A Bayesian two-stage framework for lineup-independent assessment of individual rebounding ability in the NBA

Nicholas Kiriazis [1] Christian Genest [1] Alexandre Leblanc [2]

[1]McGill University

[2]University of Manitoba

23 September, 2023

Existing measures for assessing rebounding

Existing measures for assessing rebounding

Scaled rebounding counts (per game, per 100, ...)

Existing measures for assessing rebounding

Scaled rebounding counts (per game, per 100, ...)

→ Doesn't account for number of missed shots

Existing measures for assessing rebounding

Scaled rebounding counts (per game, per 100, ...)

$\rightarrow$ Doesn't account for number of missed shots

Rebounding rates (proportion of misses)

Existing measures for assessing rebounding

Scaled rebounding counts (per game, per 100, ...)

→ Doesn't account for number of missed shots

Rebounding rates (proportion of misses)

→ Conveys information about other players on court

Existing measures for assessing rebounding

Scaled rebounding counts (per game, per 100, ...)

$\rightarrow$ Doesn't account for number of missed shots

Rebounding rates (proportion of misses)

$\rightarrow$ Conveys information about other players on court

Both these metrics only credit players who collect rebounds themselves

Existing measures for assessing rebounding

Scaled rebounding counts (per game, per 100, ...)

→ Doesn't account for number of missed shots

Rebounding rates (proportion of misses)

→ Conveys information about other players on court

Both these metrics only credit players who collect rebounds themselves

→ Engelmann (2016) suggests applying Rosenbaum (2004) APM framework to rebounding (more on this later!)

High-level ideas: rebounding

High-level ideas: rebounding

- Each player has some intrinsic, constant rebounding ability

High-level ideas: rebounding

- Each player has some intrinsic, constant rebounding ability
- Individual rebounding allocation is a function of all 10 abilities

High-level ideas: rebounding

- Each player has some intrinsic, constant rebounding ability
- Individual rebounding allocation is a function of all 10 abilities

We propose factorizing individual rebounding rates as follows:

$$\Pr(\text{A collects rebound}) =$$

$$\Pr(\text{A collects rebound} \cap \text{A's team collects rebound}) =$$

$$\underbrace{\Pr(\text{A collects rebound} \mid \text{A's team collects rebound})}_{\gamma-\textit{level}} \times$$

$$\underbrace{\Pr(\text{A's team collects rebound})}_{\beta-\textit{level}}.$$

High-level ideas: parameter reduction

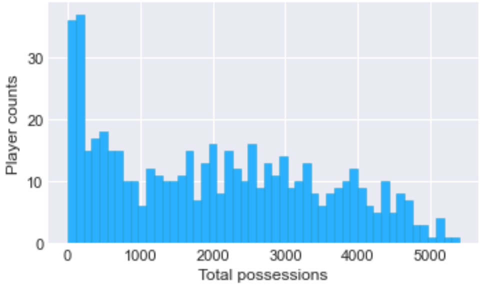


Figure: Distribution of possessions played by players in the 2020–21 NBA season.

High-level ideas: parameter reduction

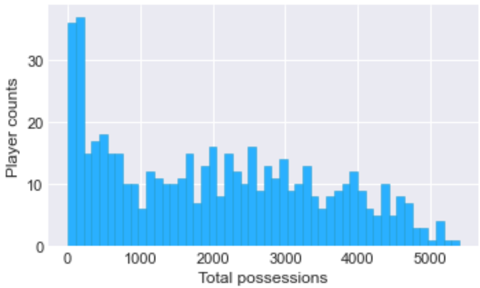


Figure: Distribution of possessions played by players in the 2020–21 NBA season.

1. Rosenbaum (2004) replaces all unusable players by a single parameter (the *Replacement Player*)

High-level ideas: parameter reduction

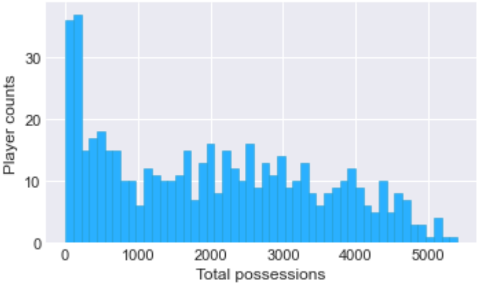


Figure: Distribution of possessions played by players in the 2020–21 NBA season.

1. Rosenbaum (2004) replaces all unusable players by a single parameter (the *Replacement Player*)
2. Sill (2010) instills an identical prior across all players

High-level ideas: parameter reduction

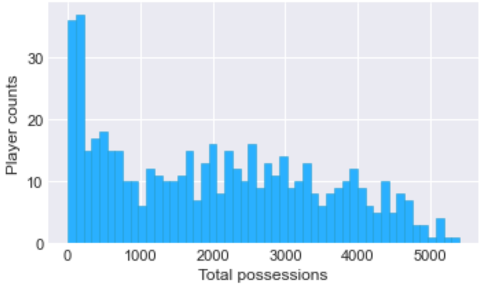


Figure: Distribution of possessions played by players in the 2020–21 NBA season.

1. Rosenbaum (2004) replaces all unusable players by a single parameter (the *Replacement Player*)
2. Sill (2010) instills an identical prior across all players

We propose a middle ground approach: group players who are similar. (See thesis for details)

Data

- Rebounding models: play-by-play data
- Technical note: not all rebounds are allocated to an individual player (referred to as *team rebounds*)

The data were accessed using the NBA API (Patel, 2023), for which the repository can be found at https://github.com/swar/nba_api

γ-level rebounding model

γ-level rebounding model

We are conditioning on the team having the rebound

γ-level rebounding model

We are conditioning on the team having the rebound

$\Rightarrow$ Treat each rebound as a multinomial random variable with 6 responses (5 possible players or the team)

γ-level rebounding model

We are conditioning on the team having the rebound

$\Rightarrow$ Treat each rebound as a multinomial random variable with 6 responses (5 possible players or the team)

- Can't use traditional multinomial regression

γ-level rebounding model

We are conditioning on the team having the rebound

$\Rightarrow$ Treat each rebound as a multinomial random variable with 6 responses (5 possible players or the team)

- Can't use traditional multinomial regression
- We have one common category across all lineups

γ-level rebounding model

We are conditioning on the team having the rebound

$\Rightarrow$ Treat each rebound as a multinomial random variable with 6 responses (5 possible players or the team)

- Can't use traditional multinomial regression
- We have one common category across all lineups

  $\rightarrow$ Model log-odds ratio between individual rebound and team rebound

γ-level rebounding model *cont.*

Therefore, for arbitrary player i and arbitrary lineup L

- Individual rebound probability: $p_i^L = \frac{e^{\gamma_i^D}}{1+e^{\gamma_1^D}+\cdots+e^{\gamma_5^D}}$
- Team rebound probability:

$$p_T^L = \frac{1}{1+e^{\gamma_1^D}+\cdots+e^{\gamma_5^D}} = 1 - p_1^L - p_2^L - p_3^L - p_4^L - p_5^L.$$

γ-level rebounding model *cont.*

Therefore, for arbitrary player i and arbitrary lineup L

- Individual rebound probability: $p_i^L = \frac{e^{\gamma_i^D}}{1+e^{\gamma_1^D}+\cdots+e^{\gamma_5^D}}$
- Team rebound probability:

$$p_T^L = \frac{1}{1+e^{\gamma_1^D}+\cdots+e^{\gamma_5^D}} = 1 - p_1^L - p_2^L - p_3^L - p_4^L - p_5^L.$$

We use MCMC with a flat prior instead of MLE

Review of APM and RAPM

Review of APM and RAPM

Rosenbaum's APM formula: $\hat{\beta} = (X^\top W^{-1} X)^{-1} X^\top Y$

- X: sparse matrix of substitution-less stints
- Y: vector of net ratings for the home team during each stint
- W: weight matrix to account for irregular number of possessions

Review of APM and RAPM

Rosenbaum's APM formula: $\hat{\beta} = (X^\top W^{-1} X)^{-1} X^\top Y$

X: sparse matrix of substitution-less stints

Y: vector of net ratings for the home team during each stint

W: weight matrix to account for irregular number of possessions

Positives:

- Box-score independent (this is what we are looking for!)

Review of APM and RAPM

Rosenbaum's APM formula: $\hat{\beta} = (X^\top W^{-1} X)^{-1} X^\top Y$

X: sparse matrix of substitution-less stints

Y: vector of net ratings for the home team during each stint

W: weight matrix to account for irregular number of possessions

Positives:

- Box-score independent (this is what we are looking for!)

Negatives:

- Assumes response contributions are linearly additive (probabilities?, diminishing returns?)
- Standard errors are large
- Doesn't distinguish between offense and defense

Review of APM and RAPM

Rosenbaum's APM formula: $\hat{\beta} = (X^\top W^{-1} X)^{-1} X^\top Y$

X: sparse matrix of substitution-less stints

Y: vector of net ratings for the home team during each stint

W: weight matrix to account for irregular number of possessions

Positives:

- Box-score independent (this is what we are looking for!)

Negatives:

- Assumes response contributions are linearly additive (probabilities?, diminishing returns?)
- Standard errors are large
- Doesn't distinguish between offense and defense

Sill's RAPM formula: $\hat{\beta} = (X^\top W^{-1} X + \lambda I)^{-1} X^\top Y$

λ: regularization parameter,

I: identity matrix.

$\Rightarrow$ Solves standard error issue problem by shrinking all players to common mean

β-level model: proposed mechanism

Logistic regression allows us to naturally oppose defensive rebounds (success) and offensive rebounds (failure):

$$p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}} \; .$$

β_i^D: defensive ability of player i,

β_j^O: offensive ability of player j,

p^L: probability of defensive rebound in lineup L.

β-level model: proposed mechanism

Logistic regression allows us to naturally oppose defensive rebounds (success) and offensive rebounds (failure):

$$p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}} \, .$$

β_i^D: defensive ability of player i,

β_j^O: offensive ability of player j,

p^L: probability of defensive rebound in lineup L.

$\rightarrow$ Natural parameter space

β-level model: proposed mechanism

Logistic regression allows us to naturally oppose defensive rebounds (success) and offensive rebounds (failure):

$$p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}} \,.$$

β_i^D: defensive ability of player i,

β_j^O: offensive ability of player j,

p^L: probability of defensive rebound in lineup L.

→ Natural parameter space

→ Diminishing returns

β-level model: proposed mechanism

Logistic regression allows us to naturally oppose defensive rebounds (success) and offensive rebounds (failure):

$$p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}.$$

β_i^D: defensive ability of player i,

β_j^O: offensive ability of player j,

p^L: probability of defensive rebound in lineup L.

→ Natural parameter space

→ Diminishing returns

→ Separates offense and defense

β-level model: proposed mechanism

Logistic regression allows us to naturally oppose defensive rebounds (success) and offensive rebounds (failure):

$$p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}\ .$$

β_i^D: defensive ability of player i,

β_j^O: offensive ability of player j,

p^L: probability of defensive rebound in lineup L.

→ Natural parameter space

→ Diminishing returns

→ Separates offense and defense

β-level model: proposed mechanism

Logistic regression allows us to naturally oppose defensive rebounds (success) and offensive rebounds (failure):

$$p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}} .$$

β_i^D: defensive ability of player i,

β_j^O: offensive ability of player j,

p^L: probability of defensive rebound in lineup L.

- $\rightarrow$ Natural parameter space
- $\rightarrow$ Diminishing returns
- $\rightarrow$ Separates offense and defense

We have two severe estimation problems: multicollinearity and unidentifiability

β-level model: multicollinearity

β-level model: multicollinearity

For rebounding, it doesn't make sense to assume everyone is equal a priori

β-level model: multicollinearity

For rebounding, it doesn't make sense to assume everyone is equal a priori

→ Individual rebounding rates probably provide a rough idea of β-ability

β-level model: multicollinearity

For rebounding, it doesn't make sense to assume everyone is equal a priori

$\rightarrow$ Individual rebounding rates probably provide a rough idea of β-ability

Individual rebounding rates should be reflected in choice of priors, but with appropriate variance

β-level model: multicollinearity

For rebounding, it doesn't make sense to assume everyone is equal a priori

$\rightarrow$ Individual rebounding rates probably provide a rough idea of β-ability

Individual rebounding rates should be reflected in choice of priors, but with appropriate variance

$\rightarrow$ We have an idea of the parameter ordering, but not of the scale

β-level model: unidentifiability

Recall the proposed model: $p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}$

β-level model: unidentifiability

Recall the proposed model: $p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}$

$\rightarrow$ We can add some constant c to every parameter (this is okay!)

β-level model: unidentifiability

Recall the proposed model: $p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}$

$\rightarrow$ We can add some constant c to every parameter (this is okay!)

More problematic case: assume we have equal defenders A and B, equal attackers C and D, but that we have only observed a subset of all possible match ups:

$$\ln\left(\frac{p_1}{1-p_1}\right) = \beta_A^D - \beta_C^O, \quad \ln\left(\frac{p_2}{1-p_2}\right) = \beta_B^D - \beta_D^O$$

β-level model: unidentifiability

Recall the proposed model: $p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}$

$\rightarrow$ We can add some constant c to every parameter (this is okay!)

More problematic case: assume we have equal defenders A and B, equal attackers C and D, but that we have only observed a subset of all possible match ups:

$$\ln(\frac{p_1}{1 - p_1}) = \beta_A^D - \beta_C^O, \quad \ln(\frac{p_2}{1 - p_2}) = \beta_B^D - \beta_D^O$$

$\rightarrow$ We can play with constants to make any player the best!

β-level model: unidentifiability

Recall the proposed model: $p^L = \frac{e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}{1 + e^{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O}}$

$\rightarrow$ We can add some constant c to every parameter (this is okay!)

More problematic case: assume we have equal defenders A and B, equal attackers C and D, but that we have only observed a subset of all possible match ups:

$$\ln\left(\frac{p_1}{1 - p_1}\right) = \beta_A^D - \beta_C^O, \quad \ln\left(\frac{p_2}{1 - p_2}\right) = \beta_B^D - \beta_D^O$$

$\rightarrow$ We can play with constants to make any player the best!

If we restrict the parameter space, we can limit how extreme any re-ordering can be

$\rightarrow$ How can we do that?

β-level model: re-phrasing the question

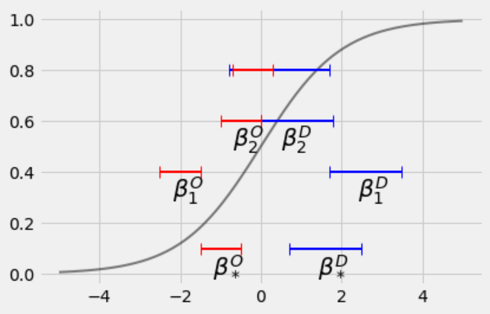


Figure: Equivalent parameterizations (1, 2, 3) to the true parameterization (*)

β-level model: re-phrasing the question

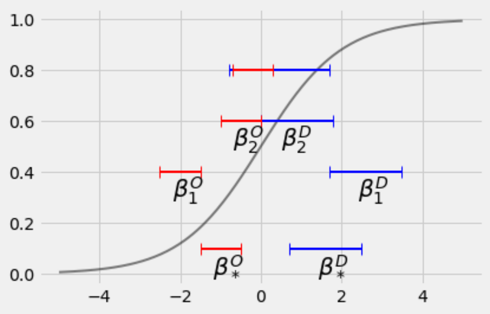


Figure: Equivalent parameterizations (1, 2, 3) to the true parameterization (*)

1. How wide do the offensive and defensive parameter ranges need to be to capture observed player variability?

β-level model: re-phrasing the question

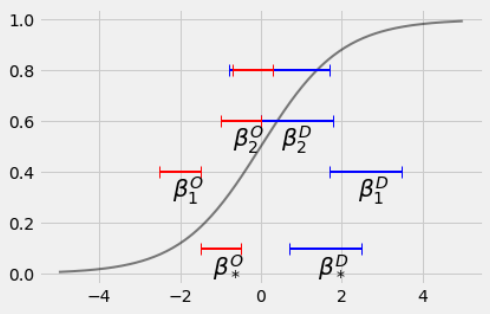


Figure: Equivalent parameterizations (1, 2, 3) to the true parameterization (*)

1. How wide do the offensive and defensive parameter ranges need to be to capture observed player variability?
2. How far apart do we need these clouds to be to express observed rebounding rates?

Answering Question 1: a thought experiment

Answering Question 1: a thought experiment

→ Suppose we could clone the best and worst β-level rebounders (β^D_{Max} and β^D_{Min})

Answering Question 1: a thought experiment

- $\rightarrow$ Suppose we could clone the best and worst β-level rebounders (β^D_{Max} and β^D_{Min})
- $\rightarrow$ Play them against an average offensive rebounding lineup

Answering Question 1: a thought experiment

→ Suppose we could clone the best and worst β-level rebounders (β^D_{Max} and β^D_{Min})

→ Play them against an average offensive rebounding lineup

→ Team defensive rebounding rate is probably less than 90% for the strong lineup, and greater than 50% for the weak lineup

Answering Question 1: a thought experiment

→ Suppose we could clone the best and worst β-level rebounders (β^D_{Max} and β^D_{Min})

→ Play them against an average offensive rebounding lineup

→ Team defensive rebounding rate is probably less than 90% for the strong lineup, and greater than 50% for the weak lineup

→ What is the "narrowest" support that allows for that to happen?

Answering Question 1: a thought experiment

→ Suppose we could clone the best and worst β-level rebounders (β^D_{Max} and β^D_{Min})

→ Play them against an average offensive rebounding lineup

→ Team defensive rebounding rate is probably less than 90% for the strong lineup, and greater than 50% for the weak lineup

→ What is the "narrowest" support that allows for that to happen?

$$\min \; \beta^D_{Max} - \beta^D_{Min}$$
$$\text{s.t.} \quad \frac{e^{5\beta^D_{\mathrm{Max}}}}{1+e^{5\beta^D_{\mathrm{Max}}}} - \frac{e^{5\beta^D_{\mathrm{Min}}}}{1+e^{5\beta^D_{\mathrm{Min}}}} \geq 0.4$$
$$\beta^D_{Max} \geq \beta^D_{\mathrm{Min}}$$

Answering Question 1: a thought experiment *(cont.)*

We can easily solve this using the Lagrangian multiplier

Answering Question 1: a thought experiment *(cont.)*

We can easily solve this using the Lagrangian multiplier

→ We find that $\beta^D_{Max} - \beta^D_{Min} = 0.340$

Answering Question 1: a thought experiment *(cont.)*

We can easily solve this using the Lagrangian multiplier

→ We find that $\beta^D_{Max} - \beta^D_{Min} = 0.340$

→ This is a very similar scale to individual rebounding rates!

Answering Question 1: a thought experiment *(cont.)*

We can easily solve this using the Lagrangian multiplier

→ We find that $\beta^D_{Max} - \beta^D_{Min} = 0.340$

→ This is a very similar scale to individual rebounding rates!

→ We similarly find that $\beta^O_{Max} - \beta^O_{Min} = 0.292$ (we allowed for a difference of 35% since offensive rebounding shows less variation)

Answering Question 1: a thought experiment *(cont.)*

We can easily solve this using the Lagrangian multiplier

→ We find that $\beta^D_{Max} - \beta^D_{Min} = 0.340$

→ This is a very similar scale to individual rebounding rates!

→ We similarly find that $\beta^O_{Max} - \beta^O_{Min} = 0.292$ (we allowed for a difference of 35% since offensive rebounding shows less variation)

We have an (approximate) upper bound on the size of each cloud!

Answering Question 2: re-defining the model

We can re-write the true parameterization in terms of any other valid parameterization, along with some shift in the "clouds":

$$p^L = \frac{e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}{1 + e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}$$

$$= \frac{\exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}{1 + \exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}$$

$$= \frac{\exp\{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O + \alpha\}}{1 + \exp\{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O + \alpha\}},$$

Answering Question 2: re-defining the model

We can re-write the true parameterization in terms of any other valid parameterization, along with some shift in the "clouds":

$$p^L = \frac{e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}{1 + e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}$$

$$= \frac{\exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}{1 + \exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}$$

$$= \frac{\exp\{\beta_1^{D} + \cdots + \beta_5^{D} - \beta_1^{O} - \cdots - \beta_5^{O} + \alpha\}}{1 + \exp\{\beta_1^{D} + \cdots + \beta_5^{D} - \beta_1^{O} - \cdots - \beta_5^{O} + \alpha\}},$$

→ Equivalent solution since it preserves ordering and has identical predictions

Answering Question 2: re-defining the model

We can re-write the true parameterization in terms of any other valid parameterization, along with some shift in the "clouds":

$$p^L = \frac{e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}{1 + e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}$$

$$= \frac{\exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}{1 + \exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}$$

$$= \frac{\exp\{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O + \alpha\}}{1 + \exp\{\beta_1^D + \cdots + \beta_5^D - \beta_1^O - \cdots - \beta_5^O + \alpha\}},$$

→ Equivalent solution since it preserves ordering and has identical predictions

→ We can force the clouds anywhere, as long as we include the α parameter to space them out accordingly

Answering Question 2: re-defining the model

We can re-write the true parameterization in terms of any other valid parameterization, along with some shift in the "clouds":

$$p^L = \frac{e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}{1 + e^{\beta_1^{*D} + \cdots + \beta_5^{*D} - \beta_1^{*O} - \cdots - \beta_5^{*O}}}$$

$$= \frac{\exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}{1 + \exp\{\beta_1^{*D} + \alpha_D + \cdots + \beta_5^{*D} + \alpha_D - (\beta_1^{*O} + \alpha_O) - \cdots - (\beta_5^{*O} + \alpha_O) + -5\alpha_D + 5\alpha_O\}}$$

$$= \frac{\exp\{\beta_1^{D} + \cdots + \beta_5^{D} - \beta_1^{O} - \cdots - \beta_5^{O} + \alpha\}}{1 + \exp\{\beta_1^{D} + \cdots + \beta_5^{D} - \beta_1^{O} - \cdots - \beta_5^{O} + \alpha\}},$$

→ Equivalent solution since it preserves ordering and has identical predictions

→ We can force the clouds anywhere, as long as we include the α parameter to space them out accordingly

→ We can (probably) learn the shift from the data

Formal β-level model formulation

We propose using the following hierarchical Bayesian framework:

$$\beta_i^D \mid DREB\%_i, \sigma \sim \mathcal{N}(DREB\%_i, \sigma^2),$$
$$\beta_j^O \mid OREB\%_j, \sigma \sim \mathcal{N}(OREB\%_j, \sigma^2),$$
$$\alpha \sim \text{improper uniform prior over } (-\infty, \infty),$$
$$Y_{L,k} \mid \beta_{i_1}^D, \ldots, \beta_{i_5}^D, \beta_{j_1}^O, \ldots, \beta_{j_5}^O, \alpha \sim \text{Bernoulli}(p^L),$$

where

$$p^L = \frac{e^{\beta_{i_1}^D + \cdots + \beta_{i_5}^D - \beta_{j_1}^O - \cdots - \beta_{j_5}^O + \alpha}}{1 + e^{\beta_{i_1}^D + \cdots + \beta_{i_5}^D - \beta_{j_1}^O - \cdots - \beta_{j_5}^O + \alpha}},$$

A simulation study

Details can be found in my thesis. The key points are:

A simulation study

Details can be found in my thesis. The key points are:

→ Sample sizes are comparable to that of one NBA season

A simulation study

Details can be found in my thesis. The key points are:

→ Sample sizes are comparable to that of one NBA season

→ We recover the original parameters by equally distributing the estimated shift amongst the ten players

A simulation study

Details can be found in my thesis. The key points are:

→ Sample sizes are comparable to that of one NBA season

→ We recover the original parameters by equally distributing the estimated shift amongst the ten players

→ We repeat a high variance and low variance version of the simulation

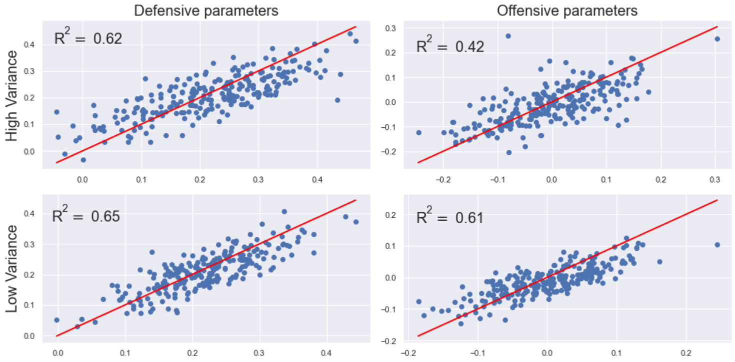
A simulation study

Details can be found in my thesis. The key points are:

→ Sample sizes are comparable to that of one NBA season

→ We recover the original parameters by equally distributing the estimated shift amongst the ten players

→ We repeat a high variance and low variance version of the simulation



Results - β-level estimates

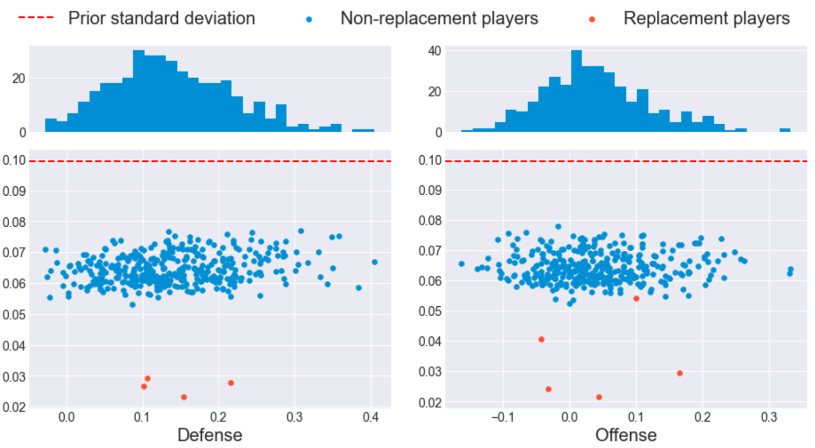


Figure: Standard deviations (y-axis) of each posterior distribution against the posterior mean (x-axis).

Results - Sanity check

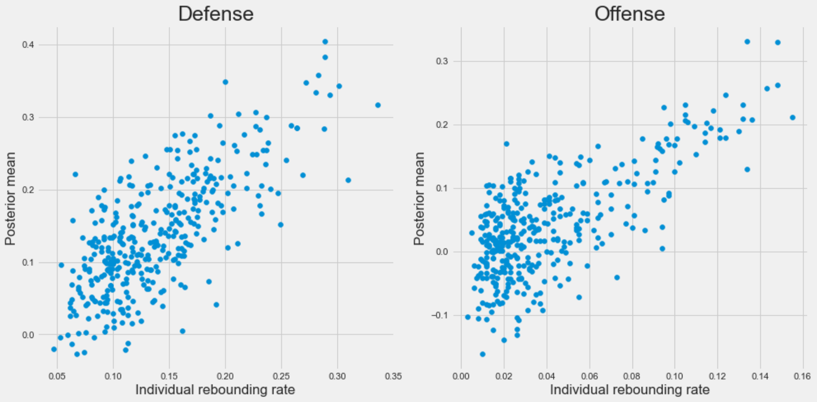


Figure: Posterior mean against prior mean.

Results - γ-level estimates

Figure: Standard deviations (y-axis) of each posterior distribution against the posterior mean (x-axis) of the γ-level parameters. Color represents the number of minutes of the player in question, and is a proxy for the number of multinomial observations used to estimate the parameter.

Results - Binned predictions

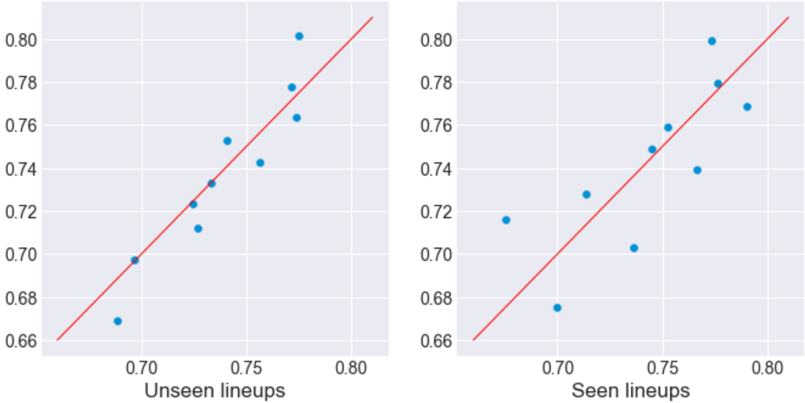


Figure: Predicted probabilities (y-axis) against observed probabilities (x-axis) for seen and unseen lineups during the 2021–22 NBA season. Note that the groups for the unseen lineups contain each about 1940 observations, and the seen lineup groups contain about 560 observations.

Results - Individual predictions

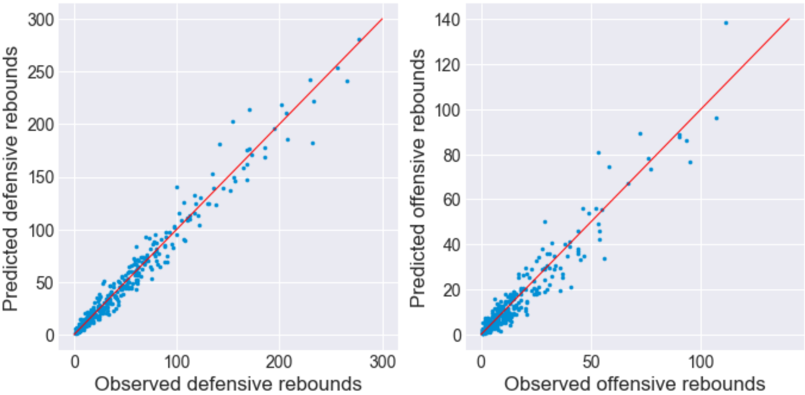
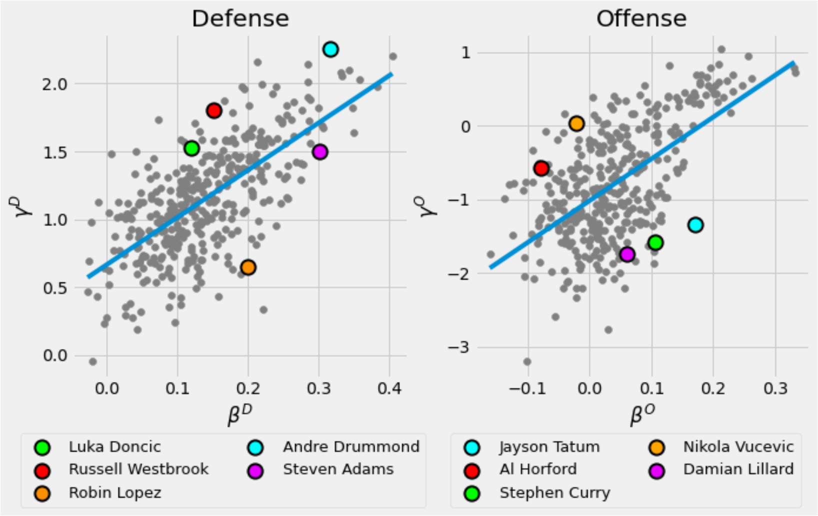


Figure: Two-stage predicted vs observed rebounding counts for individual players during the 2021–22 season.

Results - Practical example



Results - Trends

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders
  - $\rightarrow$ Forcing rotations and defense is out of position?

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders
  - → Forcing rotations and defense is out of position?
- Long-range threats are underrated offensive rebounders

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders
  - → Forcing rotations and defense is out of position?
- Long-range threats are underrated offensive rebounders
  - → Forcing long close outs?

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders
  - → Forcing rotations and defense is out of position?
- Long-range threats are underrated offensive rebounders
  - → Forcing long close outs?
- Ball-handlers are overrated defensive rebounders

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders
  - $\rightarrow$ Forcing rotations and defense is out of position?
- Long-range threats are underrated offensive rebounders
  - $\rightarrow$ Forcing long close outs?
- Ball-handlers are overrated defensive rebounders
  - $\rightarrow$ No box-out responsibilities?
  - $\rightarrow$ Teams funnel rebounds to their guards?

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders
  - → Forcing rotations and defense is out of position?
- Long-range threats are underrated offensive rebounders
  - → Forcing long close outs?
- Ball-handlers are overrated defensive rebounders
  - → No box-out responsibilities?
  - → Teams funnel rebounds to their guards?
- Some traditional centers are overrated offensive rebounders

Results - Trends

- Ball-dominant offensive engines are underrated offensive rebounders
  - → Forcing rotations and defense is out of position?
- Long-range threats are underrated offensive rebounders
  - → Forcing long close outs?
- Ball-handlers are overrated defensive rebounders
  - → No box-out responsibilities?
  - → Teams funnel rebounds to their guards?
- Some traditional centers are overrated offensive rebounders
  - → Guaranteed to collect rebounds because of proximity to basket?

Limitations

Limitations

Need to carefully think about blocked shots.

Limitations

Need to carefully think about blocked shots.

How important were the omitted covariates?

Further details

Thesis can (eventually) be found by searching by name here:
https://escholarship.mcgill.ca/

Questions and comments can be sent to me by email:
nicholas.kiriazis@mail.mcgill.ca

References

Engelmann, J. (2016). Possession-based player performance analysis in basketball (adjusted +/– and related concepts). In *Handbook of Statistical Methods and Analyses in Sports*. Chapman; Hall/CRC.

Patel, S. (2023). Nba api.

Rosenbaum, D. T. (2004). Measuring how NBA players help their teams win. Retrieved January 18, 2023, from http://www.82games.com/comm30.htm#_ftn1

Sill, J. (2010). Improved NBA adjusted +/- using regularization and out-of-sample testing. *Proceedings of the 2010 MIT Sloan Sports Analytics Conference*.