MATH 410 Project: Evaluating NBA Rebounding by Using Empirical Priors and Depedence Modelling

Nicholas Kiriazis

December 2021

Abstract

A Bayesian model is proposed to better estimate individual rebounding ability of NBA players during the 2020-2021 NBA season. Play-by-play data is used to extract positional rebounding rates for all players, and to account for small, out of position samples, a position based empirical beta prior distribution is used. Furthermore, the quantitative relationship between points scored/prevented and rebounds collected is modelled by testing the fit of multiple copula families for individual teams. Lastly, we estimate the point value of rebounds by treating the outcome of possession as the observations of a Multinoulli random variable.

1 Introduction

Rebounding has always had an important place in professional basketball: in the 1950-1951 NBA season, rebounds made their way onto official scoresheets (Basketball-Reference, n.d.), making them one of the earliest ever box score statistics to be recorded. The earliest stars of the game, such as George Mikan and Bob Pettit, were praised for their extraordinary rebounding ability, as were all great centers who followed. In the 1980s and 1990s, Dennis Rodman impacted the game almost exclusively by rebounding, and did so in a sufficiently convincing way to be inducted into the Hall of Fame. In 2016, Hassan Whiteside and Andre Drummond would both sign maximum contracts (Sportrac, n.d.), after being first and second respectively in “Rebounds per game” (RPG) that season.

This appreciation for rebounding has however led to costly mistakes by analysts as well as team managers. Throughout the 2015 NBA season, many analysts, most notably Charles Barkley, would undermine the success of the Golden State Warriors due to their preference for skill over size and rebounding, as they would go on to make the finals for five straight years, winning three championships along the way. Furthermore, after signing their maximum contracts, both Whiteside and Drummond would be traded from their respective teams, would both generate very little return, and their following contracts would both be worth the league minimum value.

These anecdotes highlight two fundamental issues with rebounding valuation: firstly, how should one go about evaluating an individual’s rebounding contributions, to avoid overvaluing players with a limited impact on team success. Secondly, although the consensus is that quality rebounding is an important asset when trying to win basketball games, there are very limited attempts to quantify the relationship between rebounding and winning.

1.1 Evaluating individual rebounding

The most popular assessment of individual rebounding is the traditional RPG metric, for which necessary data is available since as early as 1950 (“Offensive rebounds per game” and “Defensive rebounds per game” only became available in 1973, when NBA scorekeepers began distinguishing between the two types of rebounds). Although straightforward to calculate and easy to interpret, it is fundamentally flawed in that it does not account for how many rebounds were available for a given player to collect. To account for this, more sophisticated measures like "Offensive rebounding percentage" (OREB%) and "Defensive rebounding percentage" (DREB%) were established to account for the number of rebounding opportunities when evaluating individual rebounding. Furthermore, Engelmann has applied Rosenbaum's APM Framework to rebounding in order to credit players who allow for their teams to collect rebounds, but who do not collect rebounds themselves (Engelmann, 2017). More sophisticated measures to evaluate individual rebounding do exist but were built using data that has since become proprietary and are therefore limited to specific past seasons.

Although these measures capture individual rebounding ability with varying degrees of accuracy and rigor, they all share a common theme: they all measure individual rebounding production rather than individual rebounding ability. The distinction between ability and production is clearly illustrated by the following example.

During the 2018-2019 NBA season, Basketball-Reference listed Anthony Davis as a Center, and estimated his OREB% and DREB% to be approximately 9.9% and 27.7% respectively[1]. Before the start of the 2019-2020 season, Davis was traded to the Los Angeles Lakers, a team with multiple centers on the roster. To maximize lineup efficiency, the Lakers coaching staff opted to play Davis as a Power Forward (again according to Basketball-Reference). His estimated OREB% was 7.4% (a relative decrease of 25.3%) and his estimated DREB% was 22.1% (a relative decrease of 25.3%). Given that in 2020 Davis was voted an all-star, was named to the All-NBA First team, won the championship, and according to Vaci et al. (2019), was still young enough to still probably be on the upswing of his career, it is difficult to argue that Davis had simply regressed, as his rebounding statistics would suggest. A more likely explanation is that by virtue of playing Power Forward, he was no longer expected to exert as much energy rebounding.

While it is theoretically possible to measure a player's rebounding rate when occupying a specific position, we encounter the issue of small sample sizes: players tend to prefer a position, meaning their number of positional rebounding opportunities tends to be imbalanced. Furthermore, some of the most effective lineup strategies tend to be used sparingly during the regular season to minimize the risk of injury and limit visibility to opponents, which means the most meaningful games tend to follow drastically different positional strategies (for example, the original Golden State Warriors Death Lineup more than doubled their minutes per game in the playoffs, according to the NBA records, which are available at https://www.nba.com/stats/lineups/traditional/). The Bayesian model proposed in Section 3 makes it possible to predict out-of-position rebounding rates by using a position-based beta prior distribution to smooth empirical measurements.

1.2 Quantifying the relationship between rebounding and point differential

Despite there being a consensus that rebounds are important, and that a team should seek to maximize their rebounding numbers to improve their win probability, rebounds are not "worth" anything in the sense that they don't directly contribute to points. This makes it difficult to quantify how rebounding contributes to team success.

An early attempt to quantify the link between rebounds and win percentage was made by Dean Oliver (2001) by simply calculating the win percentage of teams who had outrebounded their opponents, showing that the best defensive rebounding teams win 71.7% of the time. However, by Oliver's own admission, the conclusion that defensive rebounding was the key to winning more games was ambiguous due to the strong correlation between defensive rebounds and opponent shooting percentage.

[1]Measuring OREB% and DREB% exactly requires access to Play-by-Play data, which the NBA only began recording in the 1996-1997 season. A formula exists for estimating these statistics from box score data, and for the sake of consistency across seasons, is used for all years on Basketball-Reference. The formula is not relevant in the present statistical analysis.

More recently, attempts to quantify the value of rebounds have been made by using rebounds as predictors in linear models for net rating (Myers, 2020). Although they do technically allow for a point value to be assigned to rebounds, they are limited in that they make strict distributional assumptions and are difficult to practically interpret independently of other statistics.

In this report, a semiparametric copula-based model is proposed to quantify the relationship between defensive rebounding and transition offense, while the value of the different types of opportunities are estimated with a Multinoulli random variable.

2 Data

The bulk of the data used to estimate model parameters comes from the Play-by-play logs of 1013 regular season games from the 2020-2021 NBA season (93.8% of all regular season games that year) and were collected using the nba-api. A small sample of the raw, unannotated data is found in Table 1. The intention was to use all 1080 games from that season, but significant programming logic needed to be implemented to divide each log into individual possessions, meaning that human error in scorekeeping and/or very convoluted sequences of events became very difficult to manage (for example, some games do not include the opening tip in their Play-by-play log). The 67 unparsable games did not systematically exclude any teams, so they were simply discarded.

Time	Home Description	Neutral Description	Visitor Description
12:00		Start of 1st Period (2:11 PM EST)	
12:00	Jump Ball Wood vs. Brown: Tip to Dort		
11:39			MISS Dort 26' 3PT Jump Shot
11:38	Tate REBOUND (Off:0 Def:1)		

Table 1: Example of play by play data taken from NBA.com

2.1 Positional rebounding data

Although players are technically listed at a position (either PG for Point Guard, SG for Shooting Guard, SF for Small Forward, PF for Power Forward and C for Center), the position a player occupies during any given possession depends on who else is on the court. It was assumed that every lineup consisted of a PG, SG, SF, PF and C, regardless of the players' listed positions. Therefore, the first step in the data collection process was to determine the actual on-court positions of every lineup used in the 1013 games. Listed positions were collected from the following sources:

1. ESPN.com: ESPN assigns each player a single position. This value will be referred to as a player's primary position.

2. Sports.yahoo.com: to operate their fantasy leagues, Yahoo Sports assigns players all the positions that they realistically played during the season. These positions are updated during the season, as coaching strategies evolve. The data was collected in July 2021, well after the end of the 2020-2021 season. This list of positions will be referred to as a player's possible positions.

3. Stats.nba.com: players minutes per games were used in the very rare case of positional equivalencies.

An example of the aggregated data can be found in Table 2.

NAME	ESPN	Yahoo	Minutes
De'Anthony Melton	SG	PG-SG	1045
Desmond Bane	SG	PG-SG	1519
Dillon Brooks	SF	SG-SF	1997

Table 2: Example of the aggregated position data for a few members of the Memphis Grizzlies

Using the data from the aforementioned sources, the following methodology was used to assign each player their position within each 5-man lineup:

1. Check if the five primary positions of the players fill out the five required positions. If yes, these assigned positions are used, otherwise, proceed to the next step.
2. Perform a backtracking algorithm that tries to fill out the five required positions. If exactly one lineup was found, then those assigned positions are used, otherwise, proceed to the next step.
3. There were exactly 13 combinations of possible positions in the dataset. The combinations were sorted into the following order:

   PG<PG-SG<SG<PG-SG-SF<SG-SF<SF<SG-SF-PF<SF-PF-C<PF<PG-PF-C[2]<PF-C<C

   Then given a 5-man lineup, the smallest combination was labelled PG, the second smallest was labelled SG, third smallest was labelled SF, second largest was labelled PF and the largest was labelled C. In the event of a tie, the player with more total minutes was assumed to be playing the position "closest" to their primary position, the rationale being that a player with more visibility would have more accurate positional labelling. These instances were very rare.

Examples for each method are given in Appendix A.

Although the Bayesian model only distinguishes between Guards (G), Forwards (F) and Centers (C), the use of five positions in the labelling process allowed for more granularity when dealing with unconventional lineups. After the algorithms had terminated and had attributed the five positions, PG and SG were grouped to from Guards, SF and PF were grouped to form Forwards, and Centers were left untouched.

With the positions of the players assigned, we simply counted the number of rebounds and rebounding opportunities available to each player at each position (a rebounding opportunity was simply a missed field goal by your own team when evaluating offensive rebounding or by the opposing team when evaluating defensive rebounding). Rebounds and rebounding opportunities following missed free throws were omitted because players standing along the key are at a distinct advantage. Furthermore, team rebounds were simply ignored.

2.2 Collecting team rebounding data

For team level rebounding, individual rebounding numbers were simply added up. Team rebounding opportunities were simply the total number of missed shots.

2.3 Collecting team transition rate data

The traditional definition of a transition offense opportunity is an opportunity that consists of a quick scoring attempt taking place before the defensive team is positioned (Merriam-Webster, 2021). Although spectators tend to more or less uniformly recognize transition offense when watching a game, the Play-by-play data is fundamentally limited in that there is no way to extract information about the position of the defense. Because of this, only the idea of "quickness" was accounted for when distinguishing between transition opportunities and non-transition opportunities. The following conditions were all required for a scoring attempt (i.e. an attempted field goal or an attempted free throw) to be considered a transition opportunity:

- An attempt is considered a transition opportunity if no stoppages of play occurred in between the attempt and the start of the possession (i.e. no timeouts, substitutions, non-shooting fouls...). Note that intentional take fouls were not considered a transition opportunity unless they resulted in free throws, the justification being that there is no obvious way to assign a point value to a non-shooting foul.

[2]Only Dean Wade, a 6'9 player for the Cleveland Cavaliers, had this combination of possible positions. Given his height, it was arbitrarily chosen to move him towards the end of the queue rather than towards the front.

- An attempt following a live inbound is considered a transition opportunity if it takes place within 6 seconds of the inbound.
- An attempt following a non-inbound start of possession is considered a transition opportunity if it occurs within 8 seconds of the start of the possession.

The distinction of transition and non-transition offensive opportunities produced by the above conditions tended to over-estimate the number of transition opportunities to empirical measurements made by humans in a very small sample of games. There were two main reasons for this over-estimation: firstly, some stoppages are not included in the play-by-play data (for example, non-turnover out of bounds, equipment malfunctions…). Secondly, the duration of a possession is only a proxy for it being a true transition opportunity. An argument could be made to top for a more conservative cut-off, but in that case we would simply miss some transition opportunities. The values of 6 and 8 seconds were selected after testing multiple values in a very small sample size of games that were verified by a human. Furthermore, the cutoff time following live inbounds was shortened because in the limited number human-verified games, defensive teams had more time to get into position while the ball was being inbounded, which meant that unless the offensive team shot extremely quickly, they were probably not getting a transition opportunity.

Large scale human efforts would have been required for more objective selection criteria, but the heuristic conditions also has some quantitative justification: as one would expect, the above definitions led to transition opportunities having a much higher expected point value. The values for the observed data are given in Table 8, in Section 4.

2.4 Collecting opportunity value data

For each of the 1013 games in the dataset, for both home and away teams, we distinguished three types of scoring opportunities:

1. **Transition opportunities**

   The exact definition is given in Section 2.3. One technical note it was technically possible to have multiple transition scoring opportunities on a single possession (when an offensive rebound is very quickly collected). When this happened, all shots were considered as a single transition opportunity.

2. **Second chance opportunities**

   These were defined as any scoring opportunity following an offensive rebound. We assumed that the point value of scoring chances after an offensive rebound was identically distributed regardless of whether they came after the first offensive rebound, the second offensive rebound… (hence "second" is a misnomer for this analysis).

3. **Regular opportunities**

   These were defined as the first scoring chance of possessions that did not result from a transition opportunity.

The data collection process consisted simply of identifying these opportunities in the dataset and measuring their point value. In total, there were seven values found across the 202 667 opportunities of the dataset: 0, 1, 2, 3, 4, 5 and 6. Although most possessions fall between 0 and 4 inclusively, it is technically possible to have more points because of unsportsmanlike fouls. This is extremely rare: there were only 23 5-point opportunities and 4 6-point opportunities. It was decided to simply count these as 4-point opportunities so that density plots would be more legible when fitting the Multinoulli random variables.

Furthermore, one key practical note to consider is that possessions that resulted in a turnover were omitted from the transition opportunities and the regular opportunities: by looking at the time stamps in the play-by-play data, there is no way to tell if the turnover took place early in the in a non-transition opportunity or was the result of a failed attempt at beginning the offensive transition.

Since this is only problematic when trying to compare transition opportunities to regular opportunities, it was deemed a non-issue in the case second of chance opportunities.

Also note that the computed number of opportunities is slightly larger than the true number of opportunities because occasionally, a single opportunity was labelled as both a second chance opportunity and a transition opportunity. The rationale behind allowing the overlapping of categories is that the opportunity was created by both the willingness to get out in transition and by the offensive rebounding ability.

3 Statistical Models

3.1 Bayesian model for positional rebounding

3.1.1 Prior distributions

Note: the model described in this section will apply to defensive rebounding. The exact same logic also applies to offensive rebounding, but is omitted for readability.

Let D_{ip} be a random variable denoting the number of the defensive rebounds collected by player i while playing position p.
Let DO_{ip} be a random variable denoting the number of the defensive rebounds collected by player i while playing position p.
Let p_{ip} denote the defensive rebounding rate of player i while playing position p. Obviously, p_{ip} is unknown.

We assume that conditionally on the number of rebounding opportunities and the rebounding rate, the number of rebounds collected follows a binomial distribution:

$$D_{ip} \mid DO_{ip}, p_{ip} \sim Binomial(DO_{ip}, p_{ip}) \tag{1}$$

A position-based beta prior distribution is proposed for each one of G, F, and C. The parameters of each prior distribution are estimated from a pool of players who have a significant number of defensive rebounding opportunities at the specified position. The threshold for minimal number of opportunities for selected players comes from the Normal approximation for confidence intervals of parameters in binomial distributions:

$$\hat{p} + z\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \tag{2}$$

An in depth justification for the choice of threshold, n, as well as justification for the choice of z, is given in Section 4.

The beta distribution was fit to each group of empirical defensive rebounding rates. The six parameters of the prior distributions were estimated by numerically maximizing the log-likelihood. Let X_{i} denote all the retained empiricial rebounding rates. We therefore maximize:

$$\mathcal{L}(\alpha, \beta \mid X) = (\alpha - 1)\sum_{i=1}^{N} \ln(X_i) + (\beta - 1)\sum_{i=1}^{N} \ln(1 - X_i) - N \times \ln(B(\alpha, \beta)) \tag{3}$$

Standard errors were calculated using the Fisher Information.

3.1.2 Posterior distributions

Since DO_{ip} was assumed to be a fixed value, the posterior distribution of p_{ip} can be computed using Bayes' rule as

$$f(p_{ip} \mid D_{ip}, DO_{ip}) = g(D_{ip} \mid p_{ip})f(p_{ip}) \tag{4}$$

where f is the prior density of the parameter and g the density of the data.

Furthermore, given that the beta distribution is conjugate to the binomial likelihood function (see Bolstad and Curran, 2016), it is known that the posterior distribution is given by

$$p_{ip} \sim Beta(D_{ip} + \alpha_p, DO_{ip} - D_{ip} + \beta_p), \tag{5}$$

where α_p and β_p are the parameters of the prior distribution associated with position p.

This also implies that the predictive distribution for the number of positional rebounds collected will follow a beta-binomial distribution.

3.2 Modelling the dependence between rebounding and transition

3.2.1 Practical justification for covariate selection

To model the dependence between rebounds and transition, each team's "perspective" was treated as an individual game (i.e. even though 1013 games were played, the dataset for dependence modelling contained 2026 games). Therefore, six measurements were available for each entry:

- TR_{in} := Transition rate (the proportion of opportunities that were transition opportunities)
- OTR_{in} := Opponent transition rate
- DR_{in} := Defensive rebounding rate
- ODR_{in} := Opponent defensive rebounding rate
- OR_{in} := Offensive rebounding rate
- OOR_{in} := Opponent offensive rebounding rate

For every measurement, the subscript refers to n^{th} game of the i^{th} team.

The use of rate measurements rather than raw count data has a duall purpose: it allows for model predictions to be independent of pace of play, and it eases dependence modelling by avoiding the copula unidentifiability issue described by Genest and Nešlehová (2007).

Note that since team rebounds were excluded from the data, there is not perfect dependence between DR_{in} and OOR_{in} (or ODR_{in} and OR_{in}). However, these variables are still strongly correlated, as shown in Figure 1.

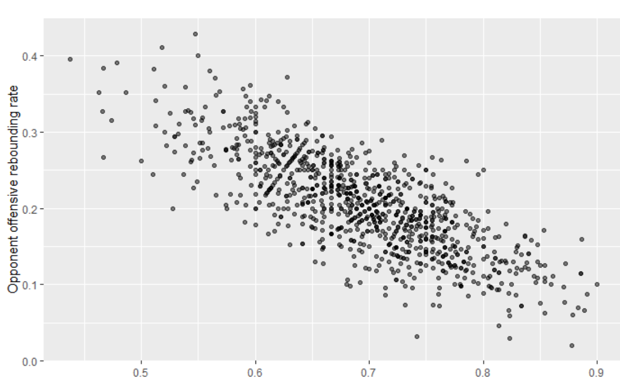


Figure 1: Scatter plot of defensive rebounding rate against opponent offensive rebounding rate. As one would expect, there is a strong correlation.

From a practical standpoint, there are two dependence patterns worth exploring:

I. Does collecting defensive rebounds allow a team to increase their number of transition offense opportunities?

II. Does aggressively pursuing offensive rebounds weaken a team's defense by giving the opposition more transition opportunities?

However, given that two teams opposing rebounding numbers are not perfectly dependent, there are in fact four quantifiable relationships in play:

A. The relationship between DR_{in} and TR_{in}

B. The relationship between OOR_{in} and TR_{in}

C. The relationship between OR_{in} and OTR_{in}

D. The relationship between ODR_{in} and OTR_{in}

Obviously, the quantifiable relationships (A.) and (B.) are both tied to the practical dependence pattern I, and similarly, the quantifiable relationships (C.) and (D.) are both tied to the practical dependence pattern II. The fact that two measurable quantities were tied to a single dependence pattern allowed for the double checking of significance: one would hope that a team with significant positive correlation in (A.) would also have significant negative correlation in (B.) and vice versa. Similarly, one would hope that teams with a significant positive correlation in (C.) would also have significant negative correlation in (D.) (and vice versa).

Therefore, only teams with "doubly" significant correlation at the 5%-level were retained. Correlation was measured with both Spearman's rho and Kendall's tau. There were no instances amongst retained teams of one measure being significant while the other was insignificant.

In the case of practical dependence pattern I, the dependence was modelled between DR_{in} and TR_{in}, as opposed to between OOR_{in} and TR_{in}. Although both should theoretically be similar, the decision was made for interpretative purposes: in a practical setting, it is more natural for a team to concern themselves with collecting defensive rebounds than it is for them to think about preventing their opponents from collecting offensive rebounds.

In the case of practical dependence pattern II, the dependence was investigated in a similar matter as pattern I, but no model was constructed. There were two reasons for this:

- The level of correlation was too small to be meaningful in practice.
- The point value of an opponent's possession obviously depends on the opponent in question, which means that the importance of a team's rebounding would need to be quantified on an opponent-by-opponent basis.

Note that dependence investigation for the pattern II is still included in Section 4 to illustrate these points.

3.2.2 Choice of dependence models

From Sklar's Theorem, it is known that the joint distribution of DR_{in} and TR_{in} can be stated in terms of copula that links the univariate marginal distributions. Furthermore, the dependence structure is invariant under monotonic transformations of the marginals. These ideas are both captured by the formal expression of the theorem. Let F_{DR} and F_{TR} denote the cdfs of DR_{in} and TR_{in} respectively. We have

$$\mathcal{C}(u_1, u_2) = \Pr[F_{DR}(DR_{in}) \leq u_1, F_{TR}(DR_{in}) \leq u_2]. \tag{6}$$

Semiparametric copulas were used to model dependence and were estimated using rank-based techniques, meaning that the shape of the marginals was of no interest as far as dependence analysis was concerned.

The fit of the following copula families was evaluated for each retained team:

- Clayton
- Frank
- Gumbel
- Gaussian

These families were selected for two main reasons:

1. The ease with which they can be constructed and the ease with which they can be interpreted make them ideal for a field where decisions are often made by non-statisticians.
2. Given that there are so few datapoints in the dataset (the average team only has 67.5 games), it was difficult to unequivocally characterize a particular dependence pattern. The variety of dependence patterns in the above copula families allows for expert opinion on how to handle tail dependence to be incorporated into the model.

Given that all the families contained a single dependence parameter, the parameter of each family was estimated by both the inversion of Kendall's tau and of Spearman's rho, as suggested by Genest and Favre (2007). Both methods produced very similar point estimates for the parameters, but occasionally, confidence intervals were of drastically different sizes due to ties in the observations. The standard errors were estimated as shown by Genest, Ghoudi and Rivest (1995).

3.3 Multinoulli model

For each team/opportunity type combination, the parameters of a Multinoulli were estimated using the MLE, which is simply the sample proportion. Standard errors were computed using the Normal approximation, given by Formula 2.

Note that the Multinoulli distribution was used rather than the multinomial distribution because rarely is the number of transition opportunities and second chance opportunities large enough to approximate the point distribution using the Normal approximation. We could therefore use the following recursive formula to get the exact point distribution:

$$\Pr(TP = i \mid TO = k) = \sum_{k=0}^{4} p_k \Pr(TX = i - k \mid TO = j - 1) \tag{7}$$

where TP denotes the total number of points scored,
TO denotes the total number of opportunities,
p_k denotes the probability of scoring k points on a single opportunity.

4 Model estimation

4.1 Bayesian model estimation

Prior distributions were constructed from the empirical distributions of rebounding rates amongst players having collected a significant number of rebounding opportunities. Therefore, the first step in the Bayesian modelling process was to determine which players should be retained for prior construction by setting a minimal number of opportunities. The threshold for retention was arbitrarily set as follows: a 95% confidence level that the player's observed rebounding rate be within 3% of his underlying rebounding rate. Although in theory stricter criteria for retention would have been preferable (especially for offensive rebounding rates, which tend to be much lower that defensive rebounding rates), in practice, the number of retained players needed to be considered. These were the strictest conditions that retained at reasonable number of players for each of the six priors.

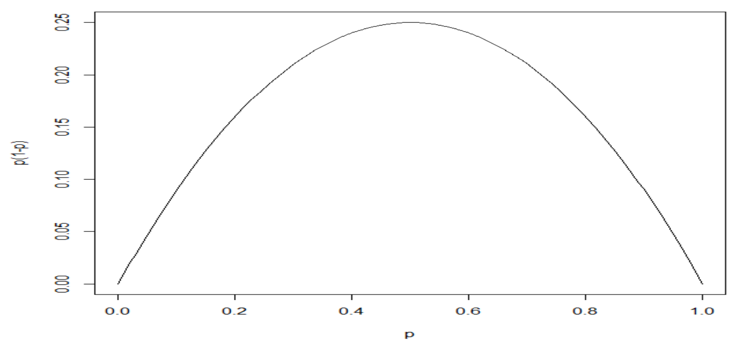
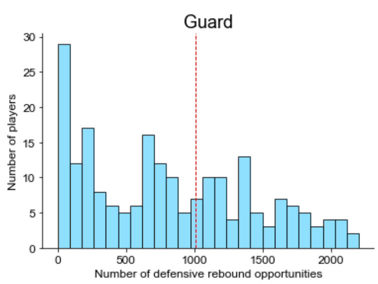
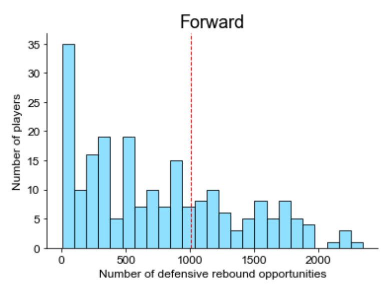


Figure 2: Plot of p against $p(1-p)$, which is proportional to the interval width. It is obvious from the plot that the maximum is found at $p = 0.5$.

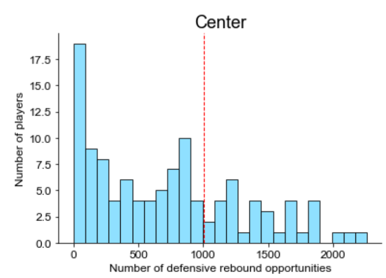
Furthermore, the confidence interval for a binomial proportion is largest when the success probability is equal to 0.5 (assuming a fixed quantile and number of attempts). This is illustrated in Figure 2. Although theoretically ideal to calculate the number of required opportunities for this "worst-case" scenario, such conservative calculations were not necessary in practice: according to Basketball-Reference, the highest ever estimated single season defensive rebounding rate was estimated to be 37.98% (recorded by Reggie Evans in the 2012-2013 season) and the highest ever estimated single season defensive rebounding rate was estimated to be 20.83% (recorded by Dennis Rodman in the 1994-1995 season). Note that these estimates do not exclude rebounds off free throws, meaning they are probably abundantly cautious upper bounds for the observations in the dataset. Using Formula 2, the minimum number of defensive rebounds was calculated to be 1006 and the minimum number of offensive rebound opportunities to was calculated to be 709. For the defensive rebounding priors, 80 guards, 69 forwards, and 33 centers were retained. For offensive rebounding priors, 114 guards, 100 forwards, and 54 centers were retained. Histograms illustrating the distribution of rebounding opportunities are included in Figure 3.



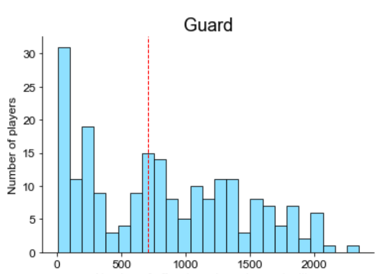
Guards - DREBS



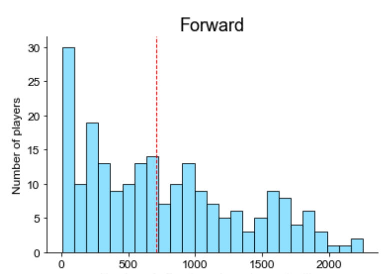
Forwards - DREBS



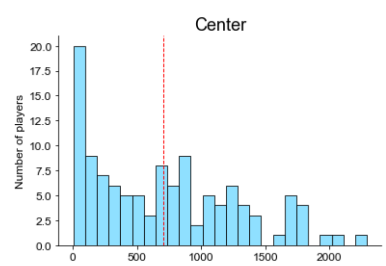
Centers - DREBS



Guards - OREBS



Forwards - OREBS



Centers - OREBS

Figure 3: Histograms of the number of rebounding attempts for players based on position and rebound type. Players to the right of the red dotted line had sufficiently many rebounding opportunities to be retained for prior estimation.

Two parameter beta distributions were fitted to each of the six groups by maximizing the log-likelihood, and their fit was evaluated with both graphical (Empirical/theoretical density plots, Q-Q plots, empirical/theoretical cdf plots, P-P plots) and formal tests using the Kolgomorov-Smirnov test and the Cramér–von Mises test, whose statistics are defined respectively as:

$$D_n = \sup_x |F_n(x) - F(x)|, \tag{8}$$

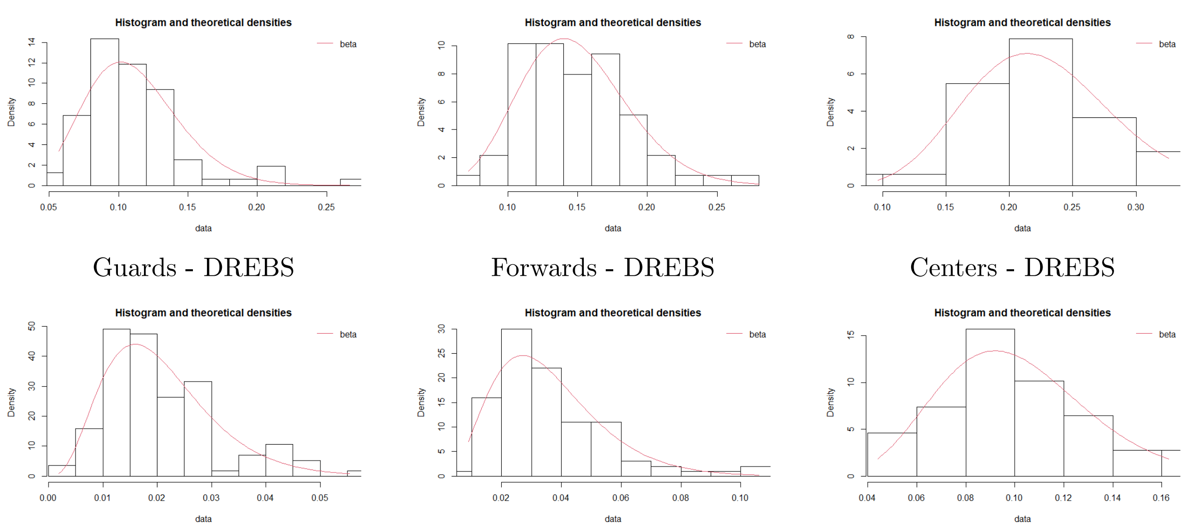
$$\omega^2 = \int_{-\infty}^{\infty} [F_n(x) - F(x)]^2 dF(x), \tag{9}$$

where F_n is the empirical cdf and F the fitted cdf.

In each of the six instances, the fitted beta distribution could not be rejected at the 5% level. Full results for formal goodness of fit testing, as well as parameters estimates and standard errors, are given in Table 3. Figure 4 contains the pdfs of the fitted distributions and histograms of the retained players rebounding rates. As one would expect, Centers tend to collect more rebounds than Forwards, who themselves tend to collect more rebounds than Guards.

Prior	α	α SE	β	β SE	D_n	KS p-value	ω^2	CVM p-value
G-OFF	4.16	0.53	196.5	26.59	0.0682	0.664	0.1212	0.491
F-OFF	3.82	0.52	102.3	14.80	0.0825	0.503	0.1323	0.448
C-OFF	9.65	1.82	86.37	16.73	0.0595	0.985	0.0241	0.991
G-DEF	9.48	1.47	75.99	12.09	0.0796	0.661	0.1215	0.490
F-DEF	12.63	2.12	72.41	12.37	0.0679	0.885	0.0504	0.874
C-DEF	12.08	2.93	41.72	10.28	0.0819	0.966	0.0366	0.952

Table 3: Parameters of the prior beta distributions, their standard errors, and the results of formal goodness of fit testing



Guards - DREBS

Forwards - DREBS

Centers - DREBS

Guards - OREBS

Forwards - OREBS

Centers - OREBS

Figure 4: Fitted densities of each of the six priors

Furthermore, the empirical first and second moments were computed and compared to the theoretical moments of the fitted distributions, which are given respectively by the following equations:

$$\bar{x} = \frac{\alpha}{\alpha + \beta} \tag{10}$$

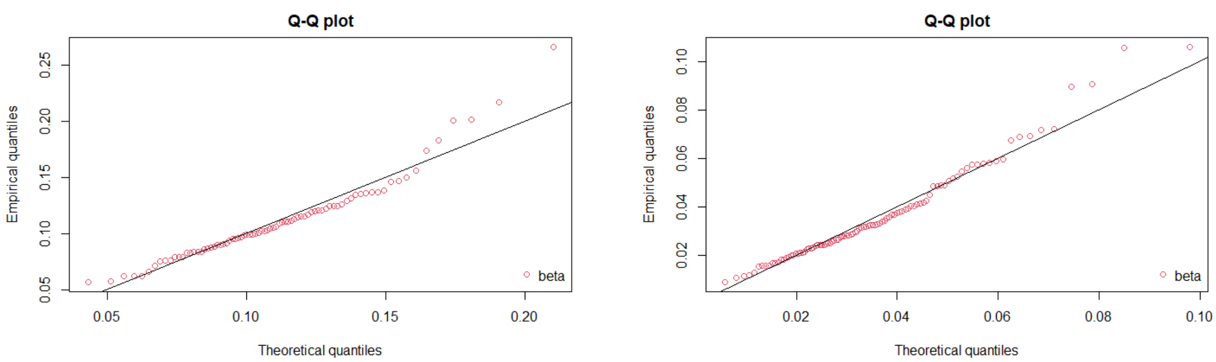
$$s^2 = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad (11)$$

Reassuringly, theoretical moments were nearly identical to the empirical moments, as shown in Table 4.

Prior	Empirical mean	Theoretical mean	Empirical variance	Theoretical variance
G-OFF	0.02073	0.02074	0.00010709	0.00010072
F-OFF	0.03596	0.03600	0.0003867	0.0003239
C-OFF	0.1005	0.1005	0.0009289	0.0009318
G-DEF	0.1108	0.1111	0.001344	0.001140
F-DEF	0.1484	0.1485	0.001561	0.001469
C-DEF	0.2247	0.2246	0.003076	0.003177

Table 4: Comparison on theoretical and empirical moments

Although the formal tests suggested that the fitted distributions were appropriate, visual inspection of the Guard defensive rebounding prior and the Forward offensive rebounding prior suggested that the right tails of the fitted distributions were too thin. The Q-Q plots illustrating this issue are included in Figure 5. All 18 plots (three per distribution) are included in Appendix B.



Guards - DREBS

Forwards - OREBS

Figure 5: The fitted priors understimate the propotional of excellent rebounds in these two cases

In the case of the guards, the poor fit of the right tail was caused almost entirely by Russell Westbrook (26.6%), Luka Doncic (21.7%), and James Harden (20.1%), the three leading defensive rebounders amongst guards. The fact that these three superstar players were not adequately accounted for by the fitted distribution was deemed a non-issue: if a guard were as capable as a rebounder as any of these players, they are probably are exceptional players, in that they are either a tremendously athletic small player who can outleap taller opponents for rebounds, or an incredibly nimble large player who can maneuver quickly enough to keep up with their smaller adversaries. In either case, the player would be receving plenty of playing time at their position (and hence plenty of rebounding opportunities), and there would therefore not be much of a difference between their empirical rebounding rate and the posterior density of their rebounding rate.

In the case of the forwards, the leading rebounders were NOT extraordinary players. The three leaders were Jarred Vanderbilt (10.6%), Thaddeus Young (10.6%), and Bruce Brown (9.1%), which makes the justification used for guards inappropriate. Given that formal testing suggested an acceptable fit, the fitted beta was still used as a prior distribution. However, this issue does have a meaningful practical interpretation that is discussed in Section 6.

One concern of the presented methodology was that some players may be receiving playing time (and hence, more rebounding opportunities) due to their rebounding ability: this would mean that by retaining players with many attempts, we would be overestimating the rebounding ability of the average player. To verify whether this was in fact the case, the correlation between rebounding percentage and number of rebounding opportunities of all players having collected at least one rebound at the given position was measured using Spearman's rho. The requirement of one rebound drastically reduced the number of ties in each of the six groups and filtered out unrealistic player positions from the data. Although some significant positive correlation was detected at the 5% level, it was deemed small enough to be safely ignored. Some more sizable negative correlation was detected, but upon further inspection, was due to some very small samples from players playing down a position (for example, Tony Bradley, led all "forwards" in offensive rebounding going 6 for 21, Thon Maker led all "guards" in offensive rebounding going 2 for 8). The fact that the correlation was negative (the idea that coaches prefer poor rebounders seems unlikely), as well as the effect being non-existent in the case of centers, was deemed appropriate justification to not account for the correlation. The six values of correlation are included in Table 5, and scatter plots of the ranks can be found in Appendix C.

Prior	Spearman's ρ	p-value
G-OFF	-0.3704	1.54×10^{-10}
F-OFF	-0.2937	4.97×10^{-8}
C-OFF	0.1311	0.0676
G-DEF	-0.1201	0.0341
F-DEF	-0.0187	0.7165
C-DEF	0.1738	0.0093

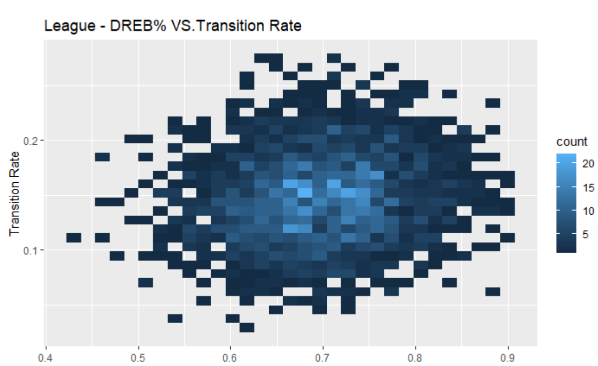
Table 5: Correlation between reboudning opprtunity volume and rebounding effectiveness. Given that the negative correlation would imply that poor rebounid-ing is coveted, it was deemed to simply be an artefact of the position labelling process

All parameter estimation for the beta distributions was done in R by maxmizing the log-likelihood, using the fitdistrplus package (Delignette-Muller and Dutang, 2015) .

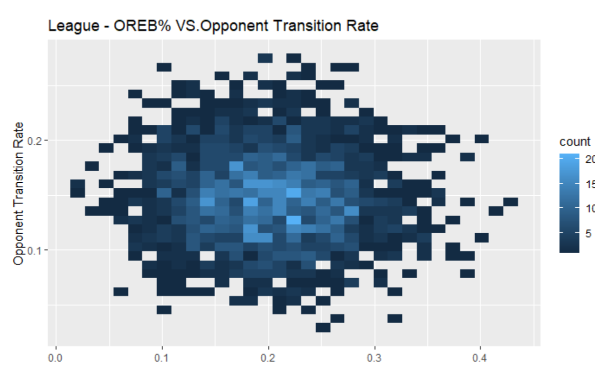
4.2 Copula fitting

4.2.1 Determining the modelling level

The first step in the dependence modelling process was to determine whether the dependence should be modelled at a league or team level. Initially, the four quantitative relationships were explored at the league scale, but no sizable significant correlation was present in any case. This point is further driven home by the plots in Figure 6. (Note that when explored at the league level, because of the artifical doubling of the dataset, this actually only represents two different quantitative relationships, hence the two scatter plots).



Relationships (A.) and (B.)



Relationship (C.) and (D.)

Figure 6: League wide trend for the correlation between rebounding and transition.

This led to the exploration of team level dependence, where some sizable significant correlation was detected. As mentioned in Section 3, the relationships (3) and (4) had significant correlation in some cases, but it was deemed too small to be of any practical use. This lack of sizable correlation was not surprising given the nature of the dataset and is discussed further in Section 6.

Relationships (1) and (2) had two teams with "doubly" significant correlation: the Portland Trailblazers and the Philadelphia 76ers. The fact that only two teams had significant dependence between their rebounding and their transition offence was incredibly surprising: conventional basketball wisdom suggests that the first step of transition offense is to collect the defensive rebound. The data used in this analysis suggests that this is generally not the case. The value of each team's Spearman's rho and Kendall's tau for each of the four relationships, as well as their p-values, can be found in Appendix D.

Furthermore, before fitting the copula, the dependence of these two teams' defensive rebounding and transition offense was further inspected with graphical methods.

4.2.2 Graphical inspection of dependence

Three types of plots were used to investigate dependence: scatter plots of the observations, chi-plots as proposed by Fisher and Switzer (1985), and Kendall plots (K-plots) as proposed by Genest and Boies (2003). The plots generated for the Portland Trailblazers will be used to illustrate the rationale behind these plots, and the plots for the Philadelphia 76ers can be found in Appendix E.

Given that chi-plots and K-plots are rather advanced when compared to more commonly used graphical assessment tools, a brief description of their construction is given.

Chi-plots

Given X_i, Y_i, bivariate observations from some distribution, the following quantities are defined:

- $H_n(X_i, Y_i) = \sum_{j\neq i} \mathbb{1}(X_j \leq X_i, Y_j \leq Y_i)/(n-1)$
- $F_n(X_i) = \sum_{j\neq i} \mathbb{1}(X_j \leq X_i)/(n-1)$
- $G_n(Y_i) = \sum_{j\neq i} \mathbb{1}(Y_j \leq Y_i)/(n-1)$

Under the assumption of independence, one would expect that $H_n(X_i, Y_i) \approx F_n(X_i)G_n(Y_i)$, the discrepancy only being due to sampling variation. In other words, large values of $|H_n(X_i, Y_i) - F_n(X_i)G_n(Y_i)|$ suggest dependence between the marginals. We therefore plot the following along the y-axis:

$$\chi_i = \frac{H_n(X_i, Y_i) - F_n(X_i)G_n(Y_i)}{\sqrt{F_n(X_i)(1 - F_n(X_i))G_n(Y_i)(1 - G_n(Y_i))}} \tag{12}$$

The denominator is simply a scaling factor that makes the difference asymptotically behave like a standard Normal random variable (Fisher and Switzer, 1985), and is related to the chi-squared statistic for 2x2 contingency tables. Along the x-axis, we plot the following quantity:

$$\lambda_i = 4 \times sgn_i \times max\{(F_n(X_i) - \frac{1}{2})^2, (G_n(Y_i) - \frac{1}{2})^2\} \text{ where } sgn_i = sgn\{(F_n(X_i) - \frac{1}{2})(G_n(Y_i) - \frac{1}{2})\} \tag{13}$$

Under the assumption of independence, λ_i follows a uniform distribution over [-1, 1], meaning that an imbalance in the number of positive and negative values of λ_i or clustering behavior in the plot would suggest dependence.

Using these definitions, one can easily observe the strong dependence suggested by the Portland Trailblazers chi-plot, which can be found in Figure 7.

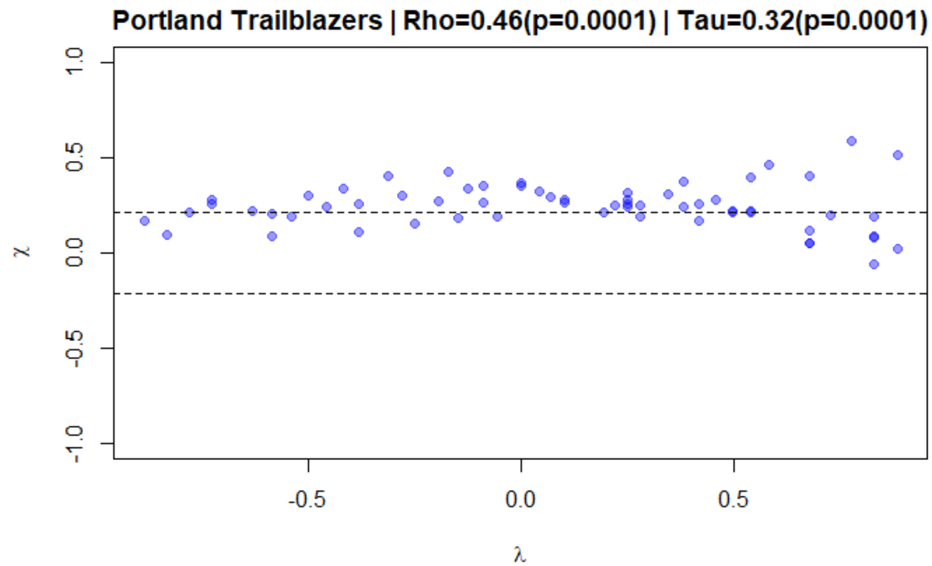


Figure 7: Given that the points are not symmetric around $\chi = 0$, and that there are more points in right side of the plot, it is likely that there is dependence in the data.

K-Plots
The idea of a K-Plot is to plot the empirical order statistics of the Kendall Distribution against the expected values of the order statistics under the assumption of independence. The derivation of the expected value of the i^{th} order statistic is given in Appendix F.

This means that a K-plot can be interpreted similarly as a Q-Q Plots, but with added benefit of being able to get a qualitative idea of how close we are to perfect positive or negative dependence. The plot in Figure 8 suggests dependence:

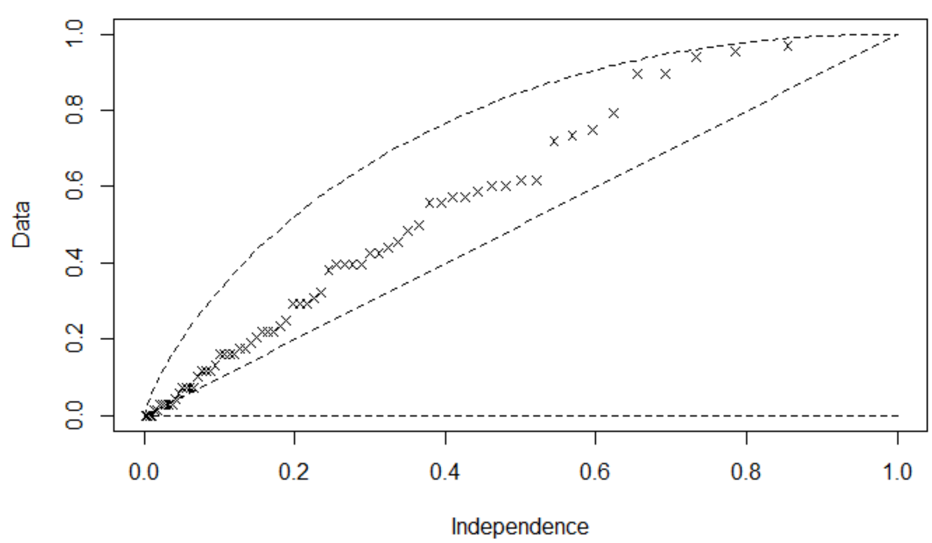


Figure 8: In the case of independence, the points should be along the identity line.

4.2.3 Parameter estimation

As mentioned in Section 3, the parameters were estimated by both the inversion of Spearman's rho and of Kendall's tau. The estimated parameter, as well as their standard errors, are given in Table 6. The equivalent table for the 76ers can be found in Appendix G.

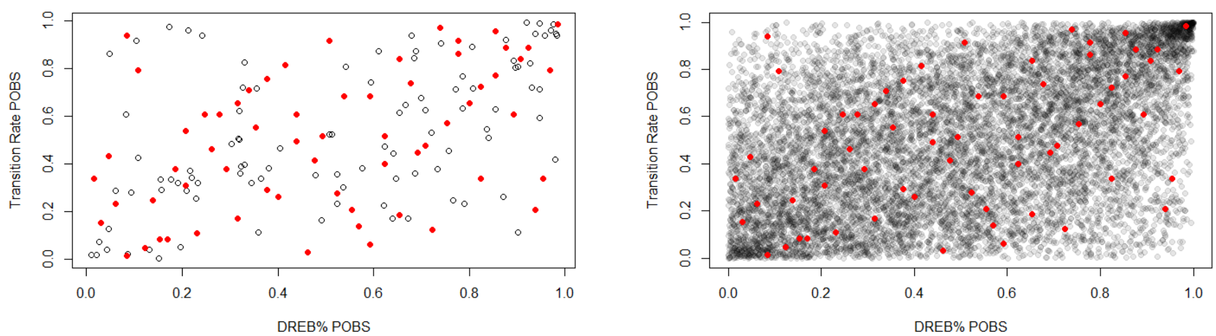
Family	θ (ρ)	SE (ρ)	$\theta(\tau)$	SE (τ)
Clayton	0.8374	0.329	0.8895	0.344
Frank	2.828	0.885	3.007	0.187
Gumbel	1.423	0.171	1.445	0.172
Gaussian	0.4441	0.112	0.4649	0.114

Table 6: Parameter estimates and their standard errors. Notice the difference in the standard error of the parameter of the Frank based on the estimation method.

The goodness of fit of each family was inspected both graphically and by formal goodness of fit testing.

Graphical goodness of fit testing

The first step in visually inspecting the models was to create a scatter plot of the empirical pseudo-observations against simulated pseudo-observations generated from the fitted copula. Figure 9 contains said plots for the fitted Gumbel copula of the Portland Trailblazers. The full set of plots for all families and for the Philadelphia 76ers can be found in Appendix H.



100 simulated pseudo-observations

10 000 simulated pseudo-observations

Figure 9: Comparison between simulated and empirical pseudo-observations. The red points are the empirical data

The second graphical tool used was the generalized K-Plot: this is exact same procedure as the K-plot described in Section 4.2.2, but we plot the empirical Kendall distribution against the Kendall distribution of C_{θ_n} rather than against the Kendall distribution under the assumption of independence. Furthermore, generalized K-plots were only used in the case of Archimedean copulas because of the explicit formula for their Kendall distribution (in terms of their generator φ), given by Genest and Favre (2007):

$$K(w) = w - \frac{\varphi(w)}{\varphi'(w)},\ w \in (0,1) \tag{14}$$

The generalized K-plot for the Gumbel copula fitted to the Portland Trailblazers is given in Figure 10. All K-plots for both teams can be found in Appendix H.

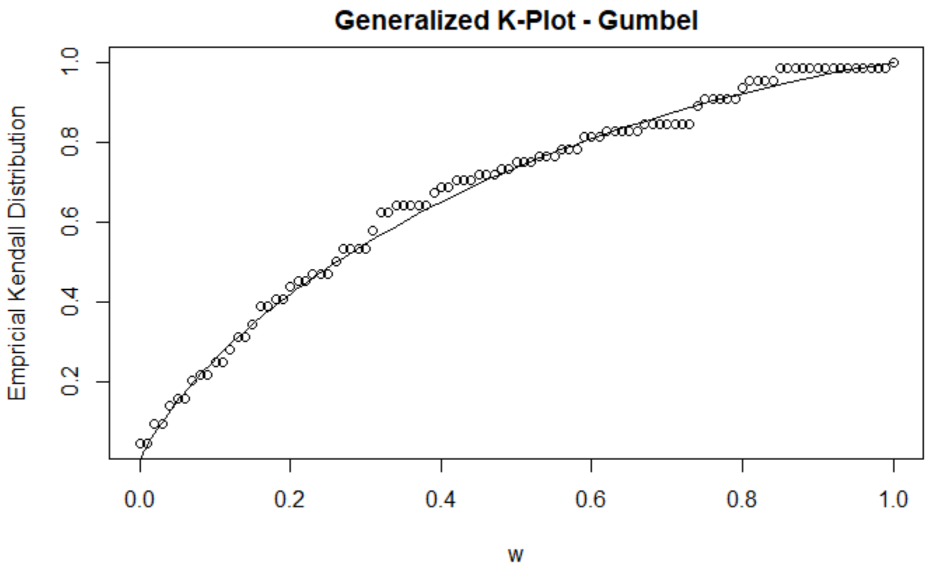


Figure 10: The solid line in the generalized K-plot represents the theoretical Kendall distribution. The fit of the Gumbel seems appropriate.

Formal goodness of fit testing
Formal goodness of fit testing was conducted using two test statistics described by Genest et al. (2009). The first, S_n, is a measure of the difference between the "empirical copula" and the fitted copula. The difference is measured using the rank-based version of the Cramér–von Mises statistic:

$$S_n = \int_{[0,1]^2} \sqrt{n}(C_n(u_1,u_2) - C_{\theta_n}(u_1,u_2))dC_n(u_1,u_2) \tag{15}$$

where C_{θ_n} is the fitted copula and $C_n(u_1,u_2) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(U_{i1} \leq u_1, U_{i2} \leq u_2)$

The second test statistic, $S_n^{(c)}$, is based on the Rosenblatt transform (which depends on a given copula), which transforms a random vector into an independent uniformly distributed random vector. Therefore, by applying the Rosenblatt transform to the pseudo-observations, if the fitted copula is in fact correct, one would expect to obtain a joint transformed distribution that is close to the independence copula. The significance of the difference can be tested using the rank-based version of the Cramér–von Mises statistic:

$$S_n^{(c)} = n\int_{[0,1]^2} (D_n(u_1,u_2) - C_{\perp}(u_1,u_2))dD_n(u_1,u_2) \tag{16}$$

where $D_n((u)) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}(\mathcal{R}_{\theta_n}(\mathbf{U}_i) \leq \mathbf{u})$ and $\mathcal{R}_{\theta_n}$ is the Rosenblatt transform associated with C_{θ_n}.

For both statistics, p-values can be obtained by simulation using the parametric bootstrap, as described by Genest et al. (2009).

In the case that the two different tests did not agree on the rejection of the null hypothesis, Table 1 of Genest et al. (2009) was consulted, and the conclusion of the test with greater power was used. Table 1 of the paper was used rather than tables 2 or 3 because Kendall's tau for the Portland Trailblazers and the Philadelphia 76ers were 0.32 and 0.31 respectively. The values of the statistics and their p-values for the Portland Trailblazers is given in Table 7.

Family	S_n	S_n p-value	$S_n^{(c)}$	$S_n^{(c)}$ p-value
Clayton	0.04105	0.04745	0.0359	0.6758
Frank	0.02448	0.4231	0.0228	0.9076
Gumbel	0.01937	0.8077	0.02610	0.8437
Gaussian	0.02489	0.4690	0.03383	0.6039

Table 7: Results from formal goodness of fit testing

4.2.4 Retained familiies

For both the Portland Trailblazers and the Philadelphia 76ers, the Clayton copula was rejected due to its poor fit (at the 5% level), while all other families were retained as potential models.

All parameter estimation, goodness of fit testing, and plotting was done using the copula package in R (Hofert et al., 2020).

4.3 Multinoulli estimation

Given that there were 3 x 30 team/type combinations of fitted distributions, each of which having 5 parameters, it was deemed more practical to include the results in Appendix I.

Instead, we include the league-wide mean of the expected value of each type of opportunity in Table 8. The sole purpose of this table is to provide justification for the transition selection criteria described in Section 2: as most basketball pundits would expect, transition opportunities tend to yield more points than regular opportunities (recall that second chance opportunities are much lower because they include turnovers in their estimation, whereas transition and regular opportunities do not).

Transition	Second Chance	Regular
1.46	0.99	1.11

Table 8: League average for expected points from each of the three opportunity types

5 Application

5.1 Small sample rebounding in the playoffs

Given that the whole premise of the model is to smooth out rebounding statistics in the case of small sample sizes, it is obviously difficult to obtain significant results from a single season's worth of playoff games. The main goal of this section is mainly to serve as a sanity check for the described smoothing. We do this by analyzing three notable changes of strategy in the 2020-2021 NBA Playoffs.

Note that in every case, the model performed as well as one could have hoped for on type of rebound and performed poorly for the other type. We only address the well-performing cases in this section, but the poor performances are covered in Section 6.

5.1.1 Case study 1: Small Ball Kevin Durant

Over the course of the regular season, 15 of Kevin Durant's 1185 defensive rebounding opportunities (1.2%) came while playing center, but over the course of the Brooklyn Nets' series against the Milwaukee Bucks, that number jumped up to 8.1% (50 of 621).

Somewhat surprisingly, Durant collected 12 of the 50 opportunities, good for a defensive rebounding rate of 24%, which was much better than his regular season overall (i.e. not handled by position) defensive rebounding rate of 18.9%, and his regular season center rebounding rate of 13.3% (2 for 15).

However, after applying the Center-based prior to his rebounding numbers, the posterior distribution for Durant's positional rebounding rate follows a $Beta(14.086, 54.72)$ distribution, good for a maximum a posteriori (MAP) estimate of 20.5%. Figure 11 illustrates the effect of the prior on his position defensive rebounding rate. We therefore had four potential models for Kevin Durant's rebounding:

1. Binomial model wth $p = 13.3\%$
2. Binomial model with $p = 18.9\%$
3. Binomial model with $p = 20.5\%$
4. Beta-binomial model with $\alpha = 14.086$ and $\beta = 54.72$

Each model was treated as a null hypothesis, and the p-value of 12/50 was calculated using a left-sided test. Furthermore, the expected value under each null hypothesis was calculated. All these results can be found in Table 9.

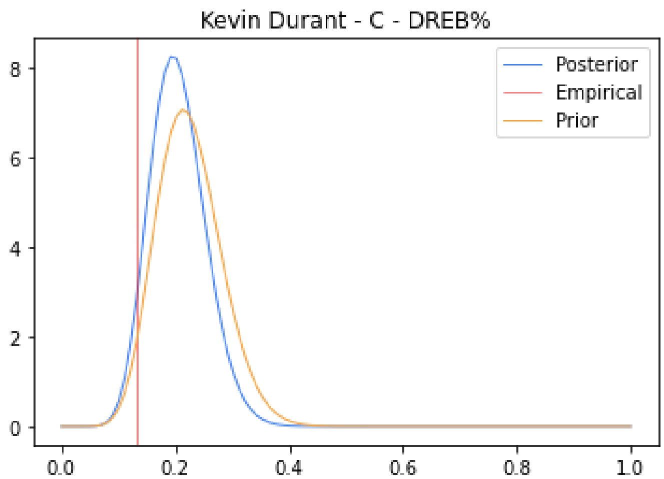


Figure 11: Durant, who was coming off a torn Achilles's tendon, was handled very delicately by the coaching staff during the regular season. His posterior rebounding rate is therefore is almost indistinguishable from that of an average center.

Model	p-value	Expectation
1	0.0284	6.65
2	0.224	9.45
3	0.321	10.24
4	0.260	10.24

Table 9: Evaluating the viability of each model for Kevin Durant

Although only model 1 can be rejected at the 5% level, the Bayesian smoothed prediction is closer to the observed data.

5.1.2 Case study 2: Replacing Jaren Jackson Jr.'s meniscus with statistics

Jaren Jackson Jr., a key player for the up-and-coming Memphis Grizzlies, played only 11 regular season games before starting against the Utah Jazz in the playoffs due to an 8-month knee injury. Unlike Durant, who rarely departs his natural forward position during the regular season, Jackson Jr.'s playoff offensive rebounding opportunities were similarly split between forward and center as they were during the regular season. However, his number of opportunities were still very low given the few games he appeared in.

In the minutes Jackson Jr. spent at center during the playoffs, he collected 7 of a potential 66 rebounds, good for a positional offensive rebounding rate of 10.6%. This was an especially pleasant surprise given overall offensive rebounding rate during the regular season was 6.3%, and his regular season positional rebounding rate at center was a measly 4.8% (7 for 145). The effect of the smoothing is illustrated in Figure

11. After applying the center-prior, Jackson Jr.'s posterior positional offensive rebounding rate followed a beta distribution with parameters α=16.65 and β=224.37. The same procedure shown in Section 5.1.1 was performed here, the results are given in Table 10.

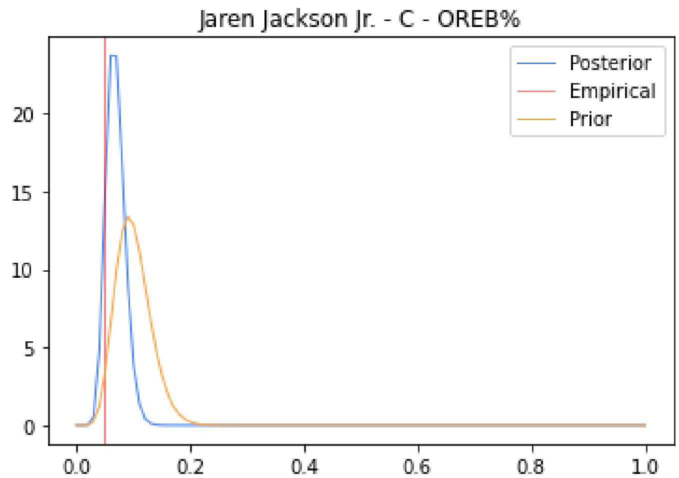


Figure 12: In the case of Jackson Jr., the smoothing is nowhere near as pronounced as it was for Durant.

Model	p-value	Expectation
1	0.0386	3.17
2	0.121	4.16
3	0.178	4.52
4	0.109	4.52

Table 10: Evaluating the viability of each model for Jaren Jackson Jr.

Again, only model 1 can be rejected at the 5% level.

5.1.3 Case study 3: Nicolas Batum fuels the upset

Nicolas Batum is the most noteworthy case of playoff strategy tinkering: over the course of the regular season, Batum had only 175 of his 1530 (11.4%) offensive rebounding opportunities come at center, but during the LA Clippers' series against the Jazz, had 126 of 155 (81.3%) opportunities come at center, and collected 6, good for a positional rebounding rate of 4.8%.

This was much better than his regular season overall rate of 3.2%, and his positional rate at center of 3.4%. After applying the Center-prior, his posterior positional offensive rebounding rate follows a beta distribution with α=15.65 and β=255.37. The MAP estimate in this instance is 5.7%. The effect of the smoothing is illustrated in Figure 13. We performed the same exercise as for Jackson Jr. and Durant, except right sided tests were performed for the Bayesian-smoothed models because they overestimated the observed rebounding rate.

(On a very unrelated note, only the LA Clippers, the team with the boldest disregard for rebounding, would go on to win their series, despite being heavy underdogs.)

Model	p-value	Expectation
1	0.259	4.28
2	0.217	4.03
3	0.418	7.18
4	0.437	7.18

Table 11: Evaluating the viability of each model for Nicolas Batum

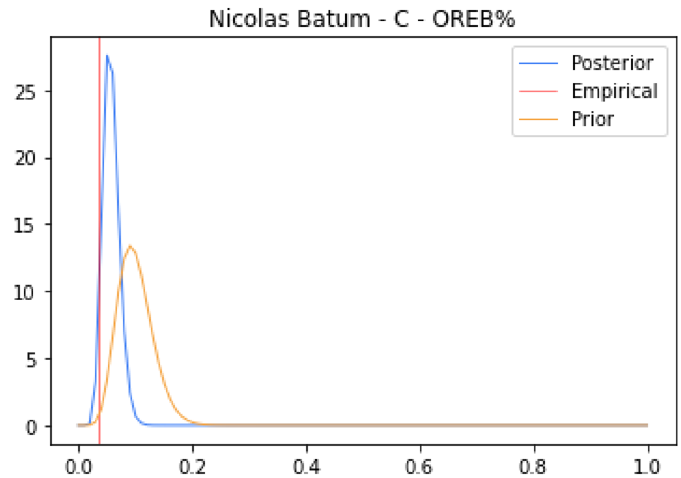


Figure 13: Again, the smoothing is a lot more subtle.

p-values for the binomial observations were computed with base R (R Core Team, 2021) and p-values for the beta-binomial were computed using the extraDistr package (Wolodzko, 2020).

5.2 Informal rebounding valuation

Given the hierarchical nature of the constructed models (Individual rebounding $\rightarrow$ Team rebounding $\rightarrow$ Transition Rate $\rightarrow$ Possession values), a formal comparison between model predictions and observations was deemed out of the scope of the project. However, we still include a very informal, "back of the envelope" type of calculation in this section just to illustrate the size of the effect.

Given that the data suggests that only the Portland Trailblazers and Philadelphia 76ers rely significantly on their defensive rebounding prowess to initiate their transition offense, we use one of these teams (Philadelphia).

Furthermore, the Gumbel copula was selected to model the dependence: the 76ers are a (anectdotally) terrible team in the offensive half-court and look to run their transition offense whenever possible. Under the heuristic assumption that every "transition-friendly" rebound results in a transition opportunity, we would expect that in larger rebounding samples, where the proportion of transition friendly rebounds should be more stable, the transition rate should also be more stable, and hence there should be greater dependence in the upper tail.

5.2.1 Defensive rebounding

The 76ers had two main options to play at center: Joel Embiid, although he is an otherworldly talent in most facets, had a great but not godly defensive rebounding rate of 26.57%. His replacement, Dwight Howard, is pedestrian at most things, but a top tier defensive rebounder, putting up an impressive defensive rebounding rate of 32.56%. Both players only played center during the season, and both had sufficiently many attempts for smoothing to be unnecessary.

In the minutes without Dwight Howard, the team collected 1865 of a possible 2757 defensive rebounds, good for a rate of 67.64% (D_E). Embiid was on the court for 1508 of these opportunities, collecting 401 defensive rebounds himself. In the purely hypothetical scenario of Howard being on the court instead of Embiid, and rebounding at his regular rate of 32.56%, how many points would have been generated by the additional transition offense opportunities he has created by the extra 90 (1955/2757, D_H) defensive rebounds he would theoretically collect?

Given that the dependence between pseudo-observations, rather than proper observations, had been modelled, producing rigorous predictions would require the fitting of distributions to the marginals. However, since this is not a rigorous prediction, the following was performed:

- "Convert" D_E and D_H into pseudo-observations by separately ranking them within the training dataset to obtain ranks R_E and R_D, and divide these both by n+1
- After producing the pseudo-responses P_E and P_H, convert them to observable responses as follows:
  $TR_E = TR_k$ where k = $\underset{1,2,\ldots n}{\arg\max}\{\frac{S_i}{(n+1)} \leq P_E\}$
  $TR_H = TR_k$ where k = $\underset{1,2,\ldots n}{\arg\min}\{\frac{S_i}{(n+1)} \geq P_H\}$
  (Recall that S_i denotes the rank of the transition rate).
  This obviously overestimates the value of the observable response for Howard and underestimates it in the case of Embiid, but it will be shown below that this is a non-issue.

In the case of Archimedean copulas with generator φ and with marginals uniformly distributed over (0, 1), the conditional expectation is given by the following expression (Crane and Van Der Hoek, 2008):

$$\mathbb{E}[Y \mid X = x] = -\varphi(x) \int_0^1 y \frac{\varphi''(C(y,x))\varphi'(y)}{[\varphi'(C(y,x))]^3} dy \tag{17}$$

This produces expected pseudo-responses of $P_E = 0.4380$ and $P_H = 0.4795$. After applying the very crude conversion method, we obtain expected observable responses of $TR_E = 0.1333$ and $TR_H = 0.1383$.

Per 100 possessions, this means that Dwight Howard would produce and additional 0.5 transition opportunities. Using the average value of the 76ers transition opportunity (1.461) and the average value of their regular opportunities (1.123), the amount of additional offense created by Howard's defensive rebounding is 0.17 points per 100 possessions, a very negligible difference, especially when considering that difference between Embiid's and Howard's responses were artificially inflated.

Howard's defensive rebounding also helps his team by preventing the opposing team's second chance points: by collecting defensive rebounds, he is preventing the opposing team from collecting offensive rebounds. Assuming again the above hypothetical setup, and in the very unlikely, best-case scenario where all of Howard's additional defensive rebounds are missed shots that would have otherwise been offensive rebounds for the opponent, a very rough value for the number of points saved can be estimated as follows:

$$\sum_{k=1}^{\infty}(D_E - D_H)^k \times \text{average missed FGs per 48 minutes} \times \\ \text{league average value of second chance opportunities} \times \frac{1}{\text{pace}} \times 100 \tag{18}$$

(Note that the summation comes from the geometric nature of rebounding. Also note that the convolutedness of the formula comes from the way the NBA records game statistics.)

Combining these two estimates (which again, by choice of team and approximation methods have been calculated so that they produce the largest values possible), we conclude that Howard's elite defensive rebounding is worth 1.75 points/100 possessions more than Embiid's good but not elite rebounding.

6 Discussion

6.1 Addressing modelling issues

6.1.1 Thin tail of the forward offensive rebounding prior

There was a common theme amongst the players found in the right tail of the Forward offensive rebounding prior: they all play with centers who are excellent 3-point shooters. It is suspected that even though these players are labelled as forwards, for all intents and purposes, they behave as centers: they occupy the space close to the basket, as the traditional center would, to allow the labelled center to stand far away from the basket and attempt 3-point shots. One potential solution to this issue would be to use a regression model to assign players both an offensive and defensive position. Such a procedure would require a significant amount of work and basketball-specific expertise to determine which measures dictate a player's position, rather than which measures are a product of their position. Furthermore, the use of a continuous variable for position would either require arbitrary division into bins for prior construction or would greatly complicate prior construction.

6.1.2 Survival bias in offensive rebounding rate

As mentioned in Section 4, there was no sizable significant correlation between a team's offensive rebounding rate and their opponent's transition rate. Although one could potentially conclude that there is therefore no opportunity cost (in the form of poor defensive positioning) associated to aggressively pursuing offensive rebounds, they must first recognize the survival bias embedded in offensive rebounding rate: OREB% was used as a proxy for the amount effort invested into pursuing offensive rebounds (and presumably, not getting back into defensive position). The problem is that by collecting an offensive rebound, a team increases the chance they score, which would seriously mitigate the opponent's ability to get out in transition (because of the additional time needed to inbound). In other words, if a team were to very aggressively pursue offensive rebounds, fail to collect many of them, and allow the opposing team to have more transition opportunities, using OREB% would suggest that team did not invest much effort into the offensive rebounds to begin with.

A more appropriate approach would be to directly model the dependence between transition and opposing offensive rebounding intent rather than opposing offensive rebounding outcome. This would require player position data, which is not publicly available.

6.1.3 Issues with the case studies

The poor accuracy of the model for Durant's offensive rebounding is probably due to the same issue described in Section 6.1.1: Durant is a great 3-point shooter and plays with the forward Bruce Brown (who has already been mentioned for his extraordinary offensive rebounding numbers).

For Jackson Jr. and Batum, their play at center was meant to mitigate the effectiveness of Rudy Gobert, one of the best centers in the league. Given that the prior was constructed using data from different calibers of opposition, it is no surprise that the model overestimated their defensive rebounding when playing against top-notch competition. However, this is probably not caused by inappropriate smoothing: consider Jonas Valanciunas, player for the Memphis Grizzlies who played exclusively at center during the regular season and collected 496 of a potential 1657 defensive rebounds (29.9%), and hence required no smoothing. In the same exact series as Jackson Jr., while also playing against Gobert, Valanciunas collected only 29 of a potential 147 defensive rebounds (19.7%). Under the null hypothesis that his playoff rebounding is identical to his regular season rebounding, this sample has a p-value of 0.00356 (when performing a left-sided test).

Given that teams match up at most four times during the regular season there is no obvious robust way to account for this type of matchup specific dependence. Instead of using models for exact predictions in the playoffs, it is suggested that it be used as an evaluation tool for positional rebounding ability.

6.2 Conclusion

Over the course of this project, it has firstly been shown that the idea that rebounding is a product of not only ability, but also intent, holds water and that the use a of a beta prior can help improve rebounding prediction accuracy when trying to account for a change in intent.

Furthermore, evidence has been produced to challenge the idea that defensive rebounding is a vital component of transition offense.

Lastly, we hope to have produced a compelling enough argument against putting enormous weight on individual rebounding numbers (especially defensive rebounding numbers) when evaluating players, as well as advise to proceed with caution when including them in all-in-one player evaluation metrics.

References

Bolstad, W. M., & Curran, J. M. (2012). *Introduction to bayesian statistics (3rd ed.).* Wiley.

Delignette-Muller, M.-L., Dutang, C. (2015). fitdistrplus: An R Package for Fitting Distributions. Journal of Statistical Software, 64(4), 1-34. http://www.jstatsoft.org/v64/i04/.

Crane, G. J., & Hoek, J. van. (2008). Conditional expectation formulae for copulas. *Australian & New Zealand Journal of Statistics, 50(1)*, 53–67. https://doi.org/10.1111/j.1467-842x.2007.00499.x.

Engelmann, J. (2016). Possession-Based Player Performance Analysis in Basketball (Adjusted +/– and Related Concepts). *Handbook of Statistical Methods and Analyses in Sports (1st ed.).* Chapman and Hall/CRC. https://doi-org.proxy3.library.mcgill.ca/10.1201/9781315166070.

ESPN Internet Ventures. (n.d.). *NBA teams.* ESPN. Retrieved November 28, 2021, from https://www.espn.com/nba/teams.

Fisher, N.I., & Switzer, P. (1985). Chi-plots for assessing dependence. *Biometrika, 72*, 253-265.

Genest, C., Ghoudi, K., & L.-P. Rivest. (1995). A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions. *Biometrika, 82(3)*, 543–552. https://doi.org/10.2307/2337532.

Genest, C., & Boies, J.-C. (2003). Detecting Dependence with Kendall Plots. *The American Statistician, 57(4)*, 275–284. http://www.jstor.org/stable/30037296.
Genest, C., & Favre, A. (2007). Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrologic Engineering*, 12, 347-368.

Genest, C., & Nešlehová, J. (2007). A Primer on Copulas for Count Data. *ASTIN Bulletin, 37(2)*, 475-515. doi:10.1017/S0515036100014963

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A Review and a power study. *Insurance: Mathematics and Economics, 44(2)*, 199–213. https://doi.org/10.1016/j.insmatheco.2007.10.005.

Glossary. Basketball-Reference. (n.d.). Retrieved November 28, 2021, from https://www.basketball-reference.com/about/glossary.html.

Hofert, M., Maechler, M. (2011). Nested Archimedean Copulas Meet R: The nacopula Package. Journal of Statistical Software, 39(9), 1-20. https://www.jstatsoft.org/v39/i09/.

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2020). copula: Multivariate Dependence with Copulas. R package version 1.0-1. https://CRAN.R-project.org/package=copula.

Kojadinovic, I., Yan, J. (2010). Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. Journal of Statistical Software, 34(9), 1-20. https://www.jstatsoft.org/v34/i09/.

Merriam-Webster. (n.d.). Fast break. In *Merriam-Webster.com dictionary*. Retrieved November 28, 2021, from https://www.merriam-webster.com/dictionary/fast%20break.

Myers, D. (2020). *About Box Plus/Minus (BPM).* Basketball-Reference. Retrieved November 28, 2021, from https://www.basketball-reference.com/about/bpm2.html.

NBA. (n.d.). *Official NBA stats*. NBA Stats. Retrieved November 28, 2021, from https://www.nba.com/stats/.

NBA active player contracts. Spotrac.com. (n.d.). Retrieved November 28, 2021, from https://www.spotrac.com/nba/contracts/.

Oliver, D. (2011). *Basketball on paper: Rules and tools for performance analysis.* Potomac Books, Inc.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Vaci, N., Cocić, D., Gula, B. *et al.* Large data and Bayesian modeling—aging curves of NBA players. *Behav Res* 51, 1544–1564 (2019). https://doi.org/10.3758/s13428-018-1183-8.

Wolodzko, T. (2020). extraDistr: Additional Univariate and Multivariate Distributions. R package version 1.9.1. https://CRAN.R-project.org/package=extraDistr.

Yahoo! (n.d.). *Yahoo Fantasy Basketball*. Yahoo! Retrieved November 28, 2021, from https://basketball.fantasysports.yahoo.com/nba.

Yan, J. (2007). Enjoy the Joy of Copulas: With a Package copula. Journal of Statistical Software, 21(4), 1-21. https://www.jstatsoft.org/v21/i04/.

Appendix A

Name	Primary position	Possible positions	**Labelling result**
Ja Morant	PG	PG	**PG**
Desmond Bane	SG	PG-SG	**SG**
Dillon Brooks	SF	SG-SF	**SF**
Jaren Jackson Jr.	PF	PF-C	**PF**
Jonas Valanciunas	C	C	**C**

Step 1 example: in the case where primary positions fill out the 5 required positions, they could simply be used to label the lineup.

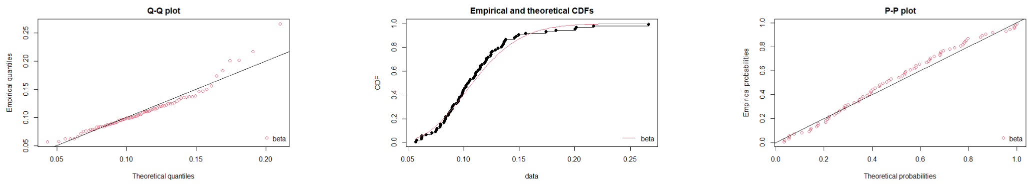
Name	Primary position	Possible positions	**Labelling result**
Desmond Bane	SG	PG-SG	**PG**
Dillon Brooks	SF	SG-SF	**SG**
Kyle Anderson	SF	SF-PF	**SF**
Jaren Jackson Jr.	PF	PF-C	**PF**
Jonas Valanciunas	C	C	**C**

Step 2 example: in the case where there is exactly one way to assign possible positions to create a valid 5-man lineup, that labelling was used.

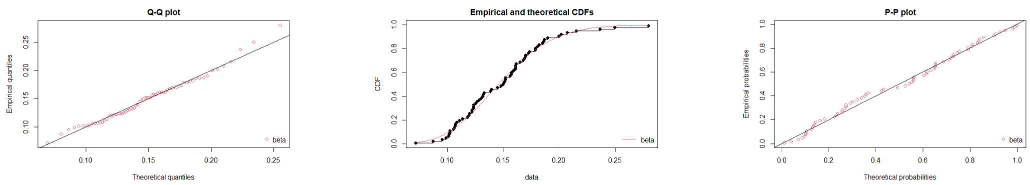
Name	Primary position	Possible positions	**Labelling result**
Ja Morant	PG	PG	**PG**
Desmond Bane	SG	PG-SG	**SG**
Grayson Allen	SG	SG-SF	**SF**
Dillon Brooks	SF	SG-SF	**PF**
Brandon Clarke	PF	SF-PF	**C**

Step 3 example: in the case where there is no possible valid labelling, the ordering of unique position combinations is used to assign positions.

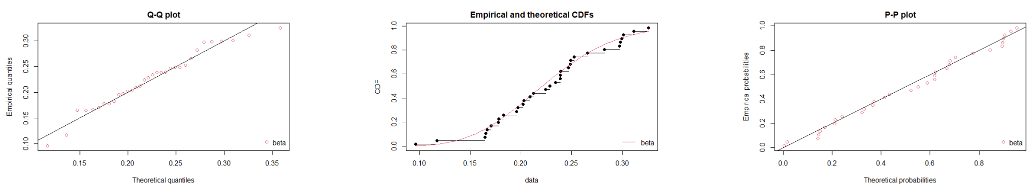
Appendix B



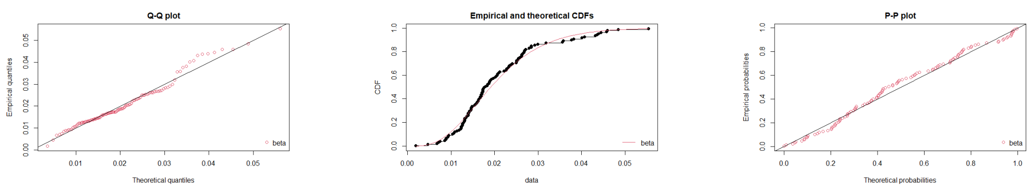
Guards - Defensive rebounding



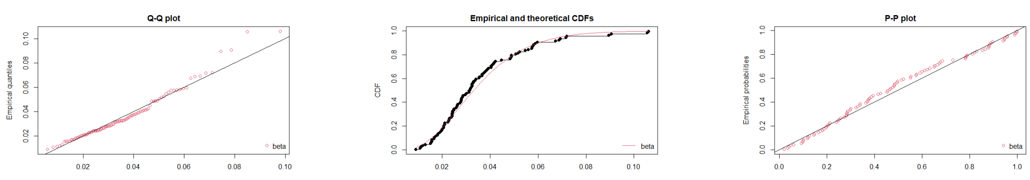
Fowards - Defensive rebounding



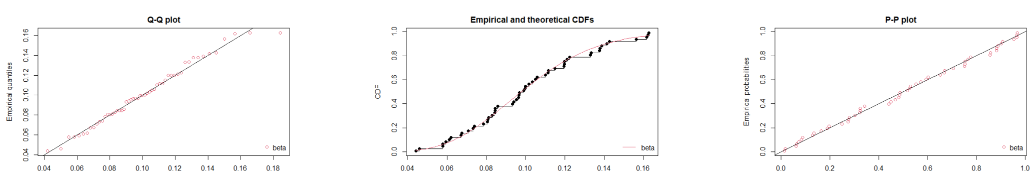
Centers - Defensive rebounding



Guards - Offensive rebounding



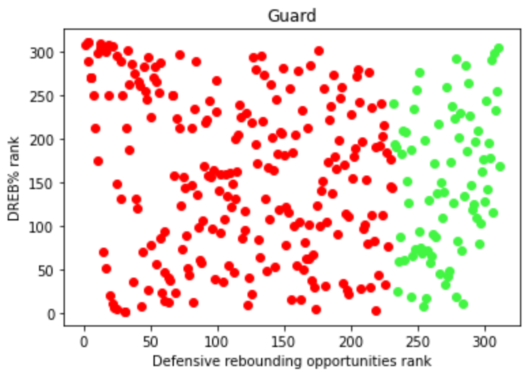
Fowards - Offensive rebounding



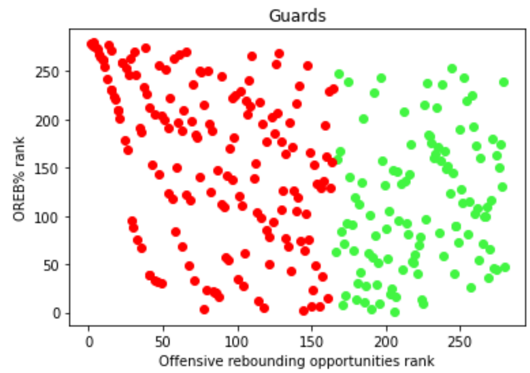
Centers - Offensive rebounding

Full graphical inspection of fit of the prior distributions.

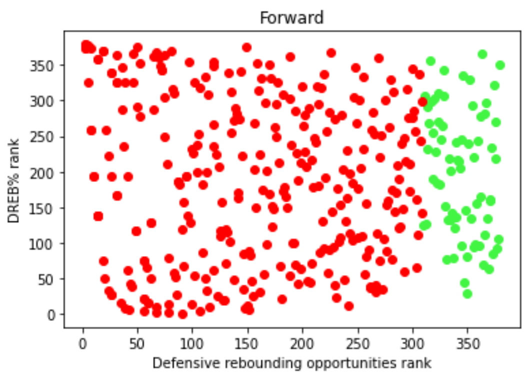
Appendix C



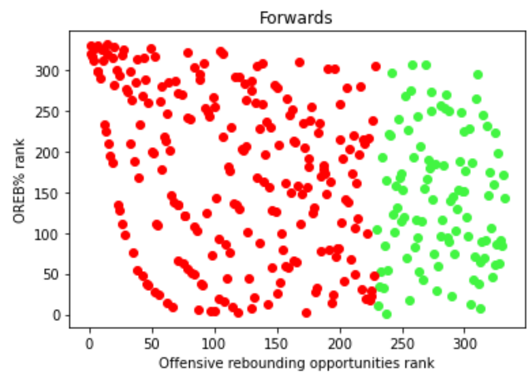
Guards - Defensive rebounding



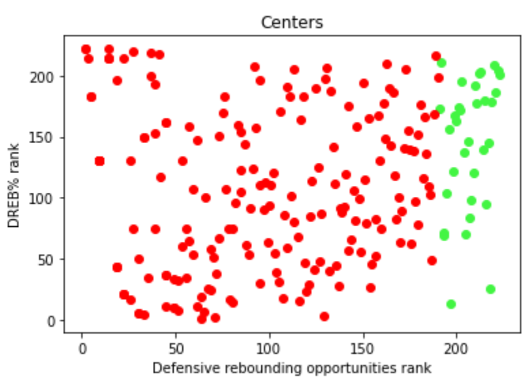
Guards - Offensive rebounding



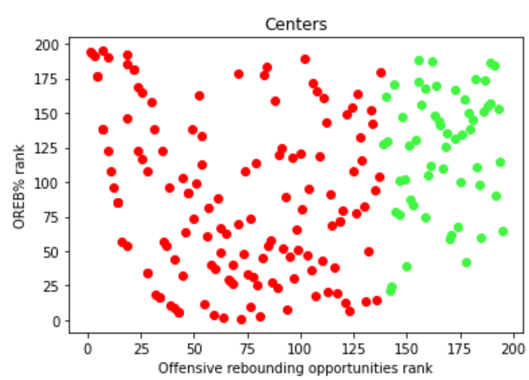
Fowards - Defensive rebounding



Fowards - Offensive rebounding



Centers - Defensive rebounding



Centers - Offensive rebounding

Verifying that there is no positive correlation between the number of rebounding opportunities and rebounding rate. Green points are the players who were retained for prior construction. Red points are players who were rejected from prior construction.

Appendix D

Team ID	Spearman's ρ	Spearman's ρ p-value	Kendall's τ	Kendall's τ p-value
1610612737	0.270739885	0.024445714	0.189905972	0.021435822
1610612738	-0.000612164	0.995957431	-0.017802983	0.827055608
1610612739	0.21106487	0.084034941	0.150860226	0.070151051
1610612740	0.214613659	0.078836053	0.142447661	0.087209898
1610612741	0.096217285	0.428152186	0.059080509	0.471445862
1610612742	0.183908993	0.130358659	0.125643329	0.1289312
1610612743	0.101114916	0.419182943	0.065044519	0.441629525
1610612744	0.102007695	0.407815481	0.057900622	0.487807111
1610612745	-0.122513548	0.327093116	-0.085506227	0.313552053
1610612746	0.231404138	0.05954605	0.158781301	0.058824155
1610612747	0.08450616	0.499915799	0.054878055	0.51713125
1610612748	-0.164106796	0.18794473	-0.121653694	0.151541804
1610612749	0.211736675	0.090415553	0.133462423	0.117999546
1610612750	0.16033313	0.181657114	0.104900847	0.198357843
1610612751	0.075576345	0.540179769	0.045072982	0.589080845
1610612752	0.17016412	0.171941419	0.1119963	0.185785616
1610612753	0.078678709	0.52681491	0.056402129	0.502037695
1610612754	-0.146020812	0.231224537	-0.090344682	0.274303862
1610612755	0.42768967	0.000423371	0.307846692	0.000355807
1610612756	0.153606782	0.211070373	0.107213894	0.198178507
1610612757	0.455141852	8.52E-05	0.315021488	0.000142669
1610612758	0.087503541	0.474637947	0.051040114	0.537482944
1610612759	0.221044649	0.067964775	0.152527863	0.065103052
1610612760	-0.043353727	0.727576448	-0.02730384	0.745303293
1610612761	0.25222007	0.036550164	0.182051835	0.02766587
1610612762	-0.036780709	0.765871759	-0.020745973	0.803443471
1610612763	-0.057311404	0.6424888	-0.038402142	0.645017081
1610612764	0.088484743	0.483352033	0.047365899	0.578879048
1610612765	0.27882255	0.022322993	0.186089525	0.027148603
1610612766	0.190308719	0.13518315	0.126445681	0.144441416

Full results for relationship (A.).

Team ID	Spearman's ρ	Spearman's ρ p-value	Kendall's τ	Kendall's τ p-value
1610612737	-0.035028094	0.775079345	-0.016705936	0.839864374
1610612738	0.173721248	0.14737645	0.12239181	0.133676624
1610612739	-0.121926308	0.321934096	-0.0913303	0.273018501
1610612740	-0.212319941	0.082166162	-0.142226204	0.088158972
1610612741	-0.06177863	0.611414003	-0.044971901	0.583876332
1610612742	-0.034708962	0.777075324	-0.028277638	0.73237565
1610612743	-0.077811994	0.534591672	-0.046304978	0.583692439
1610612744	0.056139982	0.649311994	0.037086096	0.656440971
1610612745	0.047812467	0.703034905	0.031015038	0.714765941
1610612746	-0.191327348	0.120907406	-0.132878852	0.113908805
1610612747	-0.084898864	0.497918425	-0.052668822	0.535095985
1610612748	0.242887127	0.049407605	0.173668961	0.040500296
1610612749	-0.136271735	0.279075208	-0.098503139	0.247980901
1610612750	-0.078604195	0.514669796	-0.055049595	0.499460628
1610612751	-0.106440866	0.387632529	-0.078691493	0.345848095
1610612752	-0.087075456	0.486923451	-0.05908566	0.485423466
1610612753	-0.055063728	0.658076745	-0.039189053	0.641471361
1610612754	0.143791626	0.238501907	0.091648822	0.267536771
1610612755	-0.383538899	0.001757523	-0.274000308	0.00148882
1610612756	-0.183202393	0.1348087	-0.112607972	0.176878736
1610612757	-0.299215046	0.012504892	-0.200730827	0.015509039
1610612758	-0.19808114	0.102777103	-0.135680562	0.101513624
1610612759	-0.085226131	0.48625568	-0.068581251	0.407081362
1610612760	-0.010068253	0.93555066	-0.00771675	0.92668421
1610612761	-0.186343745	0.12525824	-0.130136998	0.115254403
1610612762	0.100390412	0.415328095	0.058252427	0.484562629
1610612763	0.026508969	0.830092099	0.020318053	0.807541707
1610612764	-0.070342642	0.577659296	-0.049794573	0.559657958
1610612765	-0.069152387	0.578176127	-0.05278726	0.529990181
1610612766	-0.109450233	0.3931563	-0.071152891	0.412789966

Full results for relationship (B.).

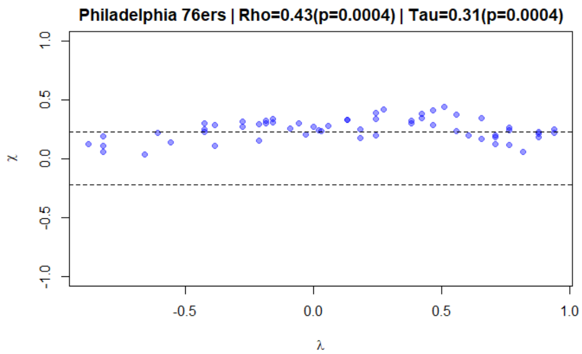
Team ID	Spearman's ρ	Spearman's ρ p-value	Kendall's τ	Kendall's τ p-value
1610612737	0.149439044	0.220367407	0.095931557	0.245815069
1610612738	-0.226369369	0.057658268	-0.147653734	0.069934644
1610612739	-0.240022908	0.048665844	-0.166740185	0.045334228
1610612740	-0.230497571	0.058618715	-0.156897327	0.059437529
1610612741	0.205446368	0.087972641	0.141578208	0.084643875
1610612742	-0.055108399	0.652895088	-0.031256697	0.705268802
1610612743	0.119576112	0.338912147	0.091015779	0.28278148
1610612744	-0.161547127	0.188134606	-0.11780292	0.157411104
1610612745	0.003111102	0.980220994	-0.005152245	0.95144183
1610612746	-0.25911808	0.034228657	-0.185244275	0.027209904
1610612747	-0.013541308	0.914064631	-0.00703566	0.9338141
1610612748	-0.080712854	0.51942099	-0.052520517	0.535216493
1610612749	0.009905753	0.937577459	0.012615259	0.882881925
1610612750	-0.171225411	0.153369691	-0.109921252	0.17686619
1610612751	-0.038849965	0.753107693	-0.018534865	0.824001253
1610612752	-0.118010607	0.345318999	-0.103068705	0.223259641
1610612753	-0.268611583	0.027958483	-0.194855578	0.020485267
1610612754	-0.036524632	0.765738954	-0.018892314	0.819623726
1610612755	-0.211094703	0.094052003	-0.153846634	0.074181515
1610612756	0.00257732	0.983358037	-0.001323627	0.987325296
1610612757	-0.11624911	0.341489317	-0.070325959	0.395433696
1610612758	-0.21189445	0.080484956	-0.138537451	0.094188012
1610612759	-0.158928719	0.192113888	-0.111588147	0.177864502
1610612760	0.232136444	0.058721767	0.153182214	0.068095149
1610612761	0.183702569	0.130798222	0.129143457	0.119931625
1610612762	-0.052271361	0.672048035	-0.031774054	0.70300775
1610612763	-0.029793612	0.809413298	-0.026937514	0.746664585
1610612764	-0.076525427	0.54459043	-0.047320185	0.578899209
1610612765	-0.095345495	0.442784049	-0.065029566	0.43884248
1610612766	-0.009965062	0.938215255	-0.011855676	0.89143005

Full results for relationship (C.).

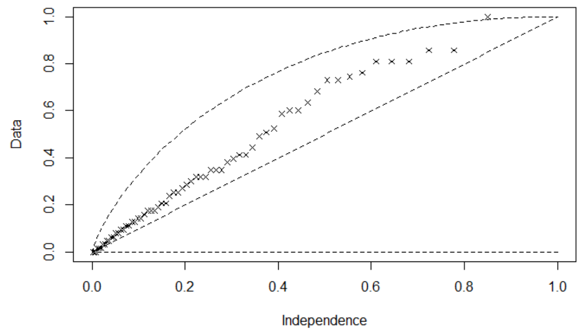
Team ID	Spearman's ρ	Spearman's ρ p-value	Kendall's τ	Kendall's τ p-value
1610612737	-0.064323105	0.599519981	-0.040641712	0.622601913
1610612738	0.249394145	0.035958447	0.171856101	0.03483965
1610612739	0.26152363	0.031221278	0.169865435	0.041484666
1610612740	0.142151957	0.247531492	0.105682816	0.205524857
1610612741	-0.288630656	0.015384846	-0.203954219	0.012954628
1610612742	0.130488808	0.285206288	0.087850967	0.288143172
1610612743	0.109285282	0.382393569	0.08294283	0.327142514
1610612744	0.2589117	0.033010411	0.171454643	0.03989301
1610612745	0.201524088	0.104674249	0.13660837	0.106023158
1610612746	0.364011018	0.002460911	0.267362935	0.00143164
1610612747	-0.02910655	0.816542795	-0.024836	0.769207156
1610612748	0.159100875	0.20195873	0.099906202	0.238296075
1610612749	0.023173451	0.854617726	0.005325587	0.950316951
1610612750	0.181275468	0.130303339	0.112971576	0.164513653
1610612751	0.170690265	0.164016582	0.106843309	0.200008049
1610612752	0.165407919	0.184419843	0.118259532	0.164360212
1610612753	0.241659432	0.048821275	0.157870808	0.060303692
1610612754	0.094202568	0.441344835	0.060476201	0.465002409
1610612755	0.260484192	0.037632283	0.188359442	0.028853933
1610612756	0.127486541	0.300190438	0.084693427	0.309312945
1610612757	0.081718997	0.504436412	0.042918554	0.604298429
1610612758	0.243973977	0.043361024	0.164806867	0.046602767
1610612759	0.369081344	0.001803676	0.258258852	0.00181131
1610612760	-0.18518223	0.133545607	-0.12409123	0.139432539
1610612761	0.005628141	0.963392256	0.006430893	0.938048456
1610612762	0.226498505	0.063262166	0.165487131	0.047592722
1610612763	0.038746074	0.753746996	0.024713161	0.766814814
1610612764	0.110172455	0.382299191	0.065637096	0.441203808
1610612765	0.150230525	0.22495818	0.105	0.211112401
1610612766	0.232082609	0.067206742	0.166580734	0.055235357

Full results for relationship (D.).

Appendix E



76ers chi-plot



76ers K-plot

The figures suggest dependence in the marginals.

Appendix F

The density of the i^{th} order statistic in a sample of size n ($W_{i:n}$) is given by

$$f_{W_{i:n}}(w) = n\binom{n-1}{i-1}[\Pr(H(X,Y) \leq w)]^{i-1} \times f_{H(X,Y)}(w) \times [1 - \Pr(H(X,Y) \leq w)]^{n-1}.$$

Under the assumption of independence,

$$\Pr(H(X,Y) \leq w) = w - w\ln(w) \implies f_{H(X,Y)}(w) = -\ln(w).$$

Therefore, the expected value of $W_{i:n}$ is given by

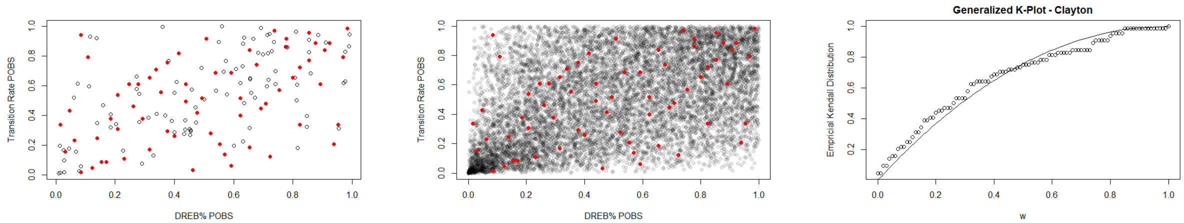
$$\mathbb{E}(W_{i:n}) = \int_0^1 n\binom{n-1}{i-1}(w - w\ln(w))^{i-1} \times (-\ln(w)) \times (1 - w + w\ln(w))^{n-i} \times w \times (-\ln(w))dw$$

Appendix G

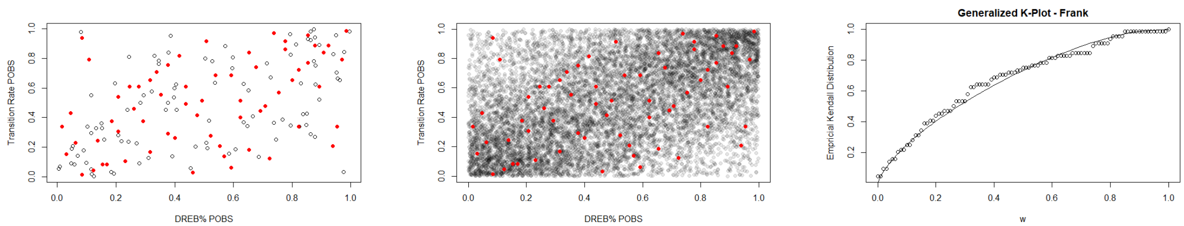
Family	θ (ρ)	SE (ρ)	$\theta(\tau)$	SE (τ)
Clayton	0.9225	0.31	0.9198	0.317
Frank	3.054	0.81	3.09	0.178
Gumbel	1.467	0.159	1.46	0.158
Gaussian	0.4721	0.098	0.4749	0.103

Parameter estimates and their standard errors for the Philadelphia 76ers.

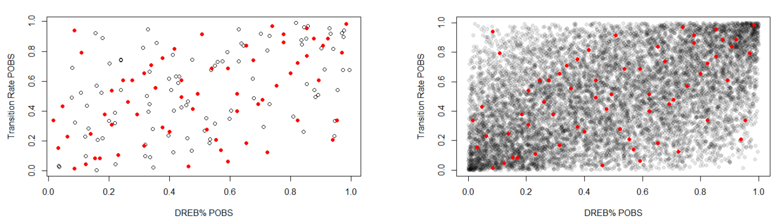
Appendix H



Clayton graphical inspection

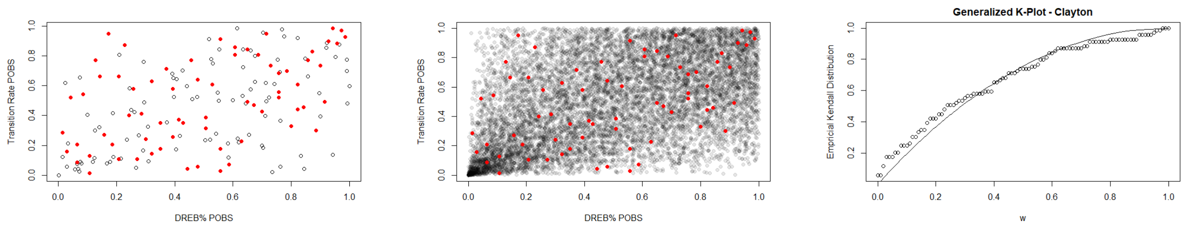


Frank graphical inspection

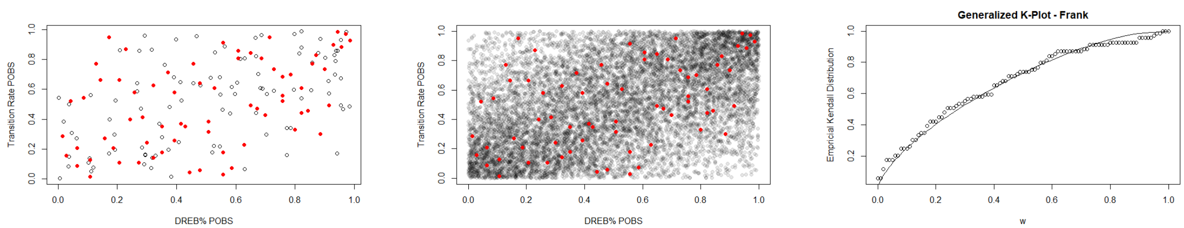


Gaussian graphical inspection

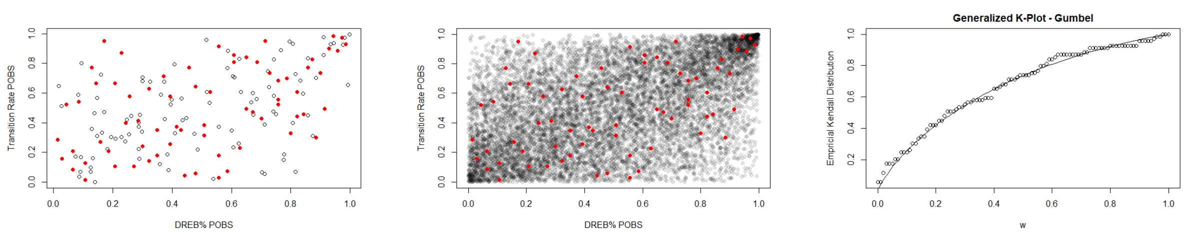
Full graphical inspection of all copula families for the Portland Traiblazers.



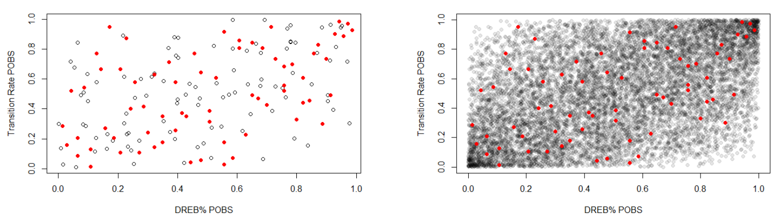
Clayton graphical inspection



Frank graphical inspection



Gumbel graphical inspection



Centers - Defensive rebounding

Full graphical inspection of all copula families for the Philaelphia 76ers.

Appendix I

Team ID	Opp. type	p_0	p_0 SE	p_1	p_1 SE	p_2	p_2 SE	p_3	p_3 SE	p_4	p_4 SE
1610612737	FB	0.324	0.015	0.036	0.006	0.479	0.016	0.157	0.012	0.004	0.002
1610612737	REG	0.502	0.007	0.026	0.002	0.321	0.007	0.15	0.005	0.001	0.001
1610612737	SC	0.516	0.016	0.031	0.006	0.361	0.015	0.092	0.009	0.0	0.0
1610612738	FB	0.313	0.015	0.05	0.007	0.447	0.016	0.189	0.012	0.002	0.001
1610612738	REG	0.512	0.007	0.03	0.002	0.296	0.006	0.162	0.005	0.001	0.0
1610612738	SC	0.562	0.015	0.037	0.006	0.313	0.014	0.087	0.009	0.001	0.001
1610612739	FB	0.296	0.015	0.071	0.008	0.495	0.016	0.138	0.011	0.0	0.0
1610612739	REG	0.523	0.007	0.036	0.003	0.312	0.007	0.128	0.005	0.001	0.001
1610612739	SC	0.56	0.016	0.036	0.006	0.319	0.015	0.084	0.009	0.001	0.001
1610612740	FB	0.308	0.014	0.058	0.007	0.463	0.015	0.164	0.011	0.006	0.002
1610612740	REG	0.497	0.007	0.037	0.003	0.332	0.007	0.132	0.005	0.002	0.001
1610612740	SC	0.529	0.016	0.039	0.006	0.357	0.015	0.073	0.008	0.002	0.001
1610612741	FB	0.316	0.015	0.041	0.006	0.444	0.016	0.196	0.013	0.003	0.002
1610612741	REG	0.506	0.007	0.029	0.002	0.316	0.007	0.148	0.005	0.0	0.0
1610612741	SC	0.533	0.017	0.023	0.005	0.316	0.016	0.126	0.011	0.001	0.001
1610612742	FB	0.323	0.016	0.062	0.008	0.428	0.017	0.185	0.013	0.002	0.002
1610612742	REG	0.494	0.007	0.033	0.003	0.306	0.006	0.164	0.005	0.002	0.001
1610612742	SC	0.541	0.017	0.029	0.006	0.316	0.016	0.113	0.011	0.001	0.001
1610612743	FB	0.326	0.016	0.05	0.007	0.426	0.016	0.197	0.013	0.001	0.001
1610612743	REG	0.486	0.007	0.027	0.002	0.332	0.007	0.153	0.005	0.002	0.001
1610612743	SC	0.533	0.017	0.018	0.004	0.366	0.016	0.084	0.009	0.0	0.0
1610612744	FB	0.341	0.015	0.053	0.007	0.393	0.016	0.212	0.013	0.001	0.001
1610612744	REG	0.506	0.007	0.028	0.002	0.292	0.006	0.172	0.005	0.002	0.001
1610612744	SC	0.572	0.018	0.026	0.006	0.286	0.017	0.117	0.012	0.0	0.0
1610612745	FB	0.362	0.017	0.049	0.008	0.439	0.017	0.148	0.012	0.002	0.002
1610612745	REG	0.527	0.007	0.034	0.003	0.272	0.006	0.165	0.005	0.002	0.001
1610612745	SC	0.548	0.017	0.025	0.005	0.316	0.016	0.111	0.011	0.0	0.0
1610612746	FB	0.325	0.016	0.05	0.007	0.441	0.017	0.18	0.013	0.003	0.002
1610612746	REG	0.494	0.007	0.023	0.002	0.311	0.007	0.17	0.005	0.002	0.001
1610612746	SC	0.519	0.017	0.02	0.005	0.347	0.016	0.112	0.011	0.002	0.002
1610612747	FB	0.309	0.015	0.074	0.008	0.464	0.016	0.148	0.012	0.004	0.002
1610612747	REG	0.51	0.007	0.033	0.003	0.314	0.007	0.141	0.005	0.002	0.001
1610612747	SC	0.52	0.017	0.036	0.006	0.334	0.016	0.11	0.011	0.0	0.0
1610612748	FB	0.316	0.015	0.051	0.007	0.465	0.016	0.169	0.012	0.0	0.0
1610612748	REG	0.5	0.007	0.03	0.002	0.311	0.007	0.156	0.005	0.003	0.001
1610612748	SC	0.585	0.018	0.027	0.006	0.284	0.017	0.101	0.011	0.003	0.002
1610612749	FB	0.342	0.014	0.053	0.007	0.399	0.014	0.204	0.012	0.003	0.001
1610612749	REG	0.488	0.007	0.031	0.003	0.321	0.007	0.158	0.005	0.002	0.001
1610612749	SC	0.543	0.017	0.029	0.006	0.326	0.016	0.101	0.01	0.001	0.001
1610612750	FB	0.332	0.014	0.05	0.007	0.48	0.015	0.138	0.011	0.001	0.001
1610612750	REG	0.524	0.007	0.031	0.002	0.285	0.006	0.158	0.005	0.002	0.001
1610612750	SC	0.551	0.016	0.036	0.006	0.321	0.015	0.091	0.009	0.0	0.0
1610612751	FB	0.345	0.014	0.054	0.007	0.402	0.015	0.199	0.012	0.0	0.0
1610612751	REG	0.475	0.007	0.027	0.002	0.331	0.007	0.165	0.005	0.002	0.001
1610612751	SC	0.517	0.018	0.025	0.006	0.344	0.017	0.115	0.011	0.0	0.0

Team ID	Opp. type	p_0	p_0 SE	p_1	p_1 SE	p_2	p_2 SE	p_3	p_3 SE	p_4	p_4 SE
1610612752	FB	0.327	0.017	0.056	0.009	0.442	0.018	0.173	0.014	0.001	0.001
1610612752	REG	0.517	0.007	0.031	0.002	0.307	0.007	0.143	0.005	0.002	0.001
1610612752	SC	0.556	0.017	0.021	0.005	0.332	0.016	0.091	0.01	0.0	0.0
1610612753	FB	0.38	0.015	0.049	0.007	0.383	0.015	0.186	0.012	0.002	0.001
1610612753	REG	0.542	0.007	0.032	0.003	0.296	0.007	0.127	0.005	0.002	0.001
1610612753	SC	0.539	0.017	0.025	0.005	0.345	0.016	0.089	0.009	0.001	0.001
1610612754	FB	0.337	0.014	0.045	0.006	0.458	0.015	0.157	0.011	0.003	0.001
1610612754	REG	0.502	0.007	0.03	0.002	0.324	0.007	0.142	0.005	0.001	0.001
1610612754	SC	0.527	0.017	0.027	0.006	0.34	0.016	0.104	0.01	0.002	0.002
1610612755	FB	0.34	0.015	0.049	0.007	0.432	0.016	0.178	0.012	0.001	0.001
1610612755	REG	0.487	0.007	0.034	0.003	0.345	0.007	0.133	0.005	0.001	0.001
1610612755	SC	0.506	0.017	0.052	0.008	0.352	0.016	0.089	0.01	0.001	0.001
1610612756	FB	0.331	0.015	0.031	0.005	0.433	0.016	0.203	0.013	0.002	0.001
1610612756	REG	0.49	0.007	0.022	0.002	0.336	0.007	0.149	0.005	0.003	0.001
1610612756	SC	0.498	0.018	0.018	0.005	0.357	0.017	0.125	0.012	0.001	0.001
1610612757	FB	0.357	0.015	0.044	0.007	0.383	0.016	0.216	0.013	0.001	0.001
1610612757	REG	0.519	0.007	0.023	0.002	0.275	0.006	0.181	0.005	0.002	0.001
1610612757	SC	0.52	0.016	0.032	0.006	0.347	0.015	0.099	0.01	0.002	0.001
1610612758	FB	0.326	0.014	0.049	0.007	0.431	0.015	0.191	0.012	0.002	0.001
1610612758	REG	0.494	0.007	0.04	0.003	0.326	0.007	0.139	0.005	0.001	0.0
1610612758	SC	0.532	0.017	0.034	0.006	0.336	0.016	0.097	0.01	0.0	0.0
1610612759	FB	0.339	0.015	0.052	0.007	0.473	0.015	0.135	0.011	0.001	0.001
1610612759	REG	0.51	0.007	0.029	0.002	0.335	0.007	0.125	0.005	0.001	0.001
1610612759	SC	0.551	0.017	0.019	0.005	0.347	0.016	0.08	0.009	0.002	0.002
1610612760	FB	0.368	0.015	0.065	0.008	0.422	0.015	0.14	0.011	0.005	0.002
1610612760	REG	0.532	0.007	0.033	0.003	0.284	0.007	0.149	0.005	0.002	0.001
1610612760	SC	0.578	0.016	0.029	0.006	0.315	0.015	0.078	0.009	0.0	0.0
1610612761	FB	0.33	0.014	0.048	0.006	0.432	0.015	0.188	0.012	0.002	0.001
1610612761	REG	0.524	0.007	0.03	0.002	0.273	0.006	0.171	0.005	0.002	0.001
1610612761	SC	0.562	0.016	0.026	0.005	0.303	0.015	0.109	0.01	0.0	0.0
1610612762	FB	0.36	0.015	0.047	0.007	0.319	0.015	0.27	0.014	0.004	0.002
1610612762	REG	0.493	0.007	0.026	0.002	0.294	0.007	0.184	0.006	0.002	0.001
1610612762	SC	0.558	0.016	0.028	0.005	0.289	0.015	0.123	0.011	0.001	0.001
1610612763	FB	0.342	0.014	0.047	0.006	0.442	0.014	0.168	0.011	0.002	0.001
1610612763	REG	0.516	0.007	0.031	0.003	0.328	0.007	0.124	0.005	0.001	0.001
1610612763	SC	0.519	0.016	0.02	0.004	0.386	0.015	0.073	0.008	0.002	0.001
1610612764	FB	0.342	0.014	0.043	0.006	0.469	0.014	0.142	0.01	0.004	0.002
1610612764	REG	0.49	0.007	0.035	0.003	0.346	0.007	0.127	0.005	0.002	0.001
1610612764	SC	0.515	0.018	0.031	0.006	0.351	0.017	0.099	0.01	0.004	0.002
1610612765	FB	0.335	0.016	0.069	0.009	0.422	0.017	0.172	0.013	0.001	0.001
1610612765	REG	0.509	0.007	0.036	0.003	0.31	0.007	0.144	0.005	0.001	0.0
1610612765	SC	0.55	0.017	0.025	0.005	0.335	0.016	0.088	0.01	0.002	0.002
1610612766	FB	0.339	0.015	0.051	0.007	0.417	0.016	0.193	0.013	0.0	0.0
1610612766	REG	0.516	0.008	0.032	0.003	0.287	0.007	0.162	0.006	0.003	0.001
1610612766	SC	0.595	0.017	0.035	0.006	0.257	0.015	0.113	0.011	0.0	0.0

All parameters for Multinoulli distributions. Recall p_k denotes the probability of scoring k points on a given opportunity.