Playing the Game: How Rookie Season Affects Basketball Career

Dan Rasay, Matthew Wong, Sari L Berger

Introduction

Basketball is not quite the American past-time, but it is still fun, especially to watch. For many, part of the fun of basketball (and other sports) is being able to follow player and team statistics, and trying to predict who is better than whom. Our paper is about using available measurements to determine if the amount of minutes played in a player's rookie season in the National Basketball Association influences their overall linear Player Efficiency Rating, or how good a player they are. This quantitative measure is useful for team coaches and management when determining the cost/benefit of deciding whether a rookie should play or not.

Literature Review

Unfortunately, basketball is woefully deprived of advanced statistics; most of the discussion around the value of players is relying on rather basic statistics. A system of evaluation that relies on more than points scored, rebounds gotten, and assists made is rather desperately needed.[i]

Some measures proposed are Offensive Rating, which is designed to measure how good a Team's offense is,[1] and Defensive Rating, which "how good a team's defense is regardless of pace."[2] Other extremely helpful statistics include what Dean Oliver called the four factors. Those factors include Effective Field Goal Percentage[3]; Turnover Percentage[4]; Offensive Rebounding Percentage;[5] and Free Throw Rate.[6] [ii]

Of course, measuring team efficiency is different than measuring players' efficiency. In the 1990s, ESPN columnist John Hollinger created a system by which someone can rank every player in the NBA based on their performance; he called it the Player Efficiency Rating (PER). Efficiency is broken down into offensive and defensive parts. Offensive efficiency is determined by points per possession, while defensive efficiency is determined by points allowed per possession. The primary units used to calculated the PER are points per shot attempt, the pure point rating, the assist ratio, the turnover rate, the rebound rate, and the usage rate. These units act as measures of performance.[iii]

Hollinger received criticisms for those claim; creators of other models claim that the metrics he uses are based more on belief regarding the players than actual statistics. Moreover, Hollinger has various weights for captured observations but never establishes that the chosen weights allow predictions on how many points a team scores or how many games the team wins.[iv] As a result, some people feel that Hollinger's metrics are not accurately representing efficiency of players, because it (like the NBA efficiency metrics), because they over-reward scoring without penalizing misses.[v]

We will be using a variant of Hollinger's equation; it uses linear weights to simplify the equation in the same fields, called the linear player efficiency rating (lPER). The lPER does account for misses by penalizing them. Our theory is that the more minutes played in the rookie season positively influences lPER rating in overall career.

Methods

Our source data was taken from the downloadable 1948 to 2008 NBA data ZIP archive on DatabaseSports.com. Individual CSV files were loaded into a Microsoft SQL Server database to transform data & calculate player lPER values for regression analysis. The source data included regular season player data at both the season and career level. Although the source data included player records dating back to 1948 only 1979+ records had all of the parameters needed to calculate lPER. Thus we limited our analysis data set to regular season player data only and players whose careers started 1979 or later and ended before 2008 (ie. last year of the source data).

Using the "RODBC" package in R we imported the necessary data into R to generate descriptive statistics/plots and run/validate our linear regressions. Each x y specification was verified by: 1) applying natural log to x, y and both x & y 2) scatter plotting residuals 3) generating a residual histogram.

The source data was transformed into two different data sets used for our regressions. The first data set was at the player career level with the unit of measure being an individual player. Initially we defined the independent variable as rookie season minutes played and the dependent variable as *career lPER*. After running and validating rookie season minutes played (MP) we also decided to look at rookie season average field goal attempts by game (AvgFGA) to determine if it also had an independent, positive effect on lPER. The second data set was at the individual season level with season and team dummy variables used to address temporal and spatial fixed effects. There is a unique observation for each player/season/team combination.

Players traded mid-season have a distinct observation for each team played on in a given season. Similar to the first data set, we used the same individual independent variables (Rookie MP & AvgFGA) but regressed against *season lPER* given the unit of measure (player/season/team). Both career and season lPER was calculated using Hein's formula below:

```
[ FGM * 85.910 + Steals * 53.897 + 3PTM * 51.757 + FTM * 46.845 + Blocks *
39.190 + Offensive_Reb * 39.190 + Assists * 34.677 + Defensive_Reb * 14.707 -
Foul * 17.174 - FT_Miss * 20.091 - FG_Miss * 39.190 - TO * 53.897 ]
* (1 / Minutes)
```

Note that additive measures have been highlighted in blue and subtractive measures in red.

Coaches and team management have the difficult job of winning as many games as possible and managing team/player morale all while developing younger talent. We chose to look at rookie season minutes played since coaches could easily incrementally increase time played which could influence the rookie player's effectiveness later in his career (lPER).

Results

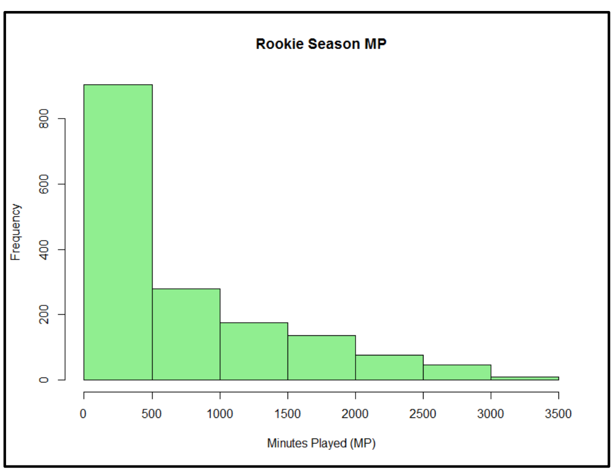
The two tables below show descriptive statistics for our two primary data sets (career and season). Values for RookieMP and RookieAvgFGA were not included in the season level table since the unit of measure was different.

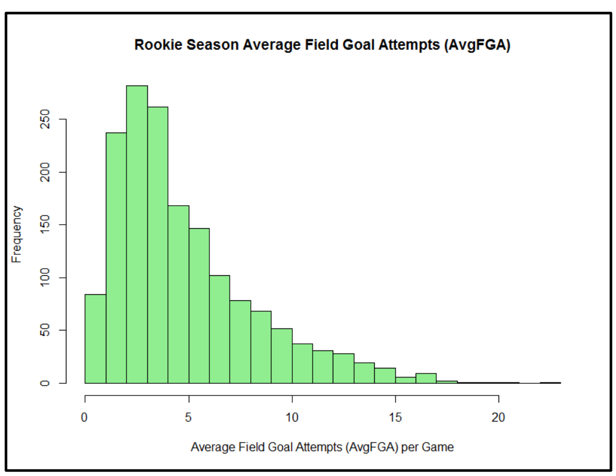
Descriptive statistics - Career level

Statistic	RookieMP	RookieAvgFGA	Career_lPER
N	1,630	1,630	1,630
Mean	693.2245	4.8160	11.1459
St. Dev.	747.3523	3.4085	7.6497
Min	1	0.0000	-56.3456
Max	3,249	22.8890	89.8000

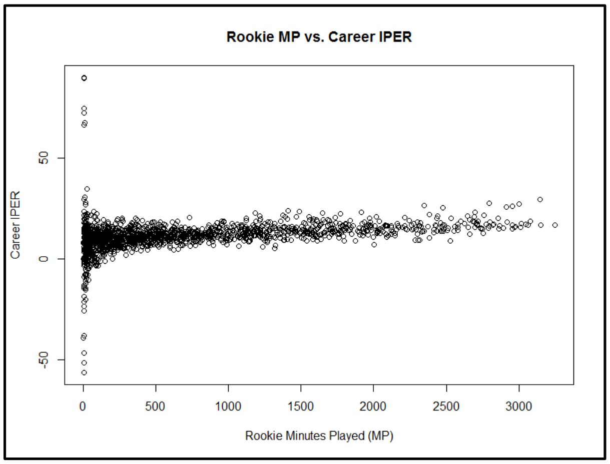
Descriptive statistics - Season level

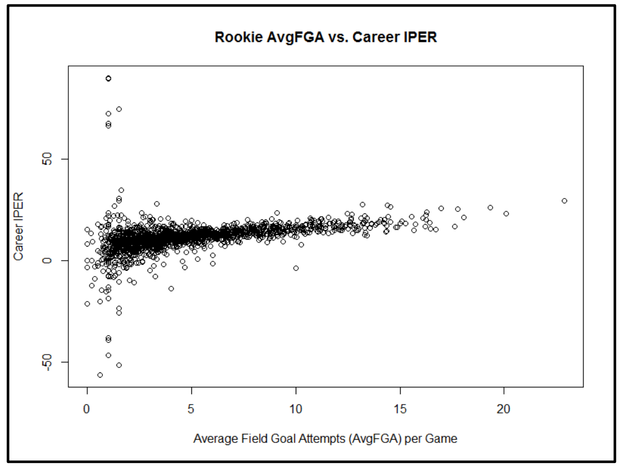
Statistic	Season_lPER
N	9,333
Mean	12.3477
St. Dev.	7.4071
Min	-193.1948
Max	89.8000



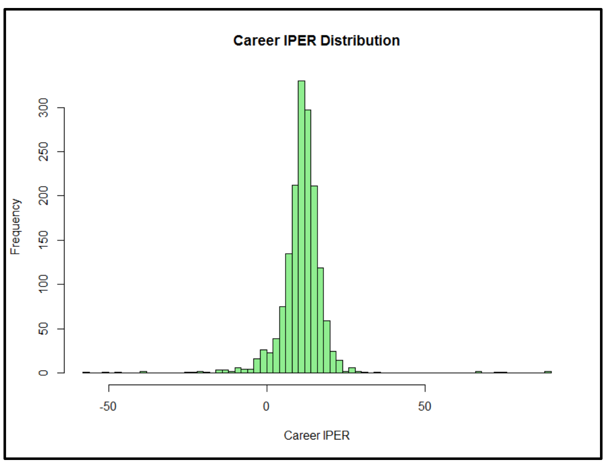


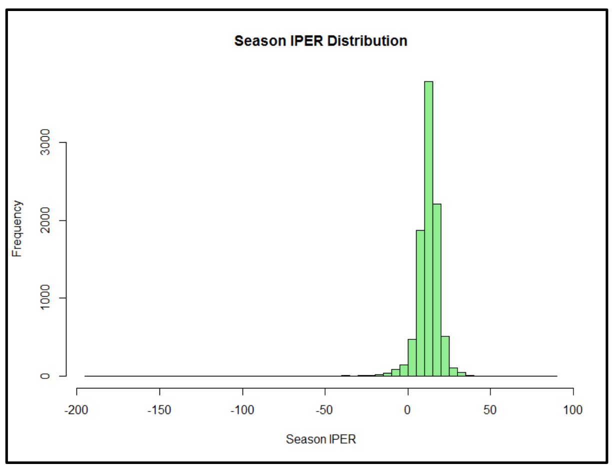
The rookie season MP histogram shows that close to half of rookies play fewer than 500 minutes in their first year. There appears to be a log normal distribution of rookie season AvgFGA.





Scatter plots for both the rookie MP and AvgFGA show a positive relationship with career lPER. While there appear to be outliers close to zero in both plots there appears to be a strong positive correlation. Rookie MP and AvgFGA have a correlation with career lPER of 0.3333 and 0.4099 respectively.





Based on their histograms both career and season lPER appear to be normally distributed.

Career Level Regression Analysis

Initially we regressed rookie MP against career lPER (regression #1 in table below) and found a statistically significant coefficient (p=0.0002, f-statistic=203.490). Given that the coefficient was relatively small we then decided to regress rookie AvgFGA against career lPER (regression #2) which yielded a larger coefficient (0.920) and was also statistically significant (p=0.051, f-statistic=328.820). In an attempt to control for simultaneity (ie. better players get more playing time and bad players get less playing time) we also performed MP and AvgFGA regressions on the middle 90% of players. We observed similar and slightly more significant results (p-value, r^2 & f statistic) with career regressions #3 & 4.

Rookie Season Minutes Played (MP) & Average Field Goal Attempts (AvgFGA) impact on Career lPER

	Dependent variable:			
	Career lPER (all players)		Career lPER (middle 90% players)	
	(1)	(2)	(3)	(4)
Rookie MP	0.003^{***}		0.002^{***}	
	(0.0002)		(0.0001)	
Rookie AvgFGA		0.920^{***}		0.741^{***}
		(0.051)		(0.025)
Constant	8.781^{***}	6.715^{***}	9.615^{***}	7.837^{***}
	(0.244)	(0.299)	(0.117)	(0.140)
Observations	1,630	1,630	1,466	1,466
R^2	0.111	0.168	0.241	0.379
Adjusted R^2	0.111	0.168	0.241	0.379
Residual Std. Error	7.214 (df = 1628)	6.980 (df = 1628)	3.203 (df = 1464)	2.897 (df = 1464)
F Statistic	203.490^{***} (df = 1; 1628)	328.820^{***} (df = 1; 1628)	465.597^{***} (df = 1; 1464)	894.808^{***} (df = 1; 1464)
Note:				*p<0.1; **p<0.05; ***p<0.01

Based on the MP coefficient for every one minute played a player's career lPER increases by 0.0034. A player's career lPER is modeled with the equation below:

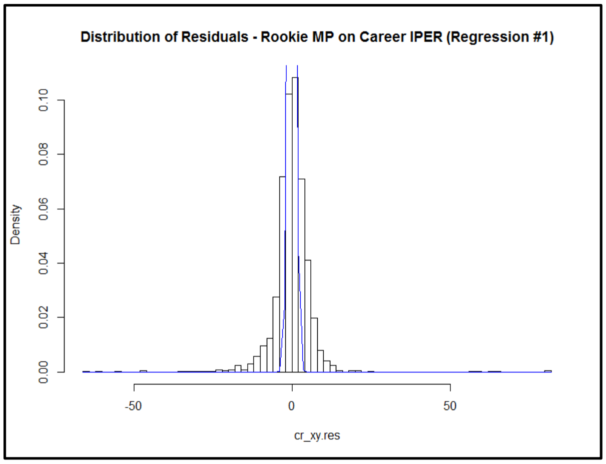
```
career lPER =  8.781 + 0.003411851 * rookie season minutes played
```

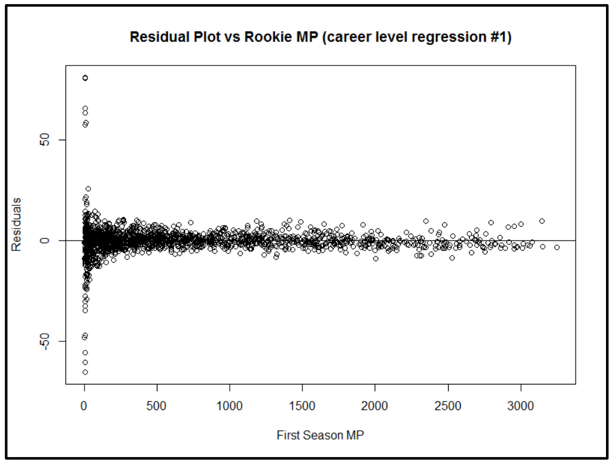
Thus if the average player who plays 693 minutes in his rookie season with a projected 11.146 career lPER plays an additional four minutes per game over an 82 game season the player's career lPER would increase by approximately 1.11 to 12.256. (All teams 82 games per season, with 12 minutes quarters per game.) Based on the AvgFGA coefficient for every additional field goal attempt per game a player's career lPER increases by 0.0.920. Thus player's career lPER is modeled with the equation below:

```
career lPER =  6.715 + 0.920 * average field goal attempt per game
```

Thus if the average player who has a AvgFGA of 4.816 in his rookie season with a projected 11.14 career lPER increase their FGA average by 1 the player's career lPER would increase by approximately 1.11 to 12.065.

Below are a residual histogram and plot against x for career regression #1.





Residuals for this regression were distributed normally and the plot of residuals against rookie has a handful of outliers close to zero but does not show any patterns.

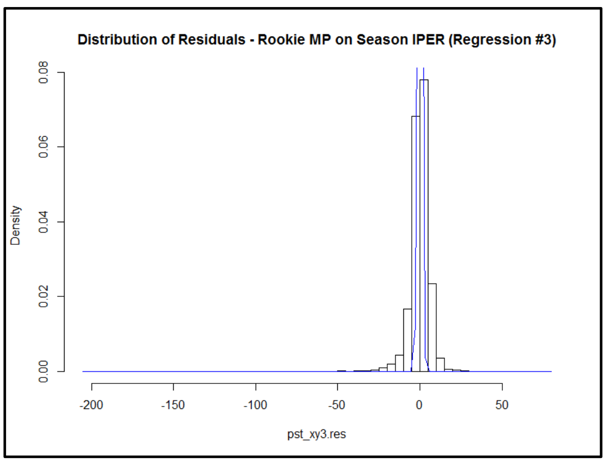
Season Level Regression Analysis

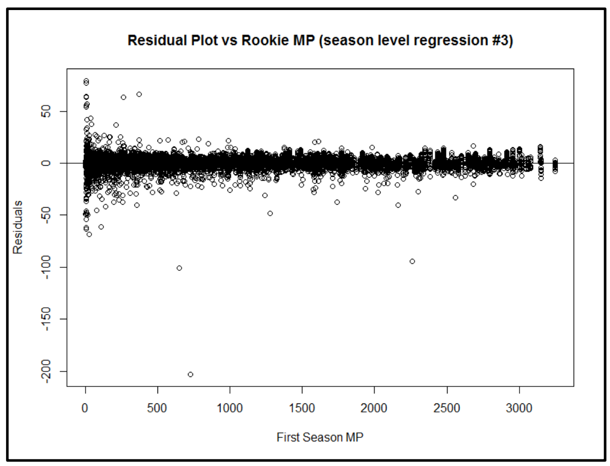
Given the career level unit of measure we were unable to use panel data to control for fixed effects. In an attempt to control for fixed effects we ran regressions with (regressions 3,4,7,8) and without (regressions 1,2,5,6) dummy variables for season and team. Since players could be traded mid-season our unit of measure had to be player/season/team and our dependent variable unit of measure became season lPER. Similar to the career level regressions, we also performed regressions using the middle 90% of players to control for simultaneity. Across the board we found $p<0.01$ statistically significant coefficients & f-statistic values.

Rookie Season Minutes Played (MP) & Average Field Goal Attempts (AvgFGA)
Impact on Season lPER

	Dependent variable:							
	Season lPER (all players)				Season lPER (middle 90% players)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Rookie MP	0.003***		0.002***		0.002***		0.002***	
	(0.0001)		(0.0001)		(0.0001)		(0.0001)	
Rookie AvgFGA		0.778***		0.752***		0.623***		0.606***
		(0.018)		(0.018)		(0.020)		(0.020)
Dummy Variables	---- n/a ----		---- Year & Team ----		---- n/a ----		---- Year & Team ----	
	Note: Coefficients for year and team dummy variables are not shown for brevity							
Constant	9.547***	7.018***	8.993***	7.029***	9.895***	7.885***	9.495***	7.879***
	(0.119)	(0.144)	(0.957)	(0.921)	(0.111)	(0.145)	(0.882)	(0.864)
Observations	9,333	9,333	9,333	9,333	8,587	8,587	8,587	8,587
R^2	0.088	0.161	0.133	0.202	0.058	0.101	0.096	0.139
Adjusted R^2	0.087	0.161	0.127	0.196	0.057	0.101	0.089	0.132
Residual Std. Error	7.076 (df = 9331)	6.783 (df = 9331)	6.922 (df = 9266)	6.640 (df = 9266)	6.374 (df = 8585)	6.224 (df = 8585)	6.268 C (df = 8520)	6.117 (df = 8520)
F Statistic	895.106*** (df = 1; 9331)	1,796.900*** (df = 1; 9331)	21.491*** (df = 66; 9266)	35.572*** (df = 66; 9266)	523.853*** (df = 1; 8585)	969.217*** (df = 1; 8585)	13.632*** (df = 66; 8520)	20.766*** (df = 66; 8520)
Note:								*p<0.1; **p<0.05; ***p<0.01

Below are a residual histogram and plot against x for season regression #3.

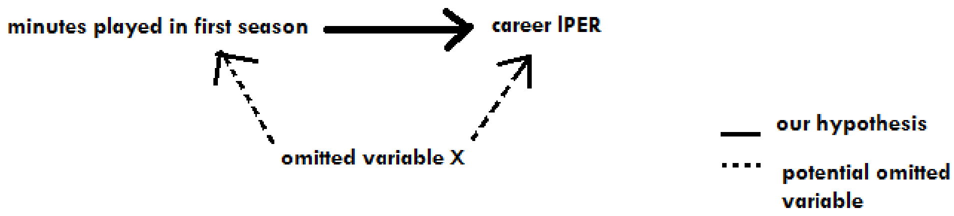




Residuals for this regression were distributed normally and the plot of residuals against rookie MP does not show any patterns.

Conclusion:

There is a statistically significant connection between a player's minutes played in the rookie season and their career lPER. Given that the f statistic values for the omitted variable bias were very significant, we cannot reject the idea that minutes played in first season are one of the primary forces behind career lPER; that said, future research should consider variables other than season and team. Factors that we did not account for that may also help explain the connection include height, coaches, college basketball and playoff experience, and injuries.



Part of the reason we did not run regressions on these dummies was lack of available data. To control for outliers and simultaneity, we refined the data by removing some of the outliers at either end that may have skewed our results.

Obviously, the degree to which someone plays will affect how good they are. Our paper worked to answer the degree to which an increased amount of play time improved a player's efficiency rating, and we found there was a significant connection. Overall, the meaning of our results is that allowing rookies to play more minutes in their rookie season will result in an overall improvement in the rest of their career. However, it could mean that a team will give more minutes to an inherently talented rookie player, which would mean that raw talent is an omitted variable that we tried to control.

What coaches want matters. On the one hand, they want to prepare their rookie players for the future, but they also want to win current games. Using the lPER gives them a method of cost benefit analysis that measures the future benefit of letting rookie players have more minutes in their first season; that benefit is tangible enough that coaches should consider giving rookies enough minutes that they will improve.

---

1 Using the formula ((points scored * 100)/possessions)
2 Using the formula ((points surrendered * 100)/possessions)
3 Which measures shooting percentage when accounting for three-pointers.
4 As measured by turnover per 100 possessions.
5 How often a team rebounds its own missed shots.
6 How well a team gets to the free throw line and converts the freebies.

END NOTES

i. Zachariah Blot, "Advanced Basketball Statistics 101," Empty the Bench, last modified December 2009, http://www.emptythebench.com/2009/12/11/advanced-basketball-statistics-101/
ii. Ibid
iii. "Calculating PER", Basketball Reference, last modified 2015 http://www.basketball-reference.com/about/per.html
iv. Ibid
v. DJ, "Do We Overvalue Rebounds?," The Wages of Wins Journal, last modified November 9, 2006 "http://wagesofwins.com/2006/11/09/do-we-overvalue-rebounds/

Works Cited

Zachariah Blot, "Advanced Basketball Statistics 101," Empty the Bench, last modified December 2009, http://www.emptythebench.com/2009/12/11/advanced-basketball-statistics-101/

"Calculating PER", Basketball Reference, last modified 2015 http://www.basketball-reference.com/about/per.html

DatabaseSports.com, http://www.databasebasketball.com/

DJ, "Do We Overvalue Rebounds?," The Wages of Wins Journal, last modified November 9, 2006 "http://wagesofwins.com/2006/11/09/do-we-overvalue-rebounds/