Big 2's and Big 3's: Analyzing How a Team's Best Players Complement Each Other

Robert Ayer
MBA 2011
Massachusetts Institute of Technology
Cambridge, MA
Email: robertayer@gmail.com

Abstract

One of the most important aspects of team construction is identifying and acquiring the most talented and productive players on your team, the players on whom a team's fortunes most rely. Teams must decide which player-types, when combined, yield the best fit. As an example, suppose there is a team, whose current best player is a scoring, shoot-first point guard. Suppose this team is looking to bring in a top-flight free agent. What type of player should this team target? Should they bring in a defense-oriented big man? Should they acquire a multi-faceted, jack-of-all-trades wing? This paper aims to answer these questions. Analyzing player data and team season data from 1977, this paper first uses clustering techniques to group players into appropriate groups, then regression to determine the degree to which the composition of a team's top 2 and top 3 players affect that team's win total, while accounting for team quality and coaching ability. This paper shows that the composition of a team's top 2 and top 3 players is a strongly statistically significant factor in the success of a team, and shows which combinations yield over-performance, and which combinations yield underperformance, relative to the team's talent and coaching quality.

1 Introduction

There have been many instances in basketball where a team, with perhaps a collection of new acquisitions, underperforms relative to the perceived talent on the team. Most observers will intuitively conclude that this team, while talented, just doesn't fit well together. Conversely, there are also many instances where a team, with perhaps relatively modest top level talent, exceeds expectations. This team, most will conclude, is put together well, i.e., the players complement each other, they "fit." Similarly, at the player level, there are many cases in which a new player on a team, perhaps acquired through free agency, who, though talented, fails to live up to expectations. This can be attributed to many things: lack of effort, erosion of skills, poor scouting. However, many times a player's underperformance is attributed to a poor fit with the team. This is to say, that it is not enough for a player to have valuable basketball skills to fully reach his potential; the player must also be on a team which is constructed in a way that is complementary to those skills.

This paper aims to provide some insight into team construction and player fit by analyzing combinations of player types (specifically, 2 and 3 man combinations, known commonly as "Big 2's or Big 3's), and determining which combinations lead to increased wins, while accounting for talent level and coaching skill. Along the way, the research uncovered some interesting insights regarding coaching ability; while those insights will be addressed briefly, this is not a focus of the paper.

The research approach is centered on the premise that success in the NBA is based on primarily 3 factors: talent, coaching and team fit. Talent and coaching are straight-forward conceptually: teams with more talent have a better chance at winning, as do teams with superior coaches. Team fit, while less straightforward and less tangible, matters as well, and although it is somewhat difficult to quantify, most observers have an intuition regarding the fit of a team. An example is presented to illustrate: Suppose there was a team whose three best players were Derrick Rose, Russell Westbrook and Rajon Rondo, three very good, and very similar players. Further suppose this team is coached by a superior coach, let's say Gregg Popovich, who is widely regarded as a superior coach, a notion also supported by data in this paper. Conventional wisdom suggests that this is a team, that while immensely talented, and with great coaching, does not fit well together and would likely underperform a similarly talented and well-coached team, but with better fitting pieces. One of the aims of the paper is to determine whether conventional wisdom regarding team fit is supported by the data.

This paper first discusses the clustering techniques used to form appropriate player groups. Collections of these player-type groups are then used as binary variables in the regression model predicting team wins[1].

2 Grouping Players [1]

To construct 2 and 3 man combinations, players need to be grouped into appropriate player-type groups. For instance, high-scoring, well-rounded wings like Kobe Bryant and Dwyane Wade will likely be grouped together, as will 3-point shooting specialists, like Steve Kerr and Dell Curry. Players were grouped according to their overall similarity on the following variables: points per game, offensive rebounds per game, defensive rebounds per game, assists per game, steals per game, blocks per game, turnovers per game, personal fouls per game, field goal attempts per game, field goals made per game, free throw attempts per game, free throws made per game, 3 pointers attempted per game, 3 pointers made per game.

Before getting to the results of the clustering, it is important to briefly discuss the clustering methodology. When thinking about clustering players, there are two intuitive, clustering structures:

1. Cluster structure that reflects both level of play and statistical similarity. For instance, this type of clustering would have multiple clusters for 3 point shooting power forwards. There would be a cluster for the all-star level players (Kevin Love, Dirk Nowitski), starter level players (Rasheed Wallace, Rashard Lewis) and a cluster for rotation/role players (Robert Horry).

2. Cluster structure that primarily reflects statistical similarity. In this case, the players listed above would ideally all fall into one, or at most, two clusters.

It would be ideal for the clusters to be structured in a way that primarily reflects statistical similarity, regardless of talent level; this is largely due to the fact that talent is already accounted for in the regression model. To achieve this end, it was more appropriate to pursue K-Means clustering, rather

---

[1] For example, if there were 4 player groups 1, 2, 3, and 4, then 1-2, 1-3, 2-3, and 3-4 would be examples of 2 man combinations, and each would be a separate binary variable in the model. A team that had an "1" and a "2" as its top 2 players, would have an 1-2 value of 1 in the model, and 0 for all of the other 2 man combinations.

than hierarchical clustering, as hierarchical clustering would have naturally formed groups into tiers. However, despite these efforts to ensure talent-agnostic groups, some clusters will naturally tend to have more star power, due to the uniqueness of star players in the NBA. The most extreme example of this is cluster 12, which almost exclusively includes legendary, Hall-of-Fame or Hall-of-Fame bound centers. This is because the statistical profiles of these centers are so profoundly unique that it is nearly impossible to include any other players in their cluster. Hall-of-Fame, or Hall-of-Fame bound 2-guards such as Michael Jordan, Kobe Bryant and Dwyane Wade, however, can be more easily placed in a cluster with good, but decidedly lesser players, such as Corey Maggette or Steve Francis, because, statistically, these lesser players do most of what the greats do, just not quite as well. Ultimately, 13 groups were selected, which seemed to generally fall into 2 talent tiers. Their statistical characteristics, brief description of representative player types, and a selection of representative players are shown below.

Figure 1: Cluster Centers (Averages by cluster)

Cluster	pts/g	oreb/g	dreb/g	asts/g	stl/g	blk/g	to/g	pf/g	fga/g	fgm/g	fta/g	ftm/g	3pa/g	3pm/g
Cluster-1	5.17	1.48	2.77	0.83	0.48	0.68	0.94	2.37	4.39	2.07	1.50	1.00	0.11	0.03
Cluster-2	21.80	1.49	3.36	4.00	1.32	0.44	2.73	2.59	17.22	8.24	6.08	4.90	1.28	0.41
Cluster-3	5.52	0.47	1.29	1.88	0.61	0.14	0.93	1.48	4.99	2.14	1.16	0.89	1.00	0.34
Cluster-4	8.85	0.66	2.23	2.06	0.77	0.24	1.17	2.00	7.61	3.24	1.64	1.28	2.90	1.09
Cluster-5	16.45	2.54	5.52	2.43	0.98	0.96	2.43	3.22	12.98	6.42	4.77	3.52	0.30	0.08
Cluster-7	16.46	1.15	3.03	8.46	2.06	0.30	3.18	2.73	13.34	6.26	4.55	3.59	1.16	0.35
Cluster-8	16.24	0.80	3.02	4.61	1.28	0.27	2.18	2.27	13.37	5.80	3.56	2.89	4.69	1.75
Cluster-9	9.39	0.64	1.90	4.68	1.20	0.18	1.91	2.22	8.17	3.68	2.26	1.76	0.87	0.27
Cluster-10	10.85	2.98	6.20	1.49	0.76	1.58	1.73	3.17	8.62	4.31	3.18	2.19	0.14	0.04
Cluster-11	16.05	1.32	3.77	2.94	1.14	0.49	2.05	2.65	13.28	5.94	3.74	2.97	3.33	1.21
Cluster-12	23.09	3.39	8.13	2.70	1.08	2.29	3.13	3.26	16.64	8.69	7.94	5.65	0.23	0.05
Cluster-13	13.31	1.27	2.50	2.38	0.98	0.34	1.77	2.36	11.19	5.25	3.33	2.61	0.71	0.21
Cluster-14	8.91	1.71	3.31	1.41	0.69	0.58	1.42	2.69	7.39	3.52	2.50	1.80	0.26	0.07

Cluster descriptions and representative players:
Cluster-1: Limited, role-playing centers: Erick Dampier, Tree Rollins
Cluster-2: High scoring, dynamic guards (mostly 2 guards, but some 3's like Grant Hill), typically not great 3 point shooters, or if they are, don't shoot very many: Kobe Bryant, Dwayne Wade, Tracy McGrady, Adrian Dantley.
Cluster-3: Somewhat limited, role-playing backcourt players: John Paxson, Jose Berea
Cluster-4: Wing 3 point shooters: Dan Marleje, Shane Battier
Cluster 5-: Dynamic, well-rounded power forwards, strong rebounding, dynamic 3's: Chris Webber, Pau Gasol, Kevin McHale

Cluster-7: High scoring, high assist, high steals, high turnover point guards, who don't shoot 3s: Kevin Johnson, Isaiah Thomas

Cluster-8: Multi-faceted, high scoring wings, with high assists for their position and are great 3 point shooters: Paul Pierce, Danny Ainge

Cluster-9: Pass first, low scoring point guards: Avery Johnson, Mark Jackson

Cluster-10: Limited 4's; very strong rebounders, defense oriented: Dennis Rodman, Ben Wallace, Buck Williams

Cluster 11: 3 point shooting bigs, i.e. stretch 4's: Rasheed Wallace, Antawn Jamison, Detlef Schrempf

Cluster 12: High scoring post players, high rebounds, high blocks: Shaquille O'Neal, Hakeem Olajuwon, David Robinson

Cluster 13: Well-rounded small forwards; generally don't shoot many 3 pointers: Luol Deng, James Worthy

Cluster 14: Role-playing big men without an exceptional skill, but contribute in several categories: Udonis Haslem, Kurt Thomas

3 Regression Model

As discussed earlier, a regression model was developed to measure the impact of the *composition* of "Big 2s" and "Big 3s" on wins, isolated from the influence of coaching and player talent:

Model 1: *Wins = Team talent + Coaching + Composition of top 2 players*
Model 2: *Wins = Team talent + Coaching + Composition of top 3 players*

4 Model Details

Several issues should be addressed before running this regression. The first is how to *quantify* team talent. Most observers have a good idea of how talented a team is, but quantifying this variable is less straightforward. The approach taken was as follows. First, each player-season was evaluated in terms of the player's efficiency that year, measured by the efficiency formula used on NBA.com[2]: *((Points + Rebounds + Assists + Steals + Blocks) - ((Field Goals Att. - Field Goals Made) + (Free Throws Att. - Free Throws Made) + Turnovers)).* Players were then classified according to their season efficiency rating, based on the percentile rank of the player's efficiency rating (compared against player-seasons across all seasons), and points given to each player according to the scoring system below, weighted by minutes played. For each team-season, their player's minutes-weighted score is summed, giving each team a metric that estimates the talent of the team.

Figure 2: Player efficiency scoring methodology

Efficiency Percentile	Points	Efficiency Percentile	Points
98th	30	50th	12.5
95th	25	40th	10
90th	22	30th	8
80th	19	20th	5
70th	16	10th	2
60th	14	0th	0

---

[2] Players with 800 or more minutes

The coaching variable was more straightforward. For each team-season, the coach who coached the majority of the team's games is assigned as a variable with a value of 1; all of the other coaches are assigned as a variable with a value of 0. For example, the 2008 Lakers would have 143 variables corresponding to all the coaches in the league since 1977. Each of these variables would be assigned a value of 0, except the variable corresponding to Phil Jackson, which would be assigned a value of 1. It should be noted that the final regression model has far fewer than 143 coaching variables, as many coaches' performance did not have a statistically significant impact on wins, either positively or negatively.

Composition of top 2 (3) players is integrated into the models as follows. Each team-season has a binary variable corresponding to all of the top 2 (3) combinations historically. Each variable will have a value of 0, except the variable corresponding to the team's top 2 players. For example, if the 2008 Lakers' top 2 players fall into Cluster 2 and Cluster 5 (subsequently denoted 2-5)[3], then their 2-5 variable will have a value of 1, all other combinations will have a value of 0. As was the case with the coaching variables, many top 2(3) combinations did not have a statistically significant impact on wins, greatly reducing the size of the model.

5 Regression Results (see Appendix A for details)

"Big 2s"
The regression showed that there were 6 combinations of Big 2's that exhibit a statistically significant impact on wins: 2-2 (regression coefficient: +3.97), 10-12 (+4.69) , 2-8 (+4.35), 8-11 (+4.75), 8-12 (+7.59) all had a significantly positive impact on wins, and 8-8 (-4.046) had a significantly negative impact. Other combinations either had very little data (for example, there was only one team in since 1977 that had a top 2 of 10-15), or when there was enough data, the impact on wins was inconclusive.

There are several takeaways from these results. First, conventional wisdom is confirmed in several cases. For instance, most observers would agree that multi-faceted small forwards who shoot 3's well would mesh quite well with high scoring, high rebounding centers that block shots (8-12, 7.59). Indeed, this Big 2 had the most positive impact on wins. Similarly, most observers would also agree that high scoring 2 guards would also fit well with multi-faceted small forwards who shoot 3's (2-8, 4.35), and that low scoring, strong rebounding, defense-oriented power forwards would fit well with the high scoring centers (10-12, 4.69).

Most observers would think that a Big 2 from the same group would not fit as well; this is partially contradicted by this analysis. While multi-faceted small forwards who shoot 3's don't fit well together (8-8, -4.046), teams with two high scoring 2-guards (2-2, 3.97) have historically over-performed their expected win total, given the team's overall talent level and coaching skill. Digging a bit further into the data, nearly all of the teams with multiple high-scoring 2 guards played at a higher than median pace; although further analysis would be required to state conclusively, this is perhaps instructive on the style of play that teams with two high-scoring 2-guards should employ.

"Big 3s"
The regression showed that there were 7 combinations of Big 3's that exhibit a statistically significant impact on wins: 2-2-5 (regression coefficient: +3.70), 2-5-8 (+3.43), 7-8-12 (+13.6), and 8-10-12

---

[3] Also, top 2 and top 3 player combinations did not consider ordering of players. So if a team's best player was in cluster 2 and the 2nd best player was in cluster 3, that was treated the same as a team whose best player was in cluster 3 and 2nd best player was in cluster 2.

(+5.43) all had a significantly positive impact on wins, and 5-5-9 (-8.47) , 2-2-7 (-4.78), and 5-8-8 (-3.61), had a significantly negative impact. As with the Big 2s, other combinations either had very little data or when there was enough data, the impact on wins was inconclusive.

Conventional wisdom is confirmed in several cases. For instance, most would expect that high-scoring 2-guards teamed with high-scoring, high turnover point guards would not fit well together (2-2-7, -4.78); the data supports this intuition, as teams with these players as their top 3 significantly underperformed their expected win total, given their talent level and coaching ability. As was the case with the Big 2s, the perception of poor fit from duplicative players is partially contradicted, again by the scoring 2-guards. While each of the Big 3's that underperformed expected wins had a duplicate player, the combination 2-2-5 over-performed expected wins.

One major takeaway from the Big 3 results is that the data shows that, cluster 8, the multi-faceted small-forwards who are very good 3-point shooters, are great players to build a team around, as long as there aren't any similar players among the most talented players on the team. Very good results occur when these small-forwards are surrounded with a variety of player-types; the Big 3's with the highest coefficients (7-8-12, and 8-10-12) both include players from cluster 8. This was true with the Big 2's as well.

Coaching

As a brief aside, the coefficients measuring the impact of the coaches are instructive, and in general, align with perceptions. For instance, Gregg Popovich was found to be the most effective coach (+23.19), followed mostly by very respected coaches. Most of the coaches with negative coefficients aren't likely to surprise most fans as well. Two cases which might challenge conventional wisdom are Mike Brown and Avery Johnson, who do very well in this analysis and are perhaps slightly underrated.

6 Conclusions

Obtaining the most talented and productive players on the team is perhaps the most important decision that NBA teams make. Constructing a team that can reach its full potential requires more than just acquiring talented players; these players have to fit well together. This analysis showed that how Big 2s and Big 3s are constructed can significantly affect expected win total, relative to the talent level on the team and the coaching skill. Conventional wisdom was confirmed in several cases. For instance, duplicative players, with one exception, generally result in a poor fit, leading to underperformance in expected wins. The analysis also showed that high-scoring point guards don't mesh well with high-scoring 2 guards. Talented, high-scoring centers fit well with more limited, defense-oriented power forwards who rebound very well, which also aligns with conventional wisdom. Unexpected results include the degree of fit when teams have two high-scoring 2 guards. While most would say that these teams would not be a good fit, the analysis showed that these teams consistently over-performed their expected wins, relative to team talent and coaching skill. Finally, the analysis highlighted the importance of multi-faceted small forwards, who shoot 3 pointers well. Teams that featured this type of player tended to over-perform expected wins by the greatest amount, as long as there was not another similar player among the top players on the team.

7 References

Shmueli, Galit, Patel, Nitin and Bruce, Peter. *Data Mining for Business Intelligence.* 3rd ed. Hoboken, New Jersy: John Wiley & Sons, Inc. Print.

Appendix A

Regression Details, Top 2 players (yellow indicates continuous variable, blue indicates binary variable

	Input variables	Coefficient	Std. Error	p-value	SS		
	Constant term	-13.412	1.976	0.000	1387694.000	Residual df	812.000
Team Talent	Team talent metric	0.217	0.008	0.000	50594.879	Multiple R-squared	0.587
	Rick Adelman	6.343	2.068	0.002	464.984	Std. Dev. estimate	8.359
	Richie Adubato	-10.993	4.870	0.024	427.505	Residual SS	56,740.215
	Larry Bird	14.743	4.842	0.002	460.849		
	Larry Brown	6.919	1.780	0.000	528.264		
	Mike Brown	16.267	4.205	0.000	931.713		
	Rick Carlisle	12.493	3.231	0.000	1220.540		
	Billy Cunningham	10.760	3.197	0.001	594.864		
	Chuck Daly	5.108	2.355	0.031	166.366		
	Don Delaney	-19.119	8.368	0.023	440.552		
	Phil Jackson	15.255	2.013	0.000	3981.902		
	Avery Johnson	20.277	4.840	0.000	1083.414		
	Phil Johnson	-9.818	4.842	0.043	355.879		
	Sidney Lowe	-7.536	4.203	0.074	347.957		
Coaches	Jack McKinney	-6.261	3.436	0.069	330.494		
	Nate McMillan	8.788	3.046	0.004	541.493		
	Dick Motta	-6.439	2.318	0.006	521.497		
	Bill Musselman	-12.991	4.840	0.008	626.120		
	Gregg Popovich	23.189	2.474	0.000	6054.770		
	Jerry Reynolds	-17.027	5.924	0.004	661.306		
	Pat Riley	7.251	1.872	0.000	1606.805		
	Doc Rivers	7.597	2.702	0.005	697.945		
	Jimmy Rodgers	-9.066	4.197	0.031	390.775		
	Bill Russell	-15.969	8.367	0.057	283.637		
	Flip Saunders	9.429	2.443	0.000	1203.330		
	Jerry Sloan	8.745	1.781	0.000	1543.577		
	Dick Vitale	-22.334	5.928	0.000	1032.815		
	Paul Westhead	-8.712	3.769	0.021	348.590		
	Dave Wohl	-8.630	4.840	0.075	254.514		

Top 2 players	2 and 2	**3.968**	**1.917**	**0.039**	**204.584**
	10 and 12	4.689	2.187	0.032	256.089
	2 and 8	4.351	1.407	0.002	572.559
	8 and 11	4.746	1.632	0.004	584.605
	8 and 12	7.594	1.927	0.000	1105.043
	8 and 8	-4.046	2.272	0.076	221.680

ANOVA

Source	df	SS	MS	F-statistic	p-value
Regression	35	80641.896	2304.0542	32.97294504	3.4307E-131
Error	812	56740.215	69.877112		
Total	847	137382.11			

Regression Details, Top 3 players

	Input variables	Coefficient	Std. Error	p-value	SS		
	Constant term	-12.336478	1.9982631	0	1387694	Residual df	811.000
Team Talent	Team talent metric	0.2152093	0.0083441	0	50594.879	Multiple R-squared	0.580
Coaches	Rick Adelman	6.5892091	2.0916212	0.0017296	464.98431	Std. Dev. estimate	8.437
	Richie Adubato	-10.057054	4.8939276	0.0404016	427.50497	Residual SS	57,727.926
	Larry Bird	14.141637	4.886034	0.0039675	460.84872		
	Larry Brown	6.3485184	1.7952193	0.0004438	528.26392		
	Mike Brown	16.764431	4.2345681	8.629E-05	931.71338		
	Rick Carlisle	14.66088	3.2165356	6.51E-06	1220.5404		
	Billy Cunningham	10.281056	3.2262204	0.0015297	594.86438		
	Chuck Daly	4.5965629	2.3764043	0.0536507	166.36594		
	Don Delaney	-19.72205	8.4456444	0.0199322	440.55173		
	Phil Jackson	15.715521	2.025667	0	3981.9016		
	Avery Johnson	18.576672	4.9176331	0.0001776	1083.4142		
	Phil Johnson	-10.340877	4.8869834	0.0348421	355.87936		
	Sidney Lowe	-8.1694126	4.2415252	0.0546686	347.95673		
	Jack McKinney	-6.7823186	3.4670584	0.0510004	330.49442		
	Nate McMillan	9.4200115	3.0184205	0.0019085	541.49286		
	Dick Motta	-5.2361803	2.3356047	0.0254094	521.49744		
	Bill Musselman	-13.541759	4.8844829	0.0057737	626.12018		
	Gregg Popovich	23.058096	2.4741864	0	6054.77		
	Jerry Reynolds	-17.550932	5.9786754	0.0034837	661.30603		
	Pat Riley	8.528367	1.8874339	7.79E-06	1606.8046		
	Doc Rivers	9.0437079	2.7017112	0.0008777	697.94507		
	Jimmy Rodgers	-9.5996275	4.2354074	0.0238496	390.77469		
	Bill Russell	-16.51387	8.4448671	0.0510861	283.63669		
	Flip Saunders	10.255658	2.4616146	3.653E-05	1203.3298		
	Jerry Sloan	8.1965294	1.7964728	6.38E-06	1543.5771		
	Dick Vitale	-22.825397	5.9834623	0.0001535	1032.8152		
	Paul Westhead	-8.3417587	3.798532	0.0285507	348.59033		
	Dave Wohl	-9.2137985	4.8846469	0.0598416	254.51369		

	2 2 5	3.697598	2.0522385	0.0721936	257.62769
	5 5 9	-8.4751492	3.4884737	0.0154748	415.99844
	2 2 7	-4.7829318	2.5910213	0.0654931	263.52786
Top 3 players	2 5 8	3.4399049	1.8947914	0.070058	237.77937
	5 8 8	-3.6058707	2.0833304	0.0841057	221.91527
	7 8 12	13.603704	5.9772806	0.0232774	366.3978
	8 10 12	5.4337397	3.294785	0.0997411	193.60146

ANOVA

Source	df	SS	MS	F-statistic	p-value
Regression	36	79654.185	2212.6162	31.084293	1.7649E-127
Error	811	57727.926	71.181166		
Total	847	137382.11			