In defense of the three-point shot

Louis Nass

Tulane University

December 17, 2019

Introduction

- Introduction of NBA three-point line 1979-80
- Trend towards larger volume of three-point shots in more recent seasons
- "NBA players don't play defense"

Project Overview

- Linear model
  - Input: o_2pm, o_3pm, d_2pm, d_3pm
  - Output: *wins*
  - Run for three data sets: 1985-88, 1995-98, 2005-08
- Analysis of data
  - Pre-processing: "normal check"
  - Correlation coefficients: which predictor influences *wins* most?
- Hypothesis test
  - $H_0 : d\_2pm = d\_3pm = 0$
  - Is defense obsolete?
- Residual vs Leverage
  - Compare and contrast R "highlighted" data points

Data Set

- Data exported from Kaggle[1]
- Records of shots made, shots attempted, steals, rebounds, wins, ... etc.
- Sample of data:

"name"	"o_2pm"	"o_3pm"	"d_2pm"	"d_3pm"	"wins"
"Atlanta Hawks"	2490	341	1717	236	30
"Boston Celtics"	2387	471	1754	212	24
"Chicago Bulls"	2566	480	1406	387	49

[1] https://www.kaggle.com/open-source-sports/mens-professional-basketball#basketball_teams.csv

Pre-processing

- We want our data to be approximately normal
- Skewed data $\rightarrow$ need to make transformation

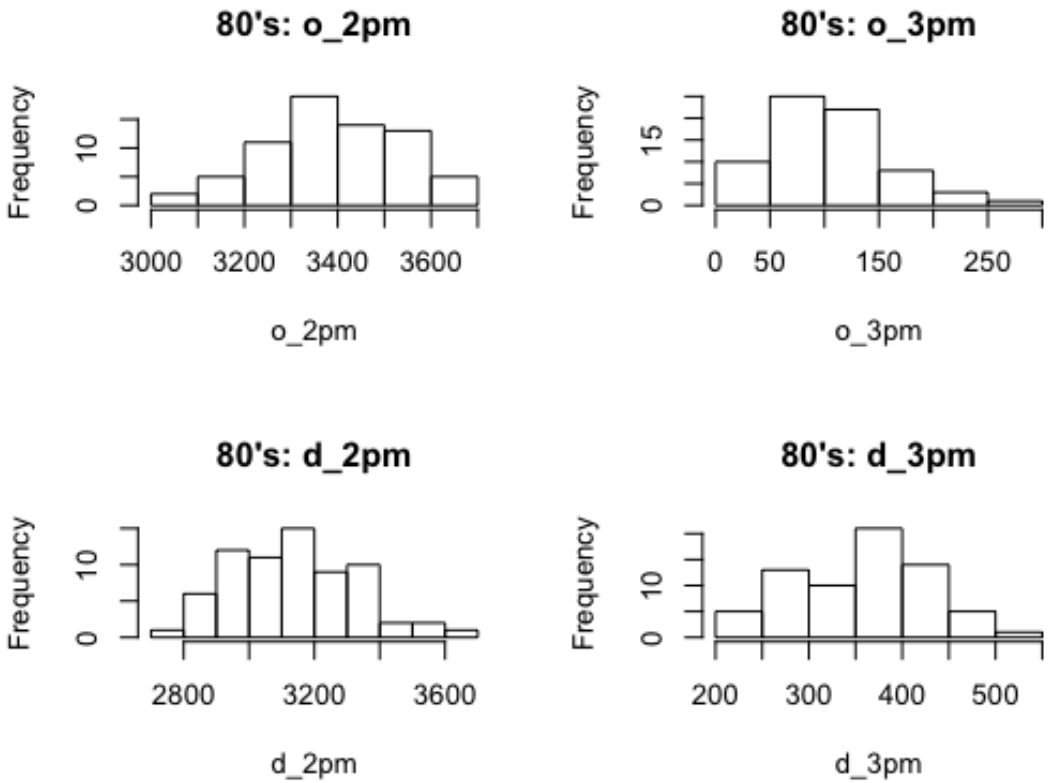

Figure: Histograms of predictor data from 1985-88 seasons. The histograms appear to be approximately normal.

Pre-processing

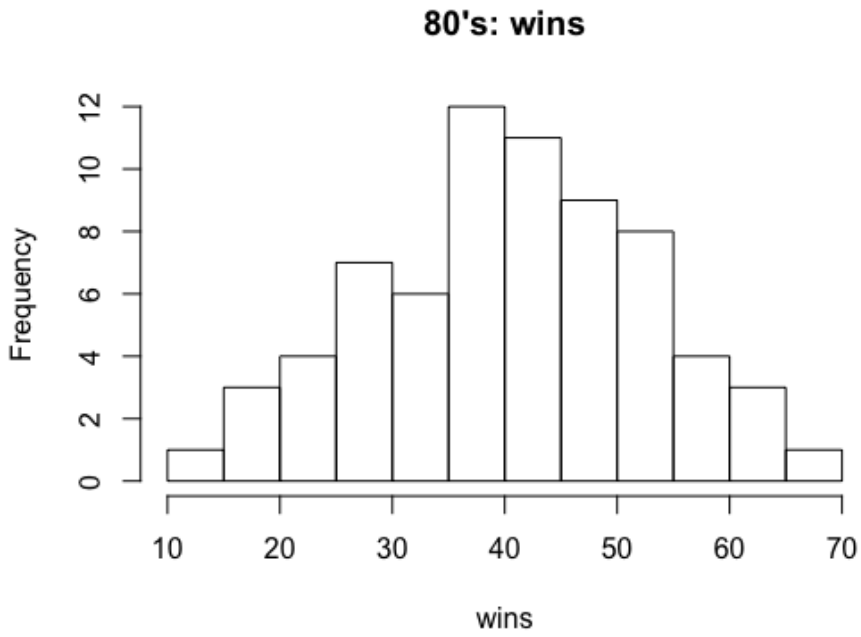

Figure: Histograms of wins from 1985-88 seasons. Again this is approximately normal.

- All data is approximately normal
- Similar figures for 1995-98 and 2005-08 data sets

Linear model

- Coefficient significance and R^2 values

Model Comparison					
Model years	$Pr(>\|t\|)$ *o_2pm*	$Pr(>\|t\|)$ *o_3pm*	Pr($>\|t\|$) *d_2pm*	Pr($>\|t\|$) *d_3pm*	R^2
1985-88	1.62e-10	**0.00186**	$1.44e-06$	**0.48015**	0.5418
1995-98	1.09e-13	1.20e-13	$< 2e-16$	$9.26e-13$	0.7725
2005-2008	$1.00e-12$	$< 2e-16$	$2.12e-12$	$< 2e-16$	0.7423

Correlation coefficients

- Which predictors have greatest effect *wins* in each data set?

Correlation coefficients to *wins*				
Data set	*o_2pm*	*o_3pm*	*d_2pm*	*d_3pm*
1985-88	**0.4232**	0.2789	-0.2845	0.1968
1995-98	0.1742	0.1903	**-0.5275**	0.1178
2005-08	0.0423	**0.4059**	0.0099	**-0.3641**

Hypothesis test

- $H_0 : d\_2pm = d\_3pm = 0$

Model Comparison				
Model years	DF	Sum of Sq	F	Pr(>F)
1985-88	2	2363.7	16.116	$2.146e-06$
1995-98	2	8993	86.812	$< 2.2e-16$
2005-2008	2	5129.2	68.23	$< 2.2e-16$

- Reject the null for each model
- Suggests that defense is not obsolete

Residual vs leverage plot

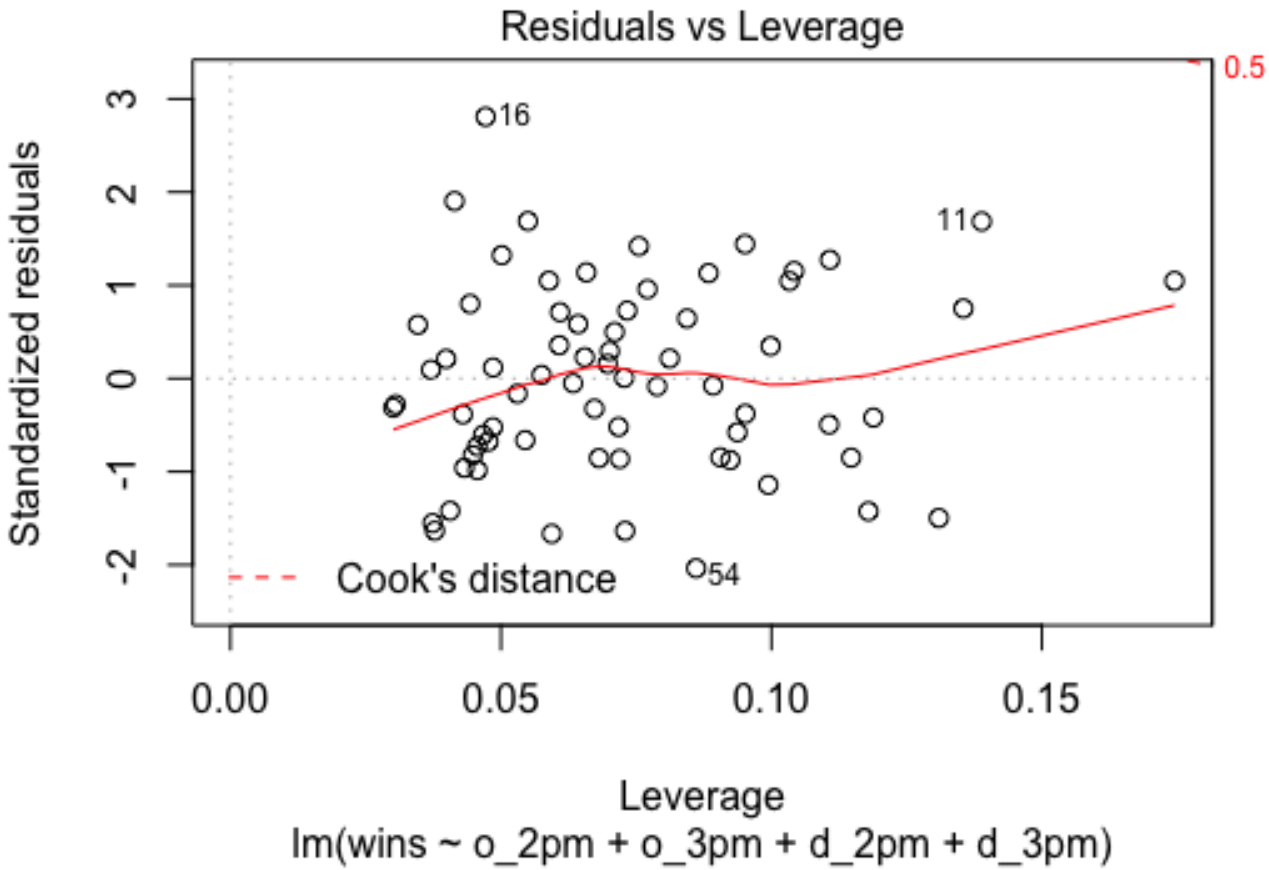

Figure: Leverage plots for 80's data set

Residual vs leverage plot

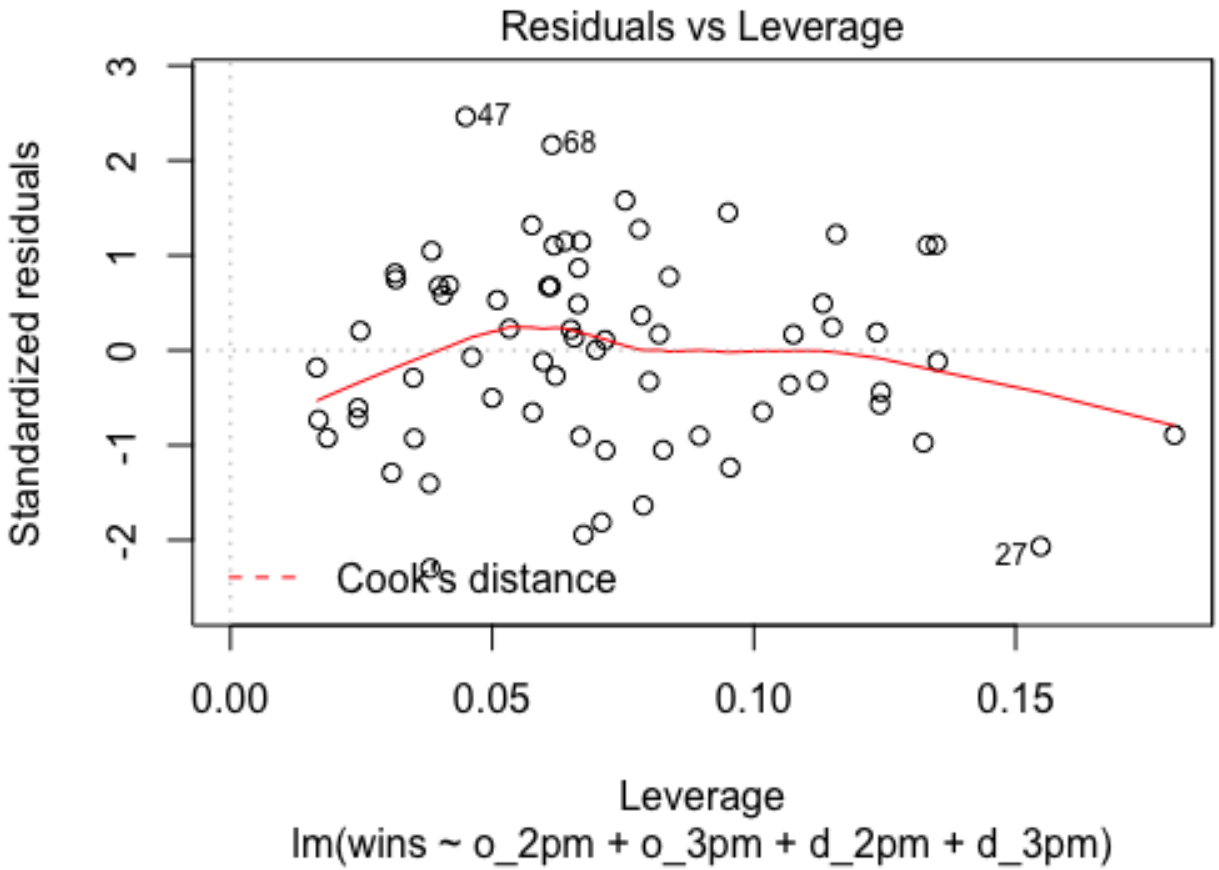

Figure: Leverage plots for 90's data set

Residual vs leverage plot

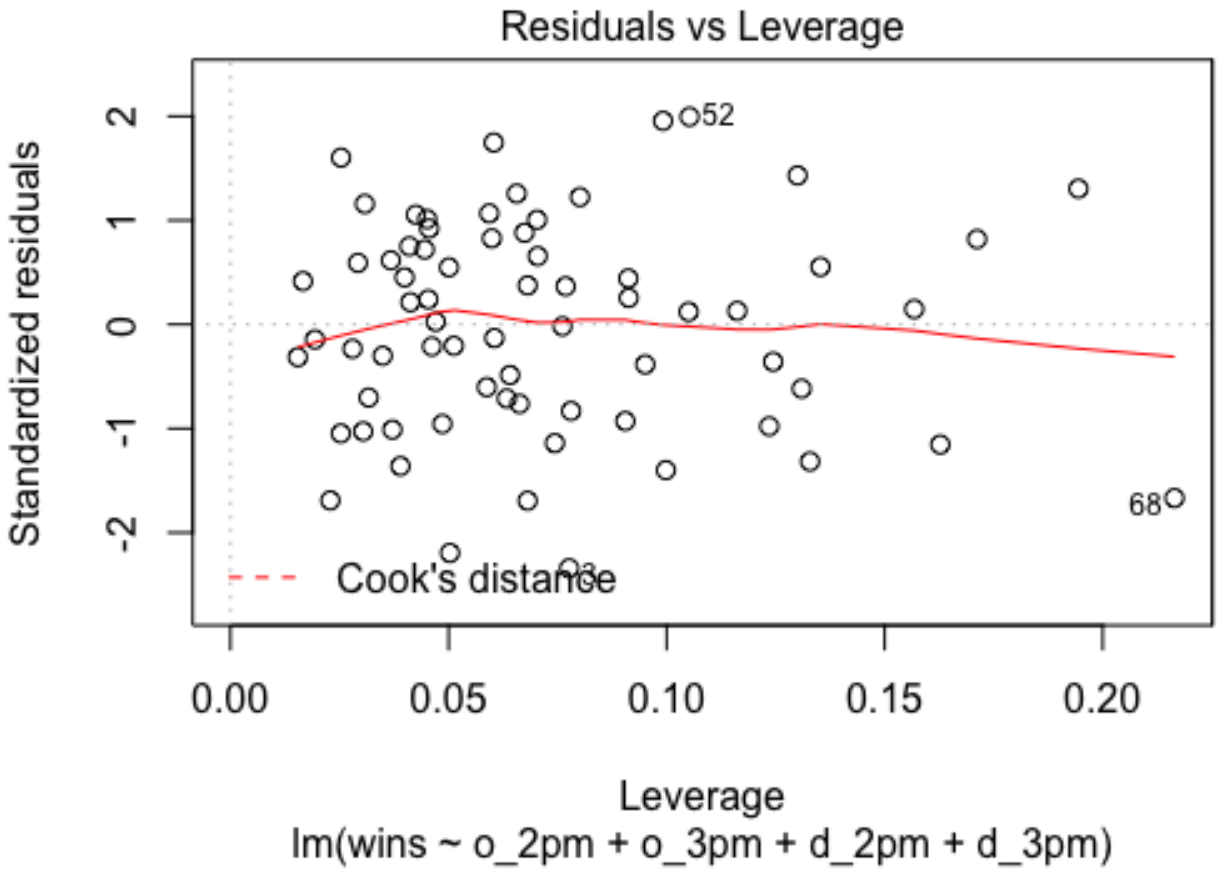

Figure: Leverage plots for 00's data set

80's data set leverage highlights						
	name	*o_2pm*	*o_3pm*	*d_2pm*	*d_3pm*	*wins*
11	Los Angeles Clippers	3324	**64**	**3588**	261	**32**
16	Philadelphia 76ers	3384	**51**	3333	282	54
54	Golden State Warriors	3372	**91**	3157	470	20
	Mean values	3390	107	3141	356	41

90's data set leverage highlights						
27	Cleveland Cavaliers	2221	483	**1258**	1230	42
47	Atlanta Hawks	2550	337	**1944**	1017	50
68	Utah Jazz	2744	**249**	1684	1122	**62**
	Mean values	2545	443	1722	1248	41
00's data set leverage highlights						
3	Chicago Bulls	2444	**560**	**1345**	1428	41
52	Denver Nuggets	2738	**569**	1624	**1710**	50
68	Utah Jazz	**2872**	**407**	**1477**	1419	54
	Mean values	2499	504	1610	1383	42

Conclusion

- Model effectively represents data for all three sets
  - $\approx$ 50% for 1985-88
  - $\approx$ 75% for 1995-98, 2005-08
- Evolution of basketball suggests greater importance for three-point shot
  - Increased correlation values, increased averages
- No statistical data suggests a dip in defense
  - Rejected null hypothesis
  - Average $\approx$ 3300 total makes allowed in all three data sets
  - Difference being magnitude of type of makes allowed

Limitations/continued study

Limitations

- Is the discussed hypothesis test the best way to determine defensive efficiency?
- Many other statistics recorded, too simple of a model?
- Limited data size

Continued Study

- Incorporate more predictors, increase the data size
- Consider trends throughout course of NBA season
- What would happen with a four-point shot?

Thank you!