Master Thesis

The Detailed Player Rating: an elaborate approach to rating NBA players

Author:
Dailinh Nguyen

Supervisors:
Prof. dr. Richard Paap
Dr. Wendung Wang

February 8, 2016

Abstract

Making correct decisions regarding player trades is important for NBA teams, in a league where every team is bound by the same salary restrictions. Therefore player rating tools can be useful for NBA teams to support these decisions. In this thesis we design a new method for rating NBA players, called the Detailed Player Rating (DPR). Our method has the added information over the current player rating methods, that it not only estimates a player's skill level, but also what aspects of basketball are most important for winning and how each player contributes to these aspects. We have shown that the DPR method shows comparable results in terms of predictive power, but provides more useful information and does not overvalue offensive skills, which most current player rating methods do.

Keywords: NBA, basketball, player ratings, two-stage regression, hierarchical Bayes, subspace prior regression.

Contents

1 Introduction

In the National Basketball Association all teams are restricted in how much salary they are allowed to allocate to players on their roster. This maximum salary amount is equal among teams and is called the *NBA salary cap.* When all teams have to operate under these same salary restrictions, it is of course important for teams to make the right decisions regarding trading players and salary to keep a competitive edge over their opponents. Statistical methods can be used by NBA teams to support these decisions, because plenty of necessary data is recorded during matches and is widely available for usage.

The need for using statistical methods in sports can be seen from an example in baseball, another popular U.S. sport. The surge of statistical methods in baseball is well documented in Michael Lewis' book *Moneyball*[1]. In the late 90's the low-ranking baseball team the Oakland Athletics first started incorporating statistical methods to supplement their decisions regarding player transactions and game strategies. They found that certain player attributes were terribly overvalued, such as the ability to hit homeruns. They found that the ability to hit a ball and reaching first base is far more important than hitting far with a lower percentage. By adjusting their game plans and transactions to these findings, they unexpectedly became a much better team, outperforming many teams with bigger budgets. However, after the book *Moneyball* was published other U.S. baseball teams became more interested in using statistical methods. With this, the competitive edge the Oakland Athletics had was gone and they were sent back into mediocrity. Nonetheless, it goes to show how a good use of statistical methods can impact a teams performance.

Baseball and basketball are very different sports, basketball is a much more dynamic sport with less breaks and stops where strategies are seemingly more important for winning. However, the use of statistical methods in basketball may help a team win more often if these methods are used in the correct way. In this thesis we will discuss some of the existing statistical methods NBA teams can use to support their decision making regarding player transactions. We will also discuss the limitations of these statistical methods and propose improvements.

1.1 Literature review

This literature review is divided into three sections. The first section reviews player rating methods using box-score data. The second section reviews player rating methods using play-by-play data. We conclude the literature review with a short summary of what we believe are some shortcomings that we could improve upon.

1.1.1 Using box-score statistics to rate players

Box-score statistics are used extensively in player rating methods, because they are widely available and easy to interpret. Box-scores are kept for each NBA match. They record for each player the minutes played, points scored and scoring percentage (for three-point shots and free throws as well), assists, rebounds, steals, blocks and fouls. These box-scores are then made widely available in newspapers or the internet, such as on www.NBA.com.

These statistics are easy to interpret, as they represent the direct output a player has on a match in terms of points scored, shooting accuracy, assists, offensive and defensive rebounds, steals and fouls. However, it is hard to interpret how much effect some of these stats have on the outcome of a match. Take for example a defensive rebound, which is the act of catching the ball after the opponent has missed a shot. Making defensive rebounds is important, as it will rob the opponent of another immediate opportunity to score. However, defensive rebounds do not always lead to points scored by the team making the defensive rebound.

So what is the effect of making a defensive rebound or any of the other statistics kept by in box-scores on the outcome of a match? We will discuss several statistics that use regression analysis to measure the effect of box-score statistics on the outcome of a match.

Player Efficiency Rating

The Player Efficiency Rating (PER) was invented by former ESPN writer and current Vice President of Basketball Operations John Hollinger [2]. The PER stat measures the per-minute performance a player has on average and can be used to compare player performances across seasons. Hollinger models player contributions to a game in terms of box-scores and adjusts them for the pace of each team. This adjustment accounts for the fact that faster paced teams have more possessions and therefore have more opportunities to score.

Wins Produced

The Wins Produced statistic measures the amount of wins a player produces per 40 minutes and was invented by sports economist David Berri [3][4]. It estimates an individual player's contribution to a win. The Wins Produced model first estimates the effect of box-score statistics on two measures of attack and defense with regression analysis. Those two measures are Offensive Efficiency (points scored per possession) and Defensive Efficiency (points conceded per possession)[5]. Then an individual player's contribution to his teams Offensive and Defensive Efficiency can be measured by looking at his box score. His contributions are then scaled back to a per-40 minute level.

Berri also accounts for the fact that good or bad teammates can deflate or inflate a player's output. For example, if a player has teammates who are all very good passers, he will be more likely to score than a player with teammates who are more selfish. Considering contributions made by teammates, he found that only defensive rebounds and assists made by teammates have a significant influence on a player's performance.

Determining box-score contributions per position

It is generally accepted that players in different positions have different roles and thus certain skills are more valued for some positions [6][7]. For example, committing turnovers is the least desirable for a point guard, the play maker, as it is his responsibility to run the ball up the court. If he commits a turnover, it will usually lead to a fast break and easy points for the opponent, because a point guard usually is the last line of defense.

Page et al. [8] used a Bayesian hierarchical model to estimate how box-score statistics affect the match outcome as measured in point differentials. Page et al. found that certain box-score statistics are more valuable for some positions than others. For example, making steals is more valuable for centers than for other positions, because their position requires them to be close to the basket. Although they did not use these results to rate players, their research can be easily extended to do just that.

1.1.2 Using play-by-play data to rate players

Play-by-play data are essentially box-score statistics kept for each possession instead of for a whole match. Using play-by-play data instead of box-score statistics to rate players, can deal with some of the disadvantages box-score based methods have. Because play-by-play data records which players are on the floor for each possession, methods using this data account for the fact that a player's contribution to a match is affected by the skill-level of his teammates and opponents. Although Berri's Wins Produced statistic[4] makes an effort to measure these effects, he estimates them using box-score statistics. This will be less accurate, as box-score statistics will have less sampling variation in terms of on-court interaction between players.

Another advantage of recording on-court presence, is that it is an extra measure to estimate a player's defensive skill. Although box-scores do record steals, blocks and rebounds, they cannot record all defensive efforts. For example, the ability to guard an opponent properly and forcing him to shoot less favorable shots is something that can be explained by on-court presence but is not recorded in the box-score statistics.

Two player rating statistics will be discussed that use play-by-play data to measure offensive and defensive skill while accounting for the skill of teammates and opponents.

Adjusted Plus-Minus

Dan Rosenbaums Adjusted Plus-Minus statistic[9] is based on the plus-minus statistic. The plus-minus statistic measures a player's impact on a game and is calculated by taking the average point differential for when the player is on the court.

Rosenbaum notes that although it is interesting to see how point differentials change when a particular player is on the court or not, the plus-minus statistic may not be accurate in some situations. He gives as example that very weak starters on strong teams will spend more time on the court with that strong unit, thus inflating his Plus-Minus rating. Therefore he uses regression analysis to account for the strength of an individuals teammates and opponents on the court.

Rosenbaum found that simply regressing the point differential on dummy variables signifying the players on the court, resulted in high standard errors for his regression coefficients. He extends this so-called "Pure Adjusted Plus-Minus" rating (PAPM), with the assumption that a player's skill level will be reflected in his box-scores statistics. Therefore he estimates the effect of box-score statistics on a player's PAPM rating, by regressing the PAPM rating on several box-score variables. Rosenbaum uses the results of this model to form the STAT rating.

He still found some strange outliers (over performers) with the STAT rating, so he formed a final rating called the Overall Adjusted Plus-Minus rating (OAPMP). This rating is a weighted average between the APMP and STAT rating, where the weights were chosen such that the standard errors for the OAPMP ratings were minimized.

Subspace Prior Regression

Dapo Omidiran's Subspace Prior Regression (SPR) statistic[10] can largely be seen as an extension from Dan Rosenbaum's OAPMP statistic[9]. Omridan's criticism on other Plus-Minus statistics, is that they don't account enough for the skill disparity between players. Omidiran notes that the NBA is a competition driven largely by star players and that better players contribute far more to success than lesser players. He therefore penalizes large player ratings to create more model sparsity and make only the best players stand out.

Omidiran found that creating model sparsity increases the predictive performance of Adjusted Plus-Minus ratings and is therefore a good model extension. Furthermore, he adds another penalty term in the regression model, that penalizes the distance between a player's rating and his box-score output. This penalty term is included, because a player's skill level should be reflected in his box-score stats.

1.1.3 Summary

When looking for player rating methods, using play-by-play data is obviously better than using box-score data, as it captures a player's defensive skill better. However, there are still some shortcomings to the current player rating methods that use play-by-play data.

The current methods are not able to make a link between a player's skill level and what type of player he is. If a team wants to acquire a new player, not only is it important to know how good that player is, but also what he specializes in. It is important to know if a player is a good passer, rebounder, shooter, on- or off-ball defender, etc. Of course, looking at box-scores can also tell you these things, but they still do not say what a player's effective contribution to a team is. For example, a player that scores many points would probably look like a good addition to most teams in the NBA. However, if his play style means that the offensive efficiency of his teammates suffers too much, it might not be a good decision to sign that player.

Another point that the current statistical methods lack, is that they cannot properly measure what kind of strategies are effective to win games. This is important for teams to know, so they can decide what types of players they should acquire. Page et al. [8] make a start with their model that estimates how important output is of several box-score statistics for each player position. However, they use box-score data to estimate these effects. When using this data, which are basically score summaries per match, their estimates do not correctly use the information that certain players spend more time playing with and against certain players with different skill levels. In other words, it does not correctly account for the skill level of opponents and teammates.

1.2 Thesis contribution

The main goal of this thesis is to find an approach to correctly estimate what kind of strategies are important for winning games, how skillful a player is and what type of player he is. Compared to the existing methods, which are only able to estimate a player's skill level, we believe this is a great improvement. This will be done in a 2-stage regression model. The first stage models the influence of the on-court presence of players on several production statistics. This stage will be estimated with a Lasso regression, to penalize the occurrence of many large contributors, as we believe there are not many star-players. This approach is influenced by Omidiran's approach for his SPR method[10]. We replace the dependent variable of Omidiran's model, score differentials, with the several production statistics.

In the second stages we model score differentials with the estimated production statistics from the first stage as explainable variables. Because we believe these effects not to be homogenous across different player positions, yet to be somewhat similar, we estimate this with a Hierarchical Bayes model (similar to the approach of Page et al. [8].) With the results from the second stage we can see what production statistics affect score differentials the most and are the most important for winning. And because we estimate the influence of the players on score differentials indirectly through the first stage, we can see which players have the largest effects on score differentials, so which players are the best and the worst. We call this player rating the Detailed

Player Rating (DPR).

We will compare the forecast accuracy of our DPR with the best player rating of the current literature, namely Omidiran's[10] Subspace Prior Regression (SPR). Lastly we will analyze NBA player salaries, to see if we can find a good way to correctly determine a player's salary according to his skill level.

What this thesis contributes to existing methods, is that current state-of-the art methods estimate player's skill levels by simply regressing score differentials on a player's on-court presence. So the current methods are only able to estimate player's skill levels. Our approach is very different, because we use a two-stage model that estimates the relation between score differentials and on-court presence of players indirectly through certain production statistics. This is something that has not been done before and provides a lot of extra useful information about what strategies are important for winning and what a player's strengths and weaknesses are. Furthermore, this thesis marks the first time where the effect of box-score statistics is estimated on score differentials using play-by-play data. Lastly, in this thesis we propose a method that tries to determine a player's salary by their statistical skill level, which has also not been done before.

2 Data

In this section we will give a short description and analysis of the data we have used to estimate player ratings. We have used play-by-play data from the 2009-2010 NBA season in our analysis. This data can be found on http://www.basketballgeek.com/ and http://www.basketballvalue.com/downloads.php. Only data from the regular season is used. In the NBA the regular season is followed by play-offs, a knock-out tournament where teams play a best-of-seven series each round. Because of this knock-out system, we believe that the data from the play-offs are much more affected by randomness and we do not use this data. Continuing with what is included in the dataset, every field goal attempt, free throw, rebound, steal, block, foul and substitution are recorded for 1211 matches between 28 teams. Each match is divided in sequences of events, such that a new event is formed every time one or more teams makes a substitution. Note that a match can contain several events featuring identical line-ups, if these events are separated by events with different line-ups. The time left in the game was also recorded for each event. Furthermore, for each event there is information which players were on the court and at which position they were playing. These positions are:

1. **Point guard**: The point guard is the playmaker of the team. He usually starts each attack and directs each play.
2. **Shooting guard**: The shooting guard position usually requires good shooting abilities and/or ball-handling abilities to set up his own shot.
3. **Small forward**: Together with the shooting guard, the small forward usually plays on the wing (for away from the basket). A good small forward is both quick and strong, so they can play both far away as well as close to the basket.
4. **Power forward**: Power forwards play close to the basket, similar to the center position. However, good power forwards should also be able to shoot from further away from the basket.
5. **Center**: The center is usually the tallest player on each team. His main skills are rebounding, scoring from under the basket and protecting the own basket from close range attempts. The center plays the closest to the basket during offensive and defensive possessions.

Additionally, for each field goal attempt there is information about the distance of the shot (distance to the basket) and whether it was a successful attempt. If the shot was made, there is also information about if it the shot was assisted and who made the assist.

Lastly, we look at some descriptive statistics regarding the score differentials. The score differentials for each event are standardized per 100 possessions. This will account for the fact that not all events have the same time length. Another option would have been to standardize the score differentials by the same unit of time. However, this approach would not account for the fact that some teams with faster paced tactics will have more possessions and thus are able to have larger score differentials.

The descriptive statistics for the score differentials can be found in table 1. A mean of 5.23 signifies that the home team will score on average 5.23 more points than the away team per 100 possessions. The fact that this mean has a positive sign can be easily explained by the fact that home teams will have a moral advantage by playing in front of their home crowd. A skewness 0.06 means that the score differentials are only slightly positively skewed, which means that the home team will only slightly more frequently win by a large margin than the away team. A kurtosis of 5.58 means that the distribution is highly peaked and has fat tails. This can also be seen by looking at the histogram of the score differentials in figure 1. We see that the distribution has a high peak around 0, but that there are also many extreme values.

Mean:	5.23
Variance:	24759.00
Skewness:	0.06
Kurtosis:	5.58

Table 1: Descriptive statistics score differentials per 100 possessions

3 Methods

In this section we formulate the models and estimation procedures of various player rating models. In section 3.1 we formulate Omidiran's[10] SPR method which will be used as a benchmark model. In section 7 we improve upon this method and formulate our Detailed Player Rating (DPR). In section 3.3 we will explain how these methods are compared in terms of forecasting power.

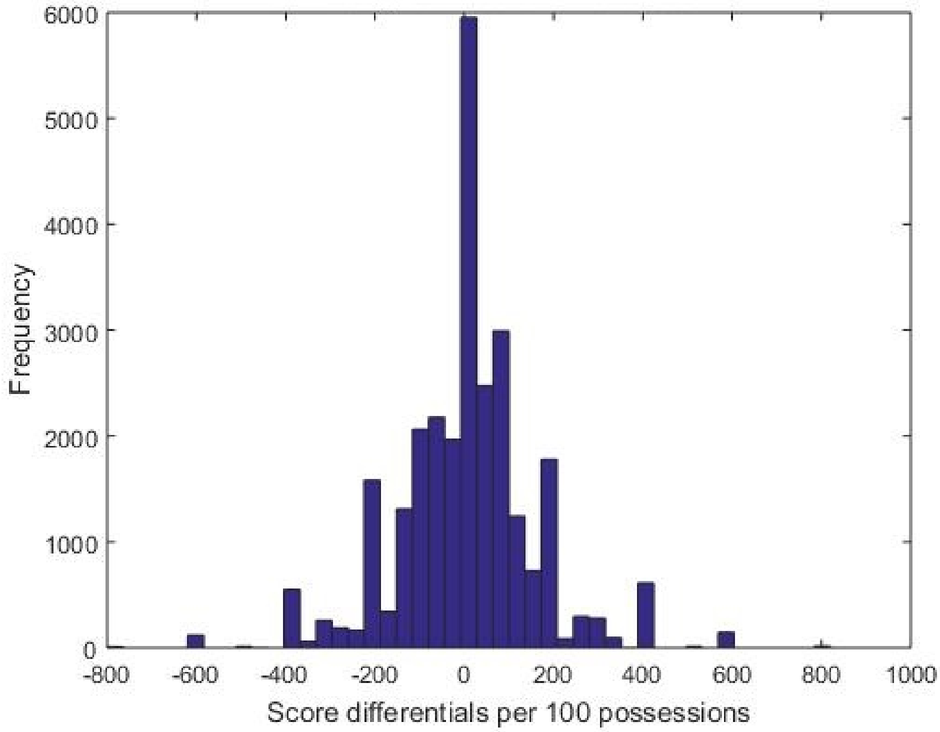


Figure 1: Histogram score differentials

3.1 Subspace Prior Regression

In this section we formulate the player rating method that has the best forecast accuracy in the current literature, namely Omidiran's[10] Subspace Prior Regression (SPR). We will use this method as a benchmark model to compare our DPR method with. The SPR method models score differentials by using the on-court presence of players as explainable variable. This captures the effect that player's have on the performance of their team. Consider a dataset of N events. Each event contains match information from matches in the 2009-2010 NBA season and is defined such that when a team makes a substitution, a new event is formed. Let y_i denote the score differential between the home and away team in event i. Let $\boldsymbol{y}$ be the $N \times 1$ vector containing the elements $\boldsymbol{y}$ for all $i \in \{1, ..., N\}$. Let M be the amount of players that are considered in our model. Let Z be an $N \times M$ matrix containing the elements $z_{i,m}$, such that

$$z_{i,m} = \begin{cases} 1 & \text{Player } m \text{ is on the floor for the home team during event i} \\ -1 & \text{Player } m \text{ is on the floor for the away team during event i} \\ 0 & \text{Player } m \text{ is not on the floor} \end{cases}$$

for $m \in \{1, ..., M\}$ and $i \in \{1, ..., N\}$. The relationship between $\boldsymbol{y}$ and Z is then modeled as

$$\boldsymbol{y} = \alpha + Z\boldsymbol{\phi} + \boldsymbol{\varepsilon} \tag{1}$$

where $\varepsilon_i \sim$N(0,σ) . The relationship between $\boldsymbol{y}$ and Z are captured in $M \times 1$ parametervector $\boldsymbol{\phi}$. Values for elements in $\boldsymbol{\phi}$ should be interpreted as the amount of points a team will score more (or less) than the opponent per 100 possessions, when the corresponding player is on the floor.

This model could simply be estimated with Least Squares, but Omidiran estimates model (1) with a penalized regression to account for two things: model sparsity and box-score information. Model sparsity basically means there is a limited amount of star players in the NBA and that the estimation procedure should account for that.

The model sparsity is accounted for by adding the penalty term $\lambda_1||\boldsymbol{\phi}||$. This ℓ_1-norm of the player ratings $\boldsymbol{\phi}$ will ensure that there will not be a large number of very good players, which seems more realistic.

The estimation procedure should also incorporate the information that can be found in box-scores, as they argue that good players will have a good numbers in terms of points scored, assists, rebounds and other box-score stats. This information is incorporated by adding the penalty term $\lambda_2||\boldsymbol{\phi} - \gamma_0 - \boldsymbol{R\gamma}||_2^2$. Here $\boldsymbol{R}$ is the $M \times S$ box-score statistics matrix, where S is the amount of box-score statistics. In parameter vector $\boldsymbol{\gamma}$, with dimensions $S \times 1$, are the weights for each box-score statistic.

The penalized regression procedure also incorporates a weighting scheme, where each weight w_i is the amount of minutes played during event i. The weights w_i are stacked in $N \times 1$ vector W.

Equation (1) is then estimated by minimizing the loss function

$$g(\alpha, \boldsymbol{\phi}, \gamma_0, \boldsymbol{\gamma}; \lambda_1, \lambda_2) = \underbrace{||\boldsymbol{y} - \alpha - Z\boldsymbol{\phi}||^2_{\boldsymbol{W}}}_{\text{Weighted Least Squares}} + \underbrace{\lambda_1||\boldsymbol{\phi}||}_{\text{Model Sparsity}} + \underbrace{\lambda_2||\boldsymbol{\phi} - \gamma_0 - \boldsymbol{R}\boldsymbol{\gamma}||^2_2}_{\text{Box-score information}} \tag{2}$$

This objective function is then minimized with the Cyclical Coordinate Descent algorithm which was found to be the fastest among other algorithms by Omidiran [10]. We refer to his paper for the formulation of this algorithm. The regularization parameters λ_1 and λ_2 are chosen through 10-fold cross-validation by Omidiran [10] . The parameters are chosen from the set

$$\Lambda := \{(2^a, 2^b)|a, b \in F\} \tag{3}$$

where

$$F := \{-10, -9, ..., 9\} \tag{4}$$

We simply choose the parameter values found by Omidiran [10], because the 10-fold cross-validation would be too computationally expensive. These parameter were found by analyzing data from the 2010-2009 NBA season, one season later than our data, but we find no reason to believe that the optimal parameters would be very different for our data. The optimal parameters found are $\lambda_1 = 2^{-10}$ and $\lambda_2 = 2^{-3}$.

3.2 Detailed Player Rating

In this section we will formulate our Detailed Player Rating (DPR) which improves upon Omidiran's[10] SPR method. The DPR method will allow us to estimate what strategies are important for winning games, estimate the skill level of players and say what their specialties are. This method is estimated in two stages. In the first stage we estimate the effect a player's on-court presence has on his team's offensive and defensive output of certain production statistics. In the second stage we estimate what strategies are most effective to winning. We do this by regressing score differentials on the estimated production statistics from the first stage. We will first formulate this two-stage model and explain the estimation results of have to be interpreted. In sections 3.2.1 and 3.2.2 we will explain the estimation procedures of the first and second stage of the DPR model respectively. Lastly we will describe the weighting scheme we used for our data in section 3.2.3.

First we formulate the first stage of the DPR model. Let Z be defined as in section 3.1. Let $x_{p,i,j}$ be the difference between the j-th production statistics of the home and away team's players, who play on the p-th position made during the i-th event for $p \in \{1, ..., P\}$, $j \in \{1, ..., K\}$ and $i \in \{1, ..., N\}$. Let $\boldsymbol{x}_{p,j}$ be the $N \times 1$ vector containing the elements $x_{p,i,j}$ with i ranging from 1 to N for a given j and p. The P positions are PG, SG, SF, PF and C and are described in section 2. The K production statistics are: field goals attempted and percentage made, three-point shots attempted and percentage made, free throws attempted and percentage made, offensive and defensive rebounds, steals, blocks, fouls and turnovers. We do not consider points made from field goals, three-point shots and free throws in our model, as it would lead to multicollinearity. With regards to the field goal related statistics, we distinguish between different ranges of shots, because players will probably shoot less efficient when shooting further away from the basket. Close range shots usually are lay-ups and dunks, while shots from further are usually jump shots. We make the distinction between 3 kinds of shots:

- Close range (between 0 ft./0 m and 8 ft./2.44 m)
- Mid range (from 8 ft./2.55 m up until 16 ft./4.88 m)
- Long range (from 16 ft/4.88 m up until the three-point line)

We also incorporate for each shot type the percentage of made shots which came from an assist. If these percentages are high, it means that these points were made because of good team play and less so by individual play. In conclusion, there are 20 production statistics that we use in our model to estimate a team's weaknesses and strengths.

Continuing with the formulation of our DPR model, in equation 5 we estimate the influence of the player's on-court presence on the difference between the output of production statistics of his team and the opponent.

$$Stage\ I:\ \boldsymbol{x_{p,j}} = \kappa + Z\boldsymbol{\theta_{p,j}} + \boldsymbol{\eta} \tag{5}$$

Note that large positive values in the estimated parameter vector $\boldsymbol{\hat{\theta}_{p,j}}$ indicate players causing a large output for the j-th production statistic of players on his team playing the p-th position (this could include himself) or causing their opponent to have a diminished output for these production statistics. We assume the error terms η_i to be independently and identically distributed with the Normal distribution.

We estimate equation (5) in the first stage for all positions $p \in \{1, ..., P\}$ and production statistics $j \in \{1, ..., K\}$. Now let $\hat{\Theta}$ be the $M \times (P \times K)$ estimated parameter matrix containing the estimated parameter vectors $\hat{\theta}_{p,j}$ for all $p \in \{1, ..., P\}$ and $j \in \{1, ..., K\}$. Furthermore let $\hat{X}$ be the $N \times (P \times K)$ matrix containing the vectors $\boldsymbol{\hat{x}_{p,j}}$ estimated in the first stage (equation (5)) for all $p \in \{1, ..., P\}$ and $j \in \{1, ..., K\}$. These results are later needed in the second stage and when we eventually form the Detailed Player Rating.

We now estimate the influence of all production statistics on score differentials in equation (6). In this second stage of our model we basically estimate which output of production statistics have the largest influence on score differentials. Or in other words, we estimate which strategies are the most effective to winning (or losing). These effects are estimated in $(P \times K) \times 1$ parameter vector $\boldsymbol{\beta}$. Large positive values in the estimated parameter vector $\boldsymbol{\hat{\beta}}$ indicate that the corresponding production statistics have a larger effect on winning games per unit. Note that the influence of the on-court presence of players on winning is measured indirectly through $\hat{X}$, which is estimated in the first stage (equation (5).) We will discuss the distribution for the error terms ε_i in section 3.2.2.

$$Stage\ II: \boldsymbol{y} = \alpha + \hat{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{6}$$

After the two stages from equations (5) and (6) have been estimated, we can now form our Detailed Player Rating (DPR). Let $\hat{\boldsymbol{\vartheta}}_m$ be the m-th row vector of matrix $\hat{\Theta}$ with dimensions $1 \times (P \times K)$. We define the DPR for player m as

$$DPR_m = \hat{\boldsymbol{\theta}}_m \hat{\boldsymbol{\beta}} \tag{7}$$

Values for DPR should be interpreted as the amount of points a team will score more than the opponent on average when player m is on the floor. This rating can be broken down to see what the strengths and weaknesses are for players. We do this in a similar fashion as how the DPR is defined, with the difference that $\hat{\boldsymbol{\theta}}_m$ and $\hat{\boldsymbol{\beta}}$ are multiplicated entrywise instead of multiplicated as vectors. This is also known as the Hadamard product. We define this breakdown as

$$\boldsymbol{DPRbreakdown_m} = \hat{\boldsymbol{\theta}}_m \circ \hat{\boldsymbol{\beta}}^{\boldsymbol{T}} \tag{8}$$

Values for elements in the $1 \times (P \times K)$ vector $\boldsymbol{DPRbreakdown_m}$ should be interpreted the same as values of DPR. This breakdown has the added information of how the players contribute to their team in terms of certain production statistics of players playing in a certain position. Note that summating al the elements of $\boldsymbol{DPRbreakdown_m}$ will result in the DPR of equation (7).

3.2.1 Estimating Stage I

In this section we will explain the estimation procedure we used to estimate the first stage of our DPR model. Equation (5) is estimated in a similar fashion as the SPR algorithm in section 3.1. We choose this algorithm, because it has proven to be effective in modeling matchdata with on-court presence of players as explainable variable. This is also done in equation (5), but with different types of match data as dependent variables, namely production statistics instead of score differentials. We remove the penalty term that incorporates box-score information, because we believe that a player's skill-level of a production statistic $\boldsymbol{x_{p,j}}$ is not reflected in the rest of his box-score stats.

$$g(\kappa, \boldsymbol{\theta_{p,j}}; \lambda) = ||\boldsymbol{x_{p,j}} - \kappa - Z\boldsymbol{\theta_{p,j}}||^2_{\boldsymbol{W}} + \lambda||\boldsymbol{\theta_{p,j}}|| \tag{9}$$

for all $p \in \{1, ..., P\}$ and $j \in \{1, ..., K\}$.

This objective function is minimized with the Cyclical Coordinate Descent algorithm. We use the same value for regularization parameter λ found by Omidiran [10], because using 10-fold cross-validation would be infeasible due to computational limitations, so we use the value $\lambda = 2^{-10}$.

3.2.2 Estimating Stage II

In this section we will explain the estimation procedure we used to estimate the second stage of our DPR model. The parameters estimated in equation (6) are interpreted as the influence that several production statistics have on score differentials.

We have included a variable in the model for each production statistics for each of the 5 possible different positions a basketball player could play. We believe these parameters are not equal over all positions, because different production statistics will not have the same effect on score differentials for the same position. However, there will be some similarity between these statistics, because they estimate the effect of the same production statistics. For these reasons we believe that a Hierarchical Bayes approach is appropriate to estimate equation (6).

To explain why Hierarchical Bayes estimation is so appropriate, we will formulate the priors and posteriors. First we consider a specification where the likelihood of the data follows a Normal distribution. This allows us to use conjugate priors on our parameters, which in turn will lead to a relatively easy estimation procedure for this complicated model, in the form of Gibbs sampling.

Later we provide a model specification where the likelihood of the data follows the Laplace distribution. The reason for this is that we have found in section 2 that the score differential data contain many extreme values.

These extreme values may influence the estimated parameters too much, so a more robust approach may be needed. When we assume the residuals to be Laplacian distributed, we basically minimize the absolute residuals, compared to minimizing the squared residuals when a Normal distribution is assumed. This method is more robust for large outliers and shares the same philosophy behind the non-Bayesian Least Absolute Deviations (LAD) regression [11].

Lastly we provide a model specification for a simplified model, which may be more parsimonious.

Normal likelihood

First we describe our model specification where the likelihood of the data follows a Normal distribution. We followed the approach from Paap [12], where the full derivation can be found for a hierarchical model where the likelihood of the data follows a Normal distribution. We will consider this procedure, as it is an Hierarchical Bayes method that fits our model specification and is relatively easy to implement.

Before we show the prior specification, note that equation (6) can also be written as

$$y_i = \alpha + \sum_{p=1}^{P} \sum_{j=1}^{K} \hat{x}_{p,j,i} \beta_{p,j} + \varepsilon_i \tag{10}$$

where $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$ with $i \in \{1, ..., N\}$, $p \in \{1, ..., P\}$ and $j \in \{1, ..., K\}$. We impose a hierarchical prior on $\beta_{p,j}$ such that the parameters are grouped per production statistic, so

$$\beta_{p,j} = \mu_j + \zeta_{j_p} \tag{11}$$

where $\zeta_{j_p} \sim N(0, \sigma_{\zeta_j}^2)$. Note that, given a production statistic j, the parameters $\beta_{p,j}$ for all positions p share the same prior mean hyperparameter μ_j. By using this specification, we acknowledge the fact that these parameters are different, yet some come from the same group. The prior specification is given by

$$\begin{aligned} &p(\beta_{p,j}) \propto N(\mu_j, \sigma^2_{\zeta_j}) \\ &p(\boldsymbol{\mu}, \alpha) \propto 1 \\ &p(\sigma^2_\varepsilon) \propto \sigma^{-2}_\varepsilon \\ &p(\sigma^2_{\zeta_j}) \propto \sigma^{-2}_{\zeta_j} \end{aligned} \tag{12}$$

We use Gibbs sampling to compute the posterior distributions for all parameters. The Gibbs sampling scheme is given by

$$\begin{aligned} &\bullet \text{ Sample } \alpha \text{ given } \{\{\beta_{p,j}\}_{p=1}^P\}_{j=1}^K, \{\mu_j\}_{j=1}^K, \sigma^2_\varepsilon, \{\sigma^2_{\zeta_j}\}_{p=1}^P, \boldsymbol{y} \\ &\bullet \text{ Sample } \sigma^2_\varepsilon \text{ given } \{\{\beta_{p,j}\}_{p=1}^P\}_{j=1}^K, \alpha, \{\mu_j\}_{j=1}^K, \{\sigma^2_{\zeta_j}\}_{p=1}^P, \boldsymbol{y} \\ &\bullet \text{ Sample } \{\mu_j\}_{j=1}^K \text{ given } \{\{\beta_{p,j}\}_{p=1}^P\}_{j=1}^K, \alpha, \sigma^2_\varepsilon, \{\sigma^2_{\zeta_j}\}_{p=1}^P, \boldsymbol{y} \\ &\bullet \text{ Sample } \{\sigma^2_{\zeta_j}\}_{p=1}^P \text{ given } \{\{\beta_{p,j}\}_{p=1}^P\}_{j=1}^K, \alpha, \{\mu_j\}_{j=1}^K, \sigma^2_\varepsilon, \boldsymbol{y} \\ &\bullet \text{ Sample } \{\{\beta_{p,j}\}_{p=1}^P\}_{j=1}^K \text{ given } \alpha, \{\mu_j\}_{j=1}^K, \sigma^2_\varepsilon, \{\sigma^2_{\zeta_j}\}_{p=1}^P, \boldsymbol{y} \end{aligned} \tag{13}$$

Before we specify these posterior distributions, consider the parameter vector $\theta = (\alpha, \{\{\beta_{p,j}\}_{p=1}^P\}_{j=1}^K, \{\mu_j\}_{j=1}^K, \sigma^2_\varepsilon, \{\sigma^2_{\zeta_j}\}_{p=1}^P)$. Then θ^{-q} stands for all parameters within θ except for q. The posterior distributions are given by

$$\begin{aligned} &p(\alpha|\theta^{-\alpha}, \boldsymbol{y}) \propto N(\sum_{i=1}^N (y_i - \sum_{p=1}^P \sum_{j=1}^K \hat{x}_{p,j,i}\beta_{p,j})/N, \sigma^2_\varepsilon/N) \\ &p(\sigma^2_\varepsilon|\theta^{-\sigma^2_\varepsilon}, \boldsymbol{y}) \propto \text{Inv-Gamma-2}(\sum_{i=1}^N (y_i - \alpha - \sum_{p=1}^P \sum_{j=1}^K \hat{x}_{p,j,i}\beta_{p,j})^2, N) \\ &p(\mu_j|\theta^{-\mu_j}, \boldsymbol{y}) \propto N(\sum_{p=1}^P (\beta_{p,j})/P, \sigma^2_{\zeta_j}/P) \\ &p(\sigma^2_{\zeta_j}|\theta^{-\sigma^2_{\zeta_j}}, \boldsymbol{y}) \propto \text{Inv-Gamma-2}(\sum_{p=1}^P (\beta_{p,j} - \mu_j)^2, P) \\ &p(\beta_{p,j}|\theta^{-\beta_{p,j}}, \boldsymbol{y}) \propto N(\tilde{\beta}, \sigma^2_\varepsilon(\hat{X}_j'\hat{X}_j + \sigma^2_\varepsilon/\sigma^2_{\zeta_j})^{-1}, N) \end{aligned} \tag{14}$$

where $\tilde{\beta} = (\hat{X}_j'\hat{X}_j + \sigma^2_\varepsilon/\sigma^2_{\zeta_j})^{-1}(\hat{X}_j'\boldsymbol{y}^* + \mu_j\sigma^2_\varepsilon/\sigma^2_{\zeta_j})$. Here $\hat{X}_j$ is defined by the $N \times P$ matrix consisting of the columns of $\hat{X}$ which correspond with the

j-th production statistic. Define $\hat{X}_{-j}$ as all columns of $\hat{X}$ except for $\hat{X}_j$. To define $\boldsymbol{y^*}$, we first define $\boldsymbol{\beta_j}$ as the $1 \times P$ parameter vector consisting of parameters of $\boldsymbol{\beta}$ which correspond with the j-th production statistic. Then $\boldsymbol{\beta_{-j}}$ is defined as the parametervector $\boldsymbol{\beta}$ without the elements of $\boldsymbol{\beta_j}$. Then $\boldsymbol{y^*} = \boldsymbol{y} - \hat{X}_{-j}\boldsymbol{\beta_{-j}}$.

Laplace likelihood

In this section we describe our model specification where the likelihood of the data follows a Laplace distribution. We consider this method because we hope it is robust, given the extreme values found in the data in section 2.

Using a Laplace likelihood gives us some complications, as we were unable to find good conjugate priors. This makes it impossible to use a Gibbs sampling scheme such as in equations (13) and (14). To explain how we worked around this issue, we will explain the model specification step by step. We start with the same hierarchical model specification as in equations (10) and (11), with the difference that $\varepsilon_i \sim Laplace(0, b)$. We use roughly the same prior specification as in equation (12). So

$$\begin{aligned}
&p(\beta_{p,j}) \propto N(\mu_j, \sigma^2_{\zeta_j}) \\
&p(\boldsymbol{\mu}, \alpha, b) \propto 1 \\
&p(\sigma^2_{\zeta_j}) \propto \sigma^{-2}_{\zeta_j}
\end{aligned} \tag{15}$$

Let $\theta = (\alpha, \{\{\beta_{p,j}\}_{p=1}^P\}_{j=1}^K, \{\mu_j\}_{j=1}^K, b, \{\sigma^2_{\zeta_j}\}_{p=1}^P)$, then the likelihood of the data is then defined as

$$p(\boldsymbol{y}|\theta) = \prod_{i=1}^{N} \Bigg(\sum_{p=1}^{P} \sum_{j=1}^{K} (2b)^{-1} \exp(-(|y_i - \alpha - \hat{x}_{p,j,i}\beta_{p,j}|)/b) \Bigg) \tag{16}$$

Let θ^{-k} be the collection of parameters in θ except for a given parameter κ. Then the marginal posterior distribution for parameter κ is given by

$$p(\kappa|\theta^{-\kappa}, \boldsymbol{y}) = \int_{\theta^{-\kappa}} \big(p(\boldsymbol{y}|\theta)p(\theta)\big) d\theta^{-\kappa} \tag{17}$$

Combining the definition of posterior distributions (17) with the definition of our likelihood function (16), we can see that using a Laplace likelihood does not affect the posterior distributions for the parameters in the second level μ_j and $\sigma^2_{\zeta_j}$ as they are given in equation (14). The other parameters however do not have a closed-form posterior distribution. These posterior distributions will therefore be computed through random walk sampling (Paap [12]). So, we will use a Metropolis-within-Gibbs sampler to draw from all posterior distributions. In appendix A we will formulate this Metropolis-within-Gibbs sampler with greater detail.

Sampling and variable selection

In this section we will provide some further information regarding the sampling procedure for the Hierarchical Bayes methods used to estimate the second stage of our DPR model in section 3.2.2. We will also describe how we have selected the variables that have to be included in the second stage of our model.

With regards to the sampling procedure, the parameters $\beta_{p,j}$'s from equation (6) are estimated by sampling a large amount of times from their respective posterior distributions. We have done this 300000 times with a thinning factor of 100 and a burn-in period of 100000.

Afterwards the 90%-confidence interval for all parameters is checked. If a confidence interval contains a zero, there cannot be said with some certainty that that parameter is not equal to zero and that variable should therefore be removed from the model. Afterwards, we estimate equation (6) once more without the removed variables. We chose this variable selection method, because it is easy to execute and adds relatively little computational time to an already computationally expensive method.

To make further inference about our estimated models, parameters have to be chosen from the posterior distributions. We do this by choosing the sample mean from all marginal posterior distributions. These parameters can be used to make residuals or forecasts.

Simplifying the model

In this section we will provide a simplification of the second stage of our DPR model. The model in equation (10) might suffer from a drawback, namely that it might be overcomplicated. This model needs $(P \times K) + 1$ variables to be estimated. In our model specifically are included 101 variables. Using too many variables may cause overfitting, which will lead to a poor predictive performance of our method. Therefore we also consider a model where we choose our variables differently.

First we pool all variables with respect to the position. So, we assume that the variables have the same effect for all positions. This decreases the amount of variables with a factor P. Lastly, we pool all variables involving close range, mid range and long range shots, since they are all two point shots. The resulting variables are two point shots attempted, two point shots percentage and two points shot percentage made from assists. The simplified model now becomes

$$y_i = \alpha + \sum_{j=1}^{K} \hat{x}^*_{j,i}\beta_j + \varepsilon_i \tag{18}$$

The variables $\hat{x}^*_{j,i}$ are estimated by estimating all variables $\hat{x}_{p,j,i}$ in equation (5) and then pooling them over positions p and pooling all two point shot related variables. Because all variables are pooled over positions p, we no longer use the hierarchical structure used in equations (10) and (11). Without this hierarchical structure, we lose the main reason for using a Bayesian method to estimate equation (18). To account for the extreme values in our data, we will use Least Absolute Deviations (LAD) regression to estimate equation (18). This LAD regression is robust for large model outliers. The LAD regression is done by finding the correct parameters β that minimize the objective function

$$g(\alpha, \beta) = \left| \sum_{i=1}^{N} \left(y_i - \alpha - \sum_{j=1}^{K} \hat{x}^*_{j,i}\beta_j \right) \right| \tag{19}$$

The parameters are estimated with a Maximum Likelihood algorithm proposed by Li and Arce. [11]. The LAD-estimator is the Maximum Likelihood estimator when a Laplace distribution is assumed for the error terms. Therefore we assume the error terms ε_i in equation (18) to be Laplacian distributed.

To perform variable selection on this model, we have several options. Dielman [13] proposes three options to perform variable selections on models estimated with Least Absolute Deviations. These are an LM-test, LR-test and Wald-test. Only the Wald-test does not require multiple model re-estimations for parameter testing. Considering how computationally expensive our estimation procedure is, we will choose the Wald-test to test for parameter significance.

3.2.3 Weighting Scheme

Not all observations in our data have the same importance, or are even relevant at all. We give different weights to events to account for 'garbage time' and 'crunch time'. 'Garbage time' occurs during a match when one team is clearly ahead and the other team has given up or is not working as hard. These events do not reflect the true skills of a team, so they should be weighted less. 'Crunch time' denotes the situation when a match is almost over and two teams are very close in score. Because many matches are usually decided in the final moments, a teams true skill comes forward during 'crunch time'. Therefore, these events should be weighted more heavily. The weighting scheme we use, was first used by Rosenbaum's[9].

Here follows an explanation of the weighting scheme. During the first three quarters, full weight is given to events where the score differential is less than 10 points, and no weight is given to events with a score differential of more than 20 points. In between the weights are phased from full to zero weight. At the start of the fourth quarter this approach is the same. However, the score differential boundaries gradually phase from 10 (20) points at the start of the fourth quarter to 3 (6) in the last minute of the fourth quarter. Finally, the weights are rescaled such that events in the fourth quarter on average have the same weight as events in the first three quarters combined.

To account for this weighting scheme, the data has to be transformed for the Bayesian methods in section 3.2.2. Let ω be an $(N \times 1)$ vector containing all weights ω_i during each observation i. For the model in equation (10), where a Normal distribution is assumed for the residuals, the data is transformed as

$$\begin{aligned} y_i^* &= \sqrt{\omega_i} * y_i \\ \hat{x}_{p,j,i}^* &= \sqrt{\omega_i} * \hat{x}_{p,j,i} \end{aligned} \tag{20}$$

When a Laplace distribution is assumed for the residuals, the data is transformed as

$$\begin{aligned} y_i^* &= \omega_i * y_i \\ \hat{x}_{p,j,i}^* &= \omega_i * \hat{x}_{p,j,i} \end{aligned} \tag{21}$$

The data for equation (5) does not have to be transformed, because the Cyclical Coordinate Descent Algorithm we use, can account for a weighting scheme. This can be found in Omidiran's [10] paper.

3.3 Model validation and comparison

We will compare the performance of our DPR statistic to the SPR statistic. We compare statistical performances by using the first 800 matches as training data to forecast the $\boldsymbol{y}$ of the last 411 matches. We compare the Root Mean Squared Prediction Error (RMSPE) and Mean Absolute Prediction Error(MAPE). Lastly we look at the fraction of games guessed right in terms of wins/losses. We can compare between our different Bayesian approaches of the second stage of our DPR approach (section 3.2.2) by comparing the Watanabe Akaike Information Criterion. When Gelman et al. [14] compared several predictive information criteria, they found this to be the best method to compare the predictive power of Bayesian methods, when cross-validation methods were infeasible to implement. The WAIC is equal to the log pointwise posterior predictive density, which is then corrected for the effective number of parameters to adjust for overfitting. So

$$\begin{aligned} WAIC &= -2(lppd - \text{effective number of parameters}) \\ &= -2(\log \prod_{i=1}^{n} p_{post}(y_i) - \sum_{i=1}^{n} var_{post}(\log\ p(y_i)\theta^s))) \end{aligned} \tag{22}$$

Considering S simulation runs have been done to obtain the posterior parameter distributions, the first term in equation (22) can be computed as

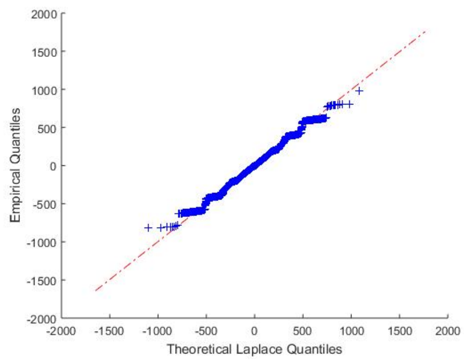
$$computed\ \log \prod_{i=1}^{n} p_{post}(y_i) = \sum_{i=1}^{N} \log \left(\frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta^s) \right) \tag{23}$$

Among Bayesian methods, the model with the lowest WAIC is preferred.

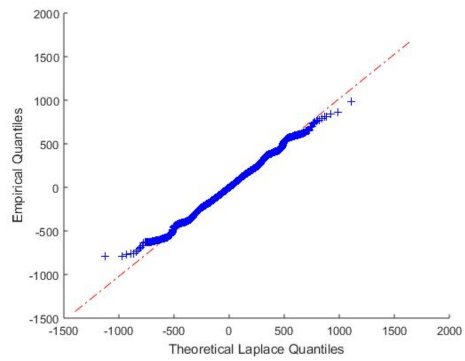
4 Comparing the DPR and SPR methods

In this section we will compare the predictive accuracy and player ranking results of our DPR method with Omidiran's[10] SPR method. Match data is used from the 2009-2010 NBA season. For the DPR method we consider the two approaches where equation (6) is estimated while the data is either assumed to be Normal or Laplacian distributed. We will refer to these methods as DPR-N (Normal) and DPR-L (Laplace). Lastly, we consider the simplified model to estimate player ratings, which is given in equations (18) and (19). We will refer to this model as DPR-S (Simplified).

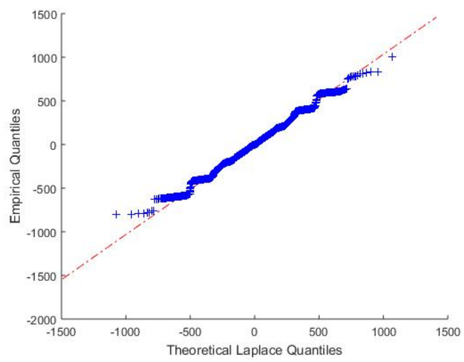
First we look at the residuals of all models to see if they follow a Laplacian distribution. In the qq-plots of figure 2 the quantiles of the residuals of all models are compared to theoretical quantiles of a Laplace distribution. For all models, except the DPR-S model, we can assume that the residuals follow a Laplace distributions, as the quantiles are plotted quite nicely on the straight line. Even though the DPR-S model is estimated with a Maximum Likelihood that assumes the residuals to be Laplacian distributed, the residuals of this model are not Laplacian distributed. This may be evidence that this model is incorrectly specified.



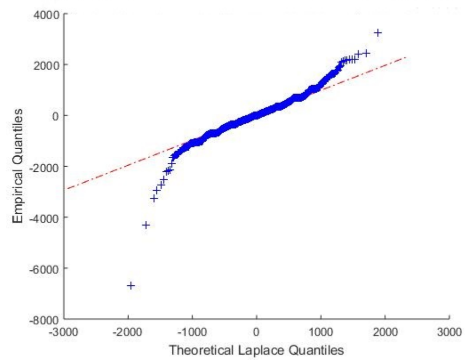
(a) QQ-plot SPR residuals



(b) QQ-plot DPR-N residuals



(c) QQ-plot DPR-L residuals



(d) QQ-plot DPR-S residuals

Figure 2: QQ-plots residuals against Theoretical Laplace Distribution

In table 2 are some descriptive statistics of the residuals of all models. What is noticeable, is that DPR-L residuals have near zero median, but not a near zero mean. The opposite holds for the SPR and DPR-N residuals. This can be explained by the fact that the SPR and DPR-N methods assume the residuals to be Normally distributed, which aims to be unbiased in the mean. While the DPR-L method assumes a Laplace residual distribution, which aims to be unbiased in the median. Because of the assumed distributions for all models, it is not surprising that the residuals for the DPR-L method has the smallest Mean Absolute Deviation. The DPR-S residuals have by far the largest MAD and variance, which gives us more evidence that this model is incorrectly specified.

	SPR	DPR-N	DPR-L	DPR-S
Mean	0.02	-0.14	2.19	6.14
Median	-2.32	-1.16	-0.03	-1.31E-8
Mean Absolute Deviaton	110.04	110.50	108.55	189.27
Variance	24518.81	23811.09	23946.51	84158.16
Skewness	0.06	0.09	0.09	-0.57
Kurtosis	5.59	5.45	5.66	25.70

Table 2: Residual Analysis

Now we look at the predictive accuracy of all methods. In table 3 are the Mean Absolute Prediction Error (MAPE), Mean Squared Prediction Error (MSPE), percentage of matches correctly predicted and Watanabe-Akaike Information Criterion (WAIC) of all methods.

First we compare the Bayesian methods DPR-N and DPR-L through the WAIC. The DPR-L has the smaller WAIC of 286727.91 against 294593.56. We conclude from this that the DPR-L method has a better predictive accuracy than the DPR-N method. In the end, the positives of the computational tractability of the DPR-N method did not outweigh the negatives of assuming the wrong distribution.

Now we compare all models. The DPR-L method outperforms the SPR method slighly in terms of MAPE and MSPE (as they are smaller). However, the SPR method has predicted a slightly higher percentage of matches. It has predicted 65.9% of the matches correctly, compared to the 64.7% of the DPR-L method. The DPR-S method has the same percentage matches correctly predicted 64.7% as the DPR-L method. However, the MAPE and MSPE are far worse for the DPR-S methods than all other methods.

We place more value on the MAPE and MSPE results than the correct percentage of matches predicted. This is because the MAPE and MSPE compare forecasts with realized observations for each event, instead of for each match, even though all methods, SPR and DPR, are estimated with data sampled per event, instead of per match.

	SPR	DPR-N	DPR-L	DPR-S
MAPE	109.84	114.86	109.23	193.55
MSPE	24602.77	25609.64	24473.15	88556.80
Corr.predicted matches %	0.659	0.620	0.647	0.647
WAIC	-	294593.56	286727.91	-

Table 3: Predictive accuracy

Finally we compare the 10 best players estimated by each method with the top 10 players according to the NBA 2009-2010 Most Valuable Player (MVP) awards. The MVP award winner is determined by the votes of several highly esteemed sports journalists and broadcasters. This data can be found on http://www.basketball-reference.com/awards/awards_2010.html. We do not consider the DPR-S method, since we have concluded from the previous results that it is incorrectly specified and we do not want to place much value in its player rankings.

In table 4 we can see that the top 10 players lists according to the MVP awards and SPR method share 7 players, although they are ranked at different positions. The MVP award top 10 list shares only 3 common players with the top 10 list according to the DPR-L method. The DPR-N list shares only 2 players with the MVP list.

From this we can conclude that the SPR ratings closely follow the judgment from journalists and sports broadcasters. This can be explained by looking at the way the SPR rating method is estimated in equation (2). The player ratings ϕ are penalized if they are too distant from $\boldsymbol{R\gamma}$. In other words the regression of the player ratings ϕ on box-scores $\boldsymbol{R}$ is also given weight. However, doing this will cause defensive skills of players to be underrated, because box-scores are unable to correctly estimate defensive skill.

It is not surprising that the top 10 players according to MVP awards is similar to the top 10 players according to the SPR method, because broadcasters and journalists will mostly base their opinion on what they see. Because offensive efforts stand out more than defensive efforts, they will also undervalue the defensive skills of players. Because our DPR methods do not use box-score data, but only play-by-play data, we are able to correctly capture both the offensive as well as defensive skill of all players. For these reasons, we prefer the DPR method over the SPR method.

Rank	MVP Awards	SPR	DPR-N	DPR-L
1	LeBron James	LeBron James	Kevin Durant	Anderson Varejao
2	Kevin Durant	Kobe Bryant	Anderson Varejao	Kevin Durant
3	Kobe Bryant	Dwyane Wade	Chris Andersen	LeBron James
4	Dwight Howard	Chris Bosh	Sean May	Kevin Love
5	Dwyane Wade	Carmelo Anthony	LeBron James	Matt Bonner
6	Carmelo Anthony	Dwight Howard	Matt Bonner	Chris Andersen
7	Dirk Nowitzki	Amare Stoudemire	David Andersen	Rudy Gay
8	Steve Nash	Tim Duncan	Caron Butler	Andrew Bogut
9	Deron Williams	Kevin Durant	Andrew Bogut	David Andersen
10	Amar'e Stoudemire	Andrew Bogut	Anthony Johnson	Tim Duncan

Table 4: Top 10 players according to the annual NBA MVP Awards and the SPR, DPR and DPR-L player rankings

5 Analyzing the DPR results

In the previous section (4) we have made several arguments for the DPR method to be favorable to the SPR method. More specifically, the DPR method where the error terms are assumed be Laplacian distributed is more favorable than the SPR method. In this section we will look at some results of our DPR method applied to the 2009-2010 NBA dataset. We will show what strategies are the most effective for teams to win and how to interpret the results for player's strengths and weaknesses.

5.1 Best strategies

From the $\beta_{p,j}$'s of equation (6) can be seen which strategies are most effective. These are given in table 5. Only the variables that are significant have been included in this table, as they are relevant.

The percentage of made shots that came from assists were only found to be significant for long range shots. This parameter is negative for shots made by players from all 5 positions. This can be explained by the fact, that this parameter is a measure for teamplay. If a team sets up an elaborate attack and they still have to settle for a long range shot, something might be wrong with that team.

The variables Steals, Fouls and Turnovers were found to be not significant for any position. Apparently, producing a unit more or less than your opponent of one of these variables will have an insignificant effect on score differentials. An explanation for this could be that the effect that fouls have on score differentials is probably already mostly captured in the attempted free throws and free throw percentage variables. Furthermore, steals and turnovers probably happen so rarely in our dataset, that their effect was unable to be estimated by our method.

We can also see that parameters for scoring percentages are smaller for shots further away from the basket. This can be explained by the fact that when teams have to settle for a long-range shot, the defending team has to do less effort defending the shot. This saves them some energy, which they can use in their counterattack.

The parameters for Three-Point shooting percentage are the largest among all shooting percentage variables, because a three-point shot is obviously worth more than a two-point shot.

There are some counterintuitive results, namely that having more Offensive Rebounds than the opponent would be detrimental to a team. This might be explained by the fact that when teams make an offensive rebound, it inherently means they missed a shot. If a team makes more offensive rebounds, this means they have missed more shots and that they have to work harder for a point.

[0]CR=Close Range shots, MR=Mid Range shots, LR=Long Range shots, 3P=3-Point shots, A=Attempted, %=Percentage, FT= Free Throws, ORB=Offensive Rebounds, DRB=Defensive Rebounds

	CR A	CR %	MR A	MR %	LR A	LR %	LR %A	3P A	3P %
PG	0.27	46.30	0.28	30.14	0.32	27.72	-3.63	0.41	50.35
SG	0.28	46.30	0.28	30.38	0.33	27.72	-3.15	0.41	49.29
SF	0.28	46.30	0.28	30.30	0.31	27.72	-2.82	0.40	50.14
PF	0.28	46.30	0.27	30.30	0.34	27.72	-2.27	0.41	48.37
C	0.27	46.30	0.27	30.01	0.31	27.72	-3.52	0.41	49.47
	FT A	FT %	ORB	DRB	Block				
PG	0.13	35.55	-0.36	0.45	0.16				
SG	0.12	37.33	-0.37	0.45	0.10				
SF	0.13	36.03	-0.37	0.45	0.15				
PF	0.12	36.34	-0.39	0.45	0.18				
C	0.12	32.50	-0.36	0.45	0.12				

Table 5: $\beta_{p,j}$'s of equation (6) for each position j and production statistic j

5.2 Best players

In table 7 are the 5 best and worst players in the 2009-2010 NBA season according to the DPR-L ranking. Anderson Varejao was found to be the best player, which is rather surprising, since he is widely considered to be a role-player on the Cleveland Cavaliers. This team is thought to be lead by superstar Lebron James, who is the third best player on our list.

Anderson Varejao has a DPR of 45.55. This number can be interpreted as the amount of points his team will outscore other team on average per 100 possessions if all the other players on the court would have a rating of 0.

Best 5 Players		Worst 5 Players	
Player	DPR	Player	DPR
Anderson Varejao	45.55	Quinton Ross	-38.58
Kevin Durant	41.16	Michael Redd	-35.73
LeBron James	40.14	Josh McRoberts	-34.11
Kevin Love	34.46	Solomon Jones	-32.76
Matt Bonner	32.27	Jeff Pendergraph	-30.93

Table 6: 5 best and worst players

We take a look at what Anderson Varejao's strengths and weaknesses are in table 7, where his DPR is broken down according to equation (8). Note that Anderson Varejao himself is a Power Forward, so all numbers for the variables corresponding with the PF position are the result of combination of his own output and his defensive skill on his direct opponent. The numbers for the variables corresponding with the other positions are the result of how he improves or diminish certain aspects of his teammates or how his on-court presence affects his non-direct opponents. Note again, that adding all elements in this breakdown will result in his original DPR.

Varejao has the largest positive elements for three-point shooting percentage for Centers and free throw percentage for Power Forwards and Shooting Guards. From this we can conclude that he either greatly influences his team, such that his teammate in the Center position will have a largely improved three-point shooting percentage, or that his opponent in the Center position will have a lowered percentage. Varejao is probably very good in shooting free throws and is able to defend his opponents without fouling them, preventing them to shoot free throws.

He has a large negative value for the variable for shooting percentage of mid-range shots. This means that he either has to stop shooting these kind of shots, because of his innefficiency, and/or because he lets his direct opponent score these points too easily.

As can be seen, a drawback of the DPR method is that we are unable tell what the mix of offensive and defensive efforts is that result in a player's contributions. It is therefore important to discuss this result with scouts and other experts to correctly identify a player's contributions.

[0] CR=Close Range shots, MR=Mid Range shots, LR=Long Range shots, 3P=3-Point shots, A=Attempted, %=Percentage, FT= Free Throws, ORB=Offensive Rebounds, DRB=Defensive Rebounds

	CR A	CR %	MR A	MR %	LR A	LR %	LR %A	3P A	3P %
PG	-0.31	-2.07	0.09	1.80	-	-1.11	-	0.87	0.29
SG	-0.71	-	0.41	5.63	0.06	2.11	-0.13	-0.49	2.45
SF	1.17	3.20	-	-	1.41	0.22	-	-1.22	-0.18
PF	-1.95	-0.67	-0.83	-4.99	-0.90	-	-0.03	-2.53	-
C	-0.17	1.77	-0.17	-	-0.56	1.03	-0.18	-	14.80

	FT A	FT %	ORB	DRB	Block
PG	0.34	-0.62	0.18	0.45	-
SG	0.75	8.49	-0.06	1.02	0.12
SF	1.11	1.40	-0.08	3.27	-0.07
PF	0.48	6.32	0.09	0.33	0.01
C	-	2.50	0.42	0.97	-

Table 7: Breakdown of Anderson Varejao's DPR

6 Salary Analysis

Now that we know which players are the best, there still remains the question how much salary they deserve. It is very important to put an accurate estimate a player's value in terms of salary because of the salary cap. This salary cap is the amount of salaries each team may use to sign players and this is equal for every team. In other words, every team has the same budget. To gain a competitive edge over opponents, teams have to make sure they do not overpay the wrong players. In this section we will try to find a method which rewards each player a fair salary. Afterwards we will analyze which players are the most over- and underpaid.

In figure 3 we have ranked the salaries from all NBA players from the 2009-2010 season. This data was taken from http://www.insidehoops.com/nbasalaries.shtml.

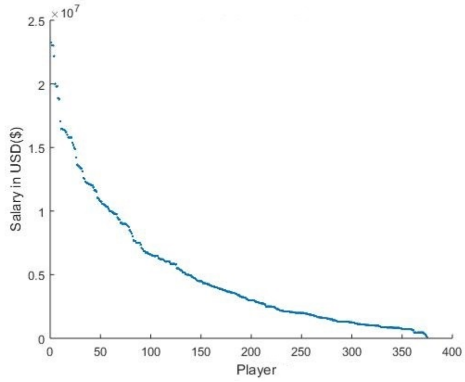


Figure 3: Ordered salaries

We believe this ranking more or less corresponds with how the teams think how much salary players deserve and how they should be ranked according to skill. As can be seen, teams do not reward players proportionally to their rank. The salary seems to scale exponentially with the rank of players. This corresponds with the thought that the league is driven by star players. So, a select few players dominate the rest of the league and their salary is awarded accordingly.

To correctly estimate a player's market value, our salary pricing method should have the same curve as the current salary market. If the player's salaries are ordered by DPR ranking, that curve should follow the same salary curve.

The observed salaries in figure 3 do not follow a completely smooth curve, because some players are still slightly over-or underpaid. These deviations from a smooth curve can be accredited to various factors, such as salary negotiating skills or image in the media. We would like to find a player's salary, without accounting for these factors. Therefore we fit a curve through this data. We will call this curve the market salary curve. We use MATLAB's curve fitting tool to easily find a correct specification for the curve. The market salary curve found was of the form

$$f(x) = a\ \exp(\mathrm{b}x) + c\ \exp(\mathrm{d}x) \tag{24}$$

where $f(x)$ are the salaries and x correspond with the players ranks. The a exp(bx) part of the equation corresponds with the salary growth curve of the whole league, while c exp(dx) corresponds with an extra salary growth for star players. The estimated parameters are a= 1689209.42, b= −0.01, c= 7632304.83 and d= −0.10. The fitted curve can be found in figure 4

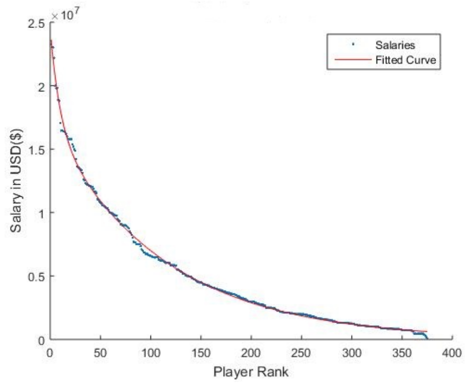


Figure 4: Fitting the market salary curve

If the players are ranked correctly according to skill, their salaries should follow this market curve. Therefore we have ranked all players according to their DPR skill level. In figure 5 we have plotted the salaries corresponding to DPR skill ranking against the market salary curve. We can see that players are generally not paid according to their skill level, because the salaries sorted by the DPR skill ranking does not follow the market curve. There does seem to be a slight declining trend in salary as the players get less skillful, as we would expect, but most players are either very over- or underpaid.

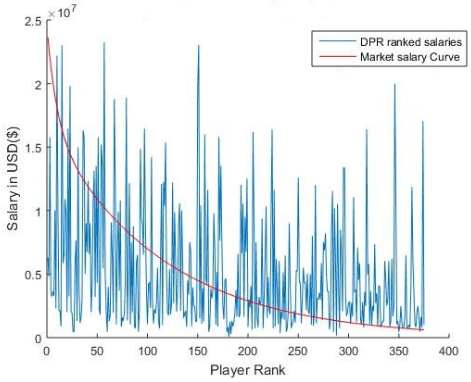


Figure 5: DPR ranked salaries vs. Market salary curve

In table 8 we listed the top 5 over- and underpaid players according to the DPR skill ranking. The most underpaid player is Kevin Durant, who should earn almost $18,000,000$ more. The reason for this is that in the 2009-2010 season, Kevin Durant was still on his rookie contract, while he was already a top 10 player. Rookie contracts can only pay a maximum amount of salary.

The rest of the top 5 underpaid players consist of players who the DPR method had ranked as top-10 players, but not according to the NBA MVP awards and SPR method.

The most overpaid player was Shaquille O'Neal, who although he was already considered by many to be on the decline this late in his career, had a massive contract. His season was not successful, partly because of injuries. The other notable overpaid players are Kobe Bryant, Kevin Garnett and Paul Pierce. These players are considered by experts among the top players. However, the experts could be wrong in these cases, because they might overvalue their offensive efforts.

Top 5 underpaid players		Top 5 overpaid players	
Player	Amount	Player	Amount
Kevin Durant	$17,988,299.37	Shaquille ONeal	$19,201,201.61
Kevin Love	$17,923,235.92	Kobe Bryant	$18,574,479.81
Matt Bonner	$17,441,230.66	Michael Redd	$16,415,989.62
Anderson Varejao	$17,316,323.42	Kevin Garnett	$15,377,454.73
Chris Andersen	$16,439,923.50	Paul Pierce	$15,296,309.41

Table 8: Top 5 over- and underpaid players according to the DPR skill ranking

7 Conclusion

In this thesis we have proposed a new method to rate NBA players. This method deals with some of the limitations of existing player rating methods. The existing methods overvalue offensive skills. This is because they use box-scores in some way or capacity, which are unable to correctly capture a players defensive skill. Furthermore, the current methods are too one-dimensional. By this we mean that they only tell you how good a player is, but not what his strengths and weaknesses are.

Our Detailed Player Rating (DPR) improves upon these methods by also estimating what strategies are effective for winning and what the strengths and weaknesses are of each player. Furthermore, we use play-by-play data to estimate our models, so we fully capture player's defensive capabilities.

The DPR method consists of a two-stage regression. The second stage is a regression of score differentials on several production statistics such as shot percentages, rebounds and blocks. For these production statistics, the difference between the production output of these statistics of the home and away team are taken. The distinction is made for all production statistics, what position the player plays that has output in the corresponding production statistic. We write this stage as a Hierarchical Bayes model with two levels, because we believe the production statistics to have different effects for different positions, yet also be very similar.

The first stage is a regression of these production statistics on the on-court presence of players. This stage is estimated with a penalized regression, to account for model sparsity. So we indirectly estimate the influence of players on score differentials through the first stage. The results of the regression in the second stage allows us to say which tactics are effective to win.

We used play-by-play match data from the 2009-2010 NBA season to compare our DPR method with Omidiran's[10] SPR method in terms of forecasting accuracy. The DPR performed slightly better in terms of Mean Absolute Prediction Error and Mean Squared Prediction Error. Among our several DPR methods, we found that the model that uses a Laplacian likelihood for the data in the second stage, outperforms a simplified model and a model that uses a Normal likelihood for the data in the second stage.

We have compared player rankings of both methods with the Most Valuable Player Award (MVP) rankings, which are awarded by highly esteemed journalists. We noticed that the SPR method ranked many of the same players as the experts did for the MVP rankings. We subsequently showed through the definition of the SPR, that this method will overvalue players with good box-score numbers and offensive skills. These offensive skills will also stand out more to the experts than defensive skills. Another shortcoming of the SPR method is that it assumes a Normal error distribution.

We have used the DPR method to analyze what strategies are effective for winning. Teams should focus on either shooting very close to the basket or shooting from 3 point range. They also should not place too much value on offensive rebounding, because having a lot of offensive rebounds inherently means that they miss a lot of shots. Therefore, they should focus more on their shooting percentage.

We have analyzed the strengths and weaknesses of the best player of the $2009 - 2010$ NBA season, according to the DPR method, namely Anderson Varejao. We found that he is an excellent free throw shooter, but should focus more on his mid-range shooting and defense of mid-range shots.

Finally, we looked at which players were over- and underpaid according to the DPR method. This was done by first ranking all salaries from high to low and subsequently fitting a curve through this data. This market salary curve should correspond with how NBA teams think the players should be paid according to their skill ranking. The salaries of the players ranked by DPR (from high to low) should follow this salary curve. We found that NBA teams do not pay their players proportionally to skill, but the salaries scales exponentially with the players rank. We also found that NBA teams severely over- and underpay some their players. We believe this is again a result from a poor judgment of offensive and defensive skill.

Overall we have found that our DPR method is good addition to the current literature. It improves upon some flaws that the current methods have, namely that they overvalue offensive skills and undervalue defensive skills. Furthermore, the DPR method provides more useful information besides a mere player rating. In terms of predictive power the DPR method has proven to be at least on par with the current best methods.

References

[1] Michael Lewis. *Moneyball: The art of winning an unfair game.* WW Norton & Company, 2004.

[2] John Hollinger. *Pro Basketball Prospectus: 2003 EDITION (Pro Basketball Forecast).* Brassey's, Washington DC, 2003.

[3] David J. Berri. A simple measure of worker productivity in the national basketball association. *The Business of sport*, 3:1–40, 2008.

[4] David J. Berri and Martin B. Schmidt. *Stumbling On Wins in Basketball.* Pearson Education, 2010.

[5] 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.

[6] S. Trninic and D. Dizdar. System of the performance evaluation criteria weighted per positions in the basketball game. *Collegium Antropologicum*, 24:217–234, 2000.

[7] S. R. Bray and L. R. Brawley. Role efficacy, role clarity, and role performance effectiveness. *Small Group Research*, 33:233–253, 2002.

[8] Garritt L. Page, Gilbert W. Fellingham, and C. Shane Reese. Using box-scores to determine a positions contribution to winning basketball games. *Journal of Quantitative Analysis in Sports*, 3.4, 2007.

[9] Dan T. Rosenbaum. Measuring how nba players help their teams win, 2004. URL http://www.82games.com/comm30.htm.

[10] Dapo Omidiran. Penalized regression models for the nba. *arXiv preprint arXiv:1301.3523*, 2013.

[11] Yinbo Li and Gonzalo R. Arce. A maximum likelihood approach to least absolute deviation regression. *EURASIP Journal on Applied Signal Processing*, 2004.

[12] Richard Paap. Bayesian econometrics. *University Lecture at Erasmus University*, 2013.

[13] Terry E. Dielman. Least absolute value regression: recent contributions. *Journal of statistical computation and simulation*, 2005.

[14] Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014.

[15] Paul Fearnhead Sherlock, Chris and Gareth O. Roberts. The random walk metropolis: linking theory and practice through a case study. *Statistical Science*, 2010.

A Metropolis-within-Gibbs sampler

In this section we will derive the Metropolis-within-Gibbs sampler mentioned in section 3.2.2. We will use this sampler to estimate the following equation

$$y_i = \alpha + \sum_{p=1}^{P}\sum_{j=1}^{K} \hat{x}_{p,j,i}\beta_{p,j} + \varepsilon_i \tag{25}$$

with $\varepsilon_i \sim Laplace(0, b)$. We choose the following prior distributions for our parameters

$$\begin{aligned} p(\beta_{p,j}) &\propto N(\mu_j, \sigma^2_{\zeta_j}) \\ p(\boldsymbol{\mu}, \alpha, b) &\propto 1 \\ p(\sigma^2_{\zeta_j}) &\propto \sigma^{-2}_{\zeta_j} \end{aligned} \tag{26}$$

We will use the following sampling scheme to obtain posterior distributions for all parameters:

$$\begin{aligned} &\bullet \text{ Sample } \alpha \text{ given } \{\{\beta_{p,j}\}_{p=1}^{P}\}_{j=1}^{K}, \{\mu_j\}_{j=1}^{K}, b, \{\sigma^2_{\zeta_j}\}_{p=1}^{P}, \boldsymbol{y} \\ &\bullet \text{ Sample } b \text{ given } \{\{\beta_{p,j}\}_{p=1}^{P}\}_{j=1}^{K}, \alpha, \{\mu_j\}_{j=1}^{K}, \{\sigma^2_{\zeta_j}\}_{p=1}^{P}, \boldsymbol{y} \\ &\bullet \text{ Sample } \{\mu_j\}_{j=1}^{K} \text{ given } \{\{\beta_{p,j}\}_{p=1}^{P}\}_{j=1}^{K}, \alpha, b, \{\sigma^2_{\zeta_j}\}_{p=1}^{P}, \boldsymbol{y} \\ &\bullet \text{ Sample } \{\sigma^2_{\zeta_j}\}_{p=1}^{P} \text{ given } \{\{\beta_{p,j}\}_{p=1}^{P}\}_{j=1}^{K}, \alpha, \{\mu_j\}_{j=1}^{K}, b, \boldsymbol{y} \\ &\bullet \text{ Sample } \{\{\beta_{p,j}\}_{p=1}^{P}\}_{j=1}^{K} \text{ given } \alpha, \{\mu_j\}_{j=1}^{K}, b, \{\sigma^2_{\zeta_j}\}_{p=1}^{P}, \boldsymbol{y} \end{aligned} \tag{27}$$

Now we will derive from which distributions we can draw. First, we define the joint posterior density function, which looks like

$$\begin{aligned} p(\beta, \boldsymbol{\mu}, \alpha, b, \sigma^2_\zeta | \boldsymbol{y}) =& (2b)^{-N} \exp(-(|\boldsymbol{y} - \alpha - \hat{X}\beta|)/b) \\ & \sigma_\zeta^{-(N+2)} \exp(-(\beta - \boldsymbol{\mu})/(2\sigma^2_\zeta)) \end{aligned} \tag{28}$$

Now we will try to find the marginal posterior density functions. We will start with the marginal posterior density function for $\boldsymbol{\mu}$, which is

$$\begin{aligned}
p(\boldsymbol{\mu}|\boldsymbol{y},\beta,\alpha,b,\sigma_\zeta^2) &= \int_\beta\int_\alpha\int_b\int_{\sigma_\zeta^2}\Big((2b)^{-N}\exp(-(|\boldsymbol{y}-\alpha-\hat{X}\beta|)/b) \\
&\qquad \sigma_\zeta^{-(N+2)}\exp(-(\beta-\boldsymbol{\mu})/(2\sigma_\zeta^2))\Big)\ \mathrm{d}\beta\ \mathrm{d}\alpha\ \mathrm{d}b\ \mathrm{d}\sigma_\zeta^2 \qquad (29)\\
&\propto \int_{\sigma_\zeta^2}\sigma_\zeta^{-(N+2)}\exp(-(\beta-\boldsymbol{\mu})/(2\sigma_\zeta^2))\sigma_\zeta^2
\end{aligned}$$

So, for a given production statistic j, the posterior density function for μ_j is normally distributed with mean $\sum_{p=1}^{P}(\beta_{p,j})/P$ and variance $\sigma_{\zeta_j}^2/P$ (see Paap [12]).

The posterior distribution for σ_ζ^2 is

$$\begin{aligned}
p(\sigma_\zeta^2|\boldsymbol{y},\beta,\boldsymbol{\mu},\alpha,b) &= \int_\beta\int_{\boldsymbol{\mu}}\int_\alpha\int_b\Big((2b)^{-N}\exp(-(|\boldsymbol{y}-\alpha-\hat{X}\beta|)/b) \\
&\qquad \sigma_\zeta^{-(N+2)}\exp(-(\beta-\boldsymbol{\mu})/(2\sigma_\zeta^2))\Big)\ \mathrm{d}\beta\ \mathrm{d}\boldsymbol{\mu}\ \mathrm{d}\alpha\ \mathrm{d}b \qquad (30)\\
&\propto \int_{\boldsymbol{\mu}}\sigma_\zeta^{-(N+2)}\exp(-(\beta-\boldsymbol{\mu})/(2\sigma_\zeta^2))\ \mathrm{d}\boldsymbol{\mu}
\end{aligned}$$

So for a given production statistic j, the posterior distribution for $\sigma_{\zeta_j}^2$ is an Inverse-Gamma distribution with location parameter $\sum_{p=1}^{P}(\beta_{p,j}-\mu_j)^2$ and P degrees if freedom (see Paap [12]).

For parameters α, β and b we were unable to find a conditional distribution of a known kind. Therefore we will use a random-walk sampler to draw these parameters. The sampling procedure is as follows. Consider the case when we are in the $(m+1)^{\text{th}}$ iteration of the random walk sampler where we try to find a posterior distribution of parameter *theta*. We will first propose a candidate value for θ^{m+1}, namely θ^* which is defined as

$$\theta^* = \theta^m + c\varepsilon \qquad (31)$$

where the random variable $\varepsilon \sim N(0, 1)$. Here c is a tuning parameter which we will explain later in this section. The acceptance probability α for candidate value θ^* is then defined as

$$\alpha = min(f_\theta(\theta^*)/f_\theta(\theta^m), 1) \tag{32}$$

If $\alpha < u$, we set $\theta^{m+1} = \theta^*$, but if $\alpha < u$ we set $\theta^{m+1} = \theta^m$. Here u is a random uniform variable distributed between 0 and 1. The probability functions f_θ for parameters β, α and b are

$$\begin{aligned} f_\beta(\beta|\boldsymbol{\mu}, \alpha, b, \sigma_\zeta^2, \boldsymbol{y}) =& \exp(-(|\boldsymbol{y} - \alpha - \hat{X}\beta|)/b)\exp(-(\beta - \boldsymbol{\mu})/(2\sigma_\zeta^2)) \\ f_\alpha(\alpha|\beta, \boldsymbol{\mu}, b, \sigma_\zeta^2, \boldsymbol{y}) =& \exp(-(|\boldsymbol{y} - \alpha - \hat{X}\beta|)/b) \\ f_b(b|\beta, \boldsymbol{\mu}, \alpha, \sigma_\zeta^2, \boldsymbol{y}) =& (2b)^{-N}\exp(-(|\boldsymbol{y} - \alpha - \hat{X}\beta|)/b) \end{aligned} \tag{33}$$

These function were chosen by taking all relevant terms from joint posterior density in equation (28).

We will perform a simulation of 200000 iterations, with a burn-in sample of 100000 iterations and a thinning factor of 100. We initialize the parameters θ^0 as the Least Absolute Deviations estimate of equation (25). Tuning parameter c in equation (31) is initialized with value 1. During the burn-in period we tune c such that the acceptance rate of the random walk sampler is between 0.2 and 0.3, which Sherlock and Roberts. [15] found to be the optimal acceptance rate for multivariate problems. We do this by decreasing c with 10% when the acceptance rate drops below 0.2, and increasing c with 10% when the acceptance rate goes above 0.3 during the burn-in period.