Predicting NBA Players’ Time on the Court Based on Their Pick in The Draft

Noam Benkler and Christian Zaytoun

ABSTRACT:

Basketball is a unique sport in the fact that it is both high scoring sport and each team is only allowed to have five players compete at a time. The overall value of a player is based on aggregate ability rather than isolated plays. In this paper we analyze how well the NBA Draft is able to correlate to the overall utility of a player. This utility is measured by average minutes played per game. It follows the idea that the best players need to be on the court as much as possible. Looking at individual career statistics of 280 players picked in the 2010-2015 drafts we were able to determine that there is a significant association between where a player is picked in the draft and how many minutes he will average per game throughout his NBA career using variable like the position and average points, rebounds, and assists per game.

INTRODUCTION:

The National Basketball Association (NBA) maked nearly 7.4 billion dollars of revenue in the 2016-2017 season. A number that has been growing at a rate of around fifteen percent over the last five seasons. Its athletes, a total of 491 on opening day rosters of the 2017-2018 season, earn an average of 6.2 million dollars per year from team contracts alone which is predicted to break the 10 million dollar average by the 2020-2021 season. The NBA has become the highest paying professional league in the United States. Its viewership is higher than ever, its average ticket price has increased around 5% since the 2011-2012 season. The average franchise valuation has blown up from around \$300 million to nearly \$1 billion in the last eight years, and the average revenue for a franchise has increased nearly 20% since the 2010-2011 season.

These numbers are of course all outputs of many different factors. There was the professional sports world version of a strike in the NBA just before the 2011-2012 season. This lockout was due to discrepancies in the revisions of the collective bargaining agreement between the NBA players and the NBA franchise owners. The agreements made ranged from practice schedules to injury protocols to compensation and taxes. Other factors in this NBA growth could include advancements in media technology. Now following a team throughout the season is more accessible than ever with apps on mobile devices and a broadening to access of Network TV. However, one conclusion we can make by the historical league revenue is that the game of basketball has never been as good a product as it is now.

Through our research we look at where the product of the NBA comes from, it's players. The NBA does not follow your typical hiring process. It, just like all US professional sports, has a draft for all potential new athletes who wish to join professional basketball in the United States. The draft is a way to equalize competitiveness within the league. Essentially, the best team gets to pick last in the draft, and the worst team gets to pick first. Therefore, the worst team will in theory get the best player and thus boost their competitiveness in the next season.

This research paper will discuss the question of how well the draft system does in maintaining its ideology of competitiveness in the NBA. In attempts to do this we will analyze how the draft pick of a player relates to the number of minutes he averages per game. We use number of minutes per game as a way of measuring the effectiveness and value of a player. Unlike many sports, basketball is less about strategic execution and more about aggregate output to win a game. For example, a kicker can have significant contribution in a football game and yet only be included in 1% of the plays. In basketball there is a true positive relationship between impact on a game and number of minutes played in that game.

DATA:

Our data looks at every player drafted in the NBA between the 2010 draft (first season being 2010-2011) and the 2015 draft (first season being 2015-2016) along with their accumulated statistics while in the NBA. These statistics record the cumulative performance measures accrued in all NBA regular and postseason games played in the careers of each player drafted between 2010 and 2015 and include all games up to the most recent regular 2017-2018 season (last update on April 13th, 2018). No playoff statistics for the 2017-2018 season have been incorporated. Individual statistics are recorded live during each game and then certified afterwards. The data set used in this research project came from Sports Reference LLC, a company dedicated to publishing the most complete set of sports statistics for general use. Sports Reference LLC uses the official book of the NBA to compile its statistics.

Table 1: Summary Statistics of NBA Data for Players Drafted between 2010 and 2015

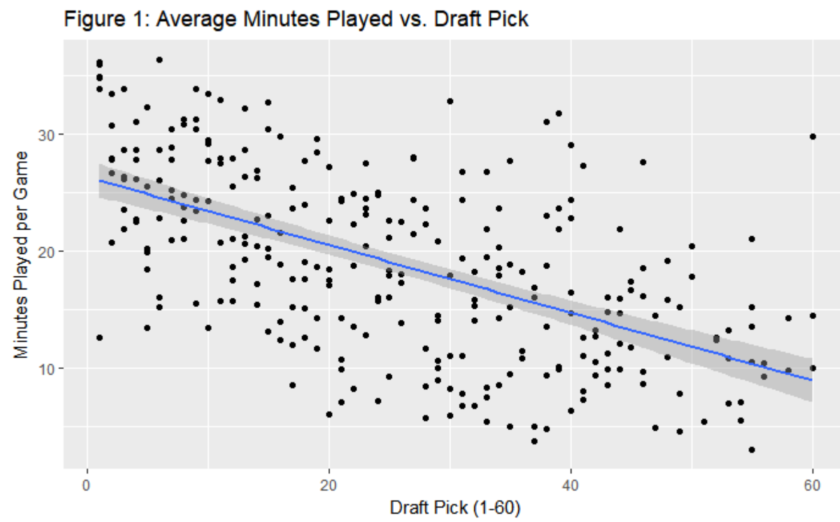
Statistic	N	Mean	St. Dev.	Min	Max
Year	309	2,012.430	1.671	2,010	2,015
Rd	309	1.424	0.495	1	2
Pk	309	27.061	16.023	1	60
Age	309	20.976	1.410	18.208	27.148
From	309	2,013.686	1.804	2,011	2,018
To	309	2,016.793	1.897	2,011	2,018
G	309	214.987	158.930	1	613
MP	309	17.763	8.433	2.000	36.400
PTS	309	7.106	4.959	0.000	23.400
TRB	309	3.223	2.271	0.000	13.400
AST	309	1.432	1.303	0.000	9.200
STL	309	0.565	0.356	0.000	1.800
BLK	309	0.389	0.412	0.000	2.400
FG.	308	0.434	0.089	0.000	0.800
WS	309	9.665	12.796	-2.000	62.500
WS.48	309	0.062	0.082	-0.597	0.291

Table 1 provides a summary of of all the variables in our dataset and Table 2 provides the summary statistics for our unrestricted linear model. Each case records one of the 309 players who were drafted in between the 2010 and 2015 season omitting players who never made an appearance. Due to strong skewness in number of assists per game and total rebounds per game we decided to log those two variables. Histograms of the relevant original and log transformed variables can be seen in the appendix.

Table 2: Model Coefficients for Unrestricted Model

Variable	Coefficient	SE
Pk (Draft Pick)	-0.001	(0.011)
Age (Age when Drafted)	0.074	(0.108)
PosC-F (Position, Center Forward)	-0.094	(0.703)
PosF (Position, Forward)	1.297**	(0.645)
PosF-C (Position, Forward Center)	0.158	(0.677)
PosF-G (Position, Forward Guard)	1.777*	(0.924)
PosG (Position, Guard)	0.956	(0.767)
PosG-F (Position, Guard Forward)	2.256***	(0.808)
Year (Year Drafted)	0.233**	(0.096)
G (Games Played)	0.016***	(0.002)
PTS (Career/Game)	0.907***	(0.056)
TRB (Rebounds/Game)	0.590***	(0.139)
AST (Assists/Game)	0.430**	(0.177)
STL (Steals/Game)	4.968***	(0.647)
BLK (Blocks/Game)	0.581	(0.549)
FG. (Field Goal Percentage)	-1.144	(3.463)
WS (Win-Share Statistic)	-0.130***	(0.022)
WS.48 (Win-Share/minute played)	2.805	(3.209)
Constant	-467.926**	(192.612)
Observations	280	
R2	0.934	
Adjusted R2	0.930	
Residual Std. Error	2.107 (df = 261)	
F Statistic	206.411*** (df = 18; 261)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Figure 1 shows simple linear regression of pick in the draft to minutes played. We incorporate the other statistics as explanatory variables in our model to more clearly understand the relationship between pick in the draft and number of minutes played.



RESULTS:

We used a stepwise selection process to arrive at the final linear model of

$$Minutes\ Played = \beta_0 + \beta_1 Pick + \beta_2 Position + \beta_3 PointsScored + \beta_5 log(Rebounds) + \beta_6 log(Assists) + \beta_7 Steals + \beta_8 Age$$

where we examined the relationship between Pick and Minutes Played. The estimates of the model parameters are given in Table 3. Our results showed that all else held constant a 1 spot increase in the round in which a player was picked for the draft is associated with an 0.031 minute decrease in mean number of minutes played, and that with 99% confidence the true value of β_1 lay between -0.004 and -0.0578. Other variables that significantly affected minutes played at $\alpha < 0.01$included points scored, number of steals, log(total rebounds), log(assists), and every position other than Forward Center and Center Forward. All else held constant a player being a Forward was associated with a 1.557 minute increase in mean number of minutes played, and that with 99% confidence the true value of $\beta_2(F)$ lay between 2.970 and 0.145. All else held constant a player being a Forward Guard was associated with a 2.707 minute increase in mean number of minutes played, and that with 99% confidence the true value of $\beta_2(F - G)$ lay between 4.830 and 0.585.

Table 3: Model Coefficients

Variable Name	Coefficient	SE
Pk	-0.031***	(0.010)
PosC-F	-0.226	(0.679)
PosF	1.557***	(0.544)
PosF-C	-0.077	(0.630)
PosF-G	2.707***	(0.818)
PosG	1.937***	(0.745)
PosG-F	2.813***	(0.704)
PTS	0.751***	(0.047)
log(TRB)	3.888***	(0.386)
log(AST)	1.660***	(0.295)
STL	3.194***	(0.603)
Age	0.168*	(0.101)
Constant	2.903	(2.253)
Observations	280	
R2	0.938	
Adjusted R2	0.936	
Residual Std. Error	2.018 (df = 267)	
F Statistic	339.011*** (df = 12; 267)	
Note:	*p<0.1; **p<0.05; ***p<0.01	

All else held constant a player being a Guard was associated with a 1.937 minute increase in mean number of minutes played, and that with 99% confidence the true value of $\beta_2(G)$ lay between 0.005 and 3.869. All else held constant a player being a Guard Forward was associated with a 2.813 minute increase in mean number of minutes played, and that with 99% confidence the true value of $\beta_2(G-F)$ lay between 4.640 and 0.987. All else held constant every additional point a player scored was associated with a 0.751 minute increase in mean number of minutes played, and that with 99% confidence the true value of β_3 lay between 0.871 and 0.630. All else held constant a doubling of Total Rebounds a player got was associated with a 2.69 minute increase in mean number of minutes played, and that with 99% confidence the true value of β_5 lay between 4.889 and 2.888. All else held constant a doubling of Assists a player received was associated with a 1.151 minute increase in mean number of minutes played, and that with 99% confidence the true value of β_6 lay between 2.426 and 0.895. All else held constant every additional Steal a player had was associated with a 3.194 minute increase in mean number of minutes played, and that with 99% confidence the true value of β_7 lay between 4.759and 1.629. Finally at an α level of 0.01 we found that, all else held constant, every additional year in a players age was associated with a 0.168 minute increase in the mean number of minutes played, and that with 90% confidence the true value of β_8 lay between 0.4301644 and -0.09462437. The rest of the variables in our model, as well as our intercept, were not significantly different from 0. Finally, we found that the model explains 93.6% of the variability in Minutes Played.

DISCUSSION:

Our regression final model meets all the assumptions necessary for accurate linear regression. The model plot (Figure 2), residuals plot (Figure 3), and qq-plot (Figure 4) created from our data, showed the data conformed to a normal distribution, with independence, and constant error variance. Despite conforming to these assumptions and having p-values below 0.01 on most of our model coefficients, there were several points outside of our prediction interval (Figure 4), which make it difficult to conclude accurate predictions 100% of the time. Moreover,

Figure 2: Residual Plot of Minutes Played to Pick in the Draft

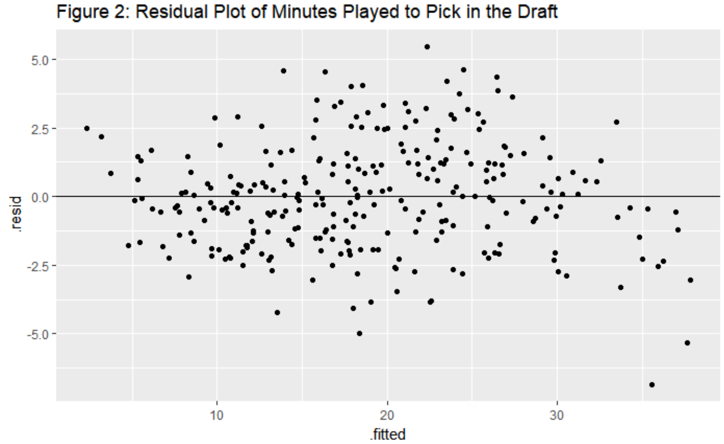
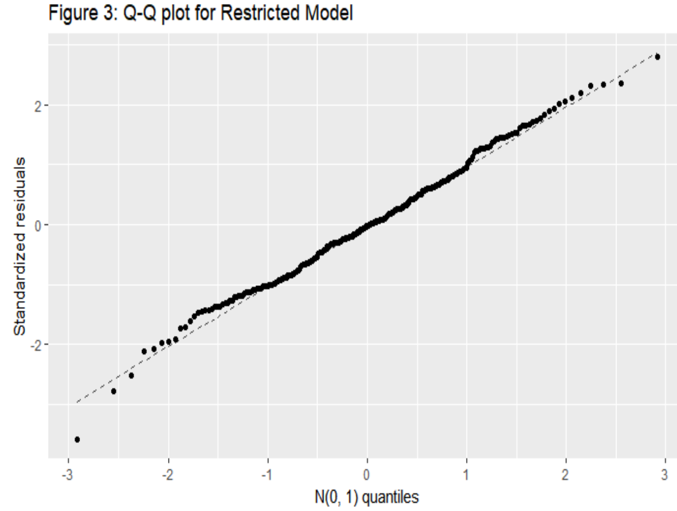
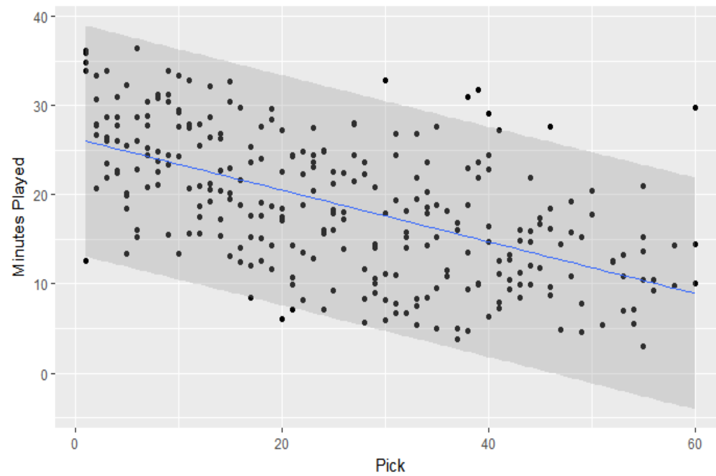


Figure 3: Q-Q plot for Restricted Model



other possible problems with the model involve the lack of normality in the other explanatory variables we used (histograms of the other variables can be seen in the appendix). However, given our model predicts 93.6% of the variation in minutes played, and tests for auto collinearity and heteroskedasticity (shown in the appendix) showed neither problem occurred in our model, it is fair to say that with some error, our model is a fair predictor of the amount of playing time a player will get given their pick in the draft.

Figure 4: Plot of Minutes Played to Pick in the Draft Prediction Interval



REFERENCES:

Badenhausen, Kurt. “The Average Player Salary And Highest-Paid In NBA, MLB, NHL, NFL And MLS.” *Forbes*, Forbes Magazine, 9 Aug. 2017, www.forbes.com/sites/kurtbadenhausen/2016/12/15/average-player-salaries-in-major-american-sports-leagues/#f87704105058.

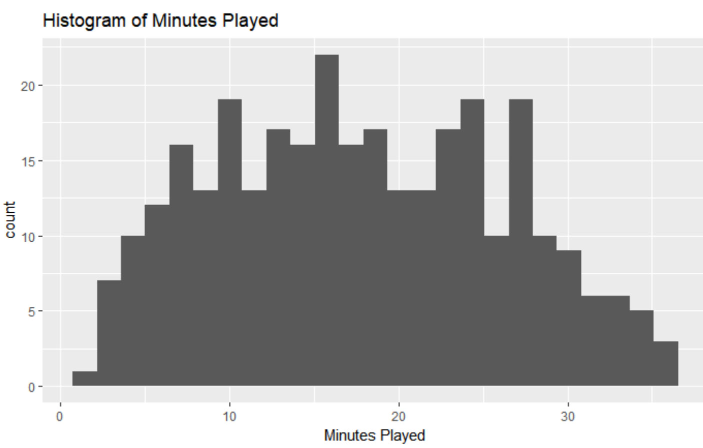
“Draft Finder.” *Basketball-Reference.com*, Sports Reference LLC, 13 Apr. 2018, www.basketball-reference.com/play-index/draft_finder.cgi.

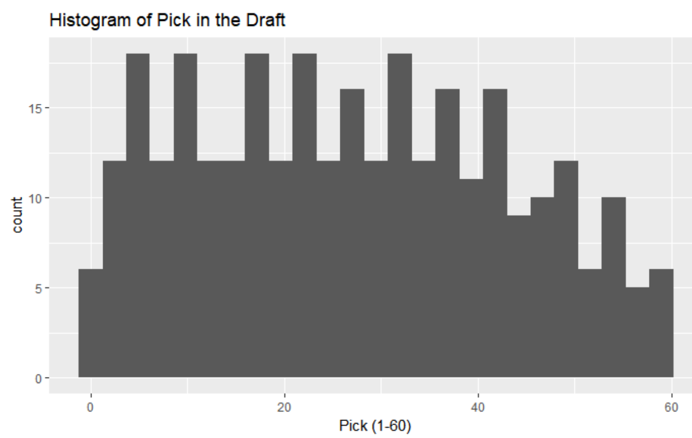
Gunnion, Lester. “Behind the Numbers: Professional Sports and the Merits of Being Big and Connected.” *Deloitte United States*, Deloitte, 7 Aug. 2015, www2.deloitte.com/insights/us/en/economy/behind-the-numbers/us-professional-leagues-sports-and-technology.html.

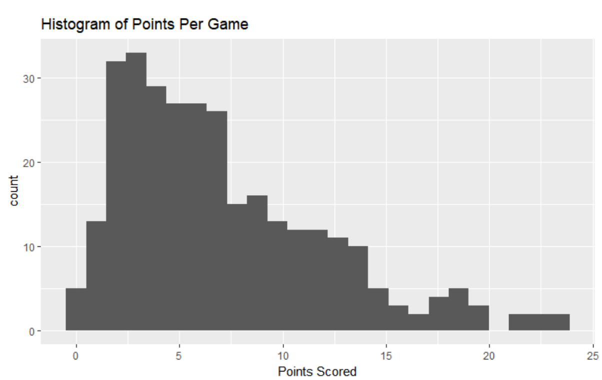
Tucker, Matt. "Pricing Out NBA Fans?" *Sonics Rising*, SBNation, 2 Oct. 2014, www.sonicsrising.com/2014/10/2/6868681/pricing-out-nba-fans-how-ticket-prices-compare-to-other-leagues.

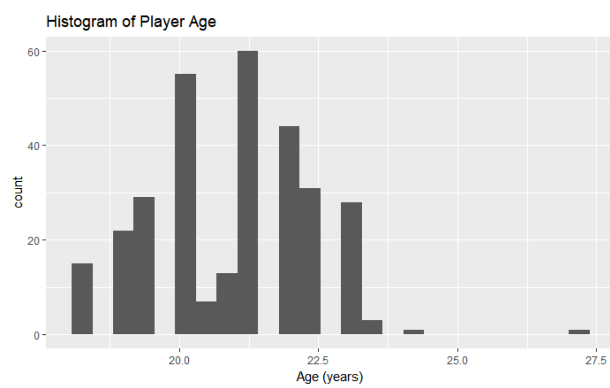
APPENDIX:

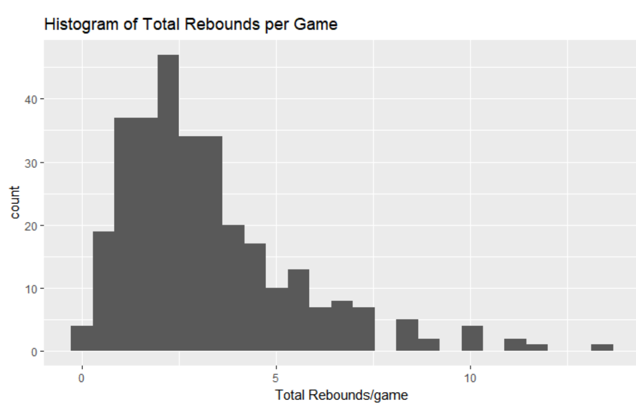
Histograms of Relevant Data:

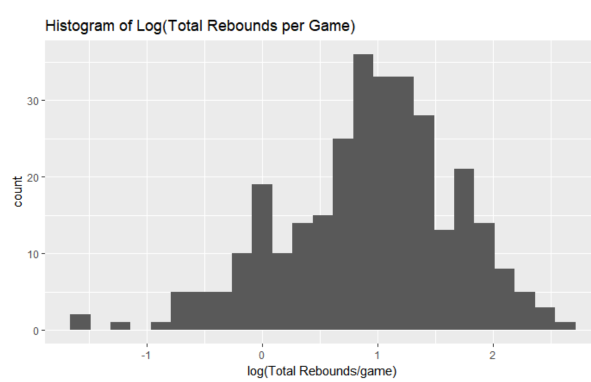


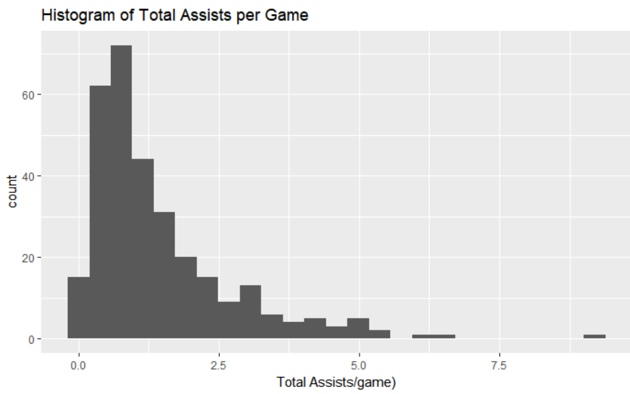


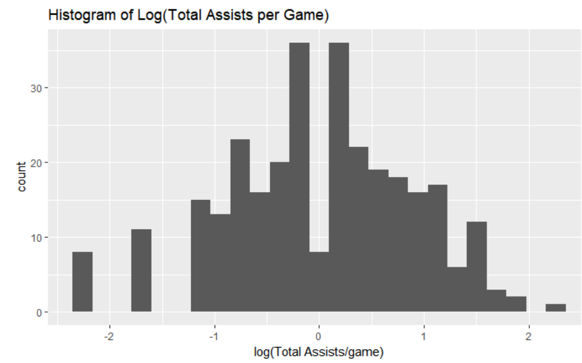


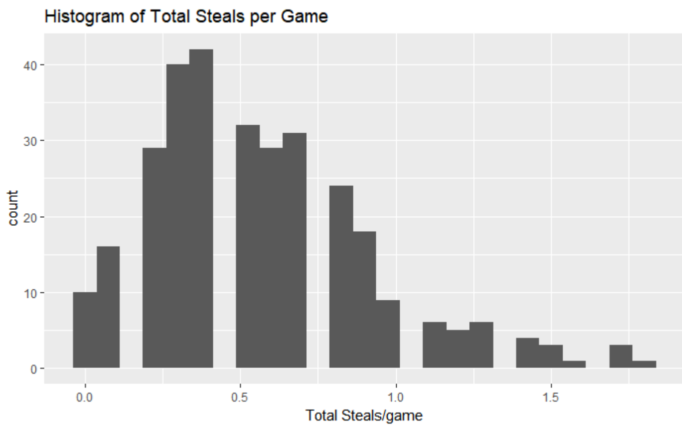












Tests for Autocorrelation in the Restricted Model: Both Show no Autocorrelation

```
1. Durbin-Watson test

data:  new.mod
DW = 1.9844, p-value = 0.427
null hypothesis: true autocorrelation = 0
alternative hypothesis: true autocorrelation is greater than 0
        Fails to reject the Null Hypothesis

2. Breusch-Godfrey test for serial correlation of order up to 1

data:  new.mod
LM test = 0.0035756, df = 1, p-value = 0.9523
null hypothesis: serial correlation order = 0
alternative hypothesis: serial correlation of order up to 1
      Fails to reject the Null Hypothesis
```

Breusch-Pagan Test for Heteroskedasticity in Restricted Model:
Indicates no significant Heteroskedasticity

```
studentized Breusch-Pagan test

data:  new.mod
BP = 13.464, df = 12, p-value = 0.3363
null hypothesis: no heteroskedasticity present in model
alternative hypothesis: heteroskedasticity greater than 0
        Fails to reject the Null Hypothesis
```