Degree: Degree Centrality
- CV: Coefficient of variation of edge weights
- HHI: Herfindahl-Hirschman Index
- GS: Games Started
- Eigen: Eigenvector Centrality (projection)
- Between: Betweenness Centrality
- N: Neighbor ERA
- Y_i: Adjusted ERA +

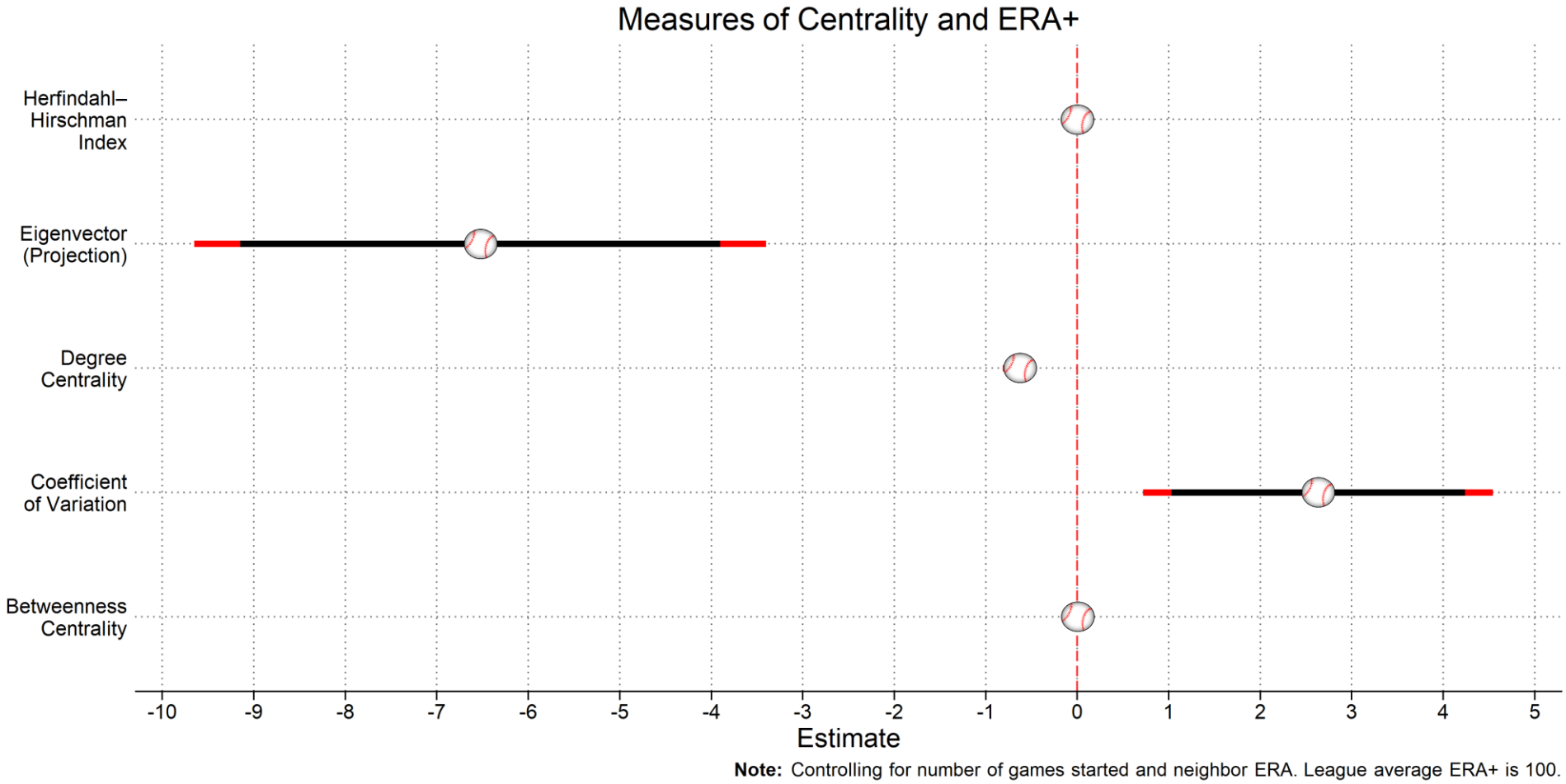
Regression Model

$$Y_i = \beta_0 + \beta_1 Degree + \beta_2 CV + \beta_3 HHI + \beta_4 Eigen + \beta_5 Between + \beta_6 N + \beta_7 GS$$

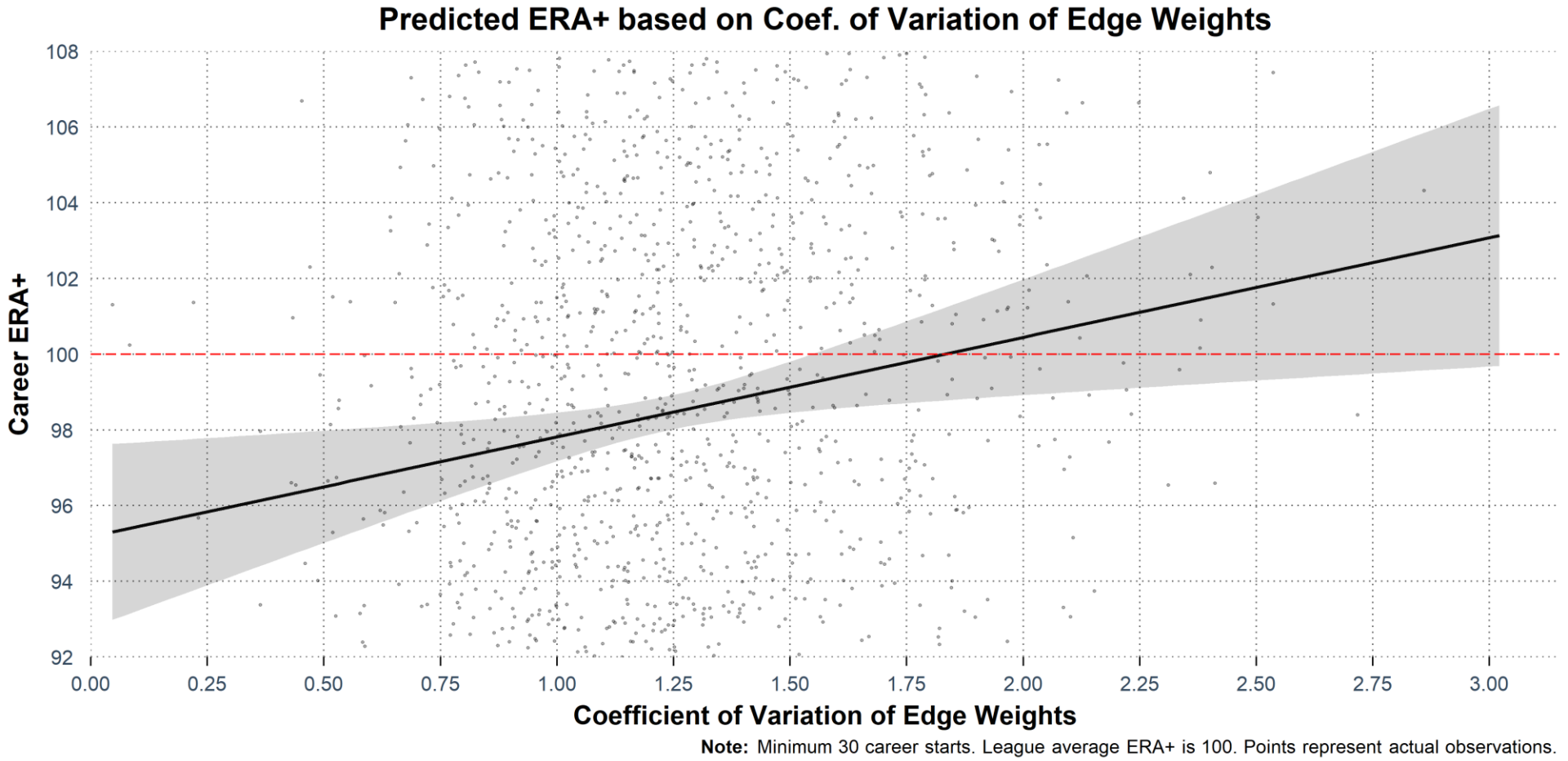
- Degree: Degree Centrality
- CV: Coefficient of variation of edge weights
- HHI: Herfindahl-Hirschman Index
- GS: Games Started
- Eigen: Eigenvector Centrality (projection)
- Between: Betweenness Centrality
- N: Neighbor ERA
- Y_i: Adjusted ERA +



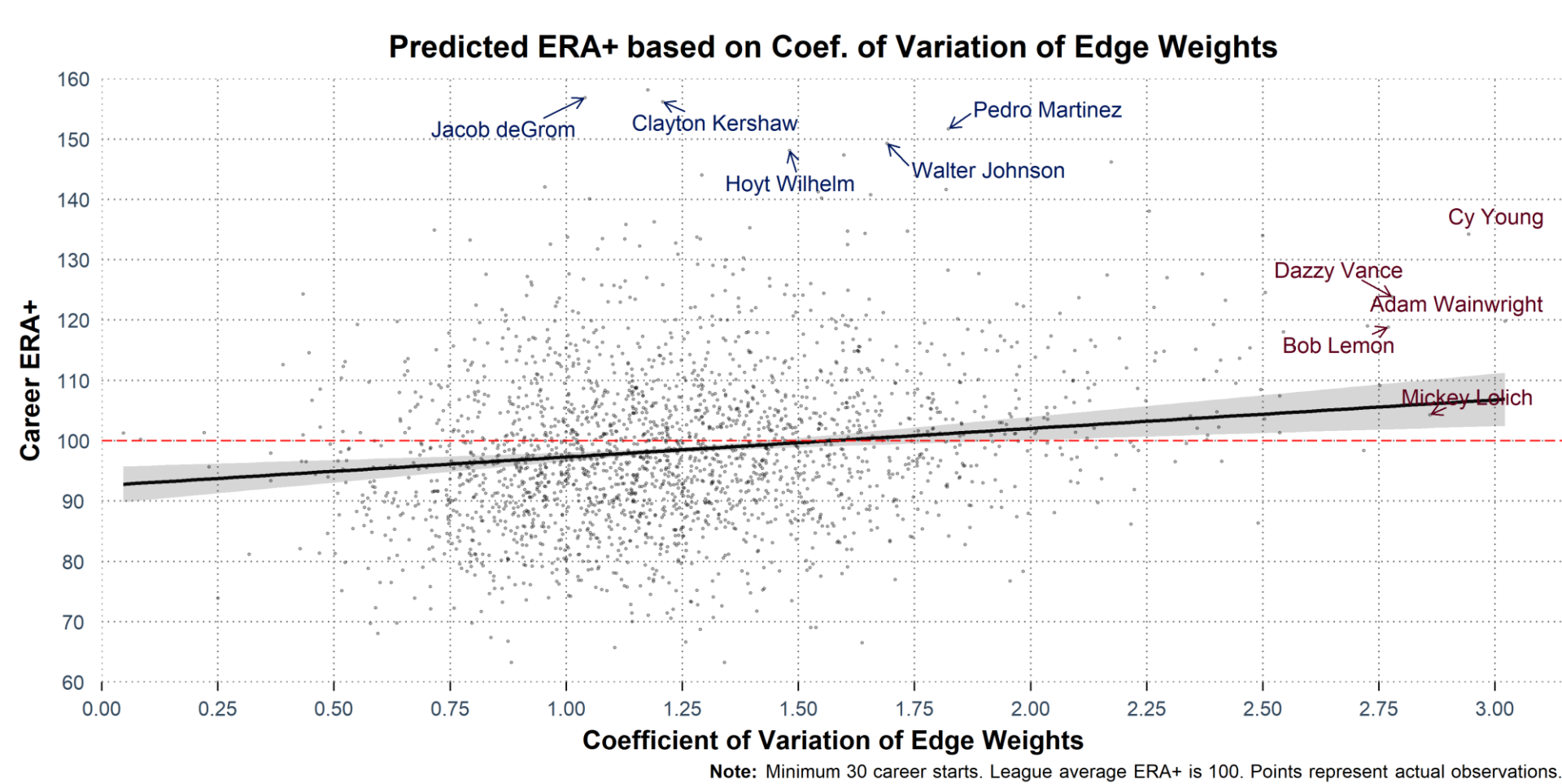
Note: Controlling for number of games started and neighbor ERA. League average ERA+ is 100.



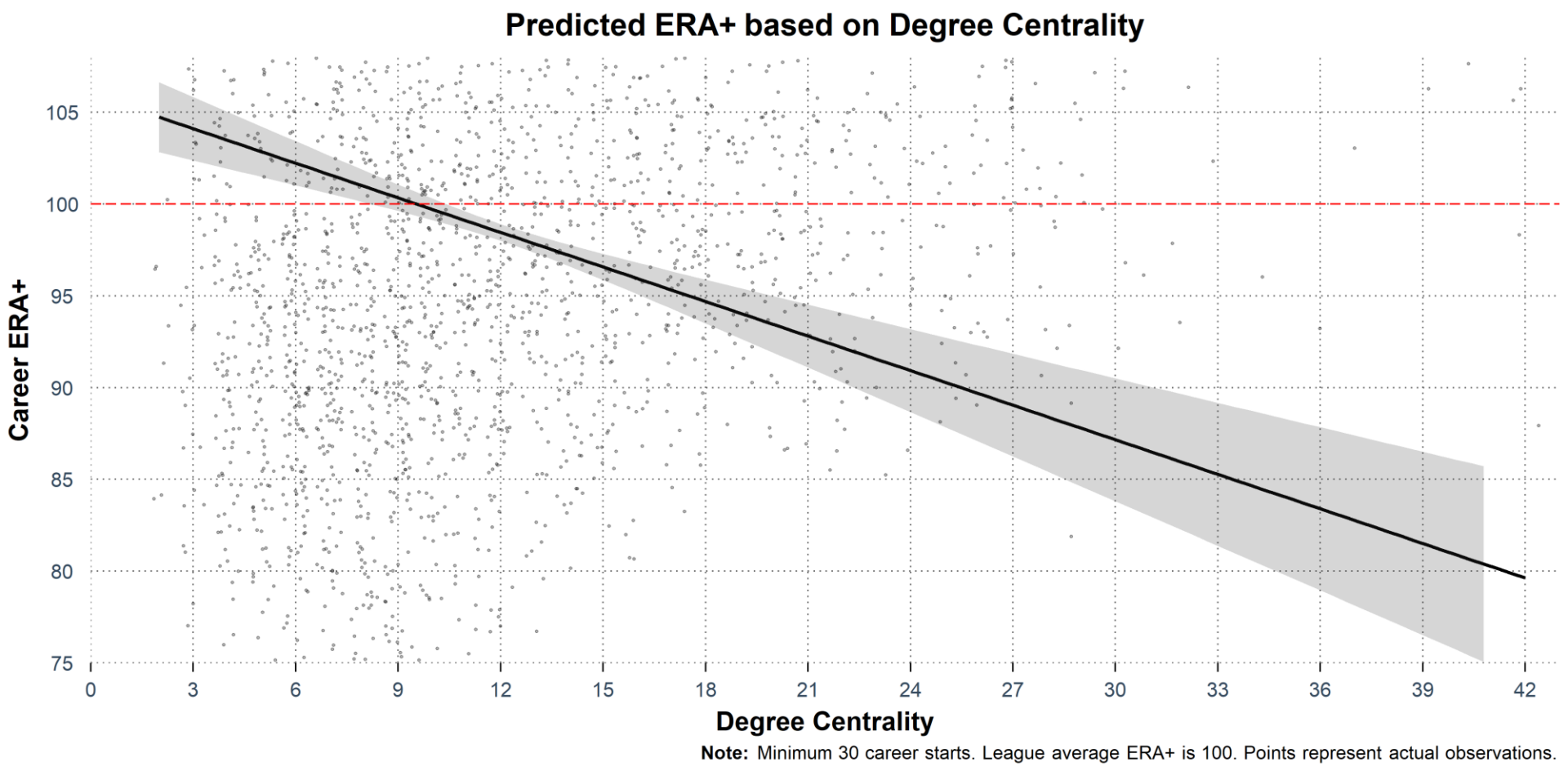
Note: Controlling for number of games started and neighbor ERA. League average ERA+ is 100.



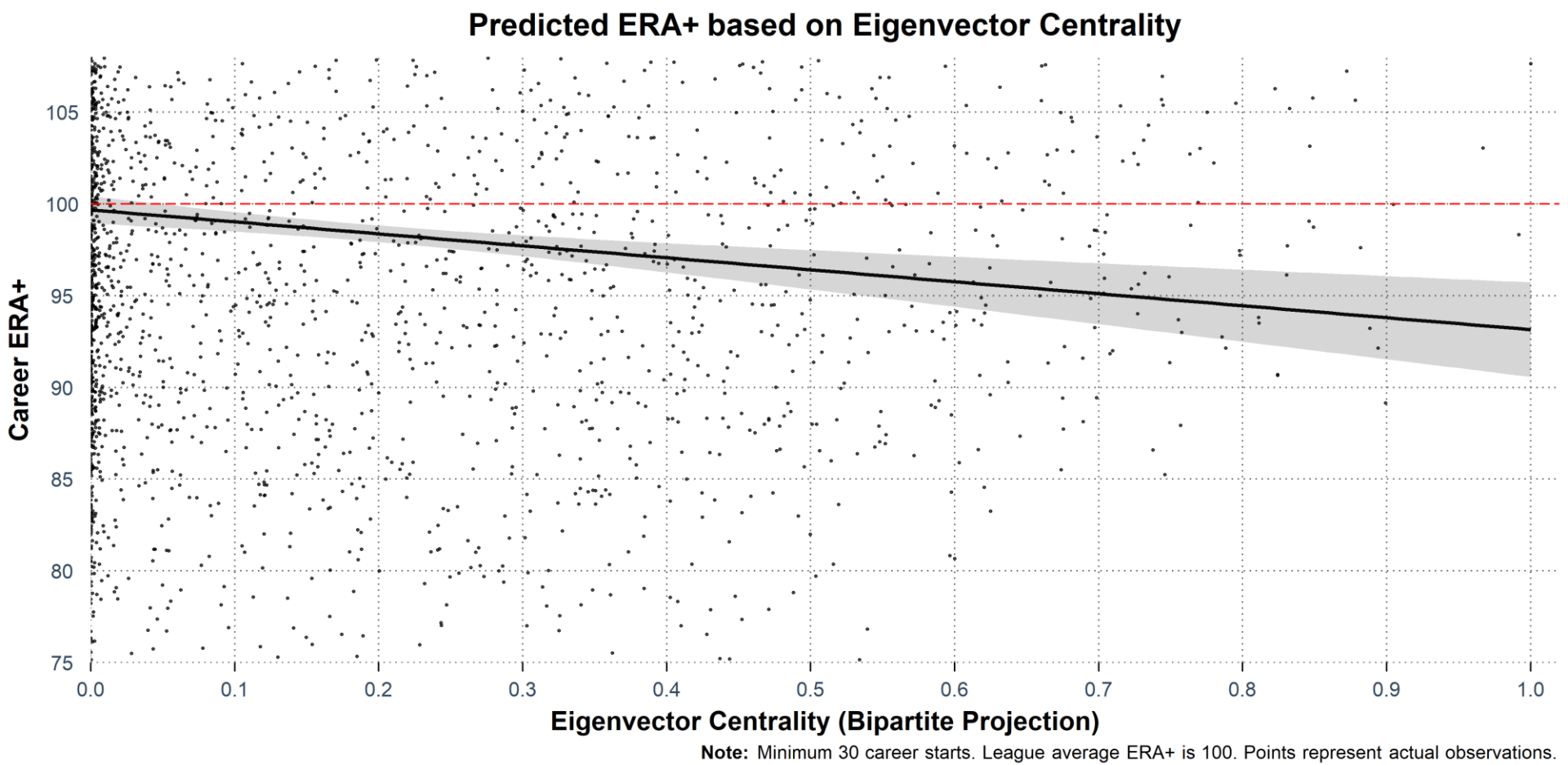
Note: Minimum 30 career starts. League average ERA+ is 100. Points represent actual observations.



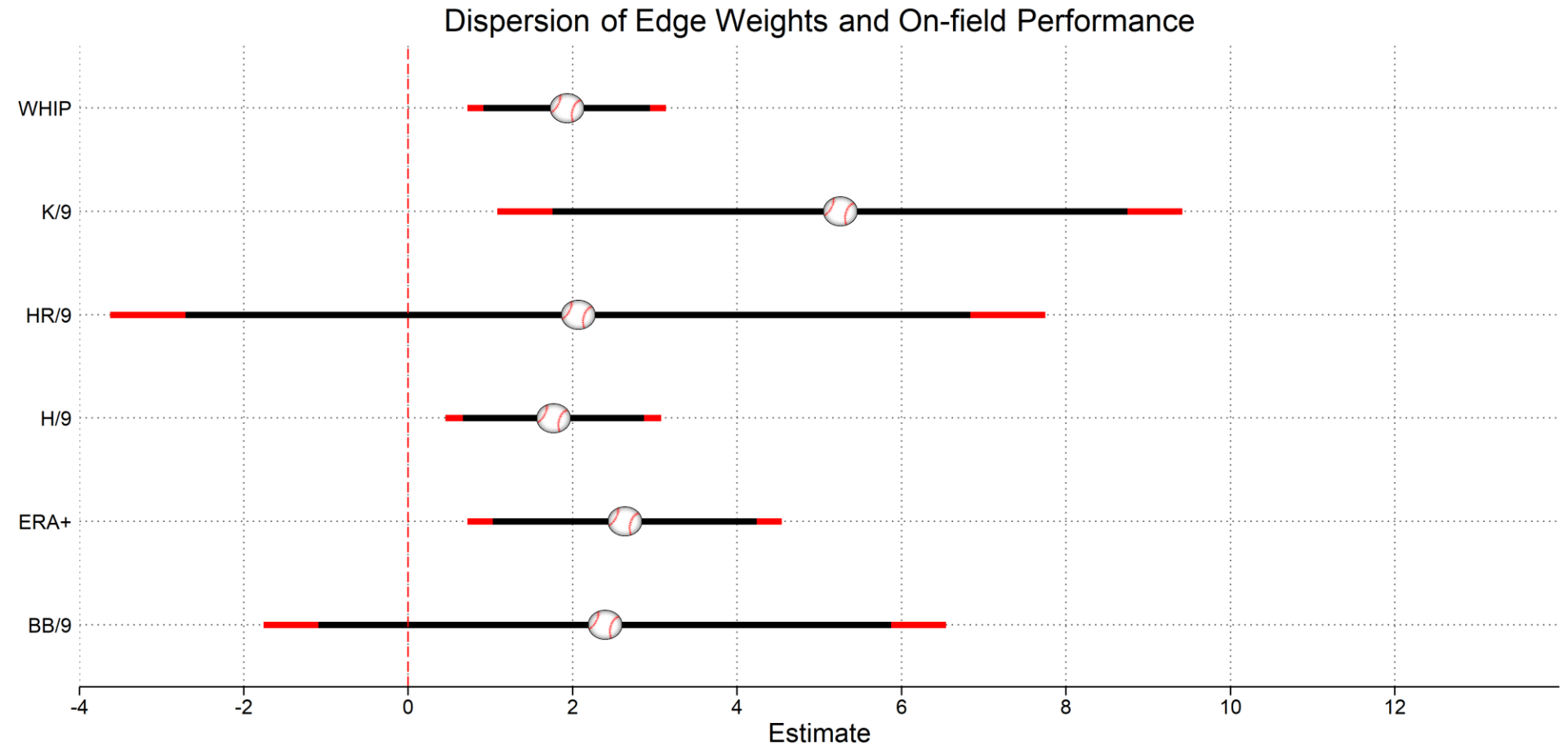
Note: Minimum 30 career starts. League average ERA+ is 100. Points represent actual observations.



Note: Minimum 30 career starts. League average ERA+ is 100. Points represent actual observations.



Note: Minimum 30 career starts. League average ERA+ is 100. Points represent actual observations.



Note: Controlling for number of games started and neighbor ERA. Statistics standardized so league average is 100.

Hypotheses Revisited

- ✓ H1: There is a negative relationship between a pitcher's degree centrality and on-field performance
- ✓ H2: Coefficient of Variation of Edge Weights will have a positive relationship with a pitcher's performance
- ✓ H3: Other standard measures of centrality (betweenness, eigenvector projection) will have little relationship with performance

Conclusions and Next Steps

- Battery connections that are more concentrated are associated with increased pitcher on-field performance
- CV and degree centrality are most illustrative of this relationship
- Addition of a time-series component to measure the battery relationship over time?