Jennies Basketball Lineup Analysis: An Economic Case Study

Peyton O. Taylor, McNair Intern

Christopher Azevedo, Ph.D Economics

University of Central Missouri

July 26, 2018

Abstract

Decision-making in team sports can be extremely complex due to the fact that individual success does not always equate to team success. Playing the right players at the right time can be very challenging. For sports like basketball, coaches try to use basic boxscore statistics, and often +/- rating systems, to evaluate players and make insightful decisions on game strategy. However, boxscores cannot capture the complexity and randomness involved in the game of basketball. This makes it difficult to identify and quantify the impact each player or player combination had on the outcome of the game. Coaches are expected to maximize the team's opportunity to win similar to economic theory that firms make decisions to maximize profits. We use data from the University of Central Missouri 2017-2018 Women's Basketball team to examine the hypothesis that coaches choose lineups that maximizes the probability of winning. Defensive and offensive output is modeled separately as a function of individual players. Offensive and defensive output is also modeled as a function of the top 10 coach-assigned player lineup combinations and the location of the game as an additional independent variable. Results showed that there is mixed statistical evidence, offensively and defensively, that coaches make decisions that maximize their probability of winning. The effects of this implication will provide valuable data for more efficient management of team talent and develop a more compressive method of measuring team performance that does not rely on any particular boxscore statistic. Though, this research does not take into consideration external factors that may influence individual player contribution such as endurance, fatigue, years of experience, or stress levels that affect student-athlete performance, the conclusions will enable programs to improve overall student-athlete experience and generate more revenue through factors such as increased fan attendance per game and longer post seasons.

Introduction

In team sports like basketball, football, soccer, or hockey, most coaches will emphasize the idea that "we" is greater than "me." How players work together in a game largely affects their overall group performance or chance of beating their opponent. Coaches must make crucial decisions on what lineup to play and how long to play it. Time is scarce, as is the pool of player talent from which a coach can choose. A basic assumption of economics is that a firm's goal is to maximize profits. Similarly for sports teams, it would be reasonable to assume that a coach's goal is to maximize their team's performance and probability of winning since "winning is associated with revenue" (Berri, Brook, Schmidt, 2007). Romer (2006) discusses how football teams' decisions in certain game situations are "systematically and overwhelmingly statistical significant" and similar to decisions firms make to maximize profits.

However, other researchers argue that decision-makers in professional sports do not behave in a way that aligns with economic optimization (Berri, et al., 2007). This paper explores two questions: (1) Can differences among player contributions and lineups be detected? and (2) Will empirical evidence confirm that a coach follows this economic theory and can this be used to predict a coach's decisions on player lineups in the game of basketball?

Background

The complex random components and situations in team sports can make it difficult to identify and quantify each variable that affects the outcome of the game. For example, only 20% of what occurs during a basketball game is tabulated in the box score (Winston, 2009). General box score statistics are limited to individual and team totals of field goal percentage, rebounds, free throws, assists, turnovers, blocks, and steals. In addition, box scores lack defensive statistics like charges, deflections, or even a secondary defender helping over. A single player's contribution is dependent on the contribution of the other four players in the lineup in the game of basketball. Therefore, these statistics cannot accurately capture all aspects of the game, making it challenging for a coach to optimize lineups to maximize the probability of winning.

Winston (2009) contends models could be improved by quantifying an individual player's contribution using an adjusted +/- rating system. A positive statistic such as a rebound would add value to a player's contribution to the game, while a negative statistic such as a turnover would depreciate a player's adjusted rating and contribution. Similar research has been conducted on player positions in the National Hockey League (Chan, Cho, & Novati, 2012). Each position was measured by a +/- composite statistic quantifying the value of each player in order for management to allocate playing time. The adjusted +/- system is limited because team sports involve multiple individuals working together simultaneously. Overall team performance cannot be based off of individual ratings. Only a "small fraction of players have significantly high box score statistical value." Therefore, these players would not necessarily "guarantee a team's performance improvement" (Vaz de Melo, Almeida, Loureiro, & Faloutsos, 2012).

Some may value the law of transitivity when evaluating player contribution, which does not hold according to Winston (2009). The law of transitivity would imply that if player X adds more value than player Y, and player Y adds more value than player Z, then player X adds more value to the team than player Z. However, different players work together better in different situations, so transitivity does not apply in this case. Thus, how does a coach decide which five players will maximize the probability of a winning outcome with a limited time constraint?

There has been limited research analyzing group contribution to overall team performance in sports. Previous research tested different optimization techniques in order to develop a more objective process in selecting specialized players in the game of cricket (Bhattacharjee & Saikia, 2016). It is common for coaches and management to believe that playing their "best" athletes yields a greater chance of winning. Empirical evidence supports the notion that as the quality of inputs, or players, increases so does output, or probability of a win (Sánchez, Castellanos, & Dopico, 2007). Because individual players have their own unique strengths and weaknesses, the opportunity cost of playing different players with different levels of skills must be evaluated. The law of diminishing returns suggests that more is not always better. For example, if too many players in a lineup specialize in offense, the lack of defense can be detrimental. Thus, coaches must take into account what would be given up to play one player over another and the impact of that decision on team performance.

Coaches strategize according to various situational factors of the game such as: the current score of the game (Sampaio, Lago, Casais, & Leite, 2010), strength of opponent and their strategy (Chiappori, Levitt, & Groseclose, 2002), and a player's changing endurance levels (Hughes, 2017). With data from the Spanish Basketball Professional League, Sampaio et al. (2010) found that the running score for a basketball game can be used as a measurement of team performance and would possibly affect players' later efforts during the game. Style of play is dependent upon the difference in score relative to the remaining time.

Team performance optimization is also dependent upon the opponent and their strategy. Previous research explores the assumption that players want to play to win and will change their strategy based on the behavior of their opponent (Chiappori et al., 2002). This study tests if economic theory can predict kicker and goalkeeper strategies to maximize performance during a penalty kick in the game of soccer. The author found that the goalie's strategy is dependent on the kicker's previous kicks, while the kicker's strategy is independent of the goalie's behavior. In team sports, it is common for a team to scout out their opponent's tendencies, strengths, or weaknesses prior to a game to gain an advantage. A coach's lineup strategy would be dependent on the opponent's behavior to maximize the team's chances of winning. Knowledge of individual players' endurance levels, as well as measurement of opponent's defensive intensity, also help in developing a substitution strategy within a set of lineups (Hughes, 2017).

These situational factors complicate decision-making. Engemann and Owyang (2009) argue that coaches may not always maximize their chances to win, but instead minimize their probability to lose. For example, in certain situations like deciding whether to try for a first down or kick a field goal on the fourth down in professional football, coaches may shift to a more conservative strategy (Romer, 2006). Going for a fourth down "more frequently would increase the probability that coaches would win games" (Berri et al., 2007).

The purpose of this case study is to utilize analytics to predict optimal basketball lineups for NCAA Division II Women's basketball teams. This case study examined statistical data from the 2017-18 University of Central Missouri Women's Basketball team in order to quantify contributions of coach-assigned Jennie basketball lineups over a given season. A secondary objective was to examine the empirical evidence on whether coaches maximize the probability of winning.

Methodology

This case study consists of a statistical comparison of coach-assigned player lineup combinations against model-ranked player combinations using Synergy software as a measurement of team performance. It was first necessary to compile a unique data set of all relevant information for each game played by the 2017-18 Jennies Basketball team. Data was retrieved from the Mid-America Intercollegiate Athletics Association website (MIAA, n.d.), StatCrew Software (StatCrew, 2018), and Synergysportstech.com (Synergy Sports, 2017) for lineups and conference rankings of each team of the MIAA 2017-18 basketball season. StatCrew is an official scoring software application (StatCrew, 2018). Synergy Sports is a database system that catalogues every possession of every game using a variety of conventional and advanced statistics.

Player movement data from the play-by-play report identifies every player substitution, time in the game it occurred, and the score at the moment of substitution. Each game was broken down by second. A multivariate linear regression approach was used to variety of models. The dependent variable for each model was the change in either home or away score for each second of the game. For games with no overtime, that is 2,400 per game. At each second of play, we observe the dependent variable ranging between 0 (no scoring) and 4 points (3 point field goal with a made free throw).

Models

Separate offensive and defensive models for both individual player and lineup contributions were utilized in this study. The models that are used to describe the individual offensive and defensive contribution of each player are specified as

Offense: $dHome = \beta_0 + \beta_{00} I_{00,t} + \beta_{01} I_{01,t} + \beta_{02} I_{02,t} + \beta_{03} I_{03,t} + \beta_{10} I_{10,t} + \beta_{13} I_{13,t} + \beta_{14} I_{14,t} + \beta_{20} I_{20,t} + \beta_{21} I_{21,t} + \beta_{22} I_{22,t} + \beta_{23} I_{23,t} + \beta_{24} I_{24,t}$

Defense: $dAway = \beta_0 + \beta_{00} I_{00,t} + \beta_{01} I_{01,t} + \beta_{02} I_{02,t} + \beta_{03} I_{03,t} + \beta_{10} I_{10,t} + \beta_{13} I_{13,t} + \beta_{14} I_{14,t} + \beta_{20} I_{20,t} + \beta_{21} I_{21,t} + \beta_{22} I_{22,t} + \beta_{23} I_{23,t} + \beta_{24} I_{24,t}$

where the dependent variable, *dHome,* is the change in the home score for each second. *dHome* can take values of zero, one, two, three and four. *dAway* is the change in the away score per second. For these models, the change in home score only reflects points scored for the Jennies and the change in away score only reflects points scored against the Jennies. I_n is an indicator variable where *n* identifies a player's jersey number, and *t* is the time in seconds of the game. β_0 represents the intercept. β_n are the parameters to be estimated. Offensively, β_n identifies the marginal impact of a player's ability to contribute to scoring. Defensively, β_n identifies the marginal impact of a player's ability to prevent the opponent from scoring. The models that are used to describe the relationship between coach-assigned player lineup combinations and their offensive and defensive contribution are specified as

Offense: $dHome = \beta_0 + \beta_2 I_{2,t} + \beta_3 I_{3,t} + \beta_4 I_{4,t} + \beta_5 I_{5,t} + \beta_6 I_{6,t} + \beta_7 I_{7,t} + \beta_8 I_{8,t} + \beta_9 I_{9,t} + \beta_{10} I_{10,t} + \beta_H HomeTRUE$

Defense: $dAway = \beta_0 + \beta_2 I_{2,t} + \beta_3 I_{3,t} + \beta_4 I_{4,t} + \beta_5 I_{5,t} + \beta_6 I_{6,t} + \beta_7 I_{7,t} + \beta_8 I_{8,t} + \beta_9 I_{9,t} + \beta_{10} I_{10,t} + \beta_H HomeTRUE$

where *dHome* and *dAway* are as described previously. β_0 represents the intercept and the estimate of the most recurrent starting lineup throughout the season. The starters, β_0, are ranked number one with the most minutes played at 102:31 minutes to the lineup number 10, or I_{10}, that played 8:24 minutes (StatCrew, 2018). β_n represents the parameters to be estimated that identify the marginal impact of a lineup's contribution either to offense or defensive where *n* designates the rank of the lineup by minutes played. The *HomeTRUE* variable is associated with the location of the game effect (home or away) parameter estimate. Parameter estimates were generated using R statistical software.

Results

There were a total of 117 lineup combinations of the thirteen players on the team over the first thirteen games of the 2017-18 season, for an average of sixteen lineup combinations per game. Table 1 and Table 2 show the estimation results from the individual player defensive and offensive model. To avoid singularity of the information matrix, one player must be eliminated from the model. Parameters that are statistically significant are identified by " *** " for a level of significance of $\alpha = 0.001$, " ** " for 0.01, " * " for 0.05, and " · " for a level of significance of 0.1. According to the individual defensive model, no players had parameters that were significant to the change in Away score. Based on the results from the offensive model in Table 2, five players had parameters that were statistically significant at a level of $\alpha = 0.05$. The parameters of six players showed to be significant at $\alpha = 0.01$, while only one player's parameter estimate was not statistically significant.

Tables 3 and 4 show the estimation results of the defensive and offensive model using the top 10 lineups for the beginning of the season as independent variables. As for the individual models, one lineup was eliminated to avoid singularity of the information matrix. The defensive model showed that the only two lineups that had significant parameters were lineups ranked 8th and 9th. All other lineup combinations were found to have parameters that were not statistically significant. Based off the offensive model in Table 4, there were also no lineup parameters that were statistically significant. Estimates associated with the game being located on the Jennies home floor were negative for defense and positive for offensive but both not significant. This would suggest that more data could be needed to determine if the location of the game is significant to lineup contribution and performance.

Discussion

Because the independent variables of the model are all dummy variables, the parameter estimates for the individual player models must be interpreted in relation to the intercept parameter estimate. The estimated model takes the form

$$dHome = 0.1417 - 0.02719I_{23} - 0.01904I_{13} - 0.02238I_{24} - 0.02015I_{02} - 0.02658I_{22} - 0.02698I_{03} - 0.01914I_{20} - 0.02063I_{10} - 0.01854I_{00} - 0.02210I_{01} - 0.03784I_{14} - 0.02837I_{21}$$

where I_{23}, I_{13}, etc. are indicator variables. It is straightforward to generate the impact of a particular player using the parameter estimates contained in Tables 1 and 2. To calculate the offensive impact of player 23, her dummy variable, I_{23}, would be a one, while all the other dummy variables would be a zero. Player 23's impact per second on the home score would take the form

$$dHome_{23} = 0.1417 - 0.02719 = 0.11451.$$

This means that player 23 would be expected to add 0.11451 points per second to the home score. Table 5 shows the offensive and defensive production for each player as well as how each player ranks against other players on the team.

The parameter estimates from the individual models, contained in Tables 1 and 2, can also be used to calculate the offensive and defensive impacts of a particular lineup. Suppose the offensive points per second of a lineup consisting of players 13, 20, 00, 03 and 22 is desired. Each of these dummy variables would be a one, while all of the other dummy variables would be a zero. This lineup's impact per second on the home score would take the form

$$dHome = 0.14178 - 0.01904 - 0.02658 - 0.02698 - 0.01914 - 0.01854 = .0315.$$

Though this method can be used to rank lineups, it is a simple model that treats all of the players as independent. It is not specifically designed to account for the various lineups.

The lineup models are specifically designed to account for the fact that there are five players on the floor at a time. The parameter estimates contained in Tables 3 and 4 can be used to estimate the offensive and defensive output of a particular lineup. Suppose the offensive output of lineup two is desired. This would take the form

$$dHome_{02} = 0.0259 + 0.0071 = 0.0330.$$

Table 6 shows each lineup's offensive and defensive impact per second compared against lineups used by the coach. Column one lists coach-assigned lineups in order of playing time, column two shows the offensive model ranking, column three shows offensive points per second for each lineup, column four shows the defensive model ranking and column five shows defensive points per second given up by each lineup.

The evidence supporting the hypothesis that coaches make decisions that maximize a team's opportunity to win is mixed. As Table 6 shows, several of the coach-assigned lineups with high numbers of playing minutes are also ranked highly by the offensive model. For example, lineup 4 is ranked second by the offensive model. However, some of the coach-assigned lineups that are used the most (starters, lineup 2 and lineup 3) are ranked in the lower half by the model. For the defensive model, several of the coach-assigned lineups with the most minutes are also ranked high by the model (starters, lineup 3, lineup 4 and lineup 5). Lineups 7 and 9 are ranked low in terms of coach-assigned minutes on the floor as well as by both models.

The results of this research cannot be used to definitively evaluate the lineup decisions of the coach. Though these models provide a ranking of lineups, they are still relatively simple. Coaches must make lineup decision in real time using a variety of information. Ranking lineups in real time requires a complex evaluation of the components of both teams on the floor down to each player's individual strengths and weakness because each team will naturally strategize to their own unique comparative advantage against the opponent.

Conclusions from this research also suggest excluding data from the last three minutes of team game. The strategy at the end of games may not necessarily reflect the same strategy a coach uses in the first, second, and third quarters. The end of games can present very unique situations that require a special strategy that is dependent upon the score, time remaining, and fouls. Since player combinations are already a product of the coach's decision-making, further research could be done to explore any shifts in the coach's utilization of various lineups throughout an entire basketball season, comparing the lineup rankings of the first half of the season to the lineup rankings of the second half of the season.

Table 2

Regression Results - Individual Offensive Model

	Estimate	T-Value
β_0	0.14178	*2.655***
β_{00}	-0.01854	*-1.635*
β_{01}	-0.02210	*-1.885*
β_{02}	-0.013958	*-1.399*
β_{03}	-0.02698	*-2.385**
β_{10}	-0.02063	*-1.851 ·*
β_{13}	-0.01904	*-1.772 ·*
β_{14}	-0.03784	*-2.054**
β_{20}	-0.01914	*-1.740 ·*
β_{21}	-0.02837	*-2.182**
β_{22}	-0.02658	*-2.379**
β_{23}	-0.02719	*-2494 **
β_{24}	-0.02238	*-1.957*

Table 1

Regression Results - Individual Defensive Model

	Estimate	T-Value
β_0	0.073338	*1.554*
β_{00}	-0.010136	*-1.012*
β_{01}	-0.011005	*-1.062*
β_{02}	-0.013958	*-1.399*
β_{03}	-0.013242	*-1.325*
β_{10}	-0.009052	*-0.919*
β_{13}	-0.010888	*-1.147*
β_{14}	-0.022216	*-1.365*
β_{20}	-0.011625	*-1.196*
β_{21}	-0.005837	*-0.508*
β_{22}	-0.005567	*-0.564*
β_{23}	-0.007887	*-0.819*
β_{24}	-0.006268	*-0.620*

Table 3

Regression Results - Lineup Defensive Model

	Estimate	T-Value
β_0	0.0260059	*4.633****
β_2	0.0042005	*0.658*
β_3	0.0003155	*0.041*
β_4	0.0028451	*0.408*
β_5	0.0040855	*0.466*
β_6	-0.0055995	*1.098*
β_7	0.0091849	*1.098*
β_8	0.0209611	*1.945 ·*
β_9	0.0274820	*1.772 ·*
β_{10}	0.0055492	*0.581*
HomeTRUE	-0.0072694	*-1.363*

Table 4

Regression Results - Lineup Offensive Model

	Estimate	T-Value
β_0	0.0258790	*3.973****
β_2	-0.0070964	*-0.958*
β_3	-0.0018422	*-0.207*
β_4	0.0104474	*1.290*
β_5	0.0082855	*0.814*
β_6	0.0001468	*0.017*
β_7	-0.0038321	*-0.395*
β_8	0.0115469	*0.923*
β_9	-0.0086116	*-0.479*
β_{10}	0.0004640	*0.042*
HomeTRUE	0.0079427	*1.283*

Table 5

	Offensive Impact on Home Score		*Defensive Impact on Away Score*	
	Model Ranking	*Points per Second*	*Model Ranking*	*Points Per Second*
β_{00}	2	0.1234	7	0.0632
β_{01}	6	0.1197	5	0.0623
β_{02}	1	0.1278	2	0.0594
β_{03}	9	0.1148	3	0.0601
β_{10}	5	0.1212	8	0.0643
β_{13}	3	0.1227	6	0.0625
β_{14}	12	0.1039	1	0.0511
β_{20}	4	0.1226	4	0.0617
β_{21}	11	0.1134	11	0.0675
β_{22}	8	0.1152	12	0.0678
β_{23}	10	0.1146	9	0.0655
β_{24}	7	0.1194	10	0.0671

Table 6

	Offensive Impact on Home Score		*Defensive Impact on Away Score*	
	Model Ranking	*Points per Second*	*Model Ranking*	*Points Per Second*
Starters	6	0.0259	2	0.0260
Lineup 2	9	0.0188	6	0.0302
Lineup 3	7	0.0240	3	0.0263
Lineup 4	2	0.0363	4	0.0363
Lineup 5	3	0.0342	5	0.0301
Lineup 6	5	0.0260	1	0.0204
Lineup 7	8	0.0220	8	0.0352
Lineup 8	1	0.0374	9	0.0470
Lineup 9	10	0.0173	10	0.0535
Lineup 10	4	0.0263	7	0.0316

References

Berri, D. J., Brook, S. L., & Schmidt, M. B. (2007). Does one simply need to score to score?. *International Journal of Sport Finance*, *2*(4), 190.

Bhattacharjee, D., & Saikia, H. (2016). An objective approach of balanced cricket team selection using binary integer programming method. *Opsearch*, *53*(2), 225-247.

Chan, T. C., Cho, J. A., & Novati, D. C. (2012). Quantifying the contribution of NHL player types to team performance. *Interfaces*, *42*(2), 131-145.

Chiappori, P. A., Levitt, S., & Groseclose, T. (2002). Testing mixed-strategy equilibria when players are heterogeneous: The case of penalty kicks in soccer. *American Economic Review*, *92*(4), 1138-1151.

Engemann, K. M., & Owyang, M. T. (2009). A winning combination? Economic theory meets sports. *The Regional Economist*, (Jan), 10-13.

Hughes, D. W. (2017). *An Approximate Dynamic Programming Approach to Determine the Optimal Substitution Strategy for Basketball* (Doctoral dissertation, George Mason University).

Mid-America Intercollegiate Athletic Association. (n.d.). Mid-America Intercollegiate Athletics Association. Retrieved December 2017, from http://www.themiaa.com/calendar.aspx?path=wbball&

Romer, D. (2006). Do firms maximize? Evidence from professional football. *Journal of Political Economy*, *114*(2), 340-365.

Sampaio, J., Lago, C., Casais, L., & Leite, N. (2010). Effects of starting score-line, game location, and quality of opposition in basketball quarter score. *European Journal of Sport Science*, *10*(6), 391-396.

Sánchez, J. M., Castellanos, P., & Dopico, J. A. (2007). The winning production function: Empirical evidence from Spanish basketball. *European Sport Management Quarterly*, *7*(3), 283-300.

StatCrew (2018). *StatCrew Software Inc.*. Cincinnati, Ohio: CBS Sports Digital.

Synergy Sports. (2017). Synergy Sports. Retrieved July 2018, from https://www.synergysportstech.com/Synergy/Sport/Basketball/web/teamsst/Video/SelectGame2.aspx

Vaz de Melo, P. O., Almeida, V. A., Loureiro, A. A., & Faloutsos, C. (2012). Forecasting in the NBA and other team sports: Network effects in action. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, *6*(3), 23.

Winston, W. L. (2009). Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football. Princeton University Press. 201, 224-232.