The San Antonio Spurs & CATALYST

Nicholas K. Neuteufel

The University of North Carolina at Chapel Hill

http://www.sportsandillumination.com

nneuteufel@gmail.com

August 14, 2014

Abstract

In this paper, I present PEAR and CATALYST—passing metrics I demonstrate to be superior to assist-based models of passing. PEAR and CATALYST build on a points created framework, solving the predicament of assists being blind to both the quality and expected value of a shot assisted. CATALYST takes passing evaluation farther by adjusting for teammate quality by using a player-independent team shooting percentage–exclusive true shooting percentage (xTS%). I begin by demonstrating the statistical and assumptive foundations of PEAR and CATALYST to be solid. I then show a variety of models that demonstrate the superiority of both PEAR and CATALYST to AST% in predicting a player's offensive value in plus/minus terms (both offensive xRAPM and RPM). These conclusions are shown to be robust to any heteroskedasticity and/or autocorrelation in the data. I also explore a variety of ways that proprietary versions of SportVU may improve PEAR/CATALYST to new heights. I end with a discussion of possible extensions of PEAR and CATALYST into intriguing innovative methods of team, line-up, and scouting analysis.

1 Introduction

Conventional passing statistics fail to describe the true impact of passing and its importance to the game of basketball. One typical metric for seeing the impact of passing is assists per game or its derivatives such as pace-adjusted assists and assists per time period interval. The two main criticisms I discuss below also apply equally to assist percentage (AST%).

One failure of the assists model is the failure to distinguish between the type of passes and assists a player makes. APG does not distinguish among assisting on open lay-ups, dunks, corner three-pointers (3s), and other types of shots. Assists thus do not tell you how many points a player creates for his team/teammates. Tony Parker, for instance, had 388 assists during the 2013-2014 NBA regular season.[1] His points created by assist, however, could range from 776 (if all of his assists were two-point field goals) to 1,164 (all 3s). This is a problem for comparing how important Tony Parker's passes are compared to Chris Paul's in terms of points created for their teammates.

Another failure of assists is that assists are blind to the quality of a player's team-mates. Take Rookie of the Year for the 2013-2014 NBA season, Michael Carter-Williams of the Philadelphia 76ers. To put it generously, MCW has not been playing with players of an MVP quality, like LeBron James and Kevin Durant. Many of his teammates were not All-Star quality or even close. This meant that MCW could have been making Tony Parker-quality passes, but his teammates did not have the offensive efficiency or skills of Parker's teammates to convert those passes into points created or even assists.

2 CATALYST: Foundations & Improvements

CATALYST was an attempt to build upon recently publicly-available SportVU player tracking data to solve the problems of conventional passing statistics. It builds upon a statistic of my own called exclusive True Shooting percentage (xTS%) to create a universal as well as contextual passing statistic.[2] I believe CATALYST is a good building block towards an incredible contextual offensive basketball statistic.

CATALYST begins with pass efficiency, a concept inspired by Phillips.[4] Pass efficiency (PE) is simply the points created by a player's passes divided by the number of passes he made. This is intended to show how "strategic" a players passes are within the team system, as the intention of all passing is to create good looks for one's teammates or to otherwise set up offensive action. The goal of an offensive system is to score as efficiently as possible, implying that direct passes to good scoring opportunities should be used if available. Redundant or unnecessary passes increase the risk of turnovers, which harm an offense's efficiency and reduce the probability of victory.

The general formula for pass efficiency is

$$PE = \frac{PCA}{Passes}$$

whereby PE is Pass Efficiency and PCA is Points Created by Assist. However publicly-available `NBA.com` data only provides that data in per-game form. Fraction division alleviates some of the issue, but the issues of significant figures and rounding errors are very real in the data presented in this report.

Recommendation 1: The Spurs can use CATALYST (and its improvements) after taking out relatively irrelevant passes, such as in-bounds or back-court passes. I call this "Half-court CATALYST." (Again, I do not have access to proprietary data, only publicly-available data.)

Recommendation 2: The Spurs should use the season/playoff totals to calculate CATALYST and its improvements/derivatives. This would take away rounding error and significant figures discrepancies.

The general formula for pass efficiency is nice but lacks two powerful uses: player comparison and universality. It is hard to compare players directly in the vacuum PE presents. Two steps were taken to make pass efficiency into the more powerful and useful CATALYST.

(1) Pass efficiency was calculated for all players in the NBA and the arithmetic mean was taken. The resulting average PE of all NBA players during the 2013-2014 NBA regular season was 0.148058 points (created by assist) per pass. Dividing a players PE by that average (0.148058) gives his PEAR—Pass Efficiency Above Replacement.[1] [2] PEAR thus is contextual—creating a context in which to view the passing of a specific player against the average NBA player

(2) PE and PEAR still rely upon teammate quality, hampering its use as a universal comparison tool. The novelty of CATALYST is weighing a players PEAR by a comparison of the league average true shooting percentage (TS%) to a players xTS%, or exclusive True Shooting percentage.[2]

xTS% is a stat of my own creation that is essentially the TS% of a player's teammate-group (in other words, the TS% of all Spurs not named Tony Parker). (See earlier citation for more on xTS%.) xTS% takes into account how well a players teammate-group shoots the ball—helping Michael Carter-Williams and hurting Chris Paul, as MCW has neither DeAndre Jordan nor Blake Griffin to pass to. The additional advantage of xTS% is not using team TS%, recognizing that a player does not usually assist himself. By comparing a player's xTS% to the league average TS%, CATALYST recognizes the value of teammate quality in creating points to be scored on a player's assists.

Thus the formula for simple CATALYST:

$$CATALYST = PEAR * \frac{LgAvgTS\%}{xTS\%}$$

One simple improvement upon CATALYST is the addition of the value of free throw assists. I call this Free Throw CATALYST. An approximation of Free Throw CATALYST is gained from simply adding the number of Free Throw assists to the Points Created by Assist(s) term in PE and PEAR, using the inherent value of FTs being 1 point.

$$PEAR_{FreeThrow} = \frac{PCA + FreeThrowAssists}{NumberOfPasses}$$

Free Throw CATALYST is then calculated as simple CATALYST is, comparing a player's PEAR using his xTS% below the league average TS%. The use of xTS% here is useful to control for teammate quality in shooting free throws. One example would be the Houston Rockets. If Jeremy Lin were to make a great entry pass to Dwight Howard in a position where the opposing team felt the need to foul Howard rather than allow him to get to the rim, Lin should be credited with the points created (if any). However, he is hurt for making the right decision because of Dwight Howard's inability to shoot FTs. There is some adjustment for this by the use of a player's xTS%, which includes both Howard's poor FT% and the proportion of FTAs of all of his true shot attempts. This adjustment, however, can be improved.

One improvement proprietary data can make to CATALYST is to include points created by secondary (aka "hockey") assists. This "extended CATALYST" can provide

[1] A better name is actually Pass Efficiency Above NBA Replacement, unless/until the NBA Development League has SportVU tracking data.

[2] Dividing was chosen to represent a player's pass efficiency as a percentage above/below the average NBA player, rather than subtracting as Wins Above Replacement (WAR) or something similar in (baseball) sabermetrics.

a more detailed picture of passing activity. (A conditional inference trees model and other machine learning-based classification analyses I have done indicate that this CATALYST would be fundamental to a whole new approach to passing statistics. I can elaborate further if the Spurs organization wishes.) "Extended CATALYST" might also help contextualize and visualize a team's passing structure and offense—helping scouts and coaches to prepare more rigorously and identify trends that could be take defensive analytics to the next level.

Perhaps the most important improvement is the use of on/off differentials in the calculation of xTS%. I call this Rotational CATALYST. Darren Collison of the LA Clippers, for example, only plays 31% of his minutes with Chris Paul.[11] However, in the simplistic version of CATALYST, Darren Collison's xTS% includes the full extent of Chris Paul's shooting efficiency above the league average, hurting Collison's CATALYST slightly. This distortion can be avoided by using player and line-up specific xTS percentages to reflect the rotational nature of teammate quality in shooting efficiency.

Below are the top 30 NBA players in CATALYST in a minimum 25-minutes/game with a qualified games played restriction up and until March 18, 2014 (keeping in line with the conference poster). Notable players missing on this list due to restrictions include Rajon Rondo, Manu Ginobli, and Russell Westbrook. (Westbrook's absence may explain part of the Durant's high status compared to popular perception.)

Rank	**CATALYST**	**Player**	**Team**
1	253.75%	Lawson	DEN
2	238.84	Curry	GSW
3	232.76	Evans	NOP
4	225.51	Durant	OKC
5	224.65	Paul	LAC
6	213.75	Iguodala	GSW
7	210.8	Ellis	DAL
8	205.27	Nelson	ORL
9	204.8	Gordon	NOP
10	201.41	Wall	WAS
11	197.76	James	MIA
12	197.44	Jennings	DET
13	197.41	Rubio	MIN
14	196.9	Teague	ATL
15	195.52	Harden	HOU
16	193.75	Stephenson	IND
17	189.67	Foye	DEN
18	183.82	Dragic	PHX
19	183.4	Thompson	GSW
20	181.76	DeRozan	TOR
21	180.39	Irving	CLE
22	174.82	Lowry	TOR
23	171.14	Hayward	UTA
24	170.14	Carter-Williams	PHI
25	169.13	Thomas	SAC
26	167.67	Waiters	CLE
27	160.16	Parker	SAS
28	159.06	Crawford	LAC
29	158.13	Felton	NYK
30	154.8	Lin	HOU

One very interesting application of CATALYST is a focus on positionally-focused analysis; that is, comparing the passing abilities of centers against other centers, forwards vs forwards, and so on. This usage can be expanded to player comparison and cluster analysis among players of the same position. (This can be further explored if the Spurs wish.)

Here are the All-NBA CATALYST teams for general positions during the 2013-2014 NBA regular season:

Team	**Guard**	**Guard**	**Forward**	**Forward**	**Center**
First	Lawson	Curry	Evans	Durant	Noah
Second	Paul	Ellis	Iguodala	James	Cousins
Third	Nelson	Gordon	Hayward	Crawford	Jefferson

3 Assumption & Formula Testing

The first step in demonstrating the validity of CATALYST—beyond its interesting rankings/results and its normative value—is to test the assumptions made within the formula. In this section, I test each part of the CATALYST formula and model the assumptions made within it and show its limitations—all of which can be considerably mitigated or eliminated with proprietary SportVU tracking data.

Table 1 shows four linear regression models of various advanced offensive statistics and points created by assist per game (PCA/Game) onto either offensive xRAPM or Offensive RPM. (Observations are the top 30 CATALYST performers and all Spurs who played at least 11 games.) All of the indicators point towards Points Created by Assist (per Game) being a much better predictor of offensive impact than assist percentage (AST%), even when controlling for other aspects of offensive activity such as shooting efficiency and offensive rebounding. This results support the first criticism of the assist approach—that assists are blind to the quality of passes and created shots from a great pass. CATALYST solves this criticism by bypassing the assist model's blindness and using Points Created by Assist instead.

Table 1: Support for the Points Created by Assist Approach

	Dependent variable:			
	Offensive xRAPM		Offensive RPM	
	(1)	(2)	(3)	(4)
USG%	−1.251**	−0.936*	−1.739***	−1.422**
	(0.566)	(0.530)	(0.567)	(0.528)
TS%	−28.053	−15.719	−44.404*	−31.889
	(24.196)	(22.763)	(24.264)	(22.666)
ORB%	−0.160	−0.135	−0.168*	−0.142
	(0.099)	(0.092)	(0.099)	(0.092)
AST%	0.082***		0.081***	
	(0.022)		(0.022)	
PCA/Game		0.178***		0.179***
		(0.038)		(0.037)
USG% & TS% interaction	2.408**	1.821*	3.248***	2.656***
	(0.988)	(0.930)	(0.990)	(0.926)
Constant	14.562	7.840	23.786*	16.951
	(13.778)	(12.959)	(13.817)	(12.904)
Observations	44	44	44	44
R^2	0.696	0.739	0.724	0.766
Adjusted R^2	0.656	0.704	0.687	0.735
Residual Std. Error (df = 38)	1.393	1.291	1.397	1.286
F Statistic (df = 5; 38)	17.380***	21.493***	19.912***	24.898***
Note:				*p<0.1; **p<0.05; ***p<0.01

In Table 2, three models are tested to see the validity of the passes per game statistic used in the denominator of PE and PEAR. The models use a dummy categorical variable considering position, broken down into "Bigs," "Point Guards," (PG) and "Wings." The first is a simple OLS linear regression using position and pace to predict passes per minute. Model 1 demonstrates that position dictates passes per minutes more than pace within the sample, though both are significant. The latter two models are linear models using random effects specifications to test the results of Model 1. Model 2 tests position as a random effect; Model 3 uses pace as the random effect. Both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) as well as the log likelihood give support to the model 1 conclusion that position overwhelms pace, though both are significant. Thus PE and PEAR can be justified on both a team and positional context basis as well as a universal basis somewhat as pace is controlled somewhat within the passes/minute framework.

Since pace is not completely controlled for within the CATALYST framework, using per-possession specifications is a good idea when implementing CATALYST and PEAR in a proprietary data environment.

Table 2: Support for passes/game? Controlled for minutes played.

	Dependent variable:		
	Passes/Minute		
	OLS	*Linear random effects*	
	(1)	(2)	(3)
Position (PG)	0.416***		0.428***
	(0.038)		(0.037)
Position (Wing)	−0.152***		−0.158***
	(0.038)		(0.037)
Pace	−0.013**	−0.013**	
	(0.006)	(0.006)	
Constant	2.470***	2.555***	1.173***
	(0.593)	(0.615)	(0.030)
Observations	360	360	360
R^2	0.381		
Adjusted R^2	0.376		
Log Likelihood		-86.782	-78.287
Akaike Inf. Crit.		181.564	166.573
Bayesian Inf. Crit.		197.108	186.004
Residual Std. Error	0.299 (df = 356)		
F Statistic	73.039*** (df = 3; 356)		
Note:			*p<0.1; **p<0.05; ***p<0.01

In Table 3, PEAR is explored within a fixed-effects model, whereby position (the same dummy as before) is considered as the fixed effect. The table shows that these models are not too great, that there is little statistical evidence of position being a fixed effect.

Table 3: What makes PEAR? A fixed-effects model considering position as a fixed effect.

	Dependent variable:	
	PEAR	
	(1)	(2)
Pace	0.555^{*}	-2.089^{*}
	(0.327)	(1.104)
Passes/Minute		-202.353^{**}
		(80.748)
Pace & Passes Interaction		2.170^{***}
		(0.836)
Observations	360	360
R^2	0.201	0.229
Adjusted R^2	0.192	0.216
Residual Std. Error	15.926 (df = 356)	15.683 (df = 354)
Note:	*p<0.1; **p<0.05; ***p<0.01	

The two models in Table 4 consider PEAR and Free Throw PEAR within a random effects framework. Models (2) and (4) consider Position (the dummy variable with three categories: Big, PG, and Wing) as the random effect; Models (1) and (3) consider Pace as the random effect. Both sets of models support the hypothesis that positional analysis is very useful in PEAR when controlling for pace.

Table 4: What makes PEAR? Two random effects models

	Dependent variable:			
	PEAR		Free Throw PEAR	
	(1)	(2)	(3)	(4)
Passes/Game	0.008^{***}	0.008^{***}	0.009^{***}	0.009^{***}
	(0.001)	(0.001)	(0.002)	(0.001)
Position (PG)	0.626^{***}		0.662^{***}	
	(0.048)		(0.051)	
Position (Wing)	0.433^{***}		0.456^{***}	
	(0.045)		(0.048)	
Pace		0.022^{***}		0.024^{***}
		(0.007)		(0.008)
Constant	0.572^{***}	-1.218^{*}	0.607^{***}	-1.296^{*}
	(0.051)	(0.731)	(0.054)	(0.769)
Observations	360	360	360	360
Log Likelihood	-152.189	-153.071	-170.437	-171.218
Akaike Inf. Crit.	316.378	316.141	352.874	352.437
Bayesian Inf. Crit.	339.694	335.572	376.191	371.867
Note:			*p<0.1; **p<0.05; ***p<0.01	

Table 5 shows two random effects models with interaction effects between passes per game and position as well as passes per game and pace. Model 1 wins out, cementing the positional approach as more valid than a pace approach.

Table 5: What makes PEAR? Two random effects models considering interaction

	Dependent variable:	
	PEAR	
	(1)	(2)
Passes/Game	0.574***	−2.712*
	(0.082)	(1.554)
Position (PG)	0.717	
	(3.507)	
Position (Wing)	−4.545	
	(3.712)	
Passes/Game & Position (PG)	0.292***	
	(0.100)	
Passes/Game & Position (Wing)	0.740***	
	(0.128)	
Pace		−0.344
		(0.555)
Passes/Game & Pace		0.037**
		(0.016)
Constant	3.369	37.064
	(2.556)	(53.797)
Observations	360	360
Log Likelihood	-1,371.013	-1,390.720
Akaike Inf. Crit.	2,758.026	2,793.439
Bayesian Inf. Crit.	2,789.114	2,816.756
Note:	*p<0.1; **p<0.05; ***p<0.01	

The following models explore the value of my exclusive True Shooting percentage, or xTS%. Table 6 shows logistic models attempting to predict whether San Antonio wins a Larry O'Brien trophy during the Gregg Popovich era using only regular season stats. Models (2) and (4) use the offensive Four Factors Dean Oliver pioneered; Models (1) and (3) use the xTS% of Tim Duncan for all seasons prior to 2010 and the xTS% of Tony Parker for all seasons after 2010. xTS%, judging by the log likelihood and AICs of the models, beats out the use of effective Field Goal percentage (eFG%).

Table 6: Larry O'Brien Logistic Model: xTS% or eFG%?

	Dependent variable:			
	Does San Antonio Win a Ring?			
	(1)	(2)	(3)	(4)
xTS%	98.959		120.527	
	(89.126)		(102.083)	
eFG%		67.744		78.221
		(71.830)		(78.732)
ORB%	46.259	31.199	47.237	30.667
	(50.035)	(43.153)	(52.877)	(43.610)
FT/FGA Ratio	13.430	17.679	8.595	15.826
	(34.655)	(36.012)	(36.350)	(36.792)
TOV%	93.717	53.202	103.802	55.533
	(97.229)	(83.238)	(99.814)	(84.272)
Duncan/Parker? (Dummy)			−1.200	−0.692
			(2.141)	(2.059)
Constant	−82.955	−54.807	−94.921	−59.689
	(68.957)	(50.216)	(75.677)	(52.206)
Observations	17	17	17	17
Log Likelihood	-8.755	-9.191	-8.597	-9.135
Akaike Inf. Crit.	27.510	28.382	29.193	30.269
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 7 shows two linear regression models attempting (horribly) to predict the Spurs' regular season win percentage. Each use three of Dean Oliver's Four Factors on offense. Model (1) uses xTS% (same specifications as in Table 7); Model (2) uses team eFG%. Model (1) outperforms (2) significantly, though neither performs well in any meaningful sense. Considering both sets of models, there is some (though not much) evidence xTS% is useful on a team level.

Table 7: Support for xTS% on a Team Level (Regular Season Win %)

	Dependent variable:	
	Spurs Regular Season Win %	
	(1)	(2)
xTS%	1.563	
	(0.930)	
eFG%		1.301
		(1.067)
ORB%	0.658	0.502
	(0.698)	(0.745)
FT/FGA Ratio	−0.298	−0.137
	(0.580)	(0.635)
TOV%	1.388	0.974
	(1.470)	(1.560)
Constant	−0.435	−0.184
	(0.658)	(0.696)
Observations	17	17
R^2	0.240	0.165
Adjusted R^2	-0.013	-0.114
Residual Std. Error (df = 12)	0.043	0.045
F Statistic (df = 4; 12)	0.949	0.592
Note:	*p<0.1; **p<0.05; ***p<0.01	

xTS%, however, does not have too much support on an individual player level. It is not too useful in regressing onto oRAPM and oRPM. The models in Table 8 support the idea that controlling for teammate quality by using xTS% is just a good scouting tool to identify under-valued players rather than a purer metric of offensive passing ability. **These models indicate there may be a better way of controlling for teammate quality than pure xTS% or pure xeFG%.** My first guess would be line-up specific-xTS%, as recommended earlier.

Table 8: Why xTS%? Offensive xRAPM & Offensive RPM

	Dependent variable:			
	Offensive RAPM		Offensive RPM	
	(1)	(2)	(3)	(4)
CATALYST	2.267^{***}		2.266^{***}	
	(0.508)		(0.526)	
PEAR		2.311^{***}		2.339^{***}
		(0.519)		(0.534)
ORB%	-0.169	-0.171	-0.182	-0.182
	(0.114)	(0.114)	(0.118)	(0.117)
TS%	26.420^{***}	24.801^{***}	30.109^{***}	28.452^{***}
	(5.508)	(5.545)	(5.701)	(5.706)
Constant	-15.533^{***}	-14.762^{***}	-17.922^{***}	-17.188^{***}
	(3.146)	(3.108)	(3.256)	(3.198)
Observations	44	44	44	44
R^2	0.595	0.594	0.608	0.612
Adjusted R^2	0.564	0.564	0.579	0.583
Residual Std. Error (df = 40)	1.568	1.568	1.622	1.614
F Statistic (df = 3; 40)	19.565^{***}	19.526^{***}	20.682^{***}	21.042^{***}
Note:				*p<0.1; **p<0.05; ***p<0.01

4 Modeling CATALYST

To test the hypothesis that CATALYST is a good indicator or metric to gauge offensive performance, we can compare CATALYST to more general models of player value on offense. The two selected were the offensive aspects of both Jeremias Englemann's xRAPM[6] and real plus-minus (RPM).[7]

Table 9 demonstrates the value of the PEAR/CATALYST approach. In the sample of 22 wings, 21 bigs, and 25 PGs (selected for their PEAR as well as meeting large minutes and games requirements), PEAR significantly outperformed the pace-and-minutes-controlled Assist Percentage (AST%) in predicting both offensive xRAPM and RPM. These models suggest that pass efficiency can serve as a better representation of player value and offensive contribution than assists and derived statistics can. Replacing AST% with PEAR leads to an increase of 15% in r-squared when predicting oxRAPM and oRPM.

(See also Table 8 on page 14 demonstrating the value of CATALYST and/or PEAR in regressing offensive xRAPM and offensive RPM even when controlling for ORB% and TS%.)

Before one should conclude strongly from these models, one should be sure that the models do not violate global assumptions of linear regression. The PEAR models pass the five major linear regression assumptions (though not with flying colors): global test statistics, skewness, kurtosis, link function,, and heteroscedasticity.[16] (Picture from gvlma package in R; see previous citation for more.)

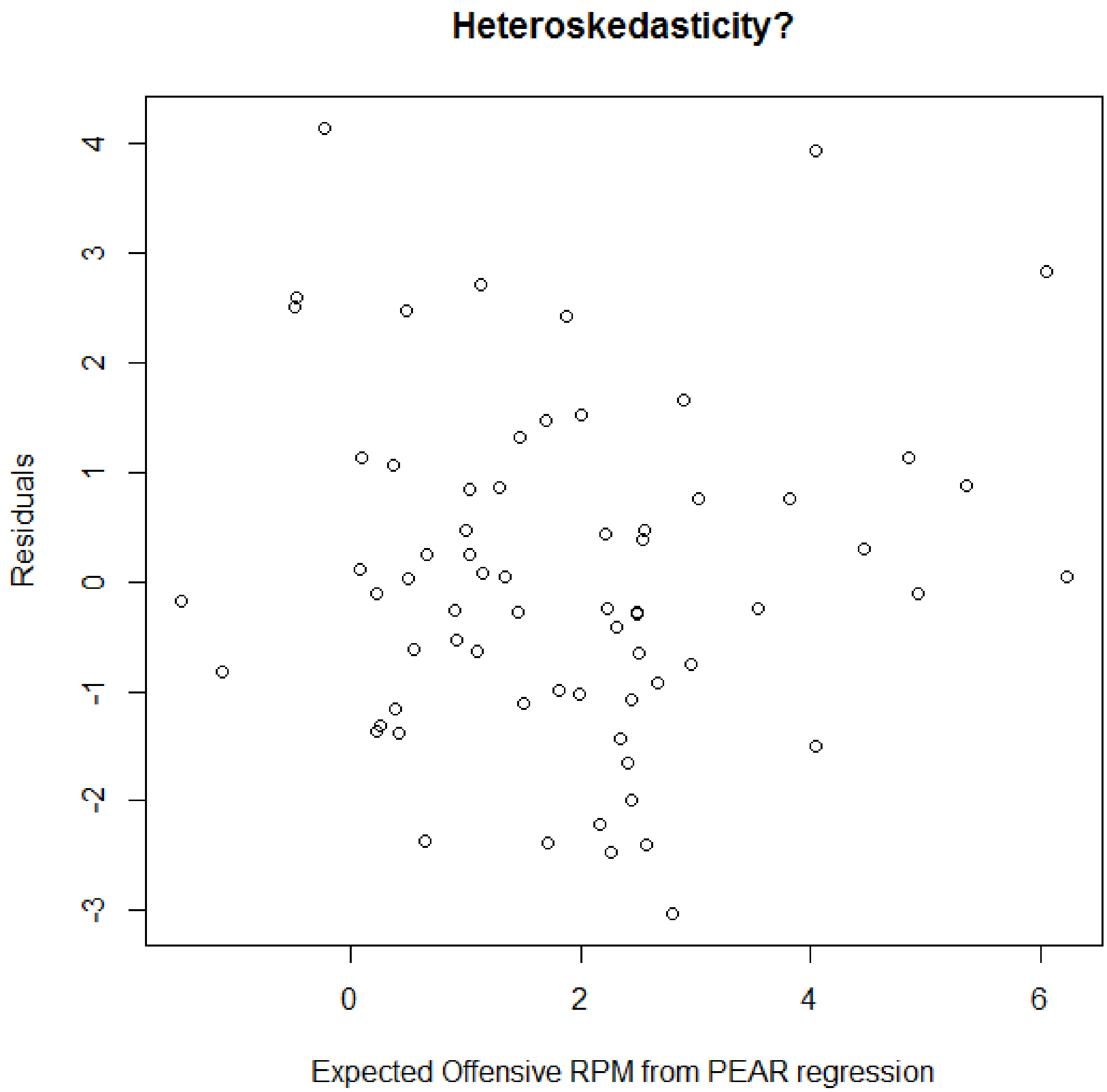
	Value	p-value	Decision
Global Stat	3.180172	0.5281	Assumptions acceptable.
Skewness	0.002577	0.9595	Assumptions acceptable.
Kurtosis	0.804163	0.3699	Assumptions acceptable.
Link Function	1.560207	0.2116	Assumptions acceptable.
Heteroscedasticity	0.813225	0.3672	Assumptions acceptable.

Table 9: The Value of PEAR in General Modeling of Plus-Minus

	Dependent variable:			
	Offensive xRAPM		Offensive RPM	
	(1)	(2)	(3)	(4)
PEAR	1.841***		1.983***	
	(0.466)		(0.470)	
AST%		0.083***		0.083***
		(0.030)		(0.031)
TS%	25.222***	24.536***	28.153***	27.266***
	(4.996)	(5.279)	(5.038)	(5.412)
ORB%	−0.120	−0.047	−0.126	−0.049
	(0.078)	(0.082)	(0.079)	(0.084)
USG%	0.108**	0.122***	0.107**	0.127***
	(0.042)	(0.044)	(0.042)	(0.046)
Position (PG)	−0.551	−0.461	−0.456	−0.248
	(0.740)	(0.853)	(0.746)	(0.874)
Position (Wing)	−1.353*	−0.297	−1.248	−0.079
	(0.742)	(0.693)	(0.748)	(0.710)
Constant	−15.880***	−15.219***	−18.177***	−17.389***
	(2.965)	(3.122)	(2.990)	(3.200)
Observations	68	68	68	68
R^2	0.511	0.455	0.551	0.482
Adjusted R^2	0.463	0.401	0.507	0.431
Residual Std. Error (df = 61)	1.538	1.625	1.551	1.666
F Statistic (df = 6; 61)	10.645***	8.477***	12.462***	9.448***
Note:				*p<0.1; **p<0.05; ***p<0.01

Simple linear regression, however, perhaps cannot simply do the trick by itself. There are two reasons to suggest this.

(1) The data might be heteroscedastic despite the analysis presented above (which only said our homoscedastic assumption was reasonable, not great). See the residual plot below for evidence of some such heteroscedasticity.



This is, however, easily fixed. One can apply heteroscedasticity-consistent (HC) standard errors (HCSEs) to the results of linear regression. This holds true even if one simply suspects heteroscedasticity.[12] Below are the results of the models using HC3, the HCSE recommended for small n models in linear regression by Long & Ervin.[12]

As Table 10 shows, PEAR still holds more statistical significance, even when using regression results using an HSCE. Thus PEAR's improvement over AST% in predicting a player's offensive plus/minus value is robust to heteroscedasticity in the data and robust overall.

Table 10: Using HSCEs in Modeling Offensive xRAPM and Offensive RPM

	Dependent variable:			
	oxRAPM —— oRPM			
	(1)	(2)	(3)	(4)
AST%	0.083***		0.083***	
	(0.030)		(0.031)	
PEAR		1.841***		1.983***
		(0.466)		(0.470)
TS%	24.536***	25.222***	27.266***	28.153***
	(5.279)	(4.996)	(5.412)	(5.038)
ORB%	−0.047	−0.120	−0.049	−0.126
	(0.082)	(0.078)	(0.084)	(0.079)
USG%	0.122***	0.108**	0.127***	0.107**
	(0.044)	(0.042)	(0.046)	(0.042)
Position (PG)	−0.461	−0.551	−0.248	−0.456
	(0.853)	(0.740)	(0.874)	(0.746)
Position (Wing)	−0.297	−1.353*	−0.079	−1.248
	(0.693)	(0.742)	(0.710)	(0.748)
Constant	−15.219***	−15.880***	−17.389***	−18.177***
	(3.122)	(2.965)	(3.200)	(2.990)
Note:				*p<0.1; **p<0.05; ***p<0.01

(2) The data is in the "zone of uncertainty" with regards to autocorrelation of the variables. The RPM model using PEAR results in a 2.425334 Durbin-Watson statistic (d). With a d_U of 1.604 and a d_L of 1.283,[17] d falls within a "zone of uncertainty" (between $4 - d_U$ and $4 - d_L$) whereby one can neither reject the null hypothesis that the error terms are not correlated nor accept an alternate hypothesis.[17]

This unfortunate predicament, however, also has an easy fix. One can use an estimator of standard errors robust to both autocorrelation as well as heteroskedasticity.[13] The results of such analysis are presented below in Table 11. PEAR's supremacy over AST% is robust to both heteroskedasticity and autocorrelation.

Table 11: Taking into account possible heteroskedasticity and autocorrelation

	Dependent variable:			
	Offensive xRAPM ——— Offensive RPM			
	(1)	(2)	(3)	(4)
TS%	24.536***	25.222***	27.266***	28.153***
	(5.278)	(5.270)	(5.581)	(5.499)
AST%	0.083***		0.083**	
	(0.030)		(0.035)	
PEAR		1.841***		1.983***
		(0.370)		(0.372)
ORB%	−0.047	−0.120	−0.049	−0.126*
	(0.077)	(0.077)	(0.086)	(0.068)
USG%	0.122**	0.108***	0.127**	0.107***
	(0.048)	(0.040)	(0.048)	(0.035)
Position (PG)	−0.461	−0.551	−0.248	−0.456
	(0.872)	(0.648)	(0.973)	(0.678)
Position (Wing)	−0.297	−1.353**	−0.079	−1.248*
	(0.611)	(0.602)	(0.659)	(0.629)
Constant	−15.219***	−15.880***	−17.389***	−18.177***
	(3.355)	(3.187)	(3.548)	(3.357)

Note: *p<0.1; **p<0.05; ***p<0.01

5 The Future of PEAR & CATALYST

Seeing that PEAR and CATALYST both have solid bases in empirical testing of their formula components and that both outperform traditional assist models in predicting players' plus/minus values in all positions, there is a bright future for the metrics within the Spurs organization.

The first element of expanded usage of PEAR and CATALYST is player analysis and evaluation as well as contract evaluation. The implications of this facet are fairly obvious and well-supported by the statistical analyses conducted within this report. If SportVU-esque passing statistics are extended to either the NBA Development League or a different NBA feeder league, the implications of CATALYST for draft evaluation are also obvious. A not-as-obvious application of PEAR and CATALYST, however, is the clustering and analysis of players by passing statistics. This is supported by the positional component of PEAR/CATALYST (demonstrated by both the random-effects and fixed-effects models). A model could be developed that could find a player's "natural" NBA position on offense if the player's college/pre-NBA career is unclear (e.g. "tweeners"). This has shown promise with just simple assist-to-turnover ratios./citelayne

The second aspect of expanding PEAR and CATALYST is line-up analysis. This is nowhere near meant to replace eye test or conventional analysis such as film study, but to supplement scouting efforts and suggest new approaches. If the proprietary data are specific (showing passing rates between different pairs and the efficiency between those pairs), then PEAR and CATALYST can help scouts and film analysts for the Spurs prioritize and focus their scouting efforts. For example, proprietary data of SportsVU disclosed by Kirk Goldsberry show that Trevor Ariza corner 3s assisted by John Wall were particularly deadly (CATALYST predicted John Wall as a top assister of high-quality shots, so this was nothing new). With metrics like CATALYST and indicators like Nylon Calculus shot charts,[14] trends can be discovered and game-planning against important sets and actions becomes much easier and more targeted. This is particularly useful for the regular season, where there is not as much dedicated time as in the post-season.

A third aspect of future PEAR and CATALYST use is in potential draft analysis. The first steps toward a CATALYST-centric view of draft analysis have been laid with `Hoop-Math.com`.[18] Hoop-Math provides data on the distribution of assists—specifically which type of shot is assisted: shots in transition, at the rim, on 2pt jumpers, and 3s. With good passing stats to provide context (namely, passes per game/100 possessions), a PEAR-similar metric can be constructed that would, most probably, bolster the Spurs's drafting models and/or analysis. This modeling approach would be strengthened with Hoop-Math's team offensive statistics detailing shots taken as well as splits in transition and non-transition.

These three ways are only the beginning of the bright futures PEAR and CATALYST hold.

6 Acknowledgments

The regression and data visualization occurred within the R environment.[5] All fixed effects models were made using the lfe package;[9] all random and mixed effects models were constructed with the lme4 package.[8] HSCE and autocorrelation-adjusted estimation support in R is based on author's modification and adaptation of the sandwich package.[13] Tables of the models provided by the package stargazer.[10] All data courtesy of `http://www.basketball-reference.com` and `NBA.com`.

References

[1] http://www.basketball-reference.com

[2] Neuteufel, Nicholas K. (2014). *Introducing exclusive True Shooting percentage.* Sports & Illumination. http://sportsandillumination.com/2014/03/22/introducing-xts/

[3] Hollinger, John (2014). 2013-2014 Hollinger Team Statistics. ESPN.

[4] Phillips, Aqeel (2014). *Dishes and Dimes Pt II–Passing Efficiency.* Princeton Sports Analytics. http://princetonsportsanalytics.com/2014/02/20/dishes-and-dimes-part-ii-passing-efficiency/

[5] R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

[6] Englemann, Jeremias (2014). xRAPM for the 2013-2014 NBA season. http://stats-for-the-nba.appspot.com/ratings/xRAPM.html

[7] Illardi, Steve (2014). The next big thing: real plus-minus. ESPN.

[8] Bates et al (2014). lme4: Linear mixed-effects models using Eigen and S4.

[9] Gaure, Simen (Dec 2013). lfe: Linear group fixed effects. *The R Journal.* 5(2):104-117.

[10] Hlavac, Marek (2014). stargazer: LaTeX code and ASCII text for well-formatted regression and summary statistics tables. R package version 5.1. http://CRAN.R-project.org/package=stargazer

[11] NBAwowy.com

[12] Long, Scott J. and Laurie H. Ervin. Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model.

[13] Lumley, Thomas and Achim Zeileis. Sandwich package for R: Robust Covariance Matrix Estimators. http://cran.r-project.org/web/packages/sandwich/sandwich.pdf

[14] Clemens, Austin. Nylon Calculus Shot Charts. http://nyloncalculus.com/shotchart/

[15] Vashro, Layne. http://twitter.com/VJL(underscore)bball/status/474635360971284481

[16] Pena, EA and Slate, EH. *Global Validation of Linear Model Assumptions.* J. American Statistical Association, 101(473):341-354, 2006.

[17] Savin, N.E. and K.J. White, "The Durbin-Watson Test for Serial Correlation with Extreme Sample Sizes or Many Regressors." *Econometrica* 45, 1977, p.1989-1996.

[18] Hoop-Math.com