

Using Box-Scores to Determine a Position's Contribution to Winning Basketball Games

1 Introduction

Basketball is a sport that is becoming increasingly popular world-wide. The National Basketball Association is generally considered to be the pinnacle of competitive basketball. In order for a team to be successful on any level, team members need to recognize their roles (i.e., what particular skills they will need to exhibit) and combine them so as to play as a single unit Bray and Brawley (2002). At the professional level, each of the five positions requires unique skills. The purpose of this paper is to determine which skills a particular position needs to optimally contribute to the teams success in the NBA.

It seems reasonable that some skills would have varying importance according to position. (Note: In this paper importance is the marginal increase in team net points per possession.) For example, a turnover from a guard could possibly be more detrimental to the outcome of the game than a turnover from the center position. A turnover from a guard can occur on the perimeter, which leads to a fast break opportunity for the opposing team. A turnover from a center generally occurs near the basket where the offense can make the transition to defense faster. Also, it seems plausible that offensive rebounds from the center position are more important to the outcome of the game than offensive rebounds from the shooting guard. A center capturing an offensive rebound usually leads to an easy basket attempt, whereas a guard gathering an offensive rebound usually results in resetting the offense, which may or may not result in another attempt at a basket. Also, each coach has a different coaching philosophy that can affect the contribution of a given position. For example, Larry Brown who was of the New York Knicks is vocal about his dislike of point guards that try to score more often than pass. Under Coach Browns system a high scoring point guard is not

desirable, and one would likely not receive much playing time. The relative abilities of the other players on the team also need to be considered. For example, a team with a dominant center may not need as much point production from other positions.

Trninic and Dizdar (2000) used nineteen performance characteristics in an attempt to determine their relative importance for each position. Ten professional basketball experts ranked the importance of each category by position. The experts had a high degree of interobserver agreement. The criteria for guards included categories such as level of defensive pressure and transition defense efficiency, while power forwards and centers had defensive and offensive rebounding efficiency and inside shots rated highly. Berri (1999) linked individual NBA player statistics to team wins via an econometric model. He was primarily interested in measuring each player's production of wins, or the player's marginal product. Dennis Rodman was found to be the highest regular season win producer for the 1997-1998 NBA regular season. Bishop and Gajewski (2004) used box-score categories along with physical characteristics to predict the potential of a collegiate basketball player to be drafted into the NBA. Using multivariate methods and logistic regression, they produced a score for each collegiate player to indicate his likelihood of being drafted. Using a cutoff score of .20, they correctly categorized undrafted players at 90% and drafted players at 78%.

In this paper, we use a hierarchical Bayesian approach to model the difference in points scored as a function of the difference of ten performance categories found in box scores of NBA games. Individual players are assumed to be a random draw from a population of individuals playing a given position. Decisions regarding the relative importance of performance category by position are made using the posterior distributions of the position parameters.

2 Data

The results of NBA games are summarized in a box-score (see Table 1.1). Through USA Today's web site, <http://www.usatoday.com/>, we were able to obtain box scores for the 1996-1997 NBA season. In the USA Today's box-scores the final score of the game, where the game was played, and each participating player's totals for thirteen performance categories

are given. They are: assists (*ast*), steals (*stl*), turnovers (*tov*), free throws made (*ftm*), free throw percentage (*ftp*), field goals made (*fgm*), field goal percentage (*fgp*), offensive rebounds (*orb*), defensive rebounds (*drb*), minutes played (*min*), personal fouls (*pf*), total points (*pts*), and total rebounds (*trb*).

These box-scores identify a players position only if the player starts. But often players that don't start have an impact on the outcome of the game. So to include all players that participate we group those players that don't start in a "bench" position. In addition, these box-scores make no distinction between a point guard and a shooting guard or between a small forward and a power forward. Since point guards and shooting guards usually have vastly different roles within the framework of the team we wanted to be able to distinguish between the two. So based on our personal recollection and using the internet as a further resource we separate the guards into point guards and shooting guards and the forwards into small forwards and power forwards.

LA LAKERS (78) AT UTAH (104)														
LA LAKERS														
PLAYER	POS	MIN	FGM	FGA	FTM	FTA	OFF	REBOUNDS		AST	PF	STL	TOV	PTS
								DEF	TOT					
C BUTLER	F	24	1	8	2	2	1	2	3	0	2	0	0	4
L ODOM	F	35	3	10	8	10	2	7	9	1	5	1	3	14
C MIHM	C	24	3	8	0	2	2	3	5	0	4	0	3	6
C ATKINS	G	19	0	2	2	2	0	1	1	1	3	0	2	2
K BRYANT	G	41	9	21	16	20	1	3	4	1	5	1	2	38
T BROWN		23	0	3	1	2	1	2	3	2	0	0	0	1
B COOK		17	1	5	1	2	0	2	2	0	4	0	0	3
J JONES		15	1	2	0	0	1	1	2	1	4	0	0	2
B GRANT		10	0	1	1	2	1	1	2	0	3	0	2	1
L WALTON		12	1	3	0	0	1	0	1	0	1	0	0	3
K RUSH		17	0	4	0	0	3	0	3	0	0	2	1	0
S VUJACIC		3	1	1	2	2	0	2	2	1	1	0	1	4
TOTALS		240	20	68	33	44	13	24	7	7	32	4	14	78
UTAH														
PLAYER	POS	MIN	FGM	FGA	FTM	FTA	OFF	REBOUNDS		AST	PF	ST	TO	PTS
								DEF	TOT					
A KIRILENKO	F	38	5	6	6	7	2	4	6	1	2	1	5	16
C BOOZER	F	34	10	13	7	10	5	6	11	3	1	0	0	27
J COLLINS	C	10	0	1	0	0	0	1	1	0	4	0	0	0
K MCLEOD	G	29	2	7	2	2	0	0	0	8	5	1	1	6
G GIRICEK	G	12	4	9	3	3	0	1	1	0	4	0	2	11
M OKUR		20	0	4	2	2	1	4	5	4	1	1	0	2
R BELL		25	4	8	2	2	0	3	3	0	5	0	1	10
M HARPRING		28	9	11	3	3	2	5	7	4	4	1	2	23
H EISLEY		19	2	7	0	0	0	0	0	3	1	0	0	4
C BORCHARDT		14	2	3	1	2	1	5	6	1	2	0	1	5
K SNYDER		9	0	3	0	0	0	3	3	0	2	0	0	0
K HUMPHRIES		2	0	3	0	0	0	0	0	0	0	0	0	0
TOTALS		240	38	75	26	31	11	32	43	24	31	4	12	104

Table 1: Typical USA Today NBA Box-Score

In 1996-1997 the NBA had 29 teams so the number of teams is $T = 29$, and the number of opponents is $O = 29$. In this study the number of players that started a game is 343. Then,

treating each teams bench as a “player”, the total number of distinct players is $Pl = 371$. Each of the five starters is assigned a position at the beginning of a game and together with the bench position gives the total number of positions to be $Po = 6$. We assume that a players positional assignment remains constant throughout the game. We understand that assuming a player plays the same position throughout the game is not a “true” representation of what actually occurs during the course of a NBA game but this is a limitation of the data. The positions are: point guard (pg), shooting guard (sg), small forward (sf), power forward (pf), center (c), and bench (b). There are box-scores for 1163 games from the 1996-1997 season so the number of games is $G = 1163$.

All categories besides rebounds and shooting percentages are standardized by possessions. The number of possessions in which a player competes needs to be estimated. To do this we use the following formulas:

$$pp = \frac{pm}{48} \times tp \quad (1)$$

where pp is the number of possessions in which a player participates, pm is the number of minutes played by the player, tp is the teams possession total, and 48 is the number of minutes in a NBA game. For the bench position we use 240 in the denominator which is the total number of minutes available. To estimate tp we use the following equation which incorporates a couple of categories found in the box-score. The equation is:

$$tp = fga - orb + tov + 0.4fta \quad (2)$$

Both equations (1) and (2) are found in Dean Oliver’s book *Basketball On Paper* (Oliver (2004)). To standardize rebounds we use individual offensive and defensive rebounding percentages. They are obtained using the following formula:

$$PlayerOR\% = \frac{PlayerOR}{\%Min \times (TeamOR + OppDR)} \quad (3)$$

$$PlayerDR\% = \frac{PlayerDR}{\%Min \times (TeamDR + OppOR)} \quad (4)$$

Where $PlayerOR$ is the total number of offensive rebounds for a player, $TeamOR$ and $TeamDR$ are the teams total number of offensive and defensive rebounds, $OppOR$ and $OppDR$ are the opponents total number of offensive and defensive rebounds and $\%MIN$

is the percent of total minutes that a player participated in a game. After standardizing the box-score categories each player is paired by position (i.e. point guards from competing teams are paired and shooting guards from competing teams are paired, etc.) for each game and the differences for the standardized box-score categories between the matched players are computed (home team player minus visiting team player). These differences become the explanatory variables. We use differences so that we might consider the talent level of the opposition. For the response variable we use the difference in the final score of the game (which will be referred to as point spread).

3 Model

We use nine of the available box-score categories as explanatory variables in the model. Also we use the equation $3pm = pts - 2fgm - ftm$ to obtain the number of three pointers made. Unfortunately with the present data it is impossible to determine how many three pointers a player attempted so 3-points shooting percentage will not be included in the model. Total rebounds and total points scored are not included because they are linear combinations of offensive rebounds and defensive rebounds and field goals made and free throws made, so including them would introduce collinearity and render the results unreliable.

A standard multiple regression model that would determine the linear relationship between point spread and the ten categories is:

$$y = \beta_0 + \beta_1 ast + \beta_2 stl + \beta_3 tov + \beta_4 3pm + \beta_5 ftm + \beta_6 ftp + \beta_7 fgm + \beta_8 fgp + \beta_9 orb + \beta_{10} drb + \epsilon \quad (5)$$

where y is the point spread, β_0 is the overall intercept, β_1 is the effect that difference in assists (ast) has on the point spread holding all else constant, etc., and $\epsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$. This model also assumes that the effects are additive. We will examine the assumptions of this model in some detail in the following paragraphs. However, the additivity assumption seems to be reasonable based on preliminary studies.

Clearly, an inadequacy in equation (5) is the assumption that each β_h , $h = 1, \dots, 10$, is the same regardless of player or position. In addition, the way in which the data are used creates dependence in the response variable for individuals on the same team, the opponent

they play, and the game in which they play, thus the assumption of independent responses is violated. We obtain a separate β estimate for each player and deal with the dependency issues by incorporating a Bayesian hierarchical model on the regression coefficients and adding an effect for the team of which the player is a member, the opponent the player's team is playing, and the game in which they are competing. Furthermore, we include an effect in model that corresponds to where the game is being played thus incorporating home court advantage. The model then becomes:

$$y_{iklmq} = \gamma_m + \tau_k + \phi_l + \nu_q + \sum_{h=1}^{10} \beta_{hi} x_h, \quad (6)$$

with $i = 1, \dots, 371, k = 1, \dots, 29, l = 1, \dots, 29, q = 1, \dots, 29$, and, $m = 1, \dots, 1163$ and x_h is the h^{th} explanatory variable and β_{hi} is the h^{th} regression coefficient for the i^{th} player.

Each of the three parameters γ, ϕ , and τ deals with a different aspect of the result. The team effect is modeled by τ_k so that players coming from the same team have the same intercept. This parameter also can help account for different coaching philosophies and talent level of teams. Both of coaching and talent level would affect a position's role in the team framework. The opponent effect is addressed with ϕ_l . The γ_m accounts for a game effect, and ν_q deals with home court advantage.

The point spread is symmetric around zero because each game contributes both a positive and negative response of the same magnitude, one to the winning and one to the losing team. NBA games are fairly competitive, which means most games would have differences close to zero and large differences would not be as common. For these reasons, a Gaussian distribution is a good choice for the likelihood of the Bayesian model. Thus, we assume:

$$y_{iklmq} \sim N(\theta_{iklmq}, \sigma^2), \quad (7)$$

where θ_{iklmq} is equal to the right side of equation (6).

In this study the difference in points can never be zero since a game never ends in a tie. Thus, we recognize that the normal likelihood cannot be exactly true. Nevertheless, we believe it is a reasonable approximation in this setting (For more justification of this choice of likelihood see Oliver (2004)).

4 Analysis Strategies

In this section we discuss the reasoning behind the assignment of the prior distributions for the parameters, computational methods, and convergence diagnostics.

The β_{hi} 's, γ_m 's, ϕ_l 's, τ_k , and ν_q 's can theoretically be either positive or negative, which leads to choosing a distribution that is defined for all real numbers. A priori there does not seem to be any reason why these parameters would take on large values opposed to small ones or visa-versa which leads to symmetry being a desirable characteristic. Thus, we select normal prior distributions for these parameters and we have:

$$\beta_{hi} \sim N(\mu_{\beta_{h,j}}, \sigma_{\beta_h}^2), \gamma_m \sim N(m_\gamma, \sigma_\gamma^2), \phi_l \sim N(m_\phi, \sigma_\phi^2), \tau_k \sim N(m_\tau, \sigma_\tau^2), \text{ and } \nu_q \sim N(m_\nu, \sigma_\nu^2).$$

Notice that the mean of the distributions for the β_{hi} 's changes according to the j^{th} position of the i^{th} player. So $\mu_{\beta_{1,1}}$ gives the estimate for the effect of point guard assists, $\mu_{\beta_{1,2}}$ returns an estimate for the effect of shooting guard assists, etc. In this way the β_{hi} 's for each player come from a "position" distribution. Thus $\mu_{\beta_{h,j}}$ is the mean of the position distribution for position j and regression coefficient h . These means will be the estimated position effect and the focus of this study. By letting each β be drawn from a position distribution we are able to "borrow strength" from all players of the same position and estimate an overall position effect. We assume $\sigma_{\beta_h}^2$ remains constant over the positions.

We use the same arguments above to assign a Gaussian distribution to the $\mu_{\beta_{h,j}}$'s and we have:

$$\mu_{\beta_{h,j}} \sim N(m_{\mu_{\beta_h}}, s_{\mu_{\beta_h}}^2).$$

That is, we assume that the means of the β 's for the $j = 5$ different positions are drawn from the same distribution.

The overall variance of the model, σ^2 , is by definition greater than or equal to zero. This necessitates a prior distribution that is always positive. The inverse gamma distribution preserves the parameter space, is very flexible in its shape, and yields closed form complete conditionals. Therefore, the inverse gamma is a logical choice for the prior on σ^2 . Using the same logic leads us to assign an inverse gamma distribution to the $\sigma_{\beta_h}^2$'s along with σ_γ^2 , σ_τ^2 , and σ_ϕ^2 . Thus, for the variance parameters we have:

$$\sigma_{\beta_h}^2 \sim IG(a_{\sigma_{\beta_h}}, b_{\sigma_{\beta_h}}), \sigma_\gamma^2 \sim IG(a_{\sigma_\gamma}, b_{\sigma_\gamma}), \sigma_\tau^2 \sim IG(a_{\sigma_\tau}, b_{\sigma_\tau}), \sigma_\phi^2 \sim IG(a_{\sigma_\phi}, b_{\sigma_\phi}), \text{ and } \sigma^2 \sim IG(a_\sigma, b_\sigma).$$

4.1 Hyperparameter Values

This section details the selection and reasoning behind the choices for hyperparameter values. Values need to be determined for: $m_{\mu_{\beta_h}}, s_{\mu_{\beta_h}}^2, m_{\gamma}, m_{\tau}, m_{\phi}, a_{\sigma_{\beta_h}}, b_{\sigma_{\beta_h}}, a_{\sigma_{\gamma}}, b_{\sigma_{\gamma}}, a_{\sigma_{\tau}}, b_{\sigma_{\tau}}, a_{\sigma_{\phi}}, b_{\sigma_{\phi}}, a_{\sigma},$ and b_{σ} .

Coming up with values for the $m_{\mu_{\beta_h}}$'s is not particularly intuitive, since each slope represents the expected change in point spread given a difference of one per possession in h^{th} category holding all other categories constant. Since this relationship is difficult to formalize even for experts, it seems reasonable that the prior specifications for these parameters should be fairly diffuse. Since an assist and a field goal made results in two points, it seems reasonable to believe that these two categories have the largest spread of possible values. So we chose priors for these two categories and assigned these values to the prior distributions of the remaining seven box-score categories.

A priori, the $\mu_{\beta_{h,j}}$'s, could be either positive or negative depending on the regression coefficient and the position. Hence, it seems reasonable that $m_{\mu_{\beta_h}} = 0$. $s_{\mu_{\beta_h}}^2$ describes the distance from zero that the $\mu_{\beta_{h,j}}$'s could plausibly assume. We focus on $s_{\mu_{\beta_1}}^2$ which is the spread of the means for assists. An average NBA game consists of 90-100 possessions depending on team and opponent. In light of this, β_{1i} could take on values as large as 100-150 if all points were scored from an assist, but this is extreme and unlikely. It is more plausible that around half of field goals made are assisted. So an upper limit somewhere in the neighborhood of 60 might be a more reasonable estimate. We chose to let $s_{\mu_{\beta_1}}^2 = 15^2$ which implies $\mu_{\beta_{1,j}}$ could plausibly be assigned values up to 60 which will in turn allow β_{1i} to take on values as large as 60. We assigned the same value to the remaining $s_{\mu_{\beta_h}}^2$'s, which complies with the desire to be diffuse for all the performance categories

$\sigma_{\beta_h}^2$ measures the variability of the β_{hi} 's. We will use moment matching to find suitable parameter values. That is, we will choose means and variances that reflected our belief about how the slopes might vary, and then find the parameters from the inverse gamma distribution that correspond to the chosen means and variances. Once again, we first consider $\sigma_{\beta_1}^2$, the spread of the assist effect. It seemed reasonable that the variability of the assist effect within a player would not be very large relative to the point spread. Choosing $E(\sigma_{\beta_1}^2) = 2^2$ and

$\text{var}(\sigma_{\beta_1}^2) = 3^2$, allows the standard deviation of β_{1i} to plausibly reach values above 10, which is a rather large point spread. This mean and variance produce inverse gamma parameters of $a_{\sigma_{\beta_1}} = \frac{34}{9}$ and $b_{\sigma_{\beta_1}} = \frac{9}{100}$. We assign the same values to the remaining β 's.

$$\sigma_{\beta_h}^2 \sim IG\left(\frac{34}{9}, \frac{9}{100}\right).$$

We assign $m_\gamma = m_\phi = m_\tau = m_\nu = 0$. This seems reasonable because the four parameters (τ_k , ϕ_l , γ_m , and ν_q) can take on either positive or negative values depending on the team, opponent, and game.

It seemed plausible that the effects for team, opponent, and homecourt (τ , ϕ , and ν) would be similar in their distributional form. Therefore, the same values were given to their hyperparameters. σ_τ^2 , σ_ϕ^2 , and σ_ν^2 are parameters that represent the variability that exists from team to team, which is probably larger than the within player variance, but large deviations from zero still seemed rather unlikely. So it seemed reasonable to find inverse gamma parameters that correspond with $E(\sigma_\tau^2) = E(\sigma_\phi^2) = E(\sigma_\nu^2) = 3^2$ and $\text{var}(\sigma_\tau^2) = \text{var}(\sigma_\phi^2) = \text{var}(\sigma_\nu^2) = 3^2$. This allows the standard deviations of τ , ϕ , and ν to plausibly take on values up to 15, which corresponds to a rather large point spread. The values of the inverse gamma distribution that correspond with the desired mean and variance are $a_{\sigma_{\beta_1}} = 11$ and $b_{\sigma_{\beta_1}} = \frac{1}{90}$. Thus we have

$$\sigma_\tau^2, \sigma_\phi^2, \sigma_\nu^2 \sim IG\left(11, \frac{1}{90}\right).$$

The effect γ is interpreted as the point spread for a particular game given that the competing teams recorded the same number of assists per possession, steals per possession, turnovers per possession, and so on. The variance of this effect, which is the within game variance, is probably smaller than that of the team and opponent effects. Once again we use moment matching to find values to the distribution of σ_γ^2 . We found values so that $E(\sigma_\gamma^2) = (\sqrt{2})^2$ and $\text{var}(\sigma_\gamma^2) = 3^2$. These values would allow the standard deviation to plausibly reach values as high as 10. Thus we chose:

$$\sigma_\gamma^2 \sim IG\left(\frac{22}{9}, \frac{173}{500}\right).$$

σ^2 is the variability of the error term. We thought it likely that on average the standard deviation would be about 6. An inverse gamma with $E(\sigma^2) = 6^2$ and $\text{var}(\sigma^2) = (2\sqrt{5})^2$ would seem to be reasonable. Again, by moment matching, we chose:

$$\sigma^2 \sim IG\left(\frac{131}{25}, \frac{3}{500}\right).$$

For a summary of hyperparameter values see table 2.

Table 2: Hyperparameter Values

Parameter	m	s^2	Parameter	a	b
$\mu_{\beta_{1j}}$	0	225	$\sigma_{\beta_1}^2$	34/9	9/100
$\mu_{\beta_{2j}}$	0	225	$\sigma_{\beta_2}^2$	34/9	9/100
$\mu_{\beta_{3j}}$	0	225	$\sigma_{\beta_3}^2$	34/9	9/100
$\mu_{\beta_{4j}}$	0	225	$\sigma_{\beta_4}^2$	34/9	9/100
$\mu_{\beta_{5j}}$	0	225	$\sigma_{\beta_5}^2$	34/9	9/100
$\mu_{\beta_{6j}}$	0	225	$\sigma_{\beta_6}^2$	34/9	9/100
$\mu_{\beta_{7j}}$	0	225	$\sigma_{\beta_7}^2$	34/9	9/100
$\mu_{\beta_{8j}}$	0	225	$\sigma_{\beta_8}^2$	34/9	9/100
$\mu_{\beta_{9j}}$	0	225	$\sigma_{\beta_9}^2$	34/9	9/100
$\mu_{\beta_{10j}}$	0	225	$\sigma_{\beta_{10}}^2$	34/9	9/100
m_τ	0		σ_τ^2	11	1/90
m_ϕ	0		σ_ϕ^2	11	1/90
m_ν	0		σ_ν^2	11	1/90
m_γ	0		σ_γ^2	22/9	173/500
			σ^2	131/25	3/500

4.2 Computation

The joint posterior distribution is highly multidimensional. In order to obtain posterior distributions we used Markov chain Monte Carlo (MCMC) simulation techniques. With our choice of likelihood and prior distributions, we have complete conditionals that are known and easy to sample from for all parameters. Because of this, we can use the Gibbs sampling algorithm as described by Gelfand and Smith (1990) to explore the posterior space and obtain draws from the posterior distribution. The complete conditionals were coded in FORTRAN and were used to obtain 25,000 posterior draws following a burn of 50,000 and a thinning

of 80. That is, every 80th draw was kept after the initial 50,000 draws were discarded until there were 25,000.

4.3 Convergence Diagnostics

Checking the convergence of Markov chains is a difficult task in models that have a large number of parameters. Time series plots were used to assess the mixing of the chains. To check convergence, we use the criteria as explained by Raftery and Lewis (1992) and their *gibbsit* function that can be used in the statistical software package R. All parameters in the model met the criteria set forth by Raftery and Lewis.

5 Results

In this section we compare the 10 box-score categories for the six positions. In the following discussion if it is not explicitly stated that the result for an effect is in the presence of all other effects then it is implicitly implied. In addition, the following results are recommendations for positions beyond what one would think is their normal roll. It is obvious, for example, that a center that solely focuses on steals and disregards his rebounding duties would be counter beneficial. Hence what follows is a discussion of the marginal effect of the ten box-score categories by position. Also, recall that each effect is a difference of the positional match up. That is, for example, if a point guard gets one more assist per 100 possessions than his opponent this is worth .3532 points in point spread holding all other positions and categories constant. If the above is not explicitly written it is implicitly implied.

5.1 Positional Performance

Our goal was to determine which skills were most important by position and the effect these skills have on the outcome of the game. In light of this, we focus our attention on the $\mu_{\beta_{h,j}}$'s.

The units of the $\mu_{\beta_{h,j}}$'s are the point spread given a difference in assists per possession, and difference in steals per possession, and so forth. Interpretation of the results in these units is difficult. (A point guard will never out rebound his positional opponent by one offensive rebound per possession). For this reason, we divide the posterior draws of the $\mu_{\beta_{h,j}}$'s by 100. Most comparisons in NBA basketball are done on a per 100 possession basis(see Oliver (2004)). Thus, we interpret the results as the point spread given the difference in assists per 100 possessions, and the difference in steals per 100 possessions, and so forth. Also, we divide both shooting percentage (free throw percentage (ftp) and field goal percentage (fgp)) and rebound percentage effects by 100 so we can interpret the effects as the average point spread given one percent increase in shooting percentage or rebounding percentage holding all other effects constant. These operations are valid because of posterior invariance.

Figures 1 and 2 provide density plots of the $\mu_{\beta_{h,j}}$'s. Table 3 contains a summary of the posterior distributions to these parameters. These figures and tables reveal some interesting associations. Of note, the posterior distributions for the “bench” position are less variable since the results for all players not starting were combined to represent this position. This, of course, provided more possessions and hence more data for this position. (The following interpretations are made in the context of the NBA. Since college, high school, and international basketball are different applying these results to leagues other than the NBA could be problematic. Although methodologies are transferable depending on available data.

For all five positions, out-assisting your opponent has a very positive impact on a basketball game. A result that was somewhat unexpected was that a small forward out-assisting his opponent on average was the most beneficial to the team. In fact, the small forward assist effect has the largest positive impact among the all the position box-score category combinations. The assist effect in general has the

Something that was not foreseen was how important it was for the center position to record more steals than his opponent. A center that gets one more steal per one hundred possessions than his opponent gives his team a 0.3793 points advantage on average. Although steals is not a perfect defensive statistic, (players with a high number of steals tend to gamble a bit on defense) it does give an indication of the relative athleticism of a player. Having an

Figure 1: Posterior Distributions for assists, steals, turnovers, 3-pointers made, and free throws made for each position

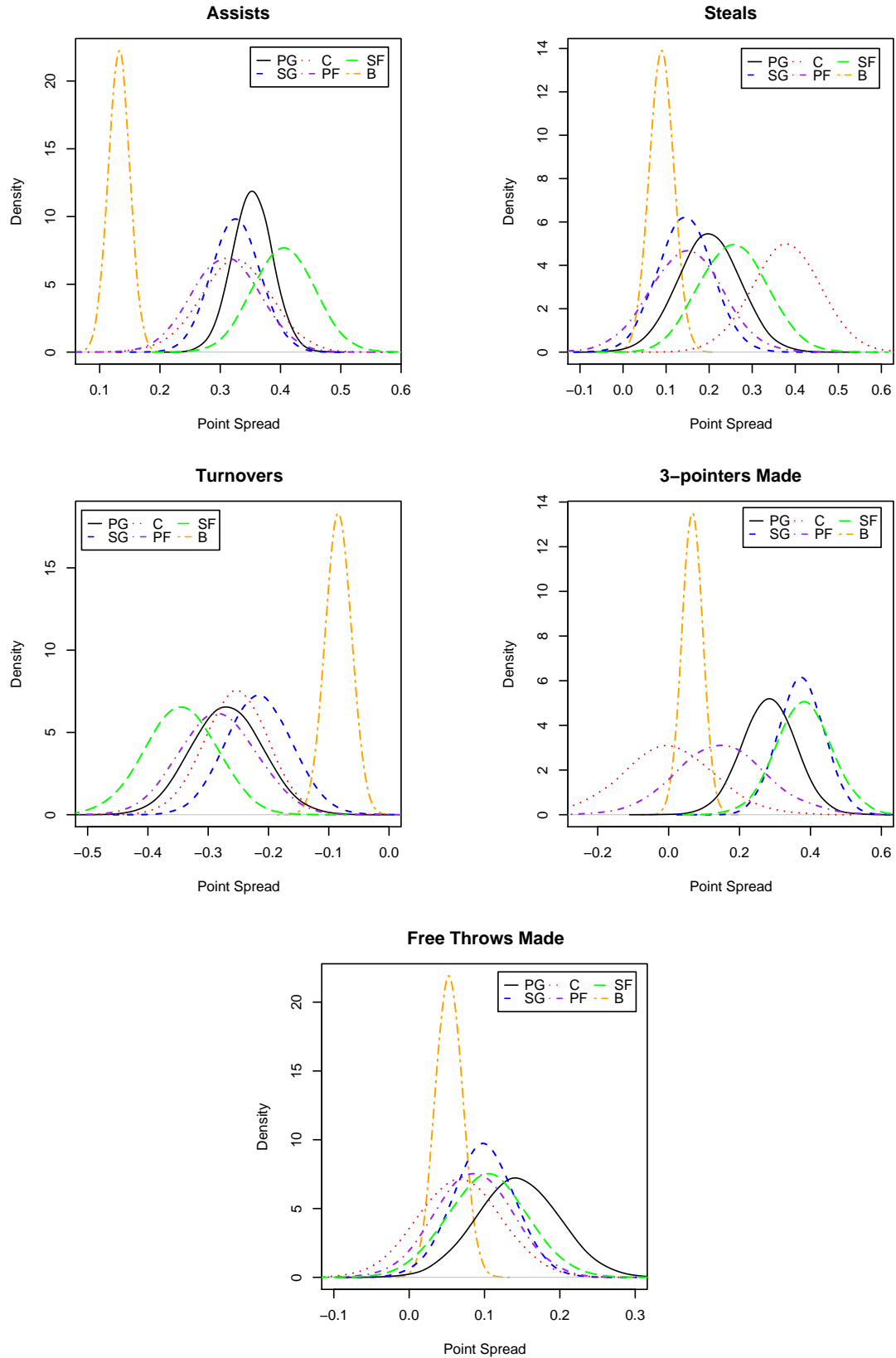


Figure 2: Plots of posterior distributions of parameters for free throws made, free throw percentage, field goals made, and field goal percentage for each position

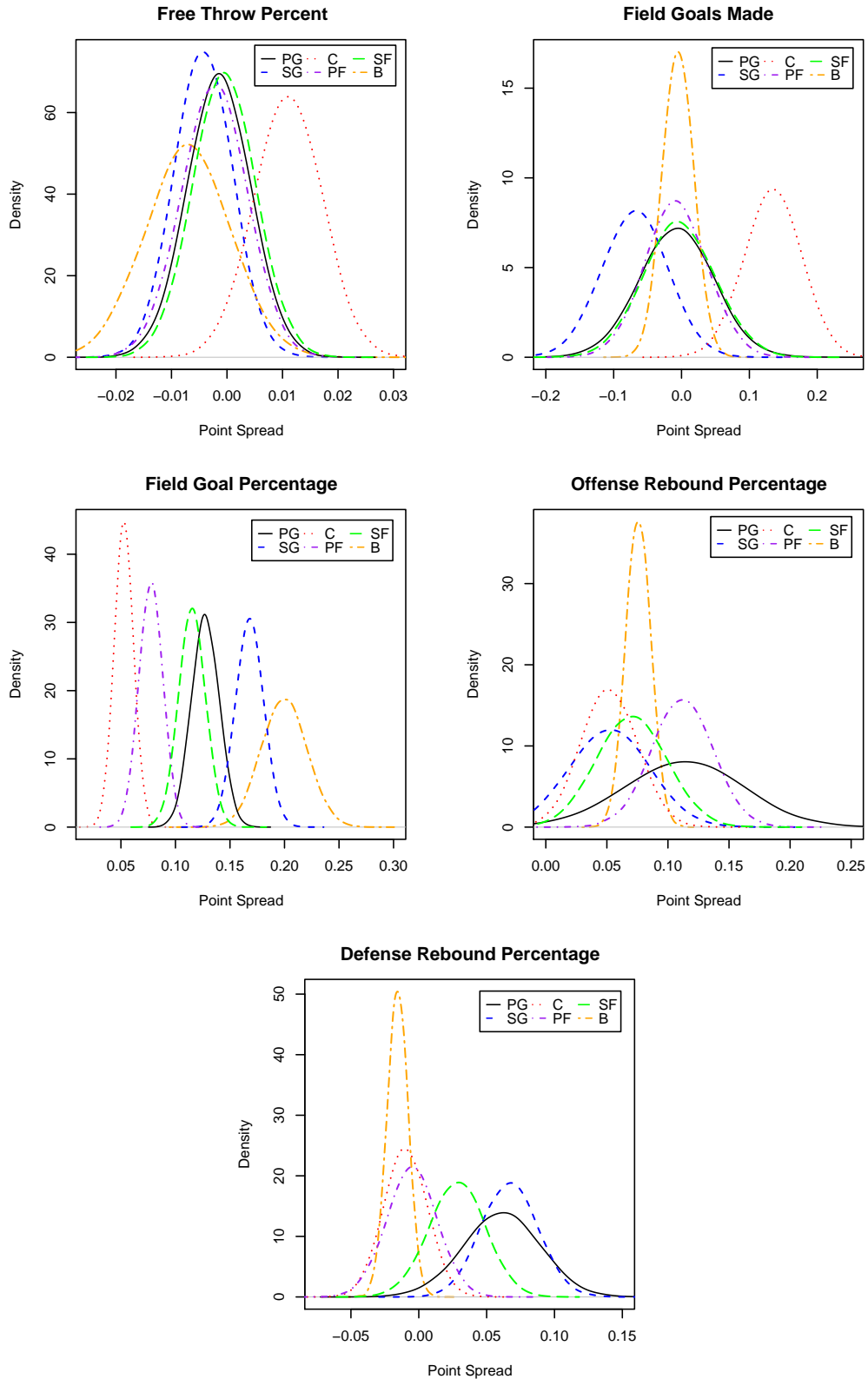


Table 3: Posterior means, standard deviations, and 95% HPD credible intervals of the positional categories

Position	Mean	StdDev	2.5% LHPD	97.5% UHPD	Position	Mean	StdDev	2.5% LHPD	97.5% UHPD
Assist(μ_{β_1})					Steals(μ_{β_2})				
Point Guard	0.3532	0.0328	0.2883	0.4175	Point Guard	0.1993	0.0712	0.0559	0.3341
Shooting Guard	0.3260	0.0396	0.2449	0.4008	Shooting Guard	0.1439	0.0625	0.0240	0.2671
Center	0.3254	0.0580	0.2160	0.4463	Center	0.3793	0.0791	0.2189	0.5302
Power Forward	0.3092	0.0552	0.2019	0.4167	Power Forward	0.1458	0.0845	-0.0203	0.3104
Small Forward	0.4054	0.0503	0.3060	0.5020	Small Forward	0.2570	0.0766	0.1073	0.4043
Bench	0.1321	0.0176	0.0970	0.1662	Bench	0.0895	0.0280	0.0347	0.1442
Turn Overs(μ_{β_3})					3-point field goals made(μ_{β_4})				
Point Guard	-0.2687	0.0586	-0.3815	-0.1531	Point Guard	0.2828	0.0757	0.1341	0.4327
Shooting Guard	-0.2153	0.0535	-0.3211	-0.1125	Shooting Guard	0.3740	0.0635	0.2474	0.4963
Center	-0.2538	0.0524	-0.3573	-0.1521	Center	-0.0014	0.1290	-0.2602	0.2509
Power Forward	-0.2836	0.0628	-0.4047	-0.1604	Power Forward	0.1495	0.1282	-0.1080	0.4021
Small Forward	-0.3454	0.0591	-0.4615	-0.2307	Small Forward	0.3810	0.0784	0.2274	0.5386
Bench	-0.0839	0.0212	-0.1251	-0.0421	Bench	0.0671	0.0287	0.0116	0.1245
Free Throws Made(μ_{β_5})					Free Throw Percentage(μ_{β_6})				
Point Guard	0.1443	0.0541	0.0386	0.2500	Point Guard	-0.0014	0.0056	-0.0123	0.0098
Shooting Guard	0.0989	0.0409	0.0204	0.1821	Shooting Guard	-0.0043	0.0051	-0.0139	0.0062
Center	0.0668	0.0544	-0.0405	0.1725	Center	0.0110	0.0060	-0.0009	0.0227
Power Forward	0.0857	0.0510	-0.0135	0.1862	Power Forward	-0.0023	0.0058	-0.0136	0.0090
Small Forward	0.1039	0.0525	-0.0018	0.2044	Small Forward	-0.0005	0.0056	-0.0112	0.0108
Bench	0.0528	0.0176	0.0176	0.0865	Bench	-0.0070	0.0075	-0.0219	0.0075
Field Goals Made(μ_{β_7})					Field Goal Percentage(μ_{β_8})				
Point Guard	-0.0071	0.0544	-0.1143	0.0993	Point Guard	0.1275	0.0126	0.1035	0.1524
Shooting Guard	-0.0691	0.0479	-0.1635	0.0236	Shooting Guard	0.1680	0.0129	0.1425	0.1929
Center	0.1345	0.0416	0.0551	0.2172	Center	0.0527	0.0088	0.0351	0.0696
Power Forward	-0.0081	0.0453	-0.0936	0.0846	Power Forward	0.0776	0.0108	0.0570	0.0990
Small Forward	-0.0033	0.0520	-0.1036	0.0996	Small Forward	0.1152	0.0122	0.0913	0.1391
Bench	-0.0051	0.0226	-0.0484	0.0399	Bench	0.1997	0.0208	0.1599	0.2411
Offensive Rebounds%(μ_{β_9})					Defensive Rebounds%($\mu_{\beta_{10}}$)				
Point Guard	0.1129	0.0482	0.0162	0.2062	Point Guard	0.0611	0.0280	0.0045	0.1140
Shooting Guard	0.0522	0.0322	-0.0118	0.1142	Shooting Guard	0.0670	0.0205	0.0260	0.1062
Center	0.0519	0.0229	0.0073	0.0966	Center	-0.0098	0.0159	-0.0407	0.0212
Power Forward	0.1116	0.0247	0.0638	0.1608	Power Forward	-0.0051	0.0183	-0.0402	0.0318
Small Forward	0.0703	0.0283	0.0140	0.1246	Small Forward	0.0283	0.0205	-0.0113	0.0689
Bench	0.0758	0.0102	0.0560	0.0959	Bench	-0.0156	0.0078	-0.0309	-0.0004

athletic center is a benefit on defense.

Obviously turnovers are detrimental to the effectiveness of any offense. This is clearly captured in the model since the effects of committing a turnover for all positions is negatively large. It is interesting that turnovers for small forwards has the most negative effect and that assists for turnovers has the most positive effect. It appears that having a small forward that can pass well and protect the basketball is highly desirable.

Free throw percentage was not a “significant” effect for any of the six positions, but free

throws made was significant for both guard positions and the bench. Thus the act of getting fouled can be beneficial.

Having a better field goal percentage than the opposition is significant for all positions, while making more field goals than the opposition is only significant for the center position. Shooting a better field goal percentage than the opposition for players that play a position that requires them to be farther from the basket has the largest effect. Both guards and small forward field goal percentages have more impact on the outcome of a game than the other positions. Thus, having a player that shoots well at these three positions is very beneficial to a team. Also, having a bench that shoots well is beneficial. Thus it is good to have all positions on the floor be an offensive threat.

Offensive rebounds are important for all positions but the shooting guard. This follows conventional thought since an offensive rebound results in another opportunity to score. In fact it is somewhat strange that offensive rebounds for a shooting guard are not significant. What is somewhat unexpected is that defensive rebounds is only significant at the guard positions. So a center out defensive rebounding his opponent is not as important as a point guard out defensive rebounding his opponent relatively speaking. But, this doesn't suggest that a center can disregard rebounds only that a point guard should make an emphasis to defensive rebound.

Another interesting observation is that out of the three categories that guarantee scored points (assists, field goals made, and free throws made) only assists are significant for all positions. And the effect for each position out assisting their opponent is by far the greatest. That is, for each position out assisting the opponent is more important than making more field goals than the opponent. This is a confirmation that having a group of players that play as a single unit increases the chances of winning a game.

In general the results from assists, steals, turnovers, and field goal percentage are significant for all six positions. These categories are representative of a well rounded basketball player. Thus having a player at each position that can perform reasonably well in all aspects of the game is desirable. This is reflection of how the NBA game has evolved in that past few years. Players that are able to perform multiple tasks and or play multiple positions are

becoming more desirable.

6 Conclusions

In summary, the point spread of a basketball game increases if all five positions out offensive rebound, out-assist, have a better field goal percentage and less turnovers than their positional opponent. These conclusions are accepted by many basketball authorities. Some trends that dont receive as much attention is that both guard positions need to focus on defensive rebounding and a point guard should help with offensive rebounding as well. Also, having a small forward that is able to perform well in all ten categories is helpful.

Results from studies of this type can help basketball coaches optimize their probability of winning basketball games by organizing practices that are customized to develop the skills which have the most impact on game outcome for each position. It also may be that these results could be used by coaches to help exploit positional match-ups in specific games. Further research along these general lines but with more detailed individual information, could be used to help coaches put optimally constructed teams together.

References

- Berri, D. (1999), "Who is "Most Valuable"? Measuring the Player's Production of Wins in the National Basketball Association," *Managerial and Decision Economics*, 20, 411–427.
- Bishop, T. and Gajewski, B. J. (2004), "Drafting a Career in Sports: Determining Underclassmen College Players' Stock in the NBA Draft," *chance*, 17, 9–12.
- Bray, S. R. and Brawley, L. R. (2002), "Role Efficacy, Role Clarity, and Role Performance Effectiveness," *Small Group Research*, 33, 233–253.
- Gelfand, A. and Smith, A. F. M. (1990), "Sampling-based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Oliver, D. (2004), *Basketball on Paper: Rules and Tools for Performance Analysis*, Dulles, VA: Brassey's Inc.

Raftery, A. E. and Lewis, S. (1992), *Bayesian Statistics*, Oxford, chap. How Many Iterations in the Gibbs Sampler?, pp. 765–776.

Trninić, S. and Dizdarević, D. (2000), “System of the performance evaluation criteria weighted per positions in the basketball game,” *Collegium antropologicum.*, 24, 217–234.