NBA Salaries: Assessing True Player Value

By: Mike Ghirardo

Advisor: Dr. Matthew Carlton

Introduction:

³7HDPZRUN LV WKH HOHPHQW RI EDVNHWEDOO PRVW GL to capture in any quantitative sense ´(Oliver). Basketball is, arguably, fundamentally different from baseball and football, with respect to the importance of teamwork. In baseball, a great hitter (no teamwork required) is going to undoubtedly make his team better. In football, although more teamwork is required than in baseball, nearly all of the offensive action is scripted, allowing for little dynamic and spontaneous interaction among teammates. Compare this to the game of basketball, where although there are certainly offensive schemes that are run, there is so much more action that UHVXOWV IURP D WHDPPDWH¶V UHVSRQVH WR HYHU\ RWKHU simultaneously. This greater interdependency between teammates inherent in the game of basketball renders individual statistics less meaningful to team success than in baseball and football.

In the past, basketball players ¶values, and therefore their compensation, have traditionally been measured in large part by their in game box-score statistics. The presumption being that more impressive individual statistics signified a greater value to the team, and therefore the greater the compensation should be to the player. In reality, players who average 20 points or 10 rebounds per game have impressive individual statistics, but may or may not be helping their teams win. This traditional system is inherently flawed. It does not adequately PHDVXUH WKH OHYHO RI FDSDELOLW\ RI D SOD\HU¶V ³teamwork´ ZKLFK DUH VR FULW the success of a basketball team. Points, rebounds, assists, steals, and blocks: what story do they really tell? A player averaging 20 points per game may be hindering his team greatly if he takes 30 or more shots on average to do so. A player with 10 assists per game might be passing up higher percentage open shot attempts in his efforts to maximize his assists, thereby hurting the team (in a way that does not show up in his personal stat line). These types of players may look like WKH\¶UH GRLQJ Zonderful things for their teams in the stat sheet. But, in reality, some of these players with impressive personal statistics may be doing more harm than good to their team; and those without great traditional statistics may be invaluable to their team by doing the myriad of difficult to measure (and therefore unrecorded) things the right way. Unrecorded plays like quality help-side defense, setting a good screen, quickly getting back on defense, blocking out your man to allow your teammate the rebound, DQG JHWWLQJ \RXU KDQGV LQ D VK all are vital to the outcome of a game. These unrecorded plays many times make an even bigger difference in the course of the game than some of the recorded ones.

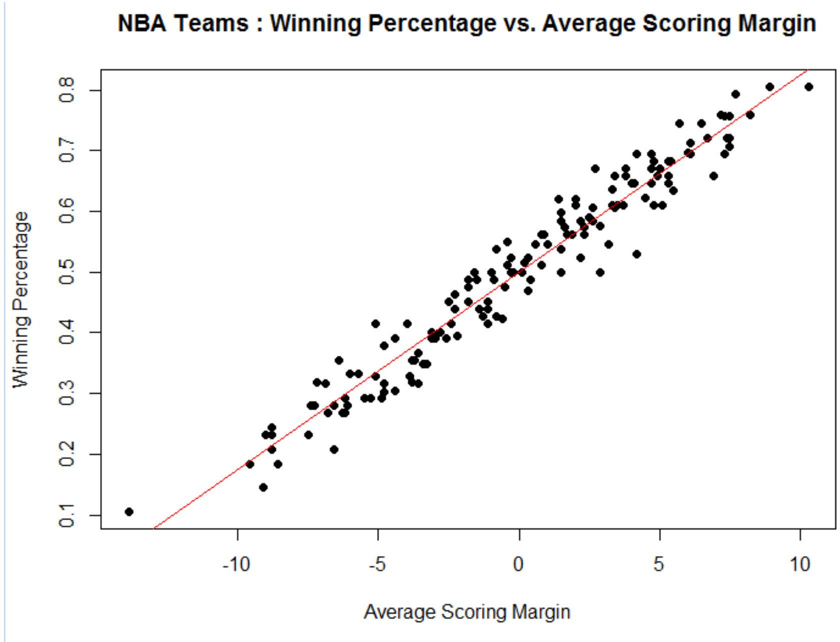
Because NBA executives now realize that previous methods and metrics were not the best way to evaluate players, they are turning to new advanced statistical methods to measure player worth. One of the major statistical metrics that has gained traction over the last decade is the traditional ³plus-minus´ statistic. The traditional plus-minus statistic captures how a team does with a player on and off the floor. For example, if a NBA player is on the court for a half and his team ³wins´ by 10, and off the court for the other half the game and his team ³loses´ by 3, his traditional plus-minus statistic would be +13. This is because when that player was on the court for that game, his team did 13 points better than when he was off of the court. There is some merit to evaluating players in this manner. First and foremost, it measures the end result ± how does <u>the team</u> do when a player is in? This approach does not presume that good personal statistics automatically translate into team success; nor does it even attempt to identify and measure the ³unrecorded things´, PHQWLRQHG LQ WKH SUHYLRXV SDUDJU ±looking only to the results of the team as a cumulative impact.

<u>Advantages of the Adjusted Plus-Minus Statistic</u>:

There is, however, a fundamental problem with the traditional plus-minus statistic: it GHSHQGV RQ WKH VWUHQJWK RI D SOD\HU¶V WHDPPDWHV D LQGLYLGXDOV¶ FRQWULEXWLRQ to team success. For example, an average player playing with four all-star teammates consistently will have an inflated traditional plus-minus statistic. Most likely, the team with an average player and 4 all-stars will do quite well, but not because of the average player. Given this obvious shortcoming, basketball upper management has begun to investigate a more advanced measure, called the <u>adjusted</u> plus-minus (³APM´) statistic. The APM measure adjusts for the other nine players on the floor. 7KLV PHDQV WKDW LW PHDVXUHV D contribution, independent of all other players on the court. The APM should therefore more accurately assess the true contribution of an individual player to the success of a team.

Another major advantage to working with the APM statistic is that it seeks to capture exactly what NBA executives should be striving for - wins! It provides insightful analysis on the players that cause teams to win, and lose. Below is a regression equation for the overall team plus-minus statistic in the NBA for all NBA teams from the 2007-2008 to the 2011-2012 seasons (NBA Standings).

Figure 1: A Linear Equation on Team Win Percentage vs. Average Point Differential




```
Coefficients:
             Estimate
(Intercept) 0.4999783
DIFF        0.0324942
```

```
Adjusted R-squared: 0.9493
```

By the adjusted R-squared from Figure 1, R Q H F D Q V H H W K D W S H U F H Q W R I S H U F H Q W D J H L V D F F R X Q W H G I R U E \ O R R N L Q J D W D W H D P ¶ V that correlation is extremely high! This means that good teams do not squeak out close games and lose big when they lose; good teams tend to win on average by a large margin and poor teams tend to lose on average by a big margin. In fact, a one-point increase in a teams ¶average point differential indicates on average that that team will win 3.25 percent more of its games. So, in an 82 game NBA season a team that increases its average point differential by one will win about 3 more games on average. This raw team plus-minus score is derived by the players on the court. The main statistic that is incorporated into the team plus-minus score is the adjusted plus minus per player, because it calculates how the point differential is affected when a player is on and off the court. For example, a center with an adjusted plus-minus of -2, and another center in free agency with an adjusted plus minus of +3 signifies a meaningful difference in impact on number of team wins, and therefore, in their relative value. NBA executives must realize that if they acquire and then play the center with the adjusted plus minus of +3 instead of their existing -2 center that their team can win about 10 more games, assuming the starting center plays 2/3 of each game on average. This is a major difference; an increase of 10 wins in the Western Conference this year could have been the difference between a non-playoff team and a 6-seed in the playoffs.

Drawbacks to the APM Statistic:

NBA teams that calculate the APM statistic in the right way have an advantage in evaluating players. However, there are major drawbacks to the APM as a predictive model as well. For one, this measure has a lot of statistical noise. This is due to the co-linearity between players. For example, an NBA player might play 80 percent of his minutes with another player. This means that the difference in their APMs is only based on the 20 percent of the time he plays when the other player is off the court. Maybe his team just happens to not play as well in this small percentage of the time when he is on the court and the other player is off the court. This can cause an individual player¶ APM statistic to fluctuate greatly from year-to-year.

Another shortcoming is that an individual¶ APM statistic might be abnormally high (or low) because of the system that he plays in. An athletic player may flourish in a run-and-gun system like the Warriors employ, but not succeed in a slower tempo style of play like the Spurs utilize. General Managers must understand these differences (and estimate their impacts) when evaluating free-agent talent, and not rely solely on APM as a predictive tool.

Moreover, the model fails to properly adjust for the interactions between players. For example, despite amassing some tremendous personal talent (Dwight Howard, Kobe Bryant, and

Steve Nash) the Los Angeles Lakers as a team struggled this year. One possible explanation for this: these all-stars and probable future hall of famers just do not play well together. As I stated in my opening paragraph, in basketball, teams are not necessarily the sum of their parts; the YDOXH RI WHDPZRUN RU WHDP FKHPLVWU\ LV DUJXDEO\ PRU other major sports. If you looked at a line-up card the Lakers should have won nearly every QLJKW EXW,W¶V WKH\ EHEDOQ\W WKH LQWHUDFWLRQ RI WZR RU W less effective than they are as an individual. On the other hand, two ordinary players can be great complements to one another ¶V VW\OH RI SOD\ DQG KDYH WKDW V\QHUJ\ contribution to team success.

Selection of the Variables:

There are a few remarks that must be made on how variables were selected in this model. For each of the 2007-2008 to 2011-2012 seasons, the only players selected were those that played greater than 500 minutes. This is not uncommon in the calculation of the APM, many basketball statisticians have set the cutoff number of minutes around 500 (basketballvalue). This helped reduce the co-linearity between the players, while still accounting for about 3/5 of the players in the NBA. Reducing the co-linearity was vital to the design because it helped keep the estimates much more stable, and the errors much smaller. Because we set the threshold at 500 minutes, we are more certain on the reasonability of our estimates.

The 500 minute cutoff also creates a reference group to which all other players are compared (Count). This is slightly counter-intuitive; one would expect the APM measure to be based on the average NBA player (a player with a 0 APM). Unfortunately, statistical models implemented to calculate the APM do not have this capability. Instead, each player is compared to the average player with less than 500 minutes. This means most of our estimates will be greater than 0, because as one would imagine the average player playing more than 500 minutes is better than the average player playing less than 500 minutes. Although this is not an ideal way to assess player value, we can still compare players in our model. A player with an APM of +8 and another player with an APM of +5, indicates the same difference (8-5=3) had a net 0 APM player been the reference group. The numbers would just decrease, possibly to +2 and -1, for example.

,Q WKH PRGHO D SOD\HU¶V FRnutes (the length of a game)V HVWLPD for the APM statistic. This was accomplished by multiplying the average points per possession (of an observation) by the average number of possessions in a game. It was important to make all estimates the length of the game in order to understand the data. The statistic is much easier to comprehend when it is interpreted on a per-game basis. Also, these results can be more easily compared to existing APM statistics, traditional plus minus statistics, and other advanced statistics that are on a per-game basis (basketballvalue). An example of an interpretation of the APM: a player with an APM of +10 is 10 points better per game than the average player that played less than 500 minutes, for that season.

Selection of the Observations:

Figure 2: The Dataset

GameID	StartTime	ElapsedSecs	HomePlayer1ID	AwayPlayer1ID
20071030HOULAL	0:48:00	455	189	308
20071030HOULAL	0:40:25	105	189	198
20071030HOULAL	0:38:40	65	189	198
20071030HOULAL	0:37:35	76	186	318
20071030HOULAL	0:36:19	19	186	318

The model is comprised by a list of each matchup of one 5-man unit against another for each of the 2007-2008 through 2011-2012 seasons. That means a period of time in the basketball game in which no substitutions are made. For each observation many variables are computed; the important ones for our model are the number of possessions, points scored, and the indicator variables for all home and away players on the court. I obtained this data from basketballvalue.com, and there are about 40,000 observations for each of the non-lockout years (every season but 2011-2012).

<u>The Model</u>:

There are 2 types of models popularly used in calculating the APM statistic: the Rosenbaum Model and the Ilardi-Barzilai model. Many believe the Ilardi-Barzilai model is slightly better because it separately DFFRXQWV IRU WKH HIIHFW RI D SOD\HU

the overall APM statistic. The model for each year is:

$$y = \beta_0 + \beta_1 X_1 + \cdots + \beta_J X_J + \delta_1 D_1 + \cdots + \delta_J D_J,$$

where J is the number of NBA players in the league with over 500 minutes (for that year). The variables in the model are defined the following way:

y = the scoring margin per 48 minutes during an observation

$$X_j \begin{cases} 1, \text{ player j is on offense during the observation;} \\ 0, \text{ player j is not playing or is on defense during the observation;} \end{cases}$$

$$D_j \begin{cases} 1, \text{ player j is on defense during the observation;} \\ 0, \text{ player j is not playing or is on offense during the observation;} \end{cases}$$

ZKHUH J. ” M ”

The observations in the model are re-coded into two lines of data. One line corresponds to the points per 48 minutes for the home team, while the other line corresponds to the points per 48 minutes for the away team. These lines are each weighted by the number of possessions in the observation. For the home team observation, the statistic is weighted by the number of home possessions, and vice versa for the away team. After splitting up the data set by home versus away, we can tell whether each player is on offense or defense. A home team player is on offense when we use the home team observation, and on defense for the away team observation, and vice versa for an away player (MacDonald).

The Ilardi-Barzilai Model can be expressed in the following way:

$$\underbrace{\begin{bmatrix} y_h \\ y_a \end{bmatrix}}_{2M\times 1} = \underbrace{\begin{bmatrix} X_h & D_a \\ X_a & D_h \end{bmatrix}}_{2M\times 2J} \underbrace{\begin{bmatrix} \beta \\ \delta \end{bmatrix}}_{2J\times 1} + \beta_0 \mathbb{1}_{2M\times 1},$$

Where

M = number of observations in a season (length of time when no substitutions are made)

J = number of players in the model (players with over 500 minutes in a particular season)

h = home, a = away

y = number of points scored by the respective teams if they played the whole 48 minutes scoring at the same rate

X = Offensive indicator variables (1 if player j is on offense, 0 otherwise), M × J matrix

Y = Defensive indicator variables (1 if player j is on defense, 0 otherwise), M × J matrix

¶V 2IIHQVLYH 3OXV 0LQXV 230 IRU DOO - SOD\HUV

/¶V 'HIHQVLYH 3OXV 0LQXV '30 IRU DOO - SOD\HUV

$1_{2M\times 1}$ PDWUL[RI ¶V IRU WKH LQWHUFHSW WHUP

As a frame of reference, a useful example will be discussed to show how the model changes one simple observation. Suppose that in a length of time when no substitutions are made, players coded 1-5 are on the court for the home team, and players coded 6-10 are on the court for the away team. This observation is re-coded to make 2 lines of data in order to run our model, one involving the points per game scored by the home team, and the other the points per game scored by the away team. Because each observation will be much less than the length of a game, each observation ¶ average points per possession is multiplied by the number of possessions in the average game. The example of one re-coded observation is given below (MacDonald).

$$X_h = [1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0 \cdots 0], D_a = [0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0 \cdots 0],$$
$$X_a = [0\ 0\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ 1\ 0 \cdots 0], D_h = [1\ 1\ 1\ 1\ 1\ 0\ 0\ 0\ 0\ 0\ 0 \cdots 0],$$

Now, imagine re-coding all the observations in a dataset (in our case about 40,000), and creating defensive and offensive indicator variables for all players (about 400) in each season. The dataset becomes 80,000 (lines of data) × 800 (defensive and offensive indicator variables for all 400 players), for every year. Because the home players on offense are the away players on defense, and vice versa, our matrix can be written as

$$\begin{bmatrix} y_h \\ y_a \end{bmatrix} = \begin{bmatrix} X_h & D_a \\ D_a & X_h \end{bmatrix} \begin{bmatrix} \beta \\ \delta \end{bmatrix} + \beta_0 \mathbb{1}_{2M \times 1}.$$

The coefficients have the following interpretation:

$_j$ = points per 48 minutes contributed by player j on offense

- /$_j$ = points per 48 minutes contributed by player j on defense

ú$_0$ = intercept,

6 R $_1$ are the points per 48 minutes that player 1 contributes to his team on offense. ±/$_1$ are the points per 48 minutes that player 1 takes away from the opponent. Therefore, a negative /$_j$ value is a good thing. It means that player j prevents his opponents from scoring that many points per 48 minutes, compared to the average player that played less than 500 minutes that season. $_j$ and - /$_j$ are the OPM and DPM statistics for player j, respectively. Again, this statistic is independent of teammates and opponents, and calculated per 48 minutes (MacDonald). The

Adj XVWHG 3OXV 0LQXV $30 LV WKH FRPELQDWLRQ RI D SOD

defense. 6R D SOD\HU¶V $30 FDQ EH F calculated as follows:

$$APM = OPM + DPM.$$

Remember that the OPM and DPM measures are compared to our reference group, the average player that played under 500 minutes.

<u>Re-Coding The Data</u>:

In order to run an Ilardi-Barzilai model, the data set must be re-coded for home and away possessions. The re-coding of the indicator variables were represented in the previous example. However, a fuller version that encompasses all necessary variables with the data set used is represented in this section. Re-coding the data into two lines main advantage is that it allows a statistician to use indicator variables to signify whether a player is on offense, defense, or off the court. If that data is not re-coded, there is no way to tell the offensive and defensive contribution of a player. Here is an example for one re-coded line of the data set:

Figure 3: Re-Coding Example

BEFORE:

PossessionsHome	PossessionsAway	HPT	APT	189	308
15	14	87	31	1	1

AFTER:

	Possessions	PT	O_189	O_308	D_189	D_308
Home	15	87	1	0	0	1
Away	14	31	0	1	1	0

After re-coding, the data set will be twice as long. For all 40,000 observations, there will be 2 rows, one for the home possessions, and the other for the away possessions. Please note that HPT is Home Point Total, the number of points the home team would score per 48 minutes if they kept scoring at the same rate. APT is the same, just for the away team. Also, note that players EHIRUH FRXOG EH LQGLFDWHG E\ ¶V DQG ¶V RQ DQG RII But, there was no way of knowing D SOD\HU¶V RIIHQVLYH DQG GHIHQVLYH FRQWULEX Now that there is a home and away row, one can label a player as playing offense or defense. This allows us to measure each SOD\HU¶V RIIHQVLYH DQG GHIHQVLYH FRQWULEXWLRQ separately.

Running the Model:

After re-coding all observations in the data set, a general linear model is run to assess HDFK SOD\HU¶V RIIHQVLYH DQG GHIHQVLYH plus-minus statistics. The explanatory variables for the general linear model are the offensive and defensive indicator variables for all J players (players with over 500 minutes for each season, individually). The response variable is the point total, and the equation is weighted on the number of possessions for each observation (basketballvalue). Thus, this equation indicates the offensive and defensive changes in point total per 48 minutes when player j is on and off the court, compared to the average player that plays less than 500 minutes. After obtaining the offensive and defensive plus minus statistics, one can easily calculate the APM by adding the two together (from the previous section).

2012-2013 APM Prediction:

After obtaining the OPM, DPM, and APM statistics for the 2007-2008 to 2011-2012 seasons, time series methods can be used to calculate the predicted APM for the next year, the 2012-2013 season. Each player that played over 500 minutes for at least one season from 2007-2008 to 2011-2012 has at least one prediction. These players have anywhere from 1 to 5 seasons with predictions. For players with 2 to 5 seasons, a simple exponential smoothing model can be used to calculate the predicted OPM and DPM statistics for the next season. This intuitively makes sense; if a player is getting better in the previous seasons this trend will probably continue. If a player is getting worse because of age or some other factor that negative trend will likely continue as well.

The other estimates to make are those in which a player only has an APM, OPM, and DPM prediction for only one season. For this case, we can see how the average player does in their second season by taking the 2011-2012 season out of our data set, and only using players with one predicted APM in the previous 4 seasons. After evaluating this, our dataset suggests that a player ¶APM did not increase or decrease dramatically, on average. Therefore, we will assume that the best estimate of a player ¶APM, DPM, and OPM if there is only one observation, is simply that single observation.

After finding the predicted OPM and DPM, the predicted APM for the year 2012-2013 was found by simply adding OPM and DPM together just like before. After this is done, we have estimates for all the players with at least one observation in the previous 5 seasons. In the 3 tables below, Tables 1,2, and 3, there are 2012-2013 APM, OPM, and DPM predictions for the top 20 players in each statistic.

2012-2013 Results:

Table 1: 2012-2013 Predicted APM

Rank	Name	Position	APM
1	Chris Paul	PG	17.93
2	Manu Ginobili	SG	17.75
3	Jameer Nelson	PG	16.02
4	Thaddeus Young	SF	15.80
5	Rajon Rondo	PG	15.32
6	Jamaal Tinsley	PG	15.12
7	Will Bynum	PG	15.01
8	James Harden	SG	14.96
9	Luc Richard Mbah a Moute	PF	14.31
10	John Wall	PG	14.26
11	Solomon Jones	PF	14.15
12	Matt Bonner	PF	13.97
13	Trevor Booker	PF	13.01
14	Derek Fisher	PG	12.80
15	Blake Griffin	PF	12.24
16	Deron Williams	PG	12.24
17	Paul Millsap	PF	12.12
18	Danny Granger	SF	12.04
19	Stephen Curry	PG	11.81
20	Kevin Durant	SF	11.65

Legend: Yellow = surpised, White = not surprised

There are some major surprises in the predicted APM statistic for the 2012-2013 season. At least some of these predictions are certainly due to the great amount of variability in the adjusted plus-minus measure. Most of the surprising players, highlighted in yellow, have had one great year according to this metric. Luc Richard Mbah a Moute, Solomon Jones, and Trevor Booker are all examples of players with one outstanding APM year possibly inflating their APM statistic prediction for the 2012-2013 season.

Table 2: 2012-2013 OPM Prediction

Rank	Name	Position	OPM
1	Jameer Nelson	PG	14.02
2	Manu Ginobili	SG	13.90
3	James Harden	SG	13.59
4	Greg Monroe	C	12.56
5	Stephen Curry	PG	12.01
6	Thaddeus Young	SF	11.88
7	Kyrie Irving	PG	11.20
8	LeBron James	SF	10.78
9	Blake Griffin	PF	10.75
10	Chris Paul	PG	10.68
11	Ersan Ilyasova	SF	10.52
12	Jamaal Tinsley	PG	10.24
13	Ryan Anderson	PF	10.13
14	Matt Bonner	PF	9.92
15	John Wall	PG	9.71
16	John Lucas	PG	9.63
17	Mike Miller	SG	9.60
18	Dirk Nowitzki	PF	9.48
19	Andre Miller	PG	9.44
20	Danny Granger	SF	9.17

Legend: Yellow = Surprised, White = Not Surprised

From Table 2, there are not many surprises for the 2012-2013 OPM prediction. The top 10 players are all known to be very good on offense. Greg Monroe is the only non-household name for avid basketball fans, but a Pistons fan would be the first to tell you that he can pass. He is arguably the best passing center in the game; thus it makes sense that he has a good OPM. There are two players in spots 11-20 that are not well known at all, Ersan Ilyasova and John Lucas. Once again, this could have just been due to the variability of OPM, like APM, but we should look into their statistics. Matt Bonner is surprising, but, he is known to be a very high percentage shooter.

Table 3: 2012-2013 DPM Prediction

Rank	Name	Position	DPM
1	Solomon Jones	PF	14.88
2	Luc Richard Mbah a Moute	PF	12.67
3	Earl Boykins	PG	8.70
4	Will Bynum	PG	8.59
5	Rajon Rondo	PG	8.53
6	Trevor Booker	PF	8.11
7	Marreese Speights	PF	7.96
8	Jose Calderon	PG	7.77
9	Mickael Pietrus	SF	7.31
10	Chris Paul	PG	7.24
11	Martell Webster	SF	6.88
12	Derek Fisher	PG	6.60
13	James Posey	SF	6.57
14	Andre Iguodala	SG	6.42
15	David West	PF	5.87
16	Anthony Johnson	PG	5.74
17	George Hill	PG	5.59
18	Rudy Fernandez	SG	5.51
19	Chuck Hayes	PF	5.46
20	Norris Cole	PG	5.41

Legend: Yellow = Surprised, White = Not Surprised

From Table 3, You will notice that there are many surprising DPM standouts. Players are rated very well by the DPM either because of the variability in the statistic, or because they are surprisingly good defenders that go unnoticed. It is interesting that there are many more surprising defensive players than offensive. This can possibly be due to the fact that there are many more intangibles on defense that go unnoticed. A player that gets a bunch of steals or blocks (good individual stats) might nonetheless be an ineffective defender. A player that gets many steals might be overly-aggressive to obtain those steals, and thereby play poor one-on-one defense by taking too many risks, resulting in easy points for the opposition on too many occassions. The DPM statistic accounts for those hard-to-see things. Getting your hands in a shooter¶s face, getting back quickly on transition defense, and playing quality help-side defense are just three aspects of defense difficult to track and measure individually, but the results of which are captured in this defensive metric.

One very interesting observation is that the predicted defensive plus-minus statistic has no centers in the top 20 observations. This is odd because many teams sign centers for their GHIHQVLYH 3SURZHVV´ Much of basketball believes that having a large presence in the center of the court will force players on the opposition to alter their inside shots, thereby reducing their field goal percentage. However, these types of players, who are usually over seven feet tall, generally do not move well. Perhaps the ability to change a few shots on the inside is not as important as DUH RWKHU DWWULEXWHV WKDW ±WKH 3VKRUWHU´ SO such as foot speed and agility.

<u>How Much Should A Player Be Paid</u>:

With the predicted plus minus statistics, we can predict how much a player should be paid if we have one other piece of information: the salary for all the players. So, I obtained the salaries from nba.com for every player in the 2012-2013 season (NBA Salaries).

After obtaining DOO SOD\HUV¶ salaries, there are several ways to assess true player monetary value. A linear approach is one such method. For the linear approach, one would add all the salaries in the data set, and divide by the number of players (the average player salary). Then, set the mean of the predicted adjusted plus minus statistic for the 2012-2013 season equal to the average player salary. After doing this, we would then set the worst player in the league (according to his APM) equal to the league minimum. Then, you have two points, the worst APM SOD\HU¶V VDODU\ APM SOD\HU¶V VDODU\ With two points one can easily set a line through the data.

An alternative to the linear approach is the APM rank approach. First, order the data so that the salaries are in descending order. Then, take the salary column out of the data set ranked in order, smallest to largest. After this, re-order the data set according to the APM statistic, worst to best. Then, combine the new dataset with the row of ordered salaries. This would ensure that the worst player in the league according to the APM would have the lowest salary, all the way to the best player in the league having the highest salary.

<u>Advantages and Disadvantages</u>:

Obviously, these are two completely different methods. There are advantages and disadvantages to each. The fact that both of these methods only attribute on court ratings is a shared shortcoming. The superstars that may be slightly over-paid if you only look at their on-court statistics many times are not over-paid because of their off the court contributions. Superstar players cause ticket prices and jersey sales to climb, causing a team to make much more money to possibly spend on other players (and provide profit to the team owners). Another way of saying this is that team wins is not the only objective of management ±it is, after all, in the business of making money. $ SOD\HU¶V LPSDFW RQ WHDP UHYHQXHV L tured perfectly by APM or any other purely on-court metric.

The linear approach is appealing in its simplicity. Further, making the salary distribution linear seems to make sense, according to our previous analysis. Team scoring margin and team win percentage seem to act in a very linear manner, therefore the APM statistic should probably be thought of as linear. It would therefore seem logical that the salary associated with the APM statistic should be linear as well.

The above notwithstanding, the linear approach does suffer from a major problem. It does not use the current salary distribution used by the NBA. Using this linear approach will systematically claim that every player paid over 12 million dollars is over-priced. This is because the league average, for players who play over 500 minutes, is around 6 million dollars, and league minimum is about 400 thousand dollars. Extrapolating linearly from these points suggest the highest paid player should make a little less than 12 million dollars (should be the same distance from 6 million as 400 thousand).

The rank method does not suffer from this problem. It utilizes the current distribution of salaries in the NBA. This is helpful because salaries will not change overnight even if player contribution truly is linear. Using the rank method enables the comparison of actual salaries with on-court APM salaries directly. $QG LW ZLOO QRW VHW DQ DUELWUDU\ ³FD approach; those making over 12 million dollars a year will not automatically be GHHPHG ³over-valued´.

7KH UDQN PHWKRG¶V PDMRU GLVDGYDQWDJH LV WKDW contribution may be. This non-linearity, however, might be a good thing because it captures superstar jersey sales off the court. Those really good on-court players should not just be paid more because they help win games, but because they directly affect the value of a team through jersey sales and ticket prices.

<u>Over and Under Valued Players</u>:

For this analysis, over and under valued players were evaluated by the rank method approach. I felt the positives in this approach outweighed the negatives. From the analysis I have ranked the 20 most under and over-valued players, in Tables 4 and 5, respectively.

Table 4: 2012-2013 Under-Paid Players

Rank	Name	Position	Salary	Under-Paid
1	Jamaal Tinsley	PG	$ 1,352,181	$ 18,092,322
2	Derek Fisher	PG	$ 261,343	$ 16,529,002
3	Solomon Jones	PF	$ 792,377	$ 16,389,623
4	Will Bynum	PG	$ 3,250,000	$ 15,750,000
5	Trevor Booker	PF	$ 1,385,280	$ 15,503,720
6	Peja Stojakovic	SF	$ 402,065	$ 14,566,185
7	Matt Bonner	PF	$ 3,630,000	$ 13,547,795
8	Luc Richard Mbah a Moute	PF	$ 4,794,192	$ 12,750,808
9	Travis Diener	PG	$ 208,802	$ 12,535,198
10	James Harden	SG	$ 5,820,417	$ 11,959,041
11	Jordan Crawford	SG	$ 1,198,680	$ 11,801,320
12	John Lucas	PG	$ 1,500,000	$ 11,700,000
13	John Wall	PG	$ 5,915,880	$ 11,629,120
14	Michael Redd	SG	$ 826,828	$ 11,573,172
15	Thaddeus Young	SF	$ 8,289,130	$ 11,463,515
16	Jameer Nelson	PG	$ 8,600,000	$ 11,348,799
17	Jeff Teague	PG	$ 2,433,077	$ 11,235,716
18	Stephen Curry	PG	$ 3,958,742	$ 11,041,258
19	Kevin Ollie	PG	$ 825,497	$ 10,474,503
20	Vince Carter	SG	$ 3,090,000	$ 10,400,000

Legend: Yellow = Old, Red = Rookie Contract, White = Under-rated

Table 4 displays the twenty most under-valued players in the 2012-2013 season according to the rank method. In this table there are three distinct types of players. One type of under-valued player are the older, seasoned veterans. Although this may not be necessarily the pure, stand alone cause of their low salary, it could be one of the many factors by which they are judged, and thought less of. Older players lose some of their athleticism with age, and this lack of athleticism could partially explain why some of these players are under-valued as well. Many times basketball viewers eyes can deceive them because they think players are less effective because of what they cannot do. The fact that these particular players F D Q ¶ W M X P S D V K L J K as fast as their younger, more inexperienced peers might make them seem less effective to the average viewer. Another recurring explanation for these older, supposedly under-valued players is the inherent random variability in the APM could be causing their under-valued measure to be far too high. The second group are those that are under-paid because they are still on their rookie contract. Because of this, known superstars are paid much less than their true value. There are three such players in this group. The third and last group consists of players that are neither rookies nor aged veterans; but are undervalued/underpaid according to the rank method. They share a common trait ±these players are not well known. The players in the table are either effective NBA players or are randomly under-valued due to the inherent variability in the APM. Teams have a tremendous advantage if they can decipher between the two, and obtain the player that is truly under-valued.

Table 5: Over-Paid Players

Rank	Name	Position	Salary	OverPaid
1	Kobe Bryant	SG	$ 27,849,149	$ 21,199,149
2	Joe Johnson	SG	$ 19,752,645	$ 17,955,045
3	Dirk Nowitzki	PF	$ 20,907,128	$ 14,857,128
4	Amare Stoudemire	PF	$ 19,948,799	$ 13,548,799
5	Pau Gasol	PF	$ 19,000,000	$ 12,301,435
6	Marc Gasol	C	$ 13,891,359	$ 11,650,909
7	Carlos Boozer	PF	$ 15,000,000	$ 11,520,000
8	Hedo Turkoglu	SF	$ 11,815,850	$ 11,303,217
9	Rudy Gay	SF	$ 16,460,538	$ 11,084,778
10	Tyson Chandler	C	$ 13,604,188	$ 10,656,388
11	Carmelo Anthony	SF	$ 19,444,503	$ 10,544,503
12	DeAndre Jordan	C	$ 10,532,977	$ 10,312,490
13	Monta Ellis	PG	$ 11,000,000	$ 10,000,000
14	Emeka Okafor	C	$ 13,490,000	$ 9,926,400
15	Ben Gordon	SG	$ 12,400,000	$ 9,259,571
16	Andre Iguodala	SG	$ 14,968,250	$ 9,147,833
17	Derrick Rose	PG	$ 16,402,500	$ 9,143,540
18	Al Jefferson	C	$ 15,000,000	$ 8,908,637
19	Richard Jefferson	SF	$ 10,164,000	$ 8,811,819
20	Luol Deng	SF	$ 13,305,000	$ 8,805,000

Table 5 displays the twenty most over-paid players according to the rank method. Obviously, these must be well-paid players. The twentieth most over-valued person is making over eight million dollars too much, which means every player on the list must be making at least that amount. Most of these players are considered superstars. Some of those over-paid are known to be less effective than what they used to be. The older players are over-paid many times because they signed a multi-year contract worth what the player should have been making at that time, but not now. Other players are simply over-valued. People still consider them superstars, and they are not aged. These players can be over-paid for many reasons. Maybe, the variability associated with the APM randomly under-rated their value for the 2012-2013 season (a problem with the measurement tool). Or alternatively the tool is largely right, and management for whatever reason is vastly overpaying these particular players when it comes to their contribution to team wins. No doubt that for some of the mega stars, management places much additional value on them by attributing greater ticket prices, number of tickets sold, merchandise revenues and food and beverage revenues to their presence alone. For these cases there might be an interesting divergence of interests between owners/management and the team coach. The owners/managers might be principally interested in maximizing short term revenues and profits (keep the superstar player on the team, even if you have to pay him more than his court FRQWULEXWLRQ VXJJHVWV ZKLOH WKH F±RWDKFDKW¶V¶VSKUR ZFKLHSD evaluated and his worth/salary is determined. Coach might think he could do much better if he did not have to accommodate the overpaid superstar and instead could spend this money on other lesser known players that could actually do much more to help the team win.

<u>Consistently Under-Valued</u>:

So, which players should general managers try and sign in free agency? Many factors weigh into whether or not a team should pursue a player. Detailed analysis must be examined to determine if the player is needed, fits the offensive and defensive style of play of a particular team, and has positive off-court characteristics. How will this player impact team chemistry, will he increase ticket prices, jersey sales,and team recognition, what is his cost; all are important factors. +DYLQJ VDLG WKDW , ZRXOG DUJXH pursuing a new WKDW WKH *0¶V S player should boil down to one thing - will this player give us a greater chance of winning games (for his cost) than all other alternatives? Winning teams over the long run do end up making more money for their owners; because winning teams result in more games, more TV revenue, more sales, more fans, etc. So, if the GM focuses on winning games ±the rest will follow.

All GMs have to deal with a limited amount of resources ±they do not have infinite dollars to spend on team payroll. Therefore, iW RQO\ PDNHV VHQVH WR VHDUFK I IRU WKHLU EXFN´ in finding players that will contribute most to team success ±and that means finding and signing those under-valued players that also fit their style of play. These are the players that consistently help their teams win without getting paid their commensurate on-court value. These players are generally deceivingly effective. Clearly, looking at one prediction in the APM statistic can be risky. The great variability will over and under-estimate because of the nature of the statistic. %XW LI D FHUWDLQ SOD\HU LV SXWWLQJ XS YH with the $30 LV SUREDEO\ VKRZLQJ WKLV SOD\HU¶V WUXH on-court worth. These are the players that GMs should seek and try to obtain. The following table displays the 20 most under valued players adjusted plus-minus statistic in the last 5 seasons.

Table 6: Under-Valued Players APM Over the Last 5 Seasons

Name	Position	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012
Jamaal Tinsley	PG	3.01	NA	15.12	NA	NA
Derek Fisher	PG	-0.65	-2.65	9.16	7.07	7.01
Solomon Jones	PF	NA	NA	0.87	14.15	NA
Will Bynum	PG	NA	15.01	NA	NA	NA
Trevor Booker	PF	NA	NA	NA	NA	13.01
Peja Stojakovic	SF	14.14	9.12	9.36	7.75	NA
Matt Bonner	PF	5.86	3.14	9.26	18.39	12.28
Luc Richard Mbah a Moute	PF	NA	0.81	2.86	NA	16.5
Travis Diener	PG	-2.61	9.6	NA	NA	NA
James Harden	SG	NA	NA	-1.2	7.3	14.97
Jordan Crawford	SG	NA	NA	NA	1.62	10.1
John Lucas	PG	NA	NA	NA	NA	10.41
John Wall	PG	NA	NA	NA	NA	14.26
Michael Redd	SG	3.57	12.68	NA	NA	-0.2
Thaddeus Young	SF	14.93	6.25	11.86	15.96	13.33
Jameer Nelson	PG	-2.99	11.57	6.14	12.77	18.9
Jeff Teague	PG	NA	NA	-0.39	2.37	10.97
Stephen Curry	PG	NA	NA	6.43	7.81	11.81
Kevin Ollie	PG	NA	8.89	NA	NA	NA
Vince Carter	SG	12.13	13.91	0.17	0.1	14.05

Legend: Yellow = Proven, Red = Unproven but Young, White = Other

Table 6 shows the APM metrics for the 20 most under valued players in the league. From the table, some players have proven their immense on-court worth because of their consistently large APM values; their 2012-2013 salary does not reflect their true on-court value (players in yellow). A few notable standouts are Thaddeus Young, Matt Bonner, and Peja Stojakovic. These players have consistently had some of the best APM statistics over the last 5 seasons. Obviously, their 2012-2013 season pay is under-compensating their projected 2012-2013 season on-court value. It would be surprising to assume that the APM measure randomly over-rated their play every season; therefore, we will assume that these players will once again be effective in the 2012-2013 season. The players in red are unproven players under the age of 28. They are players general managers should pay close attention to in the years to come and possibly go after.

According to the APM statistic, these are potentially very good, under-valued players. With a few more years we can see if this trend will continue or if the player had one very good year according to the APM because of the inherent variability. The players in white are either on their rookie contract or older, unproven players. * H Q H U D O P D Q D J H U V ¶ V K R X O G V W L O these players, but they are not as truly under-valued as those players in yellow and red.

Table 7: Other Under-Valued Notables, By Position

Name	Position	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012
Beno Udrih	PG	1.48	11.77	13.06	-6.56	16.02
Goran Dragic	PG	NA	8.69	9.03	1.32	10.34
Luke Ridnour	PG	6.63	4.33	4.62	12.44	5.68
Ramon Sessions	PG	NA	7.23	8.43	11.44	9.75
Raymond Felton	PG	5.61	8.63	13.96	12.00	6.05
Arron Afflalo	SG	1.00	9.92	3.32	15.09	6.11
Courtney Lee	SG	NA	11.00	11.10	9.64	6.45
Jamal Crawford	SG	14.13	8.41	10.10	12.69	9.78
Manu Ginobili	SG	13.99	12.53	10.15	15.66	19.10
Richard Hamilton	SG	2.42	8.41	-2.38	13.51	9.91
Chase Budinger	SF	NA	NA	12.26	5.54	NA
Ersan Ilyasova	SF	NA	NA	3.30	6.02	10.90
Kyle Korver	SF	9.61	16.91	-2.46	NA	11.09
Nicolas Batum	SF	NA	7.96	16.62	3.49	3.45
Shane Battier	SF	7.82	5.84	11.70	8.72	3.99
Amir Johnson	PF	16.79	13.64	1.90	4.58	0.21
Chris Wilcox	PF	10.08	4.02	NA	17.38	NA
Chuck Hayes	PF	11.29	0.74	7.21	4.50	12.70
Jared Jeffries	PF	5.87	8.42	4.22	12.71	NA
Paul Millsap	PF	NA	NA	12.12	8.93	14.99
Anderson Varejao	C	6.14	3.25	3.68	8.09	-4.77
Andray Blatche	C	-1.58	5.29	NA	5.25	11.33
Ronny Turiaf	C	5.56	9.75	9.02	NA	NA
Roy Hibbert	C	NA	0.10	3.11	6.65	7.43
Shaquille O'Neal	C	6.87	6.28	5.75	12.24	NA

Legend: Yellow = Most Notable, White = Other Notables

Table 7 displays other under-valued notable players, by position. One name that is V W U L N L Q J L V 6 K D T X L O O H 2 ¶ 1 H D O He is on the list despite the fact he is extremely well known, because although his APM was very good over the last few years he was known to be much less effective over that span. Because of this, a team might be willing to pay him only a few million dollars a year. His APM statistic over the last five years suggests his projected 2012-2013 worth (if he were to play) might be much more than that. The other players do not seem to share common traits. They might be under-valued for many reasons. The only attribute they share is that they are under-appreciated. People do not see the benefit that they provide, and therefore, these players are not paid what their on-court value would suggest.

The players highlighted in yellow are not necessarily the best players. Rather, they are the players who seem to have the biggest deficiencies in 2012-2013 pay for what their APM statistics indicate. Most of these players in yellow are not well known. These players are not in the media limelight. Some play in small markets like Toronto, Milwaukee, and Washington. Not many fans are aware of their names because they are undervalued, and are not televised on ESPN or TNT very often. This causes the player to slip between the cracks of basketball conversation, and therefore, become a sleeper. Teams that can value players like this efficiently might be able to pay them millions less than their on-court production would suggest.

<u>NBA All-Under Valued Team</u>:

Table 8: The NBA All-Under Valued Team

Name	Position
Starters	
Jameer Nelson	PG
Courtney Lee	SG
Thaddeus Young	SF
Matt Bonner	PF
Paul Millsap	PF
Bench	
Derek Fisher	PG
Ramon Sessions	PG
Vince Carter	SG
Ersan Ilyasova	SF
Peja Stojakovic	SF
Amir Johnson	PF
Andray Blatche	C

These are the most under-valued players, from my perspective. Some may be under-valued because of their age, others possibly because they play in small markets, or maybe because they are unathletic (by 1 % $ ¶ V   V W D Q G D U G V Regardless of the reasons why these players are not getting paid what they are worth for the 2012-2013 season; they seem to be truly under-valued. There are other players that could have easily made this list. This group is comprised by position, like a normal NBA team; that is why some point guards with great adjusted plus-minus statistics are not on the under-valued roster.

<u>Future Research</u>:

There are many more ways that one can analyze the APM statistic, and how it affects the game. The APM between the interactions of players and the prediction of the APM statistic based on traditonal statistics (Statistical Plus Minus or SPM) are two gaining the most traction. Information on both the statistical plus minus and the interactions between players can be studied much more in-depth online. The advanced basketball statistics websites basketballvalue.com and

NBAstuffer.com are where I would first look to start investigating these two measures. Also, interesting advanced metrics, some involving the APM, can be found on these websites.

Earlier, I explained that all-stars can be ineffective playing together despite great individual talent because of uncomplementary styles of play; and average NBA talent can become much more effective with great chemistry. Because of the dynamic that NBA teams are not just the sum of their parts, NBA GM¶V DUH EHFRPLQJ PRUH DQG PRUH NH in studying the interactions between players. This phenomenon has worked its way into the APM statistic, where NBA statisticians compare a group of 2 players on the court at the same time, to other 2 player combinations (or 3,4, and 5-way player interactions). A GM can use the APM interaction statistics intelligently to fit a potential new player to his existing team. Had the Lakers had great knowledge on the effects of their superstars chemistry before offering them a contract, they would probably not have signed the combination of Dwight Howard, Kobe Bryant, and Steve Nash. Also, other interesting, highly-debated player combinations like Stephen Curry and Monta Ellis could be examined. Are teams with two small guards less effective because it becomes more difficult to defend a stronger guard? All of these kinds of questions are difficult to answer. But, the APM can give an idea of the effectiveness of certain kinds of player combinations, and possibly player personalities as well. However, there are difficulties in calculating and assessing the APM interactions between players. First and foremost, the dataset would be much more difficult to manage. Every two way combination of players means there would be about 10 choose 2, or 45 variables per team, instead of 10. Higher order 3 to 5-way interactions would produce even more variables. Also, the APM results would be even more variable due to the fact that these combinations are even less independent than the individual APM statistics. The players that consistently play with similar teammates will cause even more co-linearity.

Measuring traditional statistics to calculate the APM, known as the Statistical Plus/Minus (SPM) reduces the co-linearity between players, and is easier to comprehend. Traditional statistics are nice because they are tangible. When using traditional statistics to predict the APM, manager¶V FDQ DVVHVV ZKLFK IDFWRUV PRVW HIIHFW WKH PDU performance. Measures like shooting percentage and PPG can be measured to compare the effect of both on the APM. Questions such as, are high percentage shooters more valuable than NBA scorers that do not shoot a higher percentage, can be assessed. These kinds of questions are important to NBA franchises. Understanding the possible answers to these questions have true value. The other reason this measure is gaining traction is due to its reduced co-linearity. Measuring via traditional statistics helps reduce co-linearity, and thus reduces year to year variability. The main drawback to this type of analysis is that it does not account for the numerous, unrecorded things that may effect the outcome of the game dramatically. Playing proper help-side defense, setting good screens, or getting a hand iQ D VKRRWHU¶V IDFH are all not measured. Obviously, this is an oversight. But, it is good to have a general knowledge of what affects the APM statistic that can be easily recorded and studied.

<u>Conclusion</u>:

The APM statistic is not an exhaustive approach to finding NBA talent. Many quantitative and qualitative factors, in addition to the APM, should be investigated before acquiring a player. However, the APM needs to be addressed and carefully examined along with these other factors. In any sport, basketball included, it is natural (and human) for experts to just believe what they see. However, much of what NBA scouts and upper-management see they subconsciously over or under-value, and that can lead to errant judgments. A comprehensive in-depth analysis must be conducted both statistically (with traditional and advanced measures) as well as visually. Because the APM is so variable, a basketball scout should not just believe a player will have an enduring APM of +10 merely because of one season. But, if a player continually is achieving APM numbers like +10, then it would be wise to investigate the reasons why. The APM statistic is growing quickly in popularity as a quantitative tool, and will continue to do so, because it measures better than anything else presently something that is critically important; how much does a player actually contribute to team success? $UPHG ZLWK WKLV ³D in his quiver, the smart GM can make more informed personnel decisions for his team by more effienctly allocating his limited resources. This in turn will maximize his chances of building a winning team over time.

References

³%DVNHWEDOO9DOXH FRP 'DWD)LOHV´ http://basketballvalue.com/downloads.php.

³&RXQW WKH %DVNHW $GYDQFHG 6WDWV IRU %DVNHWEDOO

http://www.countthebasket.com/blog/2008/06/03/offensive-and-defensive-adjusted-plus-minus/

/HZLV 0LFKDHO 6WDWV ³7KH $OO-6WDU´ The New York Times Magazine 15 Feb. 2009.

0DF'RQDOG %ULDQ -Based Adjusted Plus-Minus ³$ 5HJUHVVLRQ 6WDWLVWLF IRU 1+/ 3

Journal of Quantitative Analysis in Sports: Vol. 7: Iss. 3, Article 4.

³1%$ 3OD\HU 6DODULHV 2007-´ http://espn.go.com/nba/salaries/_/year/2008/

³1%$ 6WDQGLQJV´ http://espn.go.com/nba/standings/_/group/1

Oliver, Dean. Basketball on Paper, Rules and Tools for Performance Analysis. Potomac Books, Potomac Books, Inc. 2004. Washington, DC.

Appendix:

Season Code:

```
path <- "C:/Users/Samsung/Documents/"

#reading in main dataset#
fiveonfive0708 <- read.csv(eval(paste(path,"/Senior Project/NBA All 10/2007-
2008/Excel/fiveonfive0708.csv",sep="")))

#making points per home and away possession#
fiveonfive0708$PointsPerHomePoss =
fiveonfive0708$PointsScoredHome/fiveonfive0708$PossessionsHome
fiveonfive0708$PointsPerAwayPoss =
fiveonfive0708$PointsScoredAway/fiveonfive0708$PossessionsAway

#how many games are in dataset#
levels(fiveonfive0708$GameID)

#reading in the player stats file to only include players with more than 500 mins#
playerstats0708 <- read.csv(eval(paste(path,"/Senior Project/NBA All 10/2007-
2008/Excel/playerstats0708.csv",sep="")))

#taking out players with less than 500 mins#
ind <- with(playerstats0708, SimpleMin < 500)
playerstats0708 <- playerstats0708[!ind, ]

#making offensive and defensive indicator variables#
playerID0708 <- c(playerstats0708[,1])
playerID0708 <- t(playerID0708)
playerID0708Off <- paste("O",playerID0708,sep="_")
playerID0708Def <- paste("D",playerID0708,sep="_")

#creating matrix of number of observations by the number of indicator vars (offense and
defense)#
emptyplayeridmat0708 <- matrix(data=c(playerID0708), nrow=2*dim(playerID0708)[2],
ncol=dim(fiveonfive0708)[1])
emptyplayeridmat0708 <- t(emptyplayeridmat0708)
colnames(emptyplayeridmat0708) <- c(playerID0708Off,playerID0708Def)

#creating home and away rows...now data set is 2 X 2 times bigger#
fiveonfive0708Home <- cbind(fiveonfive0708, emptyplayeridmat0708)
fiveonfive0708Away <- cbind(fiveonfive0708, emptyplayeridmat0708)

#taking out observations with no home possessions#
ind <- with(fiveonfive0708Home, PossessionsHome==0)
```

```
fiveonfive0708Home <- fiveonfive0708Home[!ind, ]
#creating a common possessions and points per possessions for home and away so I can run glm
model later#
fiveonfive0708Home$Possessions <- fiveonfive0708Home$PossessionsHome
fiveonfive0708Home$PointsPerPoss <- fiveonfive0708Home$PointsPerHomePoss
# home points per 48 minutes for each stint, average points per possession of stint X average
number of possessions per game #
fiveonfive0708Home$AvePointTotal <-
fiveonfive0708Home$PointsPerHomePoss*(sum(fiveonfive0708Home$PossessionsHome)/1230
)
fiveonfive0708Home$PointsScored <- fiveonfive0708Home$PointsScoredHome


#offensive home indicator#
#finding the number of dimensions to use for the offensive indicator variables#
dim(fiveonfive0708)[2]+1
dim(playerID0708)[2]+dim(fiveonfive0708)[2]

for (i in 50:373){
#using Home players because they are on offensive for home possession rows#
fiveonfive0708Home[,i] <-
ifelse(fiveonfive0708Home[,i]==fiveonfive0708Home$HomePlayer1ID |

fiveonfive0708Home[,i]==fiveonfive0708Home$HomePlayer2ID |

fiveonfive0708Home[,i]==fiveonfive0708Home$HomePlayer3ID |

fiveonfive0708Home[,i]==fiveonfive0708Home$HomePlayer4ID |

fiveonfive0708Home[,i]==fiveonfive0708Home$HomePlayer5ID,1,0)
}

#defensive away indicator#
#finding the number of dimensions for defensive indicator variables#
dim(playerID0708)[2]+dim(fiveonfive0708)[2]+1
dim(fiveonfive0708Home)[2]-4

for (i in 374:697){
#using Away players because they are on defense for home possessions#
fiveonfive0708Home[,i] <-
ifelse(fiveonfive0708Home[,i]==fiveonfive0708Home$AwayPlayer1ID |

fiveonfive0708Home[,i]==fiveonfive0708Home$AwayPlayer2ID |

fiveonfive0708Home[,i]==fiveonfive0708Home$AwayPlayer3ID |
```

```
fiveonfive0708Home[,i]==fiveonfive0708Home$AwayPlayer4ID |

fiveonfive0708Home[,i]==fiveonfive0708Home$AwayPlayer5ID,1,0)
}

#Away rows#

#did the same thing for the home rows as the away rows#
ind <- with(fiveonfive0708Away, PossessionsAway==0)
fiveonfive0708Away <- fiveonfive0708Away[!ind, ]

fiveonfive0708Away$Possessions <- fiveonfive0708Away$PossessionsAway
fiveonfive0708Away$PointsPerPoss <- fiveonfive0708Away$PointsPerAwayPoss

#away points per 48 minutes for each stint! #
fiveonfive0708Away$AvePointTotal <-
fiveonfive0708Away$PointsPerAwayPoss*(sum(fiveonfive0708Away$PossessionsAway)/1230)
fiveonfive0708Away$PointsScored <- fiveonfive0708Away$PointsScoredAway

#offensive away indicator#

for (i in 50:373){
#using away players b/c they are on offense for away possessions#
fiveonfive0708Away[,i] <-
ifelse(fiveonfive0708Away[,i]==fiveonfive0708Away$AwayPlayer1ID |

fiveonfive0708Away[,i]==fiveonfive0708Away$AwayPlayer2ID |

fiveonfive0708Away[,i]==fiveonfive0708Away$AwayPlayer3ID |

fiveonfive0708Away[,i]==fiveonfive0708Away$AwayPlayer4ID |

fiveonfive0708Away[,i]==fiveonfive0708Away$AwayPlayer5ID,1,0)
}

#defensive home indicator#

for (i in 374:697){
#using home players b/c they are on defense for away possessions#
fiveonfive0708Away[,i] <-
ifelse(fiveonfive0708Away[,i]==fiveonfive0708Away$HomePlayer1ID |

fiveonfive0708Away[,i]==fiveonfive0708Away$HomePlayer2ID |

fiveonfive0708Away[,i]==fiveonfive0708Away$HomePlayer3ID |
```

```
fiveonfive0708Away[,i]==fiveonfive0708Away$HomePlayer4ID |

fiveonfive0708Away[,i]==fiveonfive0708Away$HomePlayer5ID,1,0)
}

#combining home and away rows to make one large dataset#
fiveonfive0708New <- rbind(fiveonfive0708Home,fiveonfive0708Away)

#finding minutes left and point differential, then taking out observations in blowout games#
#teams try and just run out the clock, not true indication of player production#
fiveonfive0708New$MinsLeft <- substr(fiveonfive0708New$StartTime,3,4)
fiveonfive0708New$PointDifferential <- abs(fiveonfive0708New$StartScoreHome-
fiveonfive0708New$StartScoreAway)
ind <- with(fiveonfive0708New, (MinsLeft=="12" | MinsLeft=="11" | MinsLeft=="10"
                        | MinsLeft=="09" | MinsLeft=="08") & PointDifferential>=30)
fiveonfive0708New <- fiveonfive0708New[!ind, ]
ind <- with(fiveonfive0708New, (MinsLeft=="07" | MinsLeft=="06" | MinsLeft=="05"
                        | MinsLeft=="04" |MinsLeft=="03" | MinsLeft=="02") &
PointDifferential>=20)
fiveonfive0708New <- fiveonfive0708New[!ind, ]
ind <- with(fiveonfive0708New, MinsLeft=="01" & PointDifferential>=15)
fiveonfive0708New <- fiveonfive0708New[!ind, ]
ind <- with(fiveonfive0708New, MinsLeft=="00" & PointDifferential>=10)
fiveonfive0708New <- fiveonfive0708New[!ind, ]

#finding the players that played over 500 minutes indicators#
playerID0708New <- c(playerID0708Off, playerID0708Def)

f <-    for (i in 1:658){
                z <- paste(playerID0708New, "+")
                print(z)
                }
m <- noquote(print(z))
m

#once you get m with no quotes copy and paste#

#run the glm#
fit.adj0708 <- glm(data= fiveonfive0708New, weights=Possessions, formula=AvePointTotal ~

        O_1 +  O_2 +  O_3 +  O_4 +  O_5 +  O_6 +  O_7 +  O_8 +  O_10 +
        O_12 + O_14 + O_15 + O_16 + O_17 + O_19 + O_20 + O_21 + O_22 +
        O_23 + O_24 + O_25 + O_26 + O_28 + O_29 + O_30 + O_31 + O_32 +
        O_33 + O_35 + O_36 + O_37 + O_39 + O_40 + O_42 + O_45 + O_46 +
        O_47 + O_48 + O_49 + O_52 + O_54 + O_55 + O_56 + O_59 + O_60 +
```

O_62 + O_63 + O_65 + O_67 + O_68 + O_69 + O_70 + O_71 + O_75 +
O_76 + O_77 + O_78 + O_79 + O_80 + O_83 + O_85 + O_86 + O_87 +
O_88 + O_89 + O_90 + O_91 + O_92 + O_93 + O_94 + O_95 + O_96 +
O_97 + O_98 + O_101 + O_103 + O_109 + O_110 + O_111 + O_112 + O_113 +
O_114 + O_116 + O_117 + O_118 + O_119 + O_121 + O_122 + O_123 + O_125 +
O_126 + O_128 + O_130 + O_132 + O_133 + O_136 + O_137 + O_138 + O_139 +
O_140 + O_141 + O_143 + O_145 + O_147 + O_148 + O_149 + O_151 + O_152 +
O_154 + O_155 + O_158 + O_159 + O_163 + O_164 + O_165 + O_166 + O_168 +
O_169 + O_170 + O_171 + O_172 + O_173 + O_174 + O_176 + O_180 + O_181 +
O_185 + O_186 + O_187 + O_189 + O_190 + O_191 + O_193 + O_194 + O_196 +
O_197 + O_198 + O_200 + O_201 + O_202 + O_203 + O_204 + O_205 + O_209 +
O_210 + O_211 + O_214 + O_215 + O_218 + O_220 + O_221 + O_222 + O_223 +
O_224 + O_226 + O_231 + O_235 + O_236 + O_237 + O_239 + O_240 + O_241 +
O_243 + O_244 + O_245 + O_246 + O_247 + O_249 + O_252 + O_253 + O_254 +
O_257 + O_258 + O_260 + O_261 + O_264 + O_265 + O_270 + O_271 + O_272 +
O_273 + O_274 + O_275 + O_276 + O_277 + O_278 + O_279 + O_280 + O_281 +
O_283 + O_284 + O_285 + O_287 + O_288 + O_289 + O_292 + O_293 + O_295 +
O_296 + O_298 + O_302 + O_305 + O_306 + O_308 + O_311 + O_312 + O_315 +
O_316 + O_318 + O_319 + O_321 + O_322 + O_324 + O_326 + O_327 + O_328 +
O_329 + O_330 + O_331 + O_337 + O_344 + O_354 + O_355 + O_356 + O_357 +
O_362 + O_366 + O_367 + O_372 + O_377 + O_383 + O_389 + O_390 + O_401 +
O_411 + O_413 + O_416 + O_417 + O_419 + O_426 + O_430 + O_444 + O_458 +
O_459 + O_468 + O_470 + O_486 + O_488 + O_489 + O_492 + O_501 + O_505 +
O_508 + O_510 + O_516 + O_518 + O_537 + O_541 + O_544 + O_551 + O_552 +
O_553 + O_557 + O_558 + O_559 + O_561 + O_565 + O_573 + O_575 + O_577 +
O_578 + O_579 + O_581 + O_588 + O_589 + O_592 + O_593 + O_594 + O_597 +
O_598 + O_601 + O_602 + O_605 + O_606 + O_607 + O_608 + O_612 + O_614 +
O_617 + O_622 + O_629 + O_640 + O_643 + O_665 + O_666 + O_667 + O_674 +
O_675 + O_678 + O_679 + O_681 + O_682 + O_685 + O_686 + O_689 + O_691 +
O_693 + O_695 + O_696 + O_699 + O_701 + O_703 + O_705 + O_707 + O_708 +
O_709 + O_711 + O_712 + O_713 + O_714 + O_715 + O_716 + O_719 + O_722 +
D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7 + D_8 + D_10 +
D_12 + D_14 + D_15 + D_16 + D_17 + D_19 + D_20 + D_21 + D_22 +
D_23 + D_24 + D_25 + D_26 + D_28 + D_29 + D_30 + D_31 + D_32 +
D_33 + D_35 + D_36 + D_37 + D_39 + D_40 + D_42 + D_45 + D_46 +
D_47 + D_48 + D_49 + D_52 + D_54 + D_55 + D_56 + D_59 + D_60 +
D_62 + D_63 + D_65 + D_67 + D_68 + D_69 + D_70 + D_71 + D_75 +
D_76 + D_77 + D_78 + D_79 + D_80 + D_83 + D_85 + D_86 + D_87 +
D_88 + D_89 + D_90 + D_91 + D_92 + D_93 + D_94 + D_95 + D_96 +
D_97 + D_98 + D_101 + D_103 + D_109 + D_110 + D_111 + D_112 + D_113 +
D_114 + D_116 + D_117 + D_118 + D_119 + D_121 + D_122 + D_123 + D_125 +
D_126 + D_128 + D_130 + D_132 + D_133 + D_136 + D_137 + D_138 + D_139 +
D_140 + D_141 + D_143 + D_145 + D_147 + D_148 + D_149 + D_151 + D_152 +
D_154 + D_155 + D_158 + D_159 + D_163 + D_164 + D_165 + D_166 + D_168 +
D_169 + D_170 + D_171 + D_172 + D_173 + D_174 + D_176 + D_180 + D_181 +
D_185 + D_186 + D_187 + D_189 + D_190 + D_191 + D_193 + D_194 + D_196 +

```
        D_197 + D_198 + D_200 + D_201 + D_202 + D_203 + D_204 + D_205 + D_209 +
        D_210 + D_211 + D_214 + D_215 + D_218 + D_220 + D_221 + D_222 + D_223 +
        D_224 + D_226 + D_231 + D_235 + D_236 + D_237 + D_239 + D_240 + D_241 +
        D_243 + D_244 + D_245 + D_246 + D_247 + D_249 + D_252 + D_253 + D_254 +
        D_257 + D_258 + D_260 + D_261 + D_264 + D_265 + D_270 + D_271 + D_272 +
        D_273 + D_274 + D_275 + D_276 + D_277 + D_278 + D_279 + D_280 + D_281 +
        D_283 + D_284 + D_285 + D_287 + D_288 + D_289 + D_292 + D_293 + D_295 +
        D_296 + D_298 + D_302 + D_305 + D_306 + D_308 + D_311 + D_312 + D_315 +
        D_316 + D_318 + D_319 + D_321 + D_322 + D_324 + D_326 + D_327 + D_328 +
        D_329 + D_330 + D_331 + D_337 + D_344 + D_354 + D_355 + D_356 + D_357 +
        D_362 + D_366 + D_367 + D_372 + D_377 + D_383 + D_389 + D_390 + D_401 +
        D_411 + D_413 + D_416 + D_417 + D_419 + D_426 + D_430 + D_444 + D_458 +
        D_459 + D_468 + D_470 + D_486 + D_488 + D_489 + D_492 + D_501 + D_505 +
        D_508 + D_510 + D_516 + D_518 + D_537 + D_541 + D_544 + D_551 + D_552 +
        D_553 + D_557 + D_558 + D_559 + D_561 + D_565 + D_573 + D_575 + D_577 +
        D_578 + D_579 + D_581 + D_588 + D_589 + D_592 + D_593 + D_594 + D_597 +
        D_598 + D_601 + D_602 + D_605 + D_606 + D_607 + D_608 + D_612 + D_614 +
        D_617 + D_622 + D_629 + D_640 + D_643 + D_665 + D_666 + D_667 + D_674 +
        D_675 + D_678 + D_679 + D_681 + D_682 + D_685 + D_686 + D_689 + D_691 +
        D_693 + D_695 + D_696 + D_699 + D_701 + D_703 + D_705 + D_707 + D_708 +
        D_709 + D_711 + D_712 + D_713 + D_714 + D_715 + D_716 + D_719 + D_722

)

#pulling out adjusted offensive and defensive numbers as well as player id#
adjusted0708 <- cbind(playerID0708[1:324],coefficients(fit.adj0708)[2:325],-
coefficients(fit.adj0708)[326:649])
colnames(adjusted0708) <- c("PlayerID","OffensiveRtg","DefensiveRtg")

#making the names first and last with no comma inbetween#
playerstats0708$PlayerTrueName <- as.character(playerstats0708$PlayerTrueName)
c <- strsplit(playerstats0708$PlayerTrueName,split=", ")
Names <- do.call(rbind, c)
playerstats0708$Name <- paste(Names[,2],Names[,1])

#pulling out new name variable and player id#
dim(playerstats0708)[2]
IDMat <- playerstats0708[,c(1,40)]

#changing to data frame, then merging#
IDMat <- as.data.frame(IDMat)
adjusted0708 <- as.data.frame(adjusted0708)
adj0708 <- merge(x=IDMat, y=adjusted0708, by="PlayerID")
adj0708$OverallRtg <- adj0708$OffensiveRtg + adj0708$DefensiveRtg

#reading in player salaries#
```

```
PlayerSalaries <- read.csv(eval(paste(path,"/Senior Project/NBA All 10/2007-
2008/Excel/PlayerSalaries.csv",sep="")))

#changing the name column to get ready for merge#
class(PlayerSalaries$NAME)
PlayerSalaries$Name <- as.character(PlayerSalaries$NAME)
d <- strsplit(PlayerSalaries$Name,split=",")
Names <- do.call(rbind, d)
PlayerSalaries$Name <- Names[,1]
PlayerSalaries$Position <- substr(Names[,2],2,3)

#only taking out necessary columns of apm and players salaries data#
PlayerSalaries <- PlayerSalaries[,c(5,6,3,4)]
adj <- adj0708[,c(2:5)]

#changing names in order to merge#
PlayerSalaries$Name <- ifelse(PlayerSalaries$Name=="Amar'e Stoudemire","Amare
Stoudemire",PlayerSalaries$Name)
PlayerSalaries$Name <- ifelse(PlayerSalaries$Name=="Nen?", "Nene Hilario",
PlayerSalaries$Name)
adj$Name <- ifelse(adj$Name=="Ronald (Flip) Murray", "Ronald Murray", adj$Name)
adj$Name <- ifelse(adj$Name=="Luc Mbah a Moute", "Luc Richard Mbah a Moute",
adj$Name)
adj$Name <- ifelse(adj$Name=="Ron Artest", "Metta World Peace", adj$Name)

#merging the two datasets#
Final0708 <- merge(x=PlayerSalaries, y=adj, by="Name")

#changing variable names to make the data look nice later on#
Final0708$DefensiveRtg0708 <- Final0708$DefensiveRtg
Final0708$OffensiveRtg0708 <- Final0708$OffensiveRtg
Final0708$OverallRtg0708 <- Final0708$OverallRtg
Final0708$SALARY <- as.numeric(gsub('\\$|,', '', Final0708$SALARY))
Final0708$Salary0708 <- Final0708$SALARY
Final0708$Team0708 <- Final0708$TEAM
Final0708 <- Final0708[,c(1,2,12,11,8,9,10)]

#saving to R workfile#
save(Final0708, file=paste(path,"Senior Project/NBA All 10/2007-
2008/Final0708.RData",sep=""))
```

Final Code:

```
path <- "C:/Users/Samsung/Documents/"

#loading all the data from the 5 seasons#
load(paste(path, "Senior Project/NBA All 10/2007-2008/Final0708.RData", sep=""))
load(paste(path, "Senior Project/NBA All 10/2008-2009/Final0809.RData", sep=""))
load(paste(path, "Senior Project/NBA All 10/2009-2010/Final0910.RData", sep=""))
load(paste(path, "Senior Project/NBA All 10/2010-2011/Final1011.RData", sep=""))
load(paste(path, "Senior Project/NBA All 10/2011-2012/Final1112.RData", sep=""))

#merging the data together to make one large dataset#
Final0709 <- merge(x=Final0708, y=Final0809, by=c("Name","Position"), all=T)
Final0710 <- merge(x=Final0709, y=Final0910, by=c("Name","Position"), all=T)
Final0711 <- merge(x=Final0710, y=Final1011, by=c("Name","Position"), all=T)
Final0712All <- merge(x=Final0711, y=Final1112, by=c("Name","Position"), all=T)

#taking out some unusual observations, players with no dpm and only one year#
ind <- with(Final0712All, rowSums(is.na(Final0712All))>21)
Final0712All <- Final0712All[!ind, ]


ind <- with(Final0712All, rowSums(is.na(Final0712All))>18)
#players that played more than one season#
Final0712 <- Final0712All[!ind, ]
#players that played only one season#
Final0712OneObs <- Final0712All[ind,]

#only taking out the overall apm for 5 seasons#
Overall0712All <- Final0712All[,c(1,2,7,12,17,22,27)]
#taking out the last year to run time series analysis#
Overall0711All <- Overall0712All[,c(1:6)]

ind <- with(Overall0711All, rowSums(is.na(Overall0711All))>=3)
#players with more than one year over 500 mins from 07-11#
Overall0711 <- Overall0711All[!ind, ]
#players with only one year over 500 mins from 07-11#
Overall0711OneObs <- Overall0711All[ind,]
ind <- with(Overall0712All, is.na(Overall0712All$OverallRtg1112))
#taking out players that don't have a rating in 11-12 year#
Overall0712New <- Overall0712All[!ind,]
#merging players who played in 11-12 and have one obs from 07-11#
Overall0712OneObs <- merge(x=Overall0711OneObs, y=Overall0712New[,c(1,2,7)],
by=c("Name", "Position"))
#changing to matrix and numeric#
as.matrix(Overall0712OneObs)
```

```
as.numeric(Overall0712OneObs$OverallRtg0708)
as.numeric(Overall0712OneObs$OverallRtg0809)
as.numeric(Overall0712OneObs$OverallRtg0910)
as.numeric(Overall0712OneObs$OverallRtg1011)
as.numeric(Overall0712OneObs$OverallRtg1112)
#setting the NA's equal to 0 so I can compare only obs in 07-11 to 11-12 season#
Overall0712OneObs[is.na(Overall0712OneObs)] <- 0

#puts the single recorded value in one column, isn't effected by 0#
Overall0712OneObs$OneRecordedValue <-
Overall0712OneObs$OverallRtg0708+Overall0712OneObs$OverallRtg0809+
        Overall0712OneObs$OverallRtg0910+Overall0712OneObs$OverallRtg1011

Overall0712OneObs$Diff <- Overall0712OneObs$OverallRtg1112 -
Overall0712OneObs$OneRecordedValue
mean(Overall0712OneObs$Diff)
# there isn't much of a difference, -.4333, I'll assume this is random#

#taking out dpm over the last 5 seasons#
Defense0712All <- Final0712All[,c(1,2,5,10,15,20,25)]

ind <- with(Defense0712All, rowSums(is.na(Overall0712All))>=4)
#more than one observation#
Defense0712 <- Defense0712All[!ind, ]
#one observation#
Defense0712One <- Defense0712All[ind, ]


pred.def <- NULL
#predicting defense for next year using basic exponential smoothing#
#for players with more than one season#
for (i in 1:362){

        series <- t(Defense0712[i,c(3:7)])
        series <- series[!is.na(series)]
        basicexpsmooth <- HoltWinters(series, beta=FALSE, gamma=FALSE)
        pred.expsmooth <- predict(basicexpsmooth, n.ahead=1, prediction.interval=TRUE)
        pred.def[i] <- pred.expsmooth[1]

}
#predicting players with only one season to just remain the same#
for (i in 363:468){
        series <- t(Defense0712One[i-362,c(3:7)])
        series <- series[!is.na(series)]
        pred.def[i] <- series + 0
}
```

```
Offense0712All <- Final0712All[,c(1,2,6,11,16,21,26)]

#the same method as defense right above#
ind <- with(Offense0712All, rowSums(is.na(Offense0712All))>=4)
Offense0712 <- Offense0712All[!ind, ]
Offense0712One <- Offense0712All[ind, ]


pred.off <- NULL

for (i in 1:362){

	series <- t(Offense0712[i,c(3:7)])
	series <- series[!is.na(series)]
	basicexpsmooth <- HoltWinters(series, beta=FALSE, gamma=FALSE)
	pred.expsmooth <- predict(basicexpsmooth, n.ahead=1, prediction.interval=TRUE)
	pred.off[i] <- pred.expsmooth[1]


}

for (i in 363:468){
	series <- t(Offense0712One[i-362,c(3:7)])
	series <- series[!is.na(series)]
	pred.off[i] <- series + 0
}

#combining one obs data with 2 or more obs to make master data#
Name <- c(Final0712[,1],Final0712OneObs[,1])
Position <- c(Final0712[,2],Final0712OneObs[,2])

#predicted apm for 2012-2013 season#
predoverallfit <- pred.off+pred.def

#apm data set#
adj <- data.frame(Name, Position,
	predoverallfit, pred.off, pred.def)

#changing data so to merge with player salary#
adj$Name <- as.character(adj$Name)
adj$Position <- as.character(adj$Position)

#player salary data, from nba.go.com#
```

```
PlayerSalaries <- read.csv(file=paste(path, "Senior Project/NBA All 10/2012-
2013/Excel/PlayerSalaries.csv", sep=""))
class(PlayerSalaries$NAME)
PlayerSalaries$Name <- as.character(PlayerSalaries$NAME)

#changing columns to be able to merge with adjusted data#
d <- strsplit(PlayerSalaries$Name,split=",")
d
Names <- do.call(rbind, d)
PlayerSalaries$Name <- Names[,1]
PlayerSalaries$Position <- substr(Names[,2],2,3)

#changing a few players name for the merge step
PlayerSalaries$Name <- ifelse(PlayerSalaries$Name=="Amar'e Stoudemire","Amare
Stoudemire",PlayerSalaries$Name)
PlayerSalaries$Name <- ifelse(PlayerSalaries$Name=="Nen?", "Nene Hilario",
PlayerSalaries$Name)
adj$Name <- ifelse(adj$Name=="Ronald (Flip) Murray", "Ronald Murray", adj$Name)
adj$Name <- ifelse(adj$Name=="Luc Mbah a Moute", "Luc Richard Mbah a Moute",
adj$Name)
adj$Name <- ifelse(adj$Name=="Ron Artest", "Metta World Peace", adj$Name)

#taking out only necessary data#
PlayerSalaries <- PlayerSalaries[,c(5,6,3,4)]

#merge adjusted and player salaries#
newmat <- merge(x=PlayerSalaries, y=adj, by=c("Name","Position"))

#change salary variable to numeric to analyze#
newmat$SALARY <- as.character(newmat$SALARY)
newmat$SALARY <- as.numeric(gsub('\\$|,', '', newmat$SALARY))

#changing colnames to look better#
colnames(newmat) <- c("Name", "Position", "Team", "Salary",
        "Predicted_APM", "Predicted_OPM", "Predicted_DPM")

#rounding to look better in a table#
newmat$Salary <- lapply(newmat$Salary,round,2)
newmat$Salary <- as.numeric(newmat$Salary)
newmat$Predicted_APM <- lapply(newmat$Predicted_APM,round,2)
newmat$Predicted_APM <- as.numeric(newmat$Predicted_APM)
newmat$Predicted_OPM <- lapply(newmat$Predicted_OPM,round,2)
newmat$Predicted_OPM <- as.numeric(newmat$Predicted_OPM)
newmat$Predicted_DPM <- lapply(newmat$Predicted_DPM,round,2)
newmat$Predicted_DPM <- as.numeric(newmat$Predicted_DPM)
```

```
#using rank method to analyze how much a player should be paid#
newermat <- newmat[order(newmat$Salary),]
salaryrank <- cbind(c(1:dim(newermat)[1]),newermat$Salary)
newermat <- newermat[order(newermat$Predicted_APM),]
newestmat <- data.frame(newermat, salaryrank[,2])

#making the rank method data look organized#
colnames(newestmat)[8] <- "Salary_APM"
newestmat$Over_Or_Under_Paid <- newestmat$Salary_APM-newestmat$Salary
truevaluemat <- newestmat[,c(1,2,3,4,9,5,6,7)]
#truevaluemat[,c(1,2,3,4,5)][order(truevaluemat$Team, truevaluemat$Over_Or_Under_Paid),]#

#top 20 players according to 2012-2013 apm, and then exporting to csv#
topPlayers <- truevaluemat[,c(1,2,6)][order(truevaluemat$Predicted_APM, decreasing = T),]
Rank <- c(1:334)
topPlayers <- cbind(Rank, topPlayers)
topPlayers <- topPlayers[c(1:20),]
write.csv(topPlayers, file = paste(path, "Senior Project/NBA All
10/Final/TopPlayers.csv",sep=""))

#top 20 players according to 2012-2013 opm, and then exporting to csv#
topOffensive <- truevaluemat[,c(1,2,7)][order(truevaluemat$Predicted_OPM, decreasing = T),]
topOffensive <- cbind(Rank, topOffensive)
topOffensive <- topOffensive[c(1:20),]
write.csv(topOffensive, file = paste(path, "Senior Project/NBA All
10/Final/TopOffensive.csv",sep=""))

#top 20 players according to 2012-2013 dpm, and then exporting to csv#
topDefensive <- truevaluemat[,c(1,2,8)][order(truevaluemat$Predicted_DPM, decreasing = T),]
topDefensive <- cbind(Rank, topDefensive)
topDefensive <- topDefensive[c(1:20),]
write.csv(topDefensive, file = paste(path, "Senior Project/NBA All
10/Final/TopDefensive.csv",sep=""))

#top 20 under-valued players according to rank method, then exporting to csv
topValue <- truevaluemat[,c(1,2,4,5)][order(truevaluemat$Over_Or_Under_Paid, decreasing =
T),]
topValue <- cbind(Rank, topValue)
#making numbers have commas#
topValue$Over_Or_Under_Paid <-
prettyNum(topValue$Over_Or_Under_Paid,big.mark=",",scientific=F)
topValue$Salary <- prettyNum(topValue$Salary,big.mark=",",scientific=F)
topValue <- topValue[c(1:20),]
write.csv(topValue, file = paste(path, "Senior Project/NBA All 10/Final/Under-
Valued.csv",sep=""))
```

```
#top 20 over-paid players in the nba, and then exporting to csv#
truevaluemat$OverPaid <- -(truevaluemat$Over_Or_Under_Paid)
bottomValue <- truevaluemat[,c(1,2,4,9)][order(truevaluemat$OverPaid, decreasing=T),]
bottomValue <- cbind(Rank, bottomValue)
#making numbers have commas#
bottomValue$OverPaid <- prettyNum(bottomValue$OverPaid,big.mark=",",scientific=F)
bottomValue$Salary <- prettyNum(bottomValue$Salary,big.mark=",",scientific=F)
bottomValue <- bottomValue[c(1:20),]
write.csv(bottomValue, file = paste(path, "Senior Project/NBA All 10/Final/Over-
Valued.csv",sep=""))

#under-valued player apm over the last 5 years, then exporting#
OtherRank <- c(1:468)
Overall0712All <- cbind(OtherRank, Overall0712All)
UnderValuedSeasons <-
Overall0712All[c(190,125,414,459,443,353,314,289,439,194,240,234,236,323,429,
                 193,220,417,264,453),]
write.csv(UnderValuedSeasons, file = paste(path, "Senior Project/NBA All 10/Final/Under-
ValuedConsistency.csv",sep=""))

#other under-valued players I picked out, Other Notables#
OtherNotables <- Overall0712All[c(43,170,291,362,370,37,97,191,296,374,83,158,275,342,407,
                 15,92,94,201,351,16,17,387,388,409),]
write.csv(OtherNotables, file = paste(path, "Senior Project/NBA All
10/Final/OtherNotables.csv",sep=""))

#NBA all-under valued team#
AllUnderValue <- Overall0712All[c(193,97,429,314,351,125,362,453,158,353,15,17),]
write.csv(AllUnderValue, file = paste(path, "Senior Project/NBA All
10/Final/NBAAllUnderValued.csv",sep=""))
```