

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)

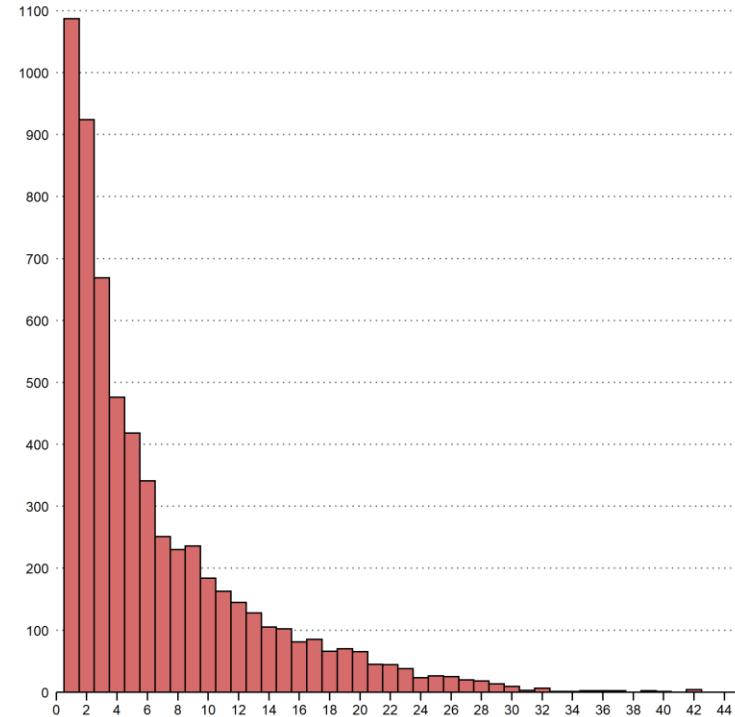
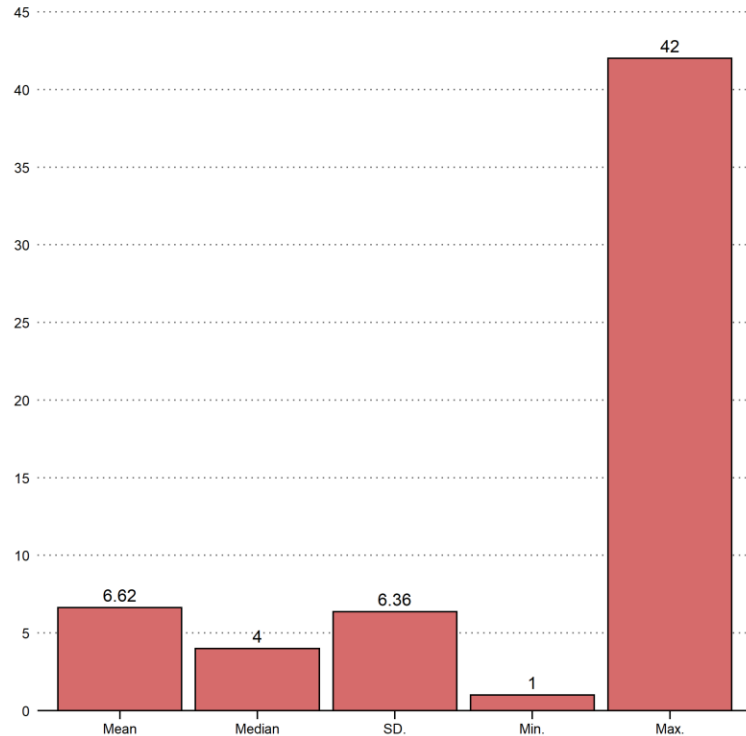
Data Sources

- 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?



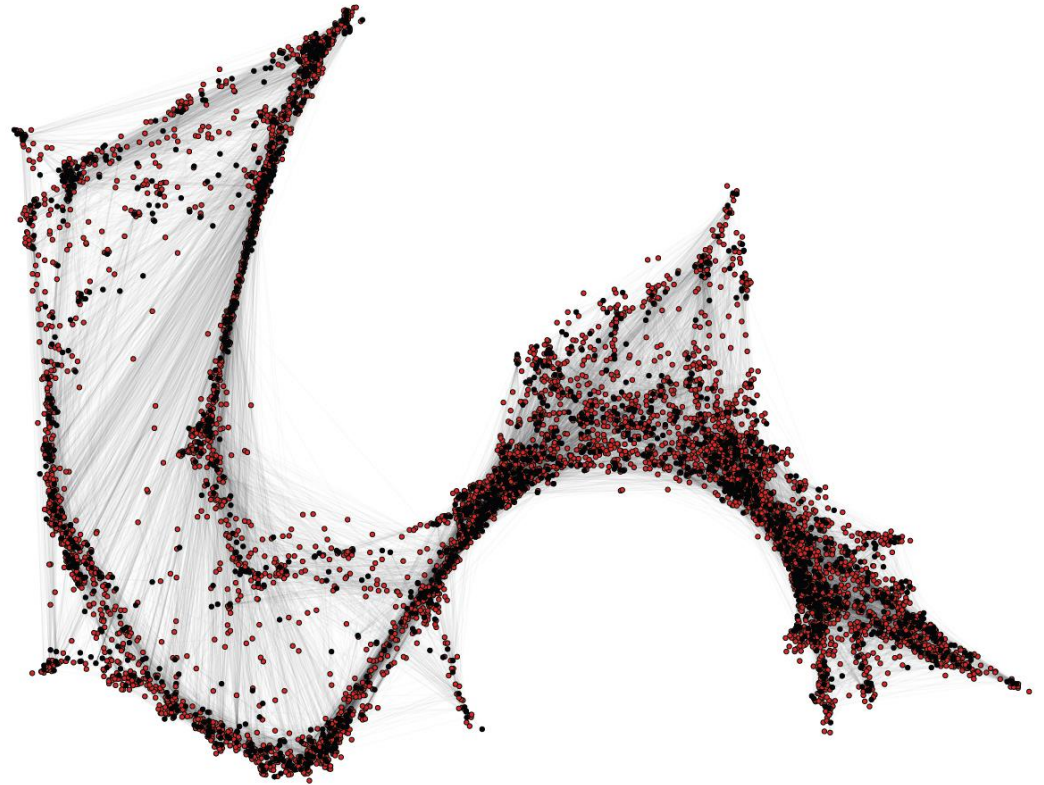
Bartolo Colon started with 42 different catchers in his career (1997-2018), tied for the MLB record.

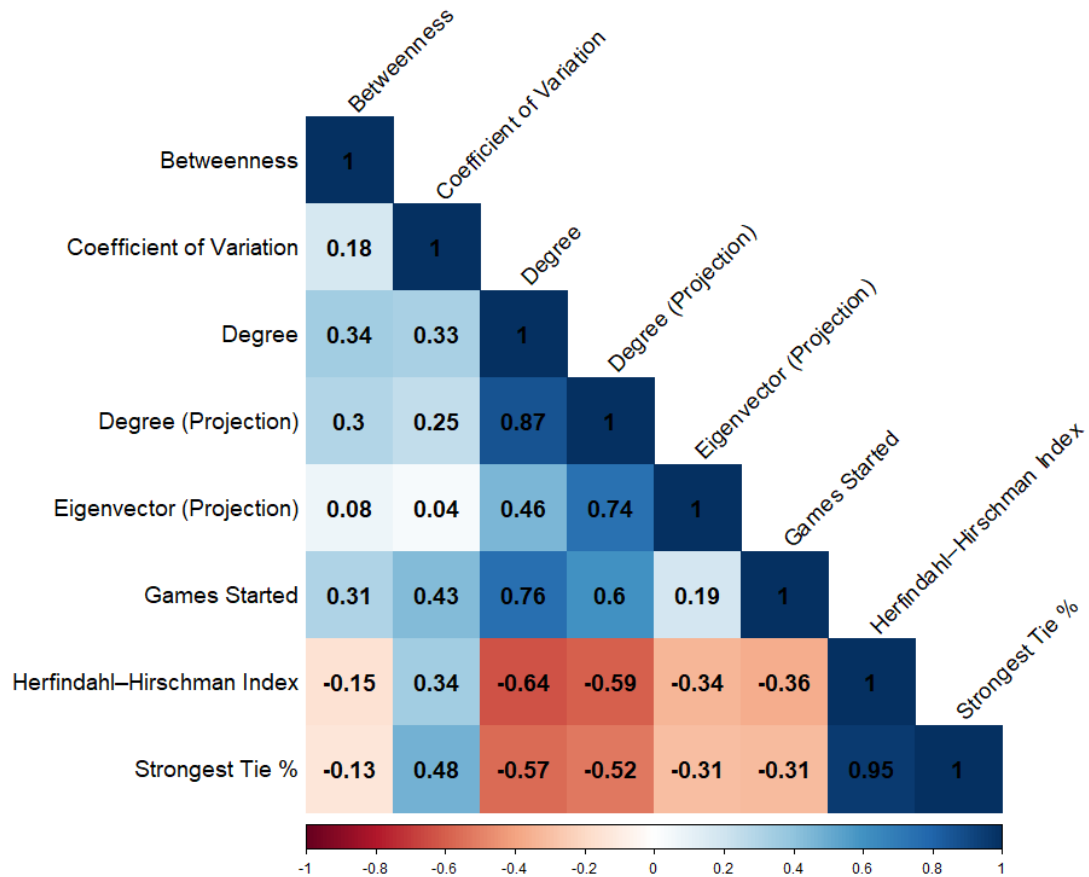
Hypotheses

- ❑ 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

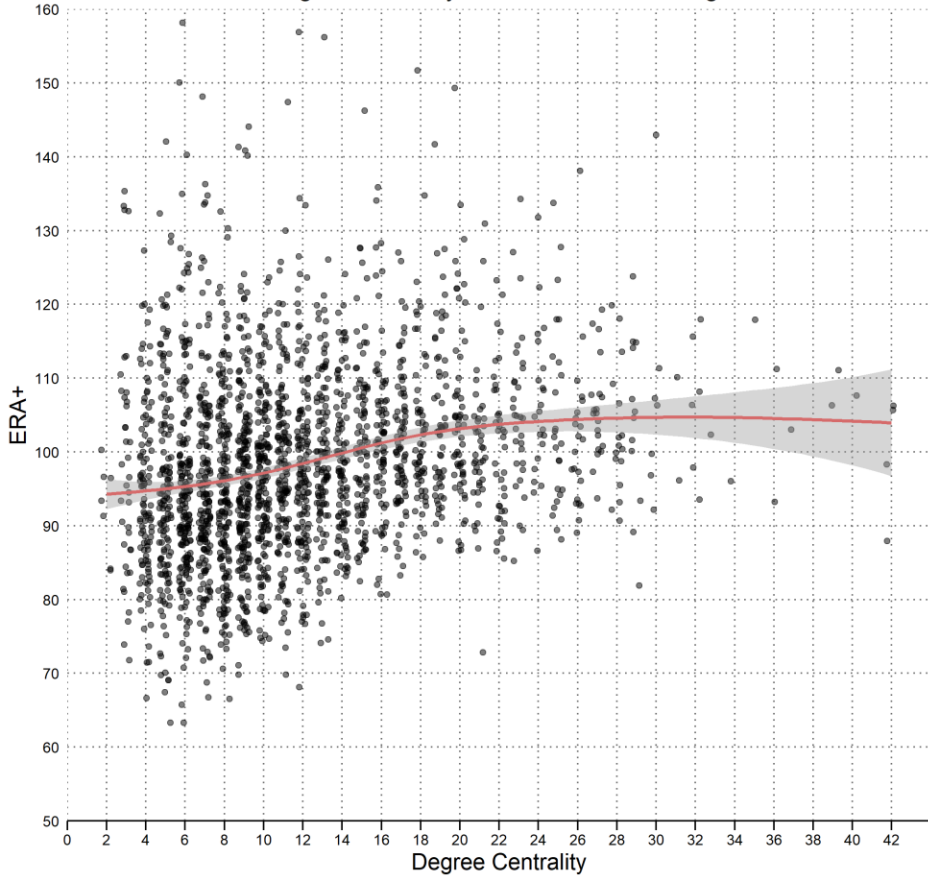
Network Characteristics

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



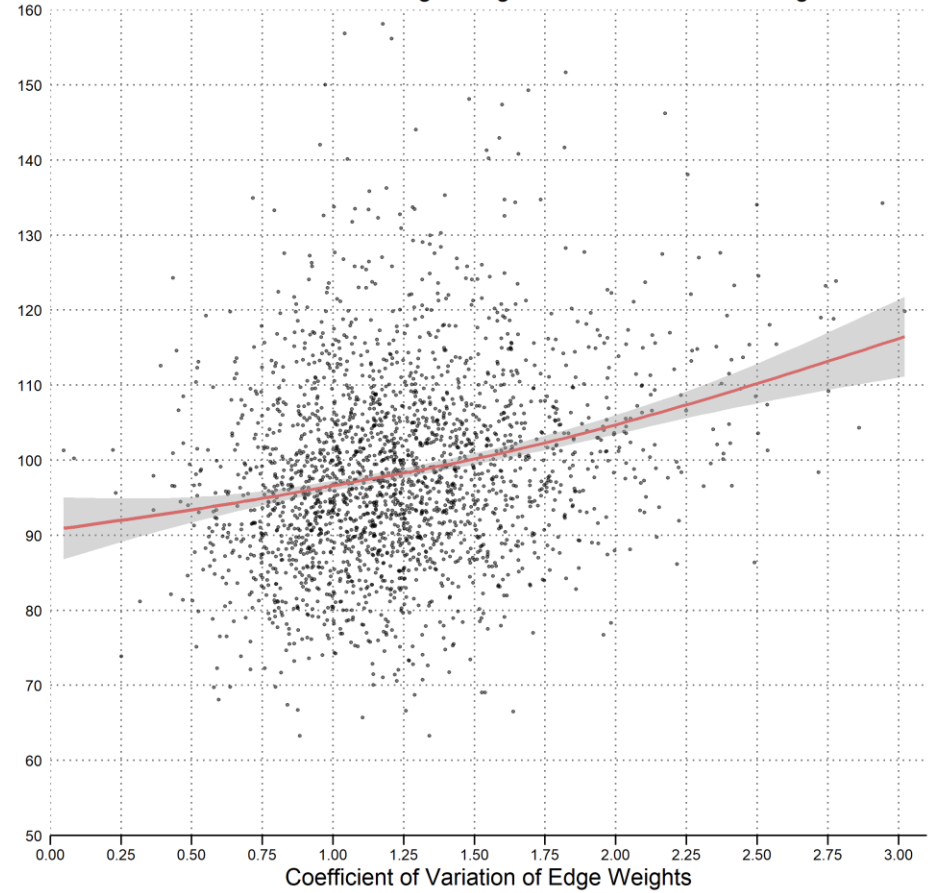


Degree Centrality and Earned Run Average



Note: Min. 30 Career Starts. League average is 100.

Coef. of Variation of Edge Weights and Earned Run Average



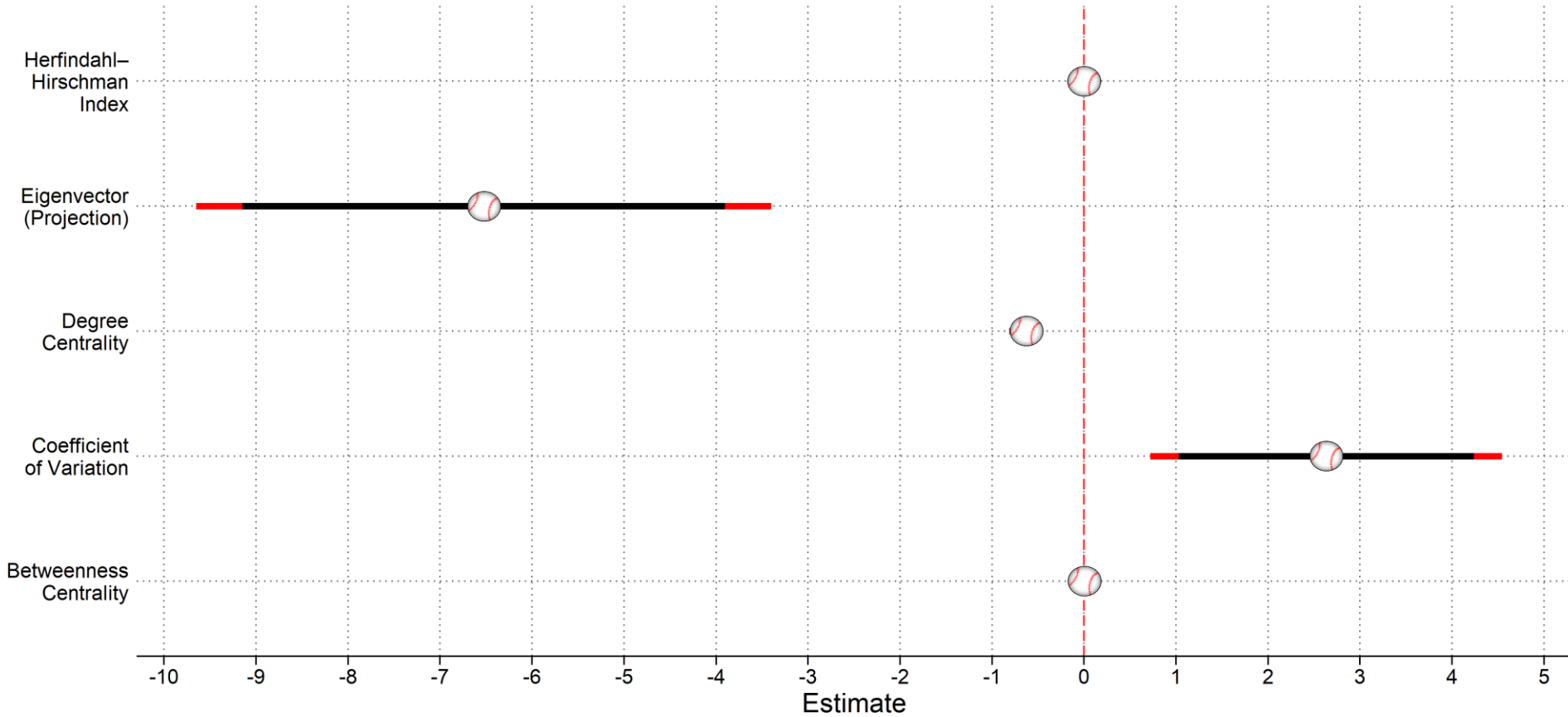
Note: Min. 30 Career Starts. League average is 100.

Regression Model

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

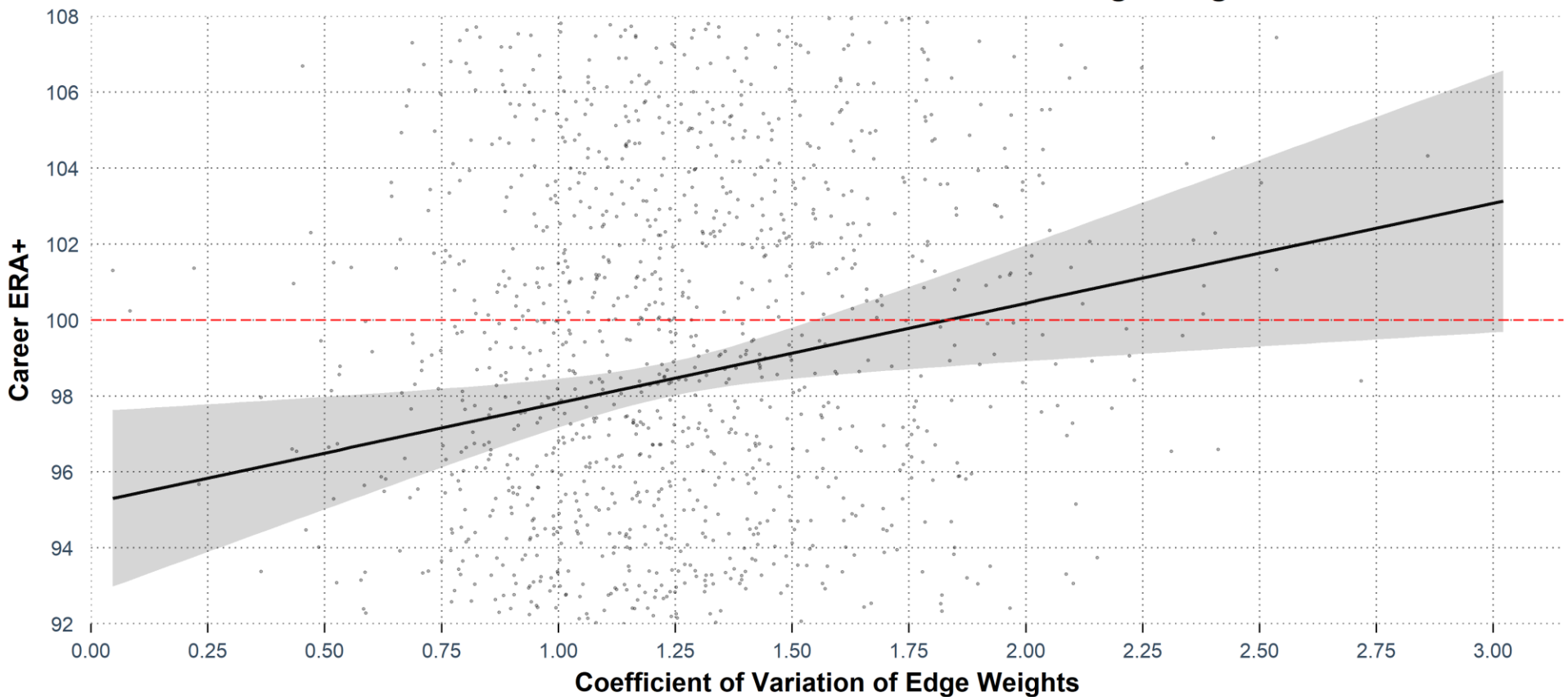
- Degree: Degree Centrality
- CV: Coefficient of variation of edge weights
- HHI: Herfindahl-Hirschman Index
- GS: Games Started
- Eigen: Eigenvector Centrality (projection)
- Between: Betweenness Centrality
- N: Neighbor ERA
- Y_i : Adjusted ERA +

Measures of Centrality and ERA+



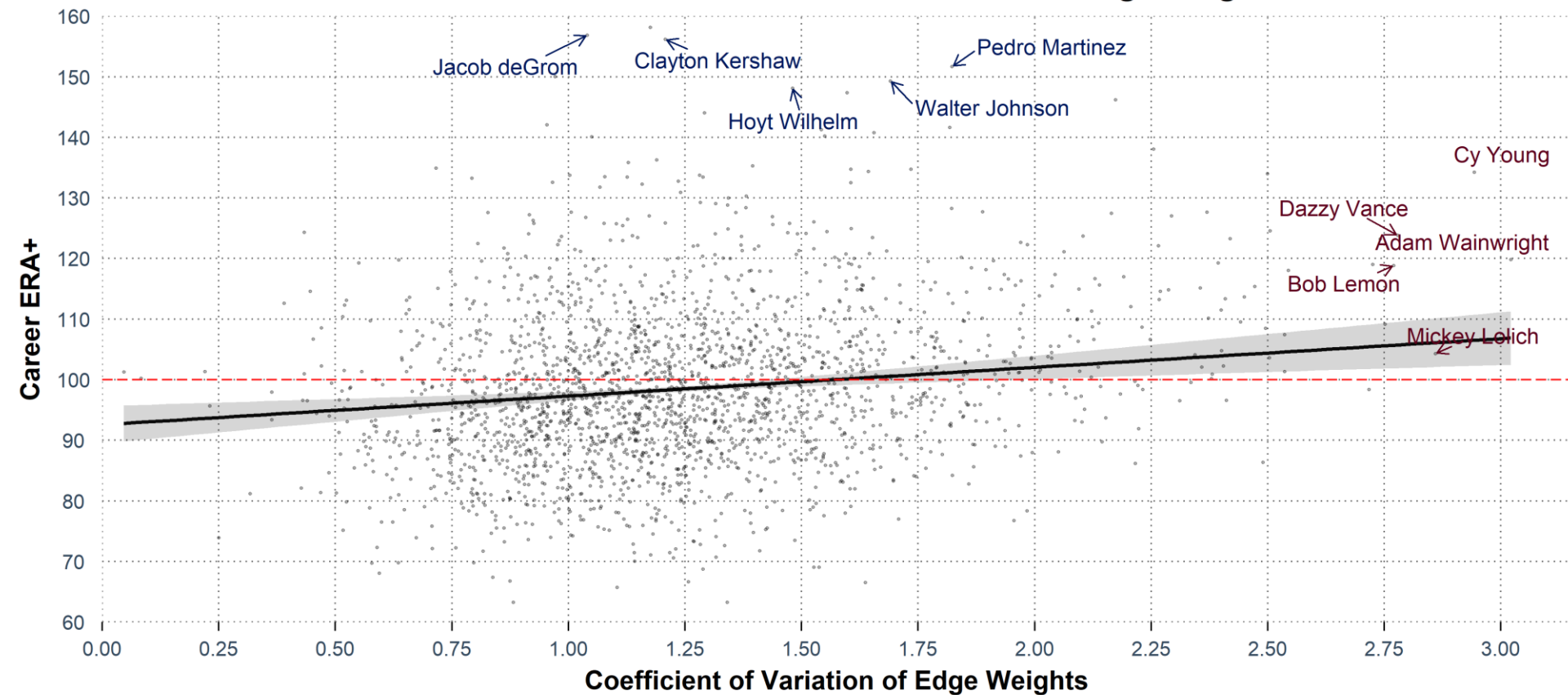
Note: Controlling for number of games started and neighbor ERA. League average ERA+ is 100.

Predicted ERA+ based on Coef. of Variation of Edge Weights



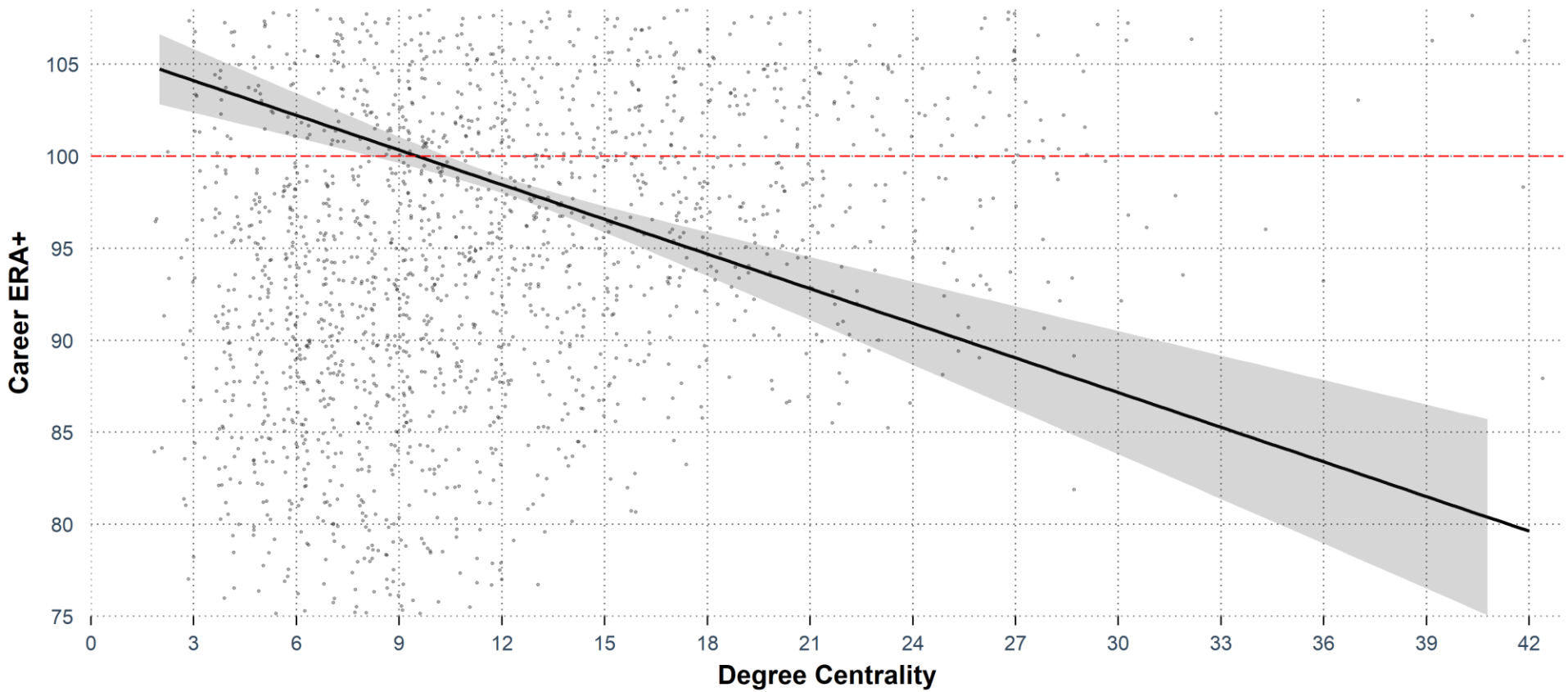
Note: Minimum 30 career starts. League average ERA+ is 100. Points represent actual observations.

Predicted ERA+ based on Coef. of Variation of Edge Weights



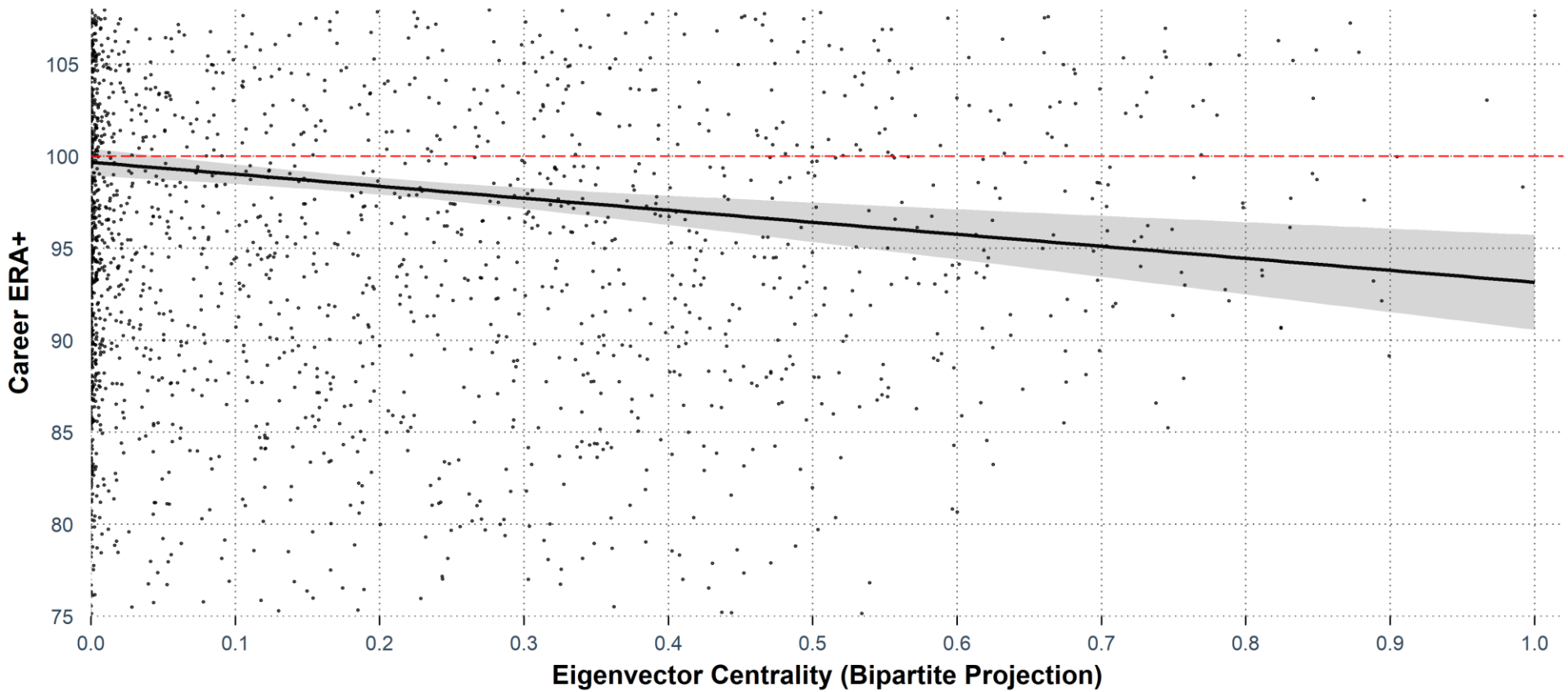
Note: Minimum 30 career starts. League average ERA+ is 100. Points represent actual observations.

Predicted ERA+ based on Degree Centrality



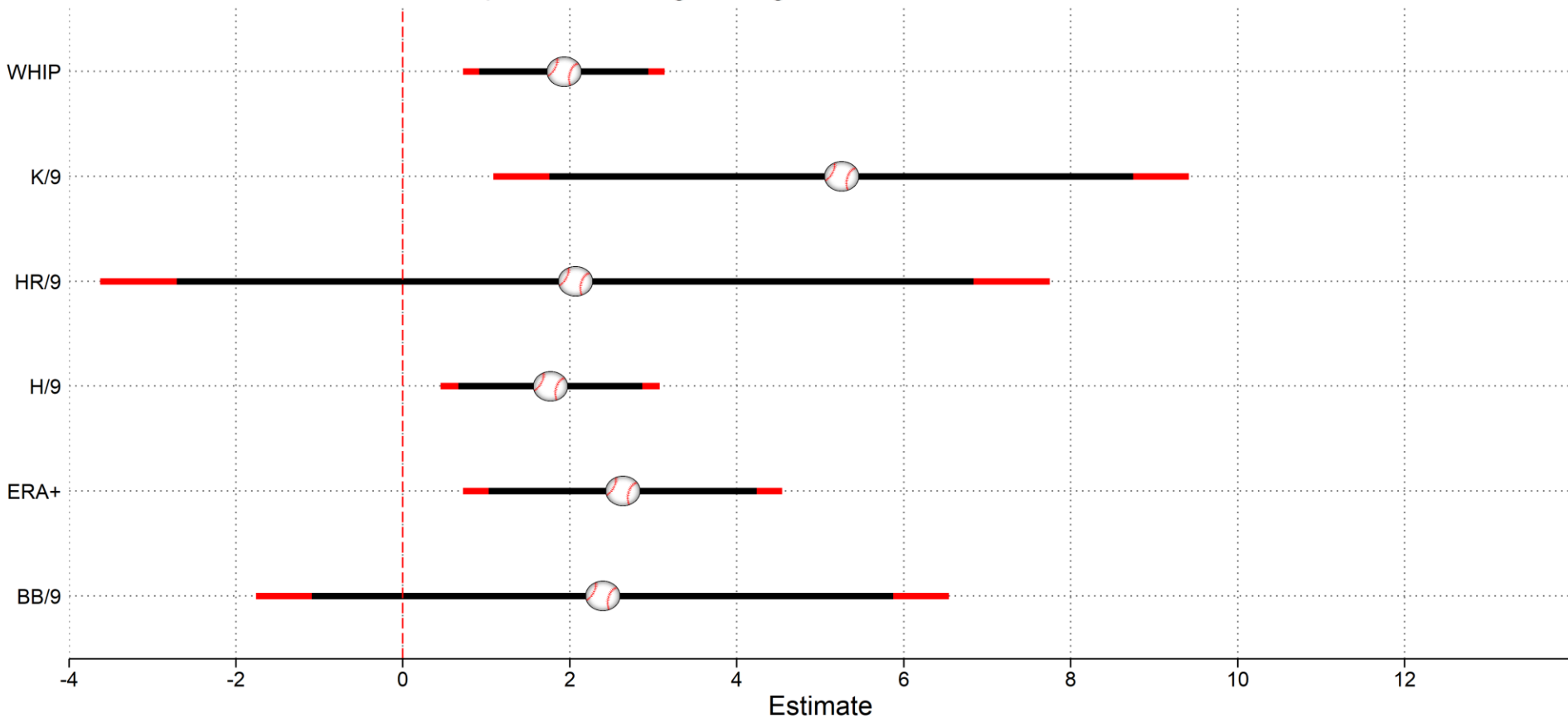
Note: Minimum 30 career starts. League average ERA+ is 100. Points represent actual observations.

Predicted ERA+ based on Eigenvector Centrality



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

Dispersion of Edge Weights and On-field Performance



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