Measuring Spatial Allocative Efficiency in Basketball

Nathan Sandholtz[1], Jacob Mortensen[1], and Luke Bornn[1]

[1]Simon Fraser University

Abstract

Every shot in basketball has an opportunity cost; one player's shot eliminates all potential opportunities from their teammates for that play. For this reason, player shot efficiency should ultimately be considered relative to the lineup. This aspect of efficiency—the optimal way to allocate shots within a lineup—is the focus of our paper. Allocative efficiency is fundamentally a spatial problem because the distribution of shot attempts within a lineup is highly dependent on court location. We propose a new metric for spatial allocative efficiency by comparing a player's field goal percentage (FG%) to his field goal attempt (FGA) rate in context of both his four teammates on the court and the spatial distribution of his shots. Leveraging publicly available data provided by the National Basketball Association (NBA), we estimate player FG% at every location in the offensive half court using a Bayesian hierarchical model. Then, by ordering a lineup's estimated FG%'s and pairing these rankings with the lineup's empirical FGA rate rankings, we detect areas where the lineup exhibits inefficient shot allocation. We estimate and visualize this inefficiency, and identify which players are responsible. Lastly, we analyze the impact that sub-optimal shot allocation has on a team's overall offensive potential, demonstrating that inefficient shot allocation correlates with reduced scoring.

Keywords: Bayesian hierarchical model, spatial data, INLA, ranking, basketball

[1]Nathan Sandholtz is PhD candidate, Department of Statistical and Actuarial Science, Simon Fraser University (e-mail: nsandhol@sfu.ca); Jacob Mortensen is PhD candidate, Department of Statistical and Actuarial Science, Simon Fraser University (e-mail: jmortens@sfu.ca); Luke Bornn is Assistant Professor, Department of Statistical and Actuarial Science, Simon Fraser University

† The first and second authors contributed equally to this work.

1 Introduction

From 2017 to 2019, the Oklahoma City Thunder faced four elimination games across three playoff series. In each of these games, Russell Westbrook attempted over 30 shots and had an average usage rate of 45.5%.[1] The game in which Westbrook took the most shots came in the first round of the 2017-18 National Basketball Association (NBA) playoffs, where he scored 46 points on 43 shot attempts in a 96-91 loss to the Utah Jazz. At the time, many popular media figures conjectured that having one player dominate field goal attempts in this way would limit the Thunder's success. While scoring 46 points in a playoff basketball game is an impressive feat for any one player, its impact on the overall game score is moderated by the fact that it required 43 attempts. Perhaps not coincidentally, the Thunder lost three out of four of these close-out games and never managed to make it out of the first round of the playoffs.

At its core, this critique is about shot efficiency. The term 'shot efficiency' is used in various contexts within the basketball analytics community, but in most cases it has some reference to the average number of points a team or player scores per shot attempt. Modern discussion around shot efficiency in the NBA typically focuses on either individual player efficiency or shot selection. The concept of shot selection efficiency is simple: 3-pointers and shots near the rim have the highest expected points per shot, so teams should prioritize these high-value shots. The idea underlying individual player efficiency is also straightforward; scoring more points on the same number of shot attempts increases a teams overall offensive potential.

However, when discussing a player's individual efficiency it is critical to do so in context of the *lineup*. Basketball is not a 1-v-1 game, but a 5-v-5 game. Therefore, when a player takes a shot, the opportunity cost is the potential shots of their four teammates. So regardless of a player's shooting statistics relative to the league at large, a certain dimension of shot efficiency can only be defined relative to the abilities of one's teammates. Applying this to the Oklahoma City Thunder example above, if Westbrook was surrounded by dismal shooters his 43 shot attempts might not only be defensible but desirable. On the other hand, if his inordinate number of attempts prevented highly efficient shot opportunities from his teammates, then he caused shots to be inefficiently distributed and ultimately decreased his team's winning potential. This aspect of efficiency—the optimal way to allocate shots within a lineup—is the primary focus of our paper.

Allocative efficiency is fundamentally a spatial problem. As illustrated in Figure 1, the distribution of shots within a lineup is highly dependent on court location. The left plot in Figure 1 shows the overall relationship between shooting frequency (x-axis) and shooting skill (y-axis), while the four plots on the right show the same relationship conditioned on various court regions. Each dot represents a player and the size of the dot is proportional to the number of shots the player took over the 2016-17 NBA regular

[1] Usage percentage is an estimate of the percentage of team plays used by a player while he was on the floor. For a detailed formula see `www.basketball-reference.com/about/glossary.html`

season. To emphasize how shot allocation within lineups is spatially dependent, we have highlighted the Cleveland Cavaliers starting lineup (consisting of Lebron James, Kevin Love, Kyrie Irving, JR Smith, and Tristan Thompson).

When viewing field goal attempts without respect to court location (left plot), Kyrie Irving appears to shoot more frequently than both Tristan Thompson and Lebron James, despite scoring fewer points per shot than either of them. However, after conditioning on court region (right plots), we see that Irving only has the highest FGA rate in the mid-range region, which is the region for which he has the highest PPS for this lineup. James takes the most shots in the restricted area and paint regions—regions in which he is the most efficient scorer. Furthermore, we see that Thompson's high overall PPS is driven primarily by his scoring efficiency from the restricted area and that he has few shot attempts outside this area. Clearly, understanding how to efficiently distribute shots within a lineup must be contextualized by spatial information.

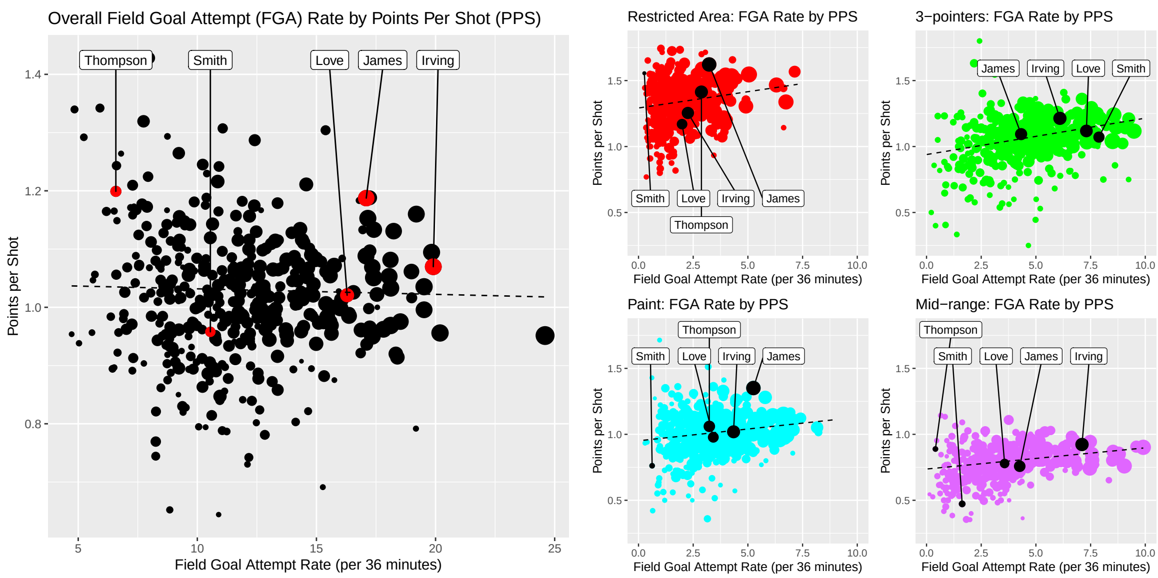

Figure 1: *Left: overall relationship between field goal attempt rate (x-axis) and points per shot (y-axis). Right: same relationship conditioned on various court regions. Each dot represents a player and the size of the dot is proportional to the number of shots the player took in the 2016-17 NBA regular season. The Cleveland Cavaliers 2016-17 starting lineup is highlighted in each plot. The weighted least squares fit of each scatterplot is overlaid in each plot by a dotted line.*

Notice that in the left panel of Figure 1, the relationship between field goal attempt (FGA) rate and points per shot (PPS) appears to be slightly negative, if there exists a relationship at all. Once the relationship between FGA rate and PPS is spatially disaggregated (see right hand plots of Figure 1), the previously negative relationship between these variables becomes positive in every region. This instance of Simpson's paradox has non-trivial implications in context of allocative efficiency which we will discuss in the following section.

The goal of our project is to create a framework to assess the strength of the relationship between shooting frequency and shooting skill spatially within lineups and to quantify the consequential impact on offensive production. Using novel metrics we develop, we quantify how many points are being lost through inefficient spatial lineup shot allocation, visualize where they are being lost, and identify which players are responsible.

1.1 Related Work

In recent years, a number of metrics have been developed to which aim to measure shot efficiency, such as true shooting percentage [1], qSQ, and qSI [2]. Additionally, metrics have been developed to quantify individual player efficiency, such as Hollinger's player efficiency rating [3]. While these metrics intrinsically account for team context, there have been relatively few studies which have looked at shooting decisions explicitly in context of lineup, and none spatially. In 2011, Goldman and Rao coined the term 'allocative efficiency' in [4], modeling the decision to shoot as a dynamic mixed-strategy equilibrium weighing both the continuation value of a possession and the value of a teammate's potential shot. They propose that a team achieves optimal allocative efficiency when at any given time, the lineup cannot reallocate the ball to increase productivity on the margin. Essentially, they argue that lineups optimize over all dimensions of an offensive strategy to achieve equal marginal efficiency for every shot. The left plot of Figure 1 is harmonious with this theory—there appears to be no relationship between player shooting frequency and player shooting skill when viewed on the aggregate. However, one of the most important dimensions the players optimize over is court location. Once we disaggregate the data by court location, (as shown in the right plots of Figure 1), we see a clear relationship between shooting frequency and shooting skill. A unique contribution of our work is a framework to assess this spatial component of allocative efficiency.

Cervone et. al.'s shot satisfaction [5] is another rare example of a shot efficiency metric that considers lineups. Shot satisfaction is defined as the expected value of a possession conditional on a shot attempt (accounting for various contextual features, such as the shot location, shooter, and defensive pressure at the time of the shot) minus the uncondtional expected value of the play. However, since shot satisfaction is marginalized over the allocative and spatial components, these factors cannot be analyzed using this metric alone. Additionally, shot satisfaction is dependent on proprietary data which limits its availability to a broad audience.

1.2 Data and Code

The data used for this project is publicly available from the NBA stats API (stats.nba.com). Shooter information and shot (x, y) locations are available through the shotchartdetail' API endpoint, while lineup information can be constructed from the playbyplayv2' endpoint. Code for constructing lineup information from play-by-play data is available at: https://github.com/jwmortensen/pbp2lineup. Ultimately this yielded a set of 224,567 shots taken by 433 players during the 2016-17 NBA regular season, which is the data that we used in this analysis. Code used to perform an empirical version of the analysis presented in this paper is also available online: https://github.com/nsandholtz/lpl.

2 Models

The foundation of our proposed allocative efficiency metrics rest on spatial estimates of both player FG% and field goal attempt (FGA) rates. In this section, we define and explain our models for each of these processes.

2.1 Estimating FG% Surfaces

Player FG% is a highly irregular latent quantity over the court space. In general, players make more shots the closer they are to the hoop, but some players are more skilled from a certain side of the court and others specialize from very specific areas, such as the corner 3-pointer. In order to capture these kinds of non-linear relationships, we implement the FG% model in [5], summarizing the spatial variation in player shooting skill by a Gaussian process represented by a low-dimensional ($D = 16$) set of deterministic basis functions. We then estimate player-specific weights for the basis functions using a Bayesian hierarchical model [6]. In this way, we maintain a feasible dimensionality for computation while allowing our model to capture nuanced spatial features that player FG% surfaces tend to exhibit.

We model the logit of $\pi_j(\mathbf{s})$, the probability that player j makes a shot at location $\mathbf{s}$, as a linear model:

$$\log\left(\frac{\pi_j(\mathbf{s})}{1-\pi_j(\mathbf{s})}\right) = \boldsymbol{\beta}'\mathbf{x} + Z_j(\mathbf{s}) + \epsilon. \tag{1}$$

Here $\boldsymbol{\beta}$ is a 4×1 vector of covariate effects and $\mathbf{x}$ is a 4×1 vector of observed covariates for the shot containing an intercept, player position, shot distance, and the interaction of player position and shot distance. $Z_j(\mathbf{s})$ is a Gaussian process which accounts for the impact of location on player j's shot make probability and ϵ is a mean-zero Gaussian error term. We model $Z_j(\mathbf{s})$ using a functional basis representation,

$$Z_j(\mathbf{s}) = \mathbf{w}_j'\boldsymbol{\Lambda}\boldsymbol{\Psi}(\mathbf{s}), \tag{2}$$

where $\mathbf{w}_j = (\mathrm{w}_{j1}, \ldots, \mathrm{w}_{jD})'$ denotes the latent basis function weights for player j and $\boldsymbol{\Lambda}\boldsymbol{\Psi}(\mathbf{s})$ denotes the basis functions. Specifically, $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_D')'$ is a $D \times K$ matrix, where each row vector $\boldsymbol{\lambda}_d$ represents the projection of the dth basis function onto a triangular mesh with K vertices over the offensive half court (more details on the construction of $\mathbf{\Lambda}$ follow below). We use the mesh proposed by [5], which was selected specifically for modeling offensive spatial behaviour in basketball. $\mathbf{\Psi}(\mathbf{s}) = (\psi_1(\mathbf{s}), \dots, \psi_K(\mathbf{s}))'$ is itself a vector of basis functions where each $\psi_k(\mathbf{s})$ is 1 at mesh vertex k, 0 at all other vertices, and values at the interior points of each triangle are determined by linear interpolation between vertices (see [7] for details). Finally, we assume $\mathbf{w}_j \sim \mathcal{N}(\boldsymbol{\omega}_j, \mathbf{\Sigma}_j)$, which makes (2) a Gaussian process with mean $\boldsymbol{\omega}_j'\mathbf{\Lambda}\mathbf{\Psi}(\mathbf{s})$ and covariance function $\text{Cov}(\mathbf{s}_1, \mathbf{s}_2) = \mathbf{\Psi}(\mathbf{s}_1)'\mathbf{\Lambda}'\mathbf{\Sigma}_j\mathbf{\Lambda}\mathbf{\Psi}(\mathbf{s}_2)$.

The bases of shot taking behavior, $\mathbf{\Lambda}$, are computed through a combination of smoothing and non-negative matrix factorization (NMF) [8]. Using integrated nested Laplace approximation (INLA) as the engine for our inference, we first fit a log Gaussian Cox Process (LGCP) [9] independently to each player's point process defined by the (x, y) locations of their made shots using the aforementioned mesh.[2] Each player's estimated intensity function is evaluated at each vertex, producing a K-dimensional vector for each of the $L = 433$ players in our data. These vectors are gathered (by rows) into the $L \times K$ matrix $\mathbf{P}$, which we then factorize via NMF:

$$\mathbf{P} \approx \left(\underset{L\times D}{\mathbf{B}}\right)\left(\underset{D\times K}{\mathbf{\Lambda}}\right). \tag{3}$$

This yields $\mathbf{\Lambda}$, the deterministic bases we use in (2). Collectively, these bases comprise a comprehensive set of shooting tendencies, as shown in Figure 2. Additionally, the NMF in (3) provides player-specific loadings onto these bases, $\mathbf{B}$, which we use in constructing a conditionally autoregressive (CAR) prior on the basis weights, $\mathbf{w}_j$ [10].

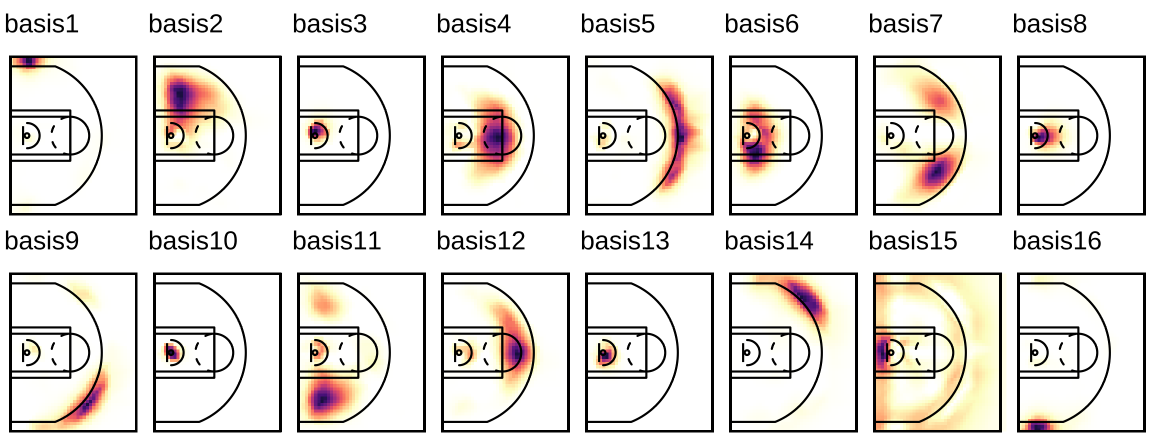

Figure 2: *Deterministic bases resulting from the the non-negative matrix factorization of* $\mathbf{P}$.

The purpose of using a CAR prior on the basis weights is to shrink the FG% estimates of players with similar shooting characteristics toward each other. This is integral for obtaining realistic FG% estimates in areas where a player took a low volume of shots. With only a handful of shots from an area, a player's empirical FG% can often be extreme

[2]Players who took less than 5 shots in the regular season are treated as "replacement players."

(e.g. near 0% or 100%). The CAR prior helps to regularize these extremes by borrowing strength from the player's neighbors in the estimation.

In order to get some notion of shooting similarity between players, we calculate the Euclidean distance between the player loadings contained in $\mathbf{B}$ and, for a given player, define the 5 players with the closest player loadings as his neighbors. We enforce symmetry in the nearest-neighbors relationship by assuming that if player j is a neighbor of player ℓ, then player ℓ is also a neighbor of player j, which results in some players having more than 5 neighbors. These relationships are encoded in a player adjacency matrix $\mathbf{H}$ where entry (j, ℓ) is 1 if player ℓ is a neighbor of player j and 0 otherwise. We can specify the CAR prior on $\mathbf{w}_j$ as

$$(\mathbf{w}_j | \mathbf{w}_{-(j)}, \tau^2) \sim \mathcal{N}\left(\frac{1}{n_j} \sum_{\ell : H_{j\ell}=1} \mathbf{w}_\ell, \frac{\tau^2}{n_j} \mathbf{I}_D\right) \tag{4}$$

$$\tau^2 \sim \text{InvGam}(1, 1).$$

where n_j is the total number of neighbors for player j.

Lastly, we set a $\mathcal{N}(\mathbf{0}, 0.001 \times \mathbf{I})$ prior on $\boldsymbol{\beta}$, and fit the model using INLA. This yields a model that varies spatially and allows us to predict player-specific FG% at any location in the offensive half court. In order to get high resolution FG% estimates, we partition the court into 1ft by 1ft grid cells (yielding a total of $M = 2350$ cells) and denote player j's FG% at the centroid of grid cell i as ξ_{ij}. The estimated projection $(\widehat{\boldsymbol{\xi}}_j)$ for Lebron James is depicted in Figure 3.

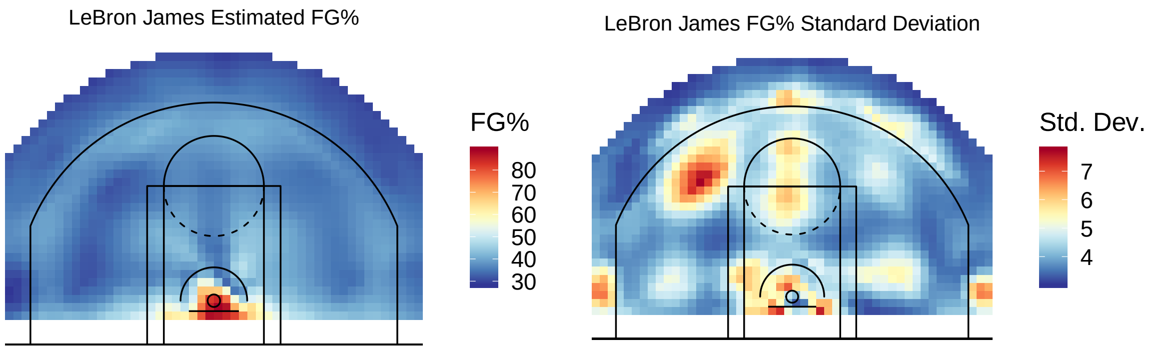

Figure 3: *Lebron James 2016-17 FG% posterior mean (left) and posterior standard deviation (right) projected onto the offensive half court. The prediction surfaces shown here and throughout the figures in this paper utilize projections onto a spatial grid of 1 ft by 1 ft cells.*

In order to have sufficient data to reliably estimate these surfaces, we assume that player FG%'s are lineup independent. We recognize this assumption may be violated in some cases, as players who draw significant defensive attention can improve the FG% of their teammates by providing them with more unguarded shot opportunities. We do not incorporate defensive pressure in our model since data on defender locations are not available publicly, but this is a promising direction for future work given access to proprietary data sources.

2.2 Determining FGA Rate Surfaces

Player FGA rate varies drastically depending on the other players in the lineup. Consider the Oklahoma City Thunder in the introductory example—Russell Westbrook's teammates' attempt rates change drastically based on whether Westbrook is on or off the court. For this reason, when estimating player-specific FGA rate surfaces we also condition on lineup.

The additional sparsity introduced by conditioning a player's FGA rate on lineup does not pose a serious challenge in the estimation. This is because we view a player's decision to shoot the ball as being completely within their control and hence non-random. Therefore, while we still require a model to interpolate FGA rate over the offensive half court, we incorporate no uncertainty in the estimated surfaces, instead treating them as deterministic.

We determine a player's lineup-specific FGA rate surface following three steps. First, we filter a player's collection of shots to those taken in the lineup of interest. Next, we smooth these shot attempts via a LGCP and scale the resulting intensity function to exactly yield the player's observed number of shot attempts in that lineup. Lastly, we normalize the surface to FGA per 36 minutes, allowing us to make meaningful comparisons between lineups who differ in the number of minutes played. As with the FG% surfaces ($\boldsymbol{\xi}$), we partition the court into 1ft by 1ft grid cells and denote player j's FGA rate at the centroid of grid cell i as A_{ij}.

3 Allocative Efficiency Metrics

The models for FG% and FGA rate described in Section 2 are the backbone of the allocative efficiency metrics we introduce in this section: lineup points lost (LPL) and player LPL contribution (PLC). Before getting into the details, we emphasize that these metrics are agnostic to the underlying FG% and FGA models. It is entirely possible to implement the metrics in this section using crude estimates of FG% and FGA rate, for example, by dividing the court into discrete regions and using the empirical FG% and FGA rate within each region. Of course, with better FG% and FGA estimates, the allocative efficiency metrics will be more accurate.

LPL is the output of a two-step process. First, we redistribute a lineup's observed distribution of shot attempts according to a proposed optimum. This optimum is based on ranking the five players in the lineup with respect to their FG% and FGA rate and then redistributing the shot attempts such that the FG% ranks and FGA rate ranks match. Second, we estimate how many points could've been gained had a lineup's collection of shot attempts been allocated according to this alternate distribution. In this section we go over each of these steps in detail and conclude by describing PLC, which measures how individual players contribute to LPL.

3.1 Spatial Rankings Within Lineup

With models for player FG% and player-lineup FGA rate, we can rank the players in a given lineup (from 1 to 5) on these metrics at any spot on the court. For a given lineup, let $\boldsymbol{R}_i^{\xi}$ be a discrete transformation of $\boldsymbol{\xi}_i$—the lineup's FG% vector in court cell i—yielding each player's FG% rank relative to their four teammates. Formally,

$$R_{ij}^{\xi} = \{(n_{\xi_i} + 1) - k : \xi_{ij} \equiv \xi_i^{(k)}\}, \tag{5}$$

where n_{ξ_i} is the length of $\boldsymbol{\xi}_i$, the vector being ranked (this length will always be 5 in our case), and $\xi_i^{(k)}$ is the kth order statistic of $\boldsymbol{\xi}_i$. Since ξ_{ij} is a stochastic quantity governed by a posterior distribution, R_{ij}^{ξ} is also distributional, however its distribution is discrete, the support being the integers $\{1, 2, 3, 4, 5\}$. The distribution of R_{ij}^{ξ} can be approximated by taking posterior samples of $\boldsymbol{\xi}_i$ and ranking them via (5). Figure 11 in the Appendix shows the 20% quantiles, medians, and 80% quantiles of the resulting transformed variates for the Cavaliers starting lineup.

We can obtain ranks for FGA rates in the same manner as for FG%, except these will instead be deterministic quantities since the FGA rate surfaces, $\boldsymbol{A}$, are fixed. We define R_{ij}^{A} as

$$R_{ij}^{A} = \{(n_{A_i} + 1) - k : A_{ij} \equiv A_i^{(k)}\}, \tag{6}$$

where n_{A_i} is the length of $\boldsymbol{A}_i$ and $A_i^{(k)}$ is the kth order statistic of $\boldsymbol{A}_i$. Figure 4 shows the estimated maximum a posteriori[3] (MAP) FG% rank surfaces, $\widehat{\boldsymbol{R}}^{\xi}$, and the deterministic FGA rate rank surfaces, $\boldsymbol{R}^{A}$, for the Cleveland Cavalier's starting lineup.

[3]For the FG% rank surfaces we use the MAP estimate in order to ensure the estimates are always in the support of the transformation (i.e. to ensure $\widehat{R}_{ij}^{\xi} \in \{1, \ldots, 5\}$). For parameters with continuous support, such as $\widehat{\boldsymbol{\xi}}$, the hat symbol denotes the posterior mean.

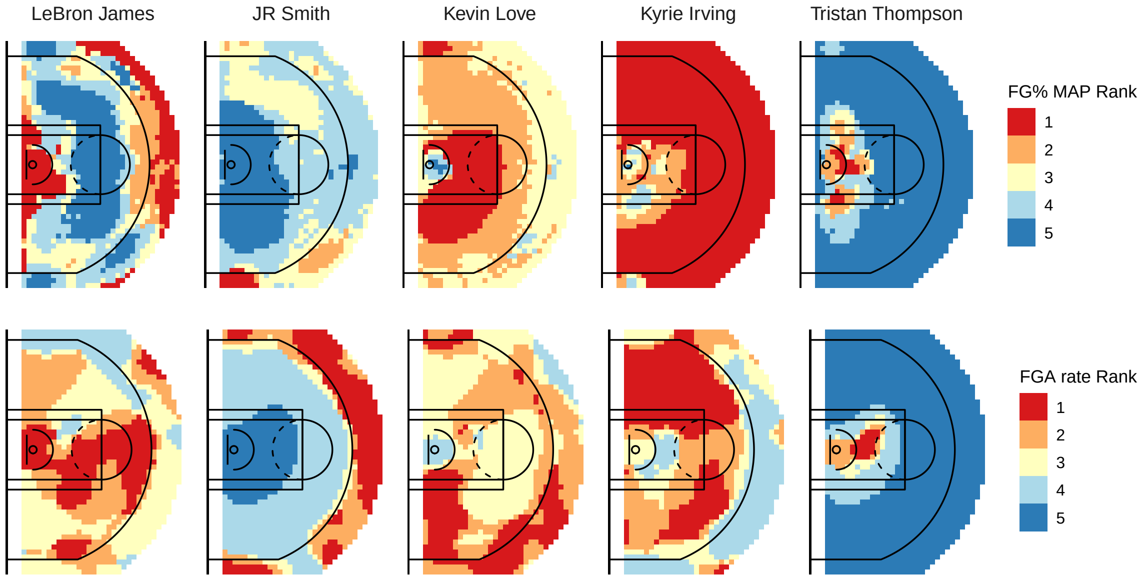

Figure 4: *Top: Estimated FG% ranks for the Cleveland Cavaliers' starting lineup. Bottom: Deterministic FGA rate ranks.*

The strong correspondence between $\widehat{\boldsymbol{R}}^{\xi}$ and $\boldsymbol{R}^{A}$ shown in Figure 4 is not surprising; all other factors being equal, teams would naturally want their most skilled shooters taking the most shots and the worst shooters taking the fewest shots in any given location. By taking the difference of a lineup's FG% rank surface from its FGA rate rank surface, $\boldsymbol{R}^{A} - \widehat{\boldsymbol{R}}^{\xi}$, we obtain a surface which measures how closely the lineup's FG% ranks match their FGA rate ranks. Figure 5 shows these surfaces for the Cavaliers' starting lineup.

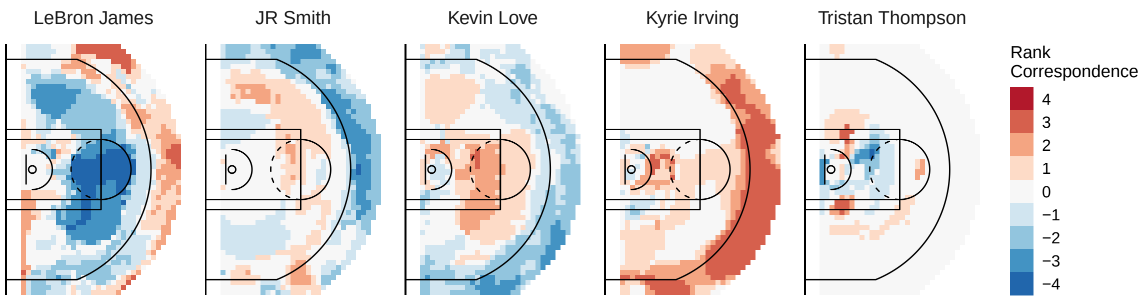

Figure 5: *Rank correspondence surfaces for the Cleveland Cavaliers' starting lineup.*

Note that rank correspondence ranges from -4 to 4. A value of -4 means that the worst shooter in the lineup is took the most shots from that location, while a positive 4 means the best shooter took the fewest shots from that location. In general, positive values of rank correspondence mark areas of potential under-usage and negative values show potential over-usage. For the Cavaliers, the positive values around the 3-point line for Kyrie Irving suggest that he may be under-utilized as a 3-point shooter. On the other hand, the negative values for Lebron James in the mid-range region suggest that he may be over-used in this area. We emphasize, however, that conclusions should be made carefully. Though inequality between the FG% and FGA ranks may be indicative of sub-optimal shot allocation, this interpretation may not hold in every situation due to confounding variables (e.g. defensive pressure, expiring shot clock, etc.).

3.2 Lineup Points Lost

By reducing the FG% and FGA estimates to ranks we compromise the magnitude of player-to-player differences within lineups. For example, if Irving, James, and Thompson are ranked one, two, and three, respectively in a given court location, then the ranked distance between Irving and James is equivalent to the ranked distance between James and Thompson, despite the fact that FG%'s for James and Irving might be much closer in value than the FG%'s for James and Thompson. Here we introduce lineup points lost (LPL), which measures deviation from perfect rank correspondence while retaining the magnitudes of player differences in FG% and FGA. LPL is defined as the difference in expected points between a lineup's actual distribution of FG attempts, $\boldsymbol{A}$, and a proposed redistribution, $\boldsymbol{A}^*$, constructed to yield perfect rank correspondence (i.e. $\boldsymbol{R}^{A^*} - \boldsymbol{R}^{\xi} = \mathbf{0}$). Formally, we calculate LPL in the ith cell as

$$\mathrm{LPL}_i = \sum_{j=1}^{5} \mathrm{v}_i \cdot \xi_{ij} \cdot \left(A_{i[g(R^{\xi}_{ij})]} - A_{ij} \right) \tag{7}$$

$$= \sum_{j=1}^{5} \mathrm{v}_i \cdot \xi_{ij} \cdot \left(A^*_{ij} - A_{ij} \right) \tag{8}$$

where, with respect to court cell i, $g(R^{\xi}_{ij}) = \{k : \; R^{\xi}_{ij} \equiv R^{A}_{ik}\}$, v_i is the point value (2 or 3) of a made shot, ξ_{ij} is the FG% for player j, and A_{ij} is player j's FG attempts (per 36 minutes). The function $g(\cdot)$ simply serves to reallocate the shot attempt vector $\boldsymbol{A}_i$ such that the best shooter always takes the most shots, the second best shooter takes the second most shots, and so forth.

We note that LPL incorporates an intentional constraint—for any court cell i, $\boldsymbol{A}^*_i$ is constrained to be a *permutation* of $\boldsymbol{A}_i$. This ensures that no single player can be reallocated every shot that was taken by the lineup. It also ensures that the total number of shots in the redistribution will always equal the observed number of shots from that location $\left(\text{i.e. } \sum_{j=1}^{5} A_{ij} = \sum_{j=1}^{5} A^*_{ij}, \text{ for all } i\right)$. The left plot of Figure 6 shows LPL (per 36 minutes) over the offensive half court for Cleveland's starting lineup.

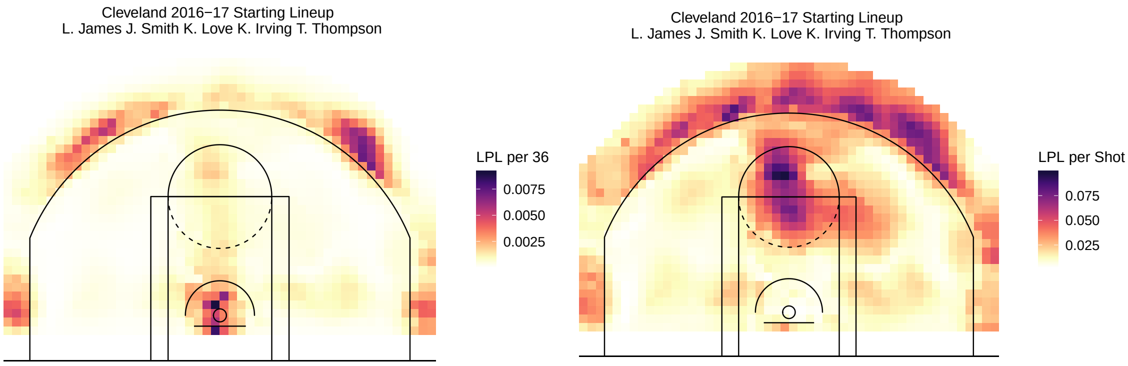

Figure 6: *LPL and LPL per shot surfaces for the Cleveland Cavaliers' starting lineup.*

Notice that LPL values are highest around the rim and along the 3-point line. These regions tend to dominate LPL values because the density of shot attempts is highest in these areas. If we re-normalize LPL with respect to the number of shots taken in each court cell we can identify areas of inefficiency that do not stand out due to low densities of shot attempts:

$$\mathrm{LPL}_i^{Shot} = \frac{\mathrm{LPL}_i}{\sum_{j=1}^{5} A_{ij}}. \tag{9}$$

This formulation gives us the average lineup points lost per shot from region i. This surface is shown in the right plot of Figure 6.

Ultimately, LPL aims to quantify the points that could have been gained had a lineup adhered to the shot allocation strategy defined by $\boldsymbol{A}^*$. However, as will be detailed in Section 4, there is not a 1-to-1 relationship between 'lineup points' as defined here, and actual points. In other words, if a lineup reduces their total LPL by 1, this doesn't necessarily correspond to a 1-point gain in their actual score. In fact, we find that a 1-point reduction in LPL corresponds to a 0.6-point gain (on average) in a team's actual score. One reason for this discrepancy may be due to a tacit assumption in our definition of LPL. By holding each player's FG% constant despite changing their volume of shots when redistributing the vector of FG attempts, we implicitly assume that a player's FG% is independent of their FGA rate. The basketball analytics community generally agrees that this assumption does not hold—that the more shots a player is allocated, the less efficient their shots become. This concept, typically referred to as the 'usage-curve' or 'skill-curve', was introduced by Oliver in [11] and has been further examined by Goldman in [4]. Incorporating player-specific usage curves into LPL could be a promising area of future work.

3.3 Player LPL Contribution

LPL summarizes information from all players in a lineup into a single surface which compromises our ability to identify how each individual player contributes to LPL. Fortunately, we can parse out each player's contribution to LPL and even distinguish between points lost due to undershooting and points lost due to overshooting. We define player j's LPL contribution (PLC)in court location i as

$$\text{PLC}_{ij} = \text{LPL}_i \times \left(\frac{A^*_{ij} - A_{ij}}{\sum_{j=1}^{5} |A^*_{ij} - A_{ij}|} \right), \tag{10}$$

where all terms are as defined in the previous section. The parenthetical term in (10) apportions LPL_i among the 5 players in the lineup proportional to the size of their individual contributions to LPL_i. Players who are reallocated more shots under $\boldsymbol{A}^*_i$ compared to their observed number of shot attempts will have $\text{PLC}_{ij} > 0$. Therefore, positive PLC values indicate potential undershooting and negative values indicate potential overshooting. As in the case of LPL, if we divide PLC by the sum of shot attempts in cell i we obtain average PLC per shot from location i:

$$\text{PLC}^{Shot}_i = \frac{\text{PLC}_i}{\sum_{j=1}^{5} A_{ij}}. \tag{11}$$

The PLC^{Shot}_i surfaces for the Cleveland Cavaliers' 2016-17 starting lineup are shown in Figure 7. We see that Kyrie Irving is potentially being under-utilized from beyond the arc and that Lebron James is potentially over-shooting from the top of the key, which is harmonious with our observations from Figure 5. However, it's worth noting that the LPL per 36 plot (left plot in Figure 6) shows very low LPL values from the mid-range region since the Cavaliers have a very low density of shots from this area. So while it may be true that conditional on a shot attempt Lebron tends to overshoot from the top of the key, the lineup shoots so infrequently from this area that the inefficiency is negligible.

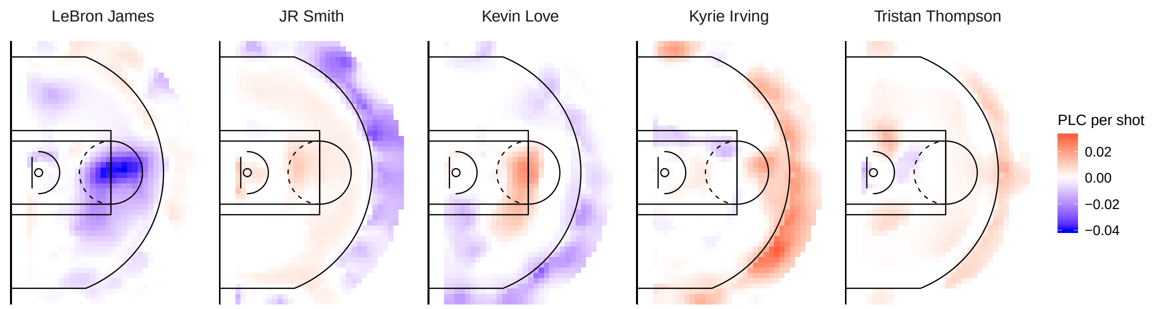

Figure 7: *PLC per shot surfaces for the Cleveland Cavaliers' starting lineup.*

As a final note, notice that for every red region (undershooting) there are corresponding blue regions (overshooting) among the other players. This highlights the fact that LPL is made up of balancing player contributions from undershooting and overshooting; for every player who overshoots, another player (or combination of players) undershoots. By nature of how LPL is constructed, there cannot be any areas where the entire lineup overshoots or undershoots. For this reason, our method does not shed light on shot selection. LPL and PLC say nothing about whether shots from a given region are efficient or not, instead they measure how efficiently a lineup adheres to optimal allocative efficiency given the shot attempts from that region.

4 Optimality - Discussion and Implications

We have now defined LPL and given the theoretical interpretation (i.e. over-use and under use), but we have not yet established that this interpretation is valid in practice. The legitimacy of using LPL as a diagnostic tool for evaluating allocative efficiency hinges on the answers to three important questions:

1. Do lineups minimize LPL?
2. Does LPL relate to offensive production?
3. Is minimizing LPL always optimal?

In this section we explore each of these questions in detail and provide analysis to support our conclusions.

4.1 Do lineups minimize LPL?

While the low values of LPL displayed in Figure 6 (cell values range from 0 to .008 and the sum over all locations in the half court is 0.68) suggest that the Cavaliers starters were minimizing LPL, without a frame of reference we cannot make this claim with certainty. The frame of reference we will use for comparison is the distribution of LPL under completely random shot allocation. In statistical terms, this comparison can be stated as a hypothesis test. We are interested in testing the null hypothesis that shot allocation is random relative to shooter skill to the alternative, that shooter skill informs shot allocation.

A permutation test allows us to test these hypotheses by comparing a lineup's observed total LPL (summing over all court locations, $\sum_i^M \text{LPL}_i$, where M is the total number of 1ft by 1ft cells in the half court) against the distribution of total LPL that we would expect under completely random shot allocation. We can simulate variates from this distribution via

$$\sum_i^M \widetilde{\text{LPL}}_i^{H_0} = \sum_i^M \sum_{j=1}^5 \text{v}_i \cdot \widehat{\xi}_{ij} \cdot \left(A^*_{ij} - \widetilde{A}_{ij}\right) \tag{12}$$

where $\widehat{\xi}_{ij}$ is player j's posterior mean FG% in cell i, $\widetilde{A}_{ij}$ is the jth element of a random permutation of the observed FGA rate vector $\boldsymbol{A}_i$, and all other symbols are defined as in (7-8). Note that a *different* random permutation is drawn for each court cell i. After simulating 500 variates from the null distribution, we approximate the one-sided p-value of the test as the proportion of variates that are less than the observed sum of LPL, $\sum_i^M \text{LPL}_i$.

Figure 8 illustrates this test for the Cleveland Cavaliers' starting lineup. The gray bars show a histogram of the variates from (12) while the red dotted line marks the actual value of total LPL, $\sum_i^M \text{LPL}_i$. Needless to say, the p-value of the test is essentially 0. We can therefore say with certainty that the Cleveland starters do not randomly allocate shots among their players—there is at least some degree of minimization that is occurring. The computational burden of performing the test precludes performing it for every lineup, but we did perform the test for each team's 2016-17 starting lineup. For every team's starting lineup, there were no variates from the null distribution less than their observed sum of LPL with one exception: 11 of the 500 variates from H_0 were less than $\sum_i^M \text{LPL}_i$ for the Sacramento Kings, yielding an approximate p-value of 0.022 for their team. Based on these results we are confident that most lineups employ shot allocation strategies that minimize LPL to some degree, though it appears that some teams do so better than others.

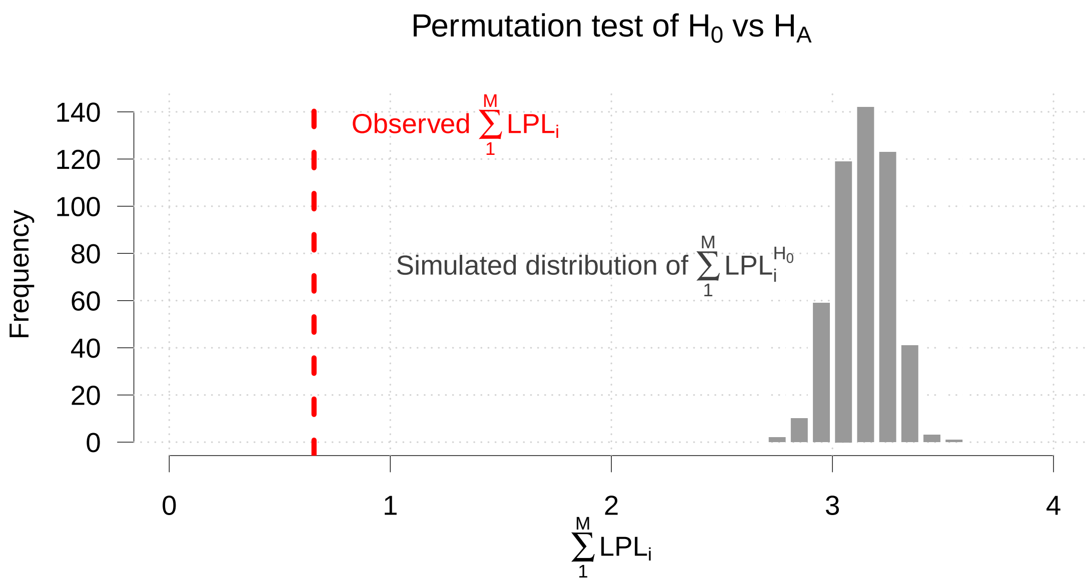

Figure 8: *Permutation test illustration for the Cleveland Cavaliers' 2016-17 starting lineup. The gray bars show a histogram of the variates from (12). The red dotted line is the actual value of total LPL. This value was less than every variate in the histogram by a wide margin.*

4.2 Does LPL relate to offensive production?

We next want to determine whether teams with lower LPL values tend to be more proficient on offense. In order to achieve greater discriminatory power, we've chosen to make this assessment at the game level. Specifically, we regress a team's total game score against their total LPL generated in that game, accounting for other relevant covariates including the team's offensive strength, the opponents' defensive strength, and home court advantage. This framework is analogous to the model proposed by Dixon and Coles in [12].

To calculate LPL at the game level, we must make a few modifications to the LPL construct. As defined in Section 3, LPL reallocates a lineup's total FGA attempts *from the entire season* (normalized to a per 36 minute rate) over a 1ft by 1ft spatial grid. Because the FGA rate surfaces are based on season totals, projecting them over this fine spatial grid is feasible. Unfortunately, this is not true for individual games. At the game level, 1ft by 1ft cells (usually) have at most 1 observed shot attempt for each lineup, hence LPL amounts to reallocating every shot to the best shooter in each location. Essentially, no other players are allocated any shots.

To avoid this issue, we calculate game LPL (GLPL) at a much coarser resolution. We divide the court into three broad court regions (restricted area, mid-range, and 3-pointers), then for a given game and lineup, we calculate GLPL in each of these court regions (indexed by c) by redistributing the lineups' observed vector of shot attempts based on a weighted average of each player's $\widehat{\boldsymbol{\xi}}_j$:

$$\mathrm{GLPL}_c = \sum_{j=1}^{5} \mathrm{v}_c \cdot f_c(\widehat{\boldsymbol{\xi}}_j) \cdot \left(A^*_{cj} - A_{cj}\right) \tag{13}$$

where

$$f_c(\widehat{\boldsymbol{\xi}}_j) = \frac{\sum_{i \in c} w_{ij} \widehat{\xi}_{ij}}{\sum_{i \in c} w_{ij}}. \tag{14}$$

In (14), w_{ij} is a weight proportional to player j's total observed shot attempts in court cell i over the regular season. The notation $\sum_{i \in c}$ simply means we are summing over all the 1ft by 1ft grid cells that are contained in court region c. Finally, for a given game g and team a, we calculate the team's total game LPL (TGLPL) by summing GLPL_c over all court regions c and all lineups ℓ:

$$\mathrm{TGLPL}_{ag} = \sum_{\ell=1}^{L_a} \sum_{c \in C} \mathrm{GLPL}^{\ell}_c \tag{15}$$

where $C = \{\text{restricted area, mid-range, 3-pointers}\}$ and L_a is the total number of team a's lineups. This process is carried out separately for the home and away teams, yielding two TGLPL observations per game and allowing us to include TGLPL as a covariate in a model of offensive proficiency.

We model team a's game score against opponent b in game g as

$$\mathrm{Score}_{abg} = \mu + \alpha_a + \beta_b + \gamma \times \mathrm{I}(\mathrm{Home}_{ag}) + \theta \times \mathrm{TGLPL}_{ag} + \epsilon_{abg} \tag{16}$$

$$\epsilon_{abg} \sim N(0, \sigma^2) \tag{17}$$

where μ represents the global average game score, α_a is team a's offensive strength parameter, β_b is team b's defensive strength parameter, γ governs home court advantage, θ is the effect of TGLPL, and ϵ_{abg} is a normally distributed error term. θ is the parameter that we are primarily concerned with. We fit this model in a Bayesian framework using Hamiltonian Monte Carlo methods implemented in Stan [13]. Our prior distributions are as follows: $\mu \sim N(100, 10^2)$; $\alpha_a, \beta_b, \gamma, \theta \sim N(0, 10^2)$; $\sigma \sim Gamma(\mathrm{shape} = 2, \mathrm{rate} = 0.2)$.

The 95% highest posterior density interval for θ is (-1.08, -0.17) and the posterior mean is -0.62. Therefore, under the posterior mean, we estimate that for each additional lineup point lost a team loses 0.62 actual points. Put differently, by shaving roughly 3 points off of their TGLPL, a team could gain an estimated 2 points in a game. Given that 10% of games were decided by 2 points or less in the 2016-17 season, this could have a significant impact a team's win-loss record and could even have playoff implications for teams on the bubble. Figure 9 shows the estimated density of actual points lost for every team's 82 games in the 2016-17 NBA regular season (i.e. density of $\widehat{\theta} \times \mathrm{TGLPL}_{ag},\ g \in \{1, \ldots, 82\}$ for each team a). Houston was the most efficient team, only losing about 1 point per game on average due to inefficient shot allocation. Washington, on the other hand, lost over 3 points per game on average from inefficient shot allocation.

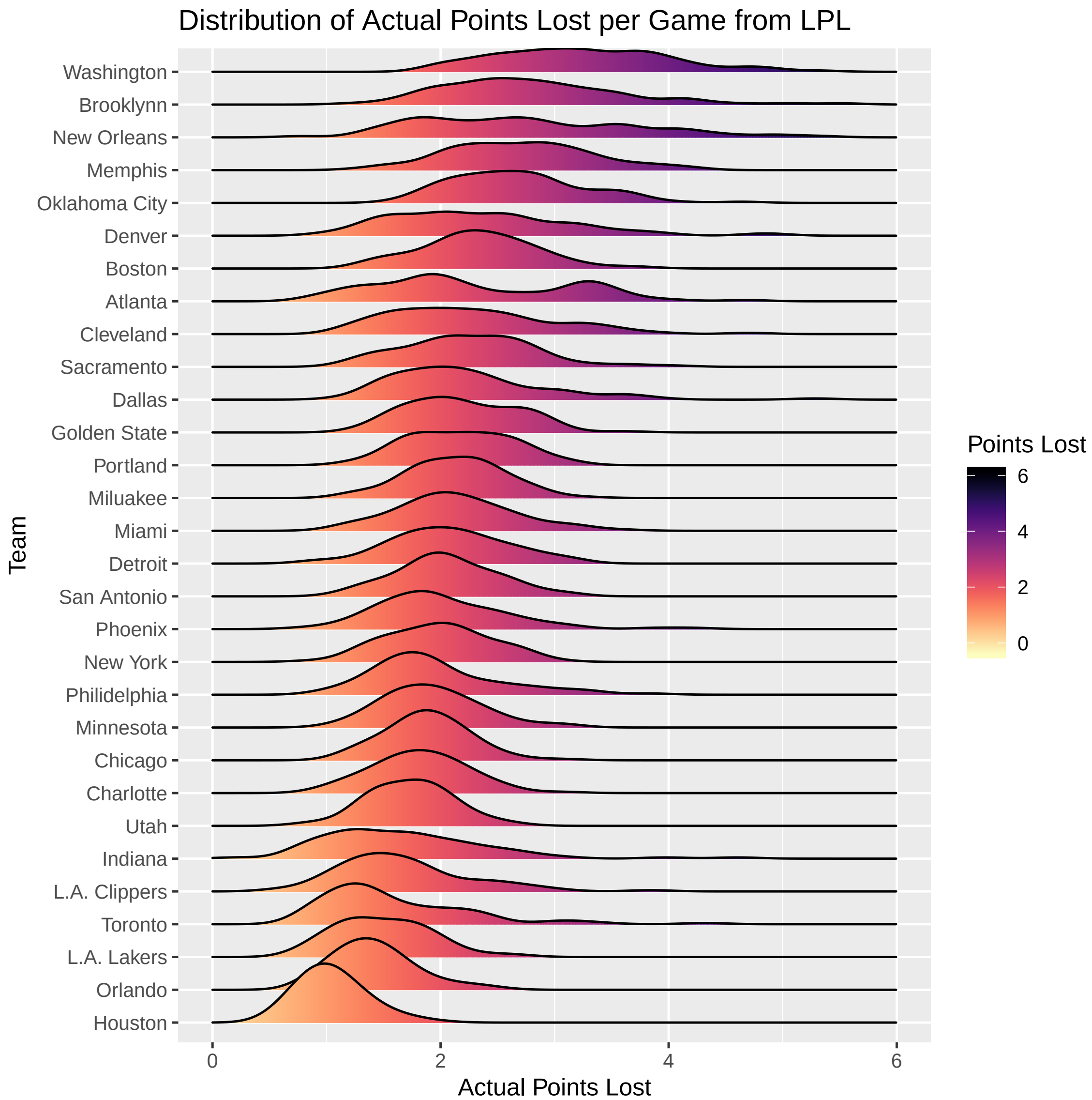

Figure 9: *Estimated density of actual points lost for every team's 82 games in the 2016-17 NBA regular season.*

4.3 Is minimizing LPL always optimal?

While we have demonstrated that lower LPL is associated with increased offensive production, we stress that LPL is a diagnostic tool that should be used to inform basketball experts rather than as a prescriptive measure that should be strictly adhered to in all circumstances. There are certain game situations where minimizing LPL may be sub-optimal.

One such situation is illustrated in Figure 10. The first panel from the left in this figure shows positive PLC values for Russell Westbrook in both the right and left corner 3-point regions, suggesting that Westbrook should be taking more shots from these areas. However, anyone who watched the Thunder play that season will know that many of these corner 3-point opportunities were created by Westbrook driving to the basket, drawing extra defenders toward him, and subsequently kicking the ball out to an open teammate in the corner. Obviously, Westbrook cannot both drive to the rim and simultaneously pass to himself in the corner. In this case, if the Thunder strictly minimized LPL it would reduce the number of these drive-and-kick plays, potentially attenuating their offensive firepower. Shot-creation is not accounted for by LPL and should be carefully considered when exploring LPL and PLC.

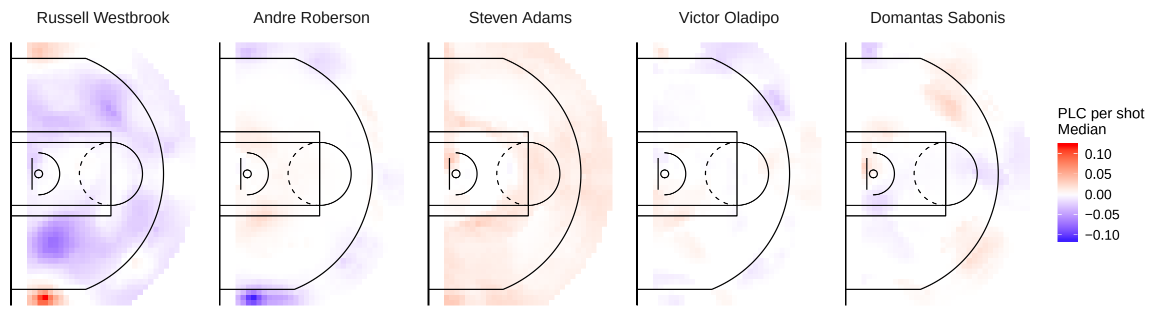

Figure 10: *Oklahoma City 2016-17 starting lineup PLC per shot surfaces.*

There are game theoretic factors to be considered as well. Rigid adherence to minimizing LPL could lead to a more predictable offense and thus make it easier to defend. D'Amour et. al. [14] demonstrated that increased unpredictability in an offense led to more shot opportunities, so having worse shooters take more shots from a given region could potentially confuse defenses leading to more open shot opportunities. Needless to say, offensive game-planning should be informed by more than LPL metrics alone.

5 Conclusion

Our research introduces novel methods to evaluate allocative efficiency spatially and shows that this efficiency has a real impact on game outcomes. The approach we take uses publicly available data and our code base is available online, allowing our methods to be immediately utilized by teams and analysts. Our methods are also simple enough that they could easily be implemented at G-league, NCAA, and international levels. The examples we have shown here provide just a few examples of how LPL can be used to help teams identify areas where they could improve shot allocation among their lineups.

There are many promising directions for future work. As mentioned previously, we do not account for usage curves in our analysis. Doing so would turn LPL into a complex constrained optimization problem, which would be fascinating challenge to tackle. There are also a number of contextual variables which we do not account for but that, with access to proprietary data, present valuable opportunities for improvement of the FG% model and LPL estimation. For example, our estimates could be improved by accounting for defensive pressure and by omitting play types that do not follow the underlying assumptions for allocative efficiency, such as fast breaks, drive-and-kick plays, and double teams. Additionally, by pairing LPL with play call data coaches could gain insight into the efficiency of specific plays. Finally, using LPL to inform player-specific shot policy changes, entire seasons could be simulated using the method in [15] to quantify the impact of specific shot allocation changes on point production. We hope that teams and researchers not only utilize but build upon the methods presented here.

6 Appendix

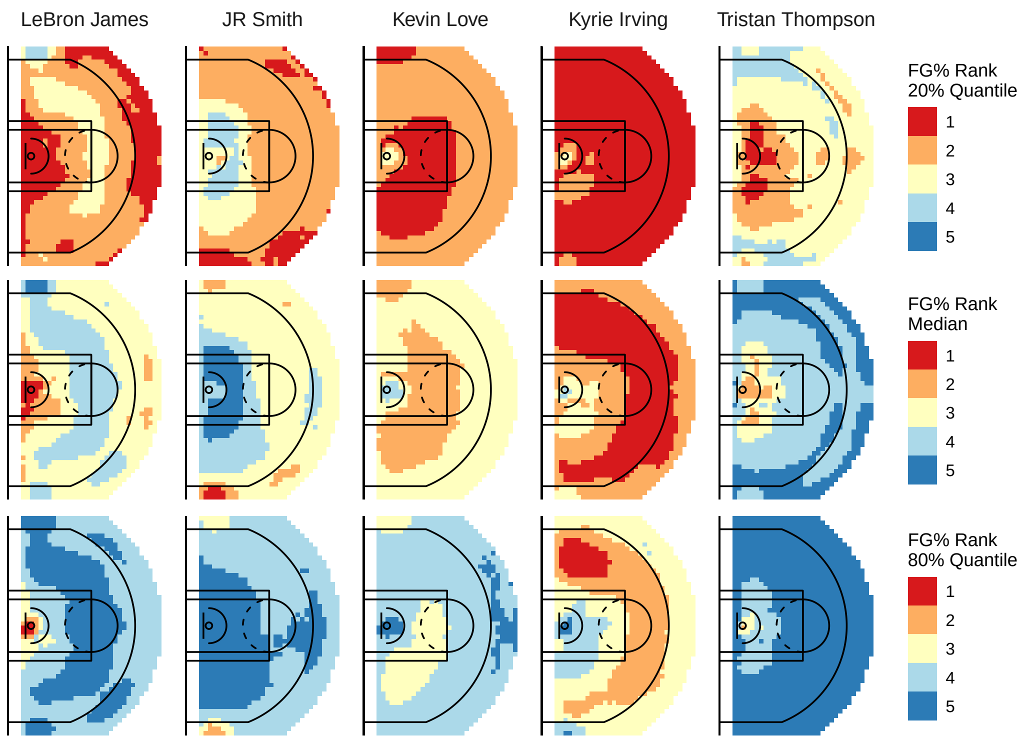

Figure 11: *Top: 20% quantiles of the Cleveland Cavaliers' starting lineup posterior distributions of FG% ranks. Middle: medians of these distributions. Bottom: 80% quantiles.*

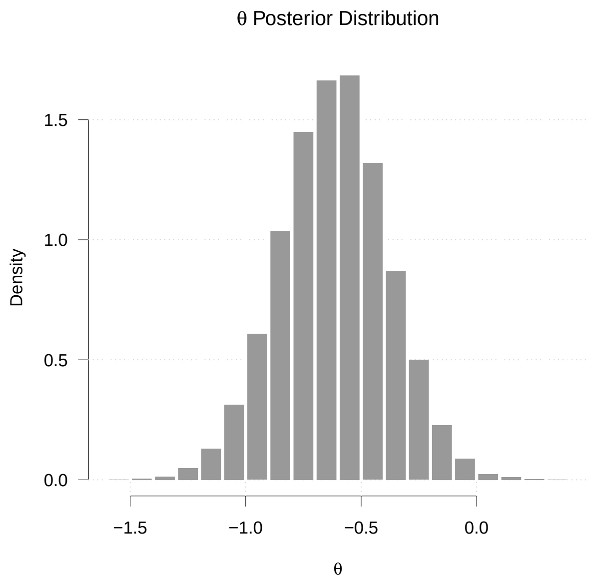

Figure 12: *Posterior distribution of the effect for TGL in model (16-17) described in Section 4.2*

.

References

[1] Justin Kubatko, Dean Oliver, Kevin Pelton, and Dan T Rosenbaum. A starting point for analyzing basketball statistics. *Journal of Quantitative Analysis in Sports*, 3(3), 2007.

[2] Yu-Han Chang, Rajiv Maheswaran, Jeff Su, Sheldon Kwok, Tal Levy, Adam Wexler, and Kevin Squire. Quantifying shot quality in the NBA. In *The 8th Annual MIT Sloan Sports Analytics Conference, Boston, MA*, 2014.

[3] Sports Reference LLC. Calculating PER.

[4] Matt Goldman and Justin M Rao. Allocative and dynamic efficiency in nba decision making. In *The 5th Annual MIT Sloan Sports Analytics Conference, Boston, MA*, 2011.

[5] Daniel Cervone, Alex DAmour, Luke Bornn, and Kirk Goldsberry. A multiresolution stochastic process model for predicting basketball possession outcomes. *Journal of the American Statistical Association*, 111(514):585–599, 2016.

[6] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

[7] Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B*, 73(4):423–498, 2011.

[8] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788, oct 1999.

[9] Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press, Boca Raton, FL, 2nd edition, 2015.

[10] Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society. Series B.*, 36(2):192–236, 1974.

[11] Dean Oliver. *Basketball on Paper: Rules and Tools for Performance Analysis*. Brassey's, Incorporated, 2004.

[12] Mark J Dixon and Stuart G Coles. Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280, 1997.

[13] Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76(1), 2017.

[14] Alexander D'Amour, Daniel Cervone, Luke Bornn, and Kirk Goldsberry. Move or die: How ball movement creates open shots in the nba. *Sloan Sports Analytics Conference*, 2015.

[15] Nathan Sandholtz and Luke Bornn. Transition tensor markov decision processes: Analyzing shot policies in professional basketball. *arXiv preprint arXiv:1812.05170*, 2018.