Nba Market evolution

Business, Economic and Financial Data Course Project

Davide Pivato - Andrea Mascari

THE TASK

Analyze and interpret the reasons behind the league salaries trend behavior and consider some relevant in and out game scenarios

SOURCE OF DATA

Basketball Reference

A website serving up basketball stats and history

Kaggle

Online Data Science community with open source datasets and projects

FINAL DATASET

01

Players

Dataset comprehensive of all the players anagraphics

02

Season_stats

Nba_stats

Datasets that contains per season stats of the players

03

Salaries_1985to2018

nba-salaries

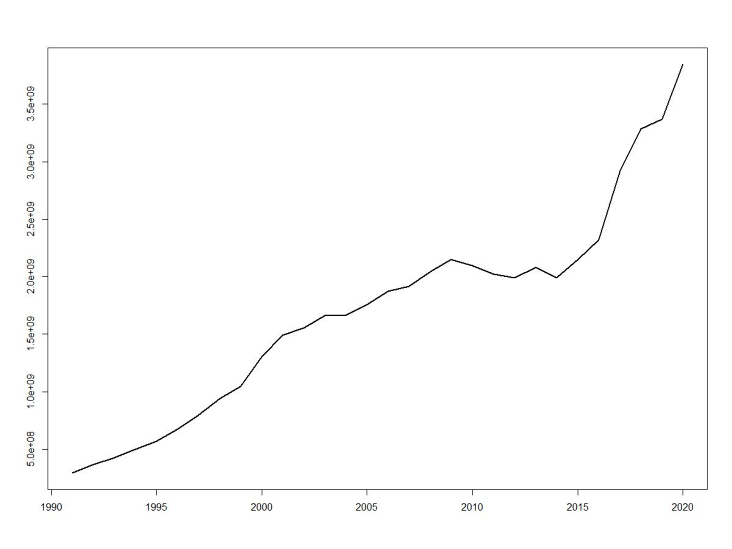
Salaries of players between 1985 and 2018

04

Salary_cap

Small dataset with the salary cap values for each year

THE LEAGUE TREND



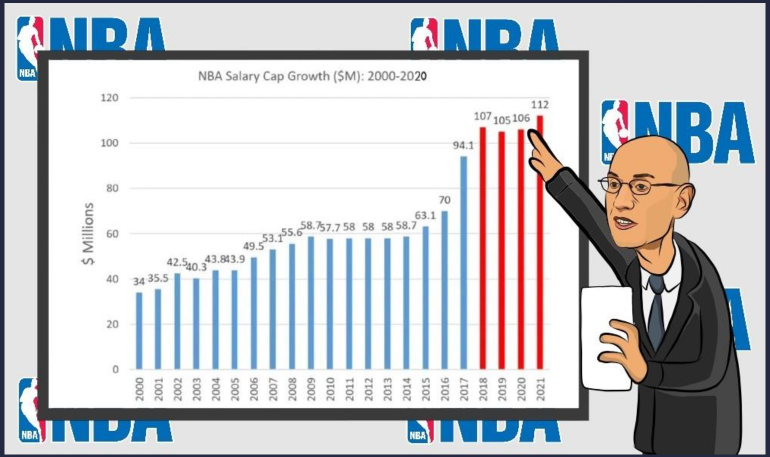
Rapid increases in specific points

Exponential growth of the salaries

How does it works?

SALARY CAP SYSTEM

- Establishes the amount of money that each team can spend
- Useful to maintain an equilibrium between the rosters

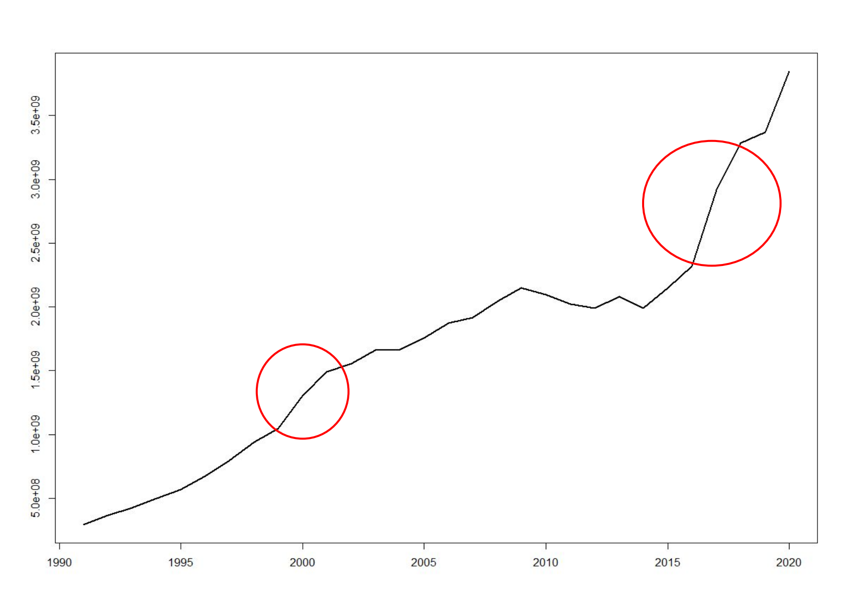


SALARY SPIKES

1999

Nba-lockout Players strike

Raises the league's minimum salary.

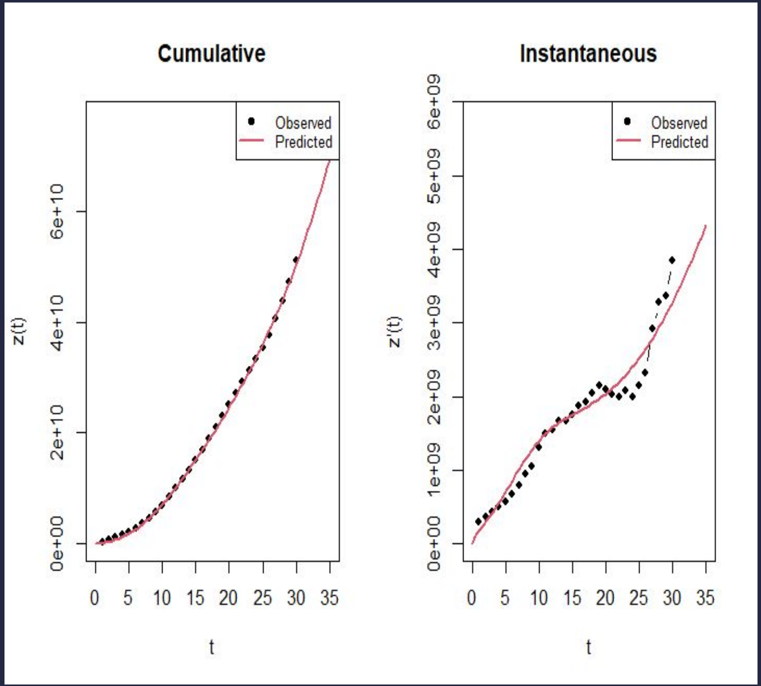


2016

$24-billion television deal signed by the NBA in 2014

from $70 million all the way up to $94.1

DYNAMIC MARKET POTENTIAL - GGM



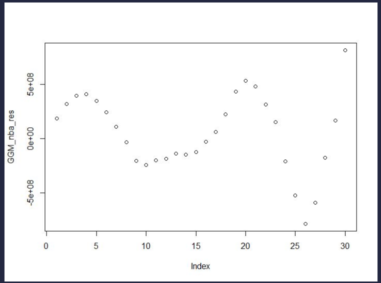
Hard to make reliable predictions before the peak of a phenomenon

Good fit!

Coefficients:

	Estimate	Std.Error	Lower	Upper	p-value	
K	3.465723e+11	1.936349e+12	-3.448602e+12	4.141747e+12	8.59e-01	
pc	6.162932e-05	6.929251e-04	-1.296479e-03	1.419738e-03	9.30e-01	
qc	1.276397e-01	2.840259e-02	7.197163e-02	1.833078e-01	1.38e-04	***
ps	3.537999e-02	5.083169e-03	2.541716e-02	4.534281e-02	2.70e-07	***
qs	1.979190e-01	5.111379e-02	9.773777e-02	2.981001e-01	6.88e-04	***

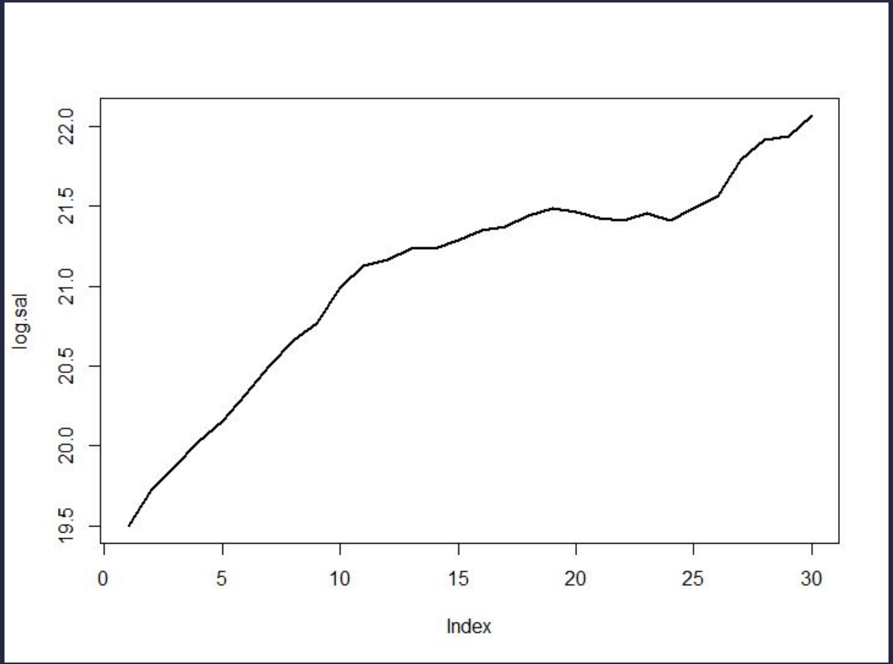
Some evident positive autocorrelation in the residuals to be explained



LINEAR MODELS PERFORM BETTER?

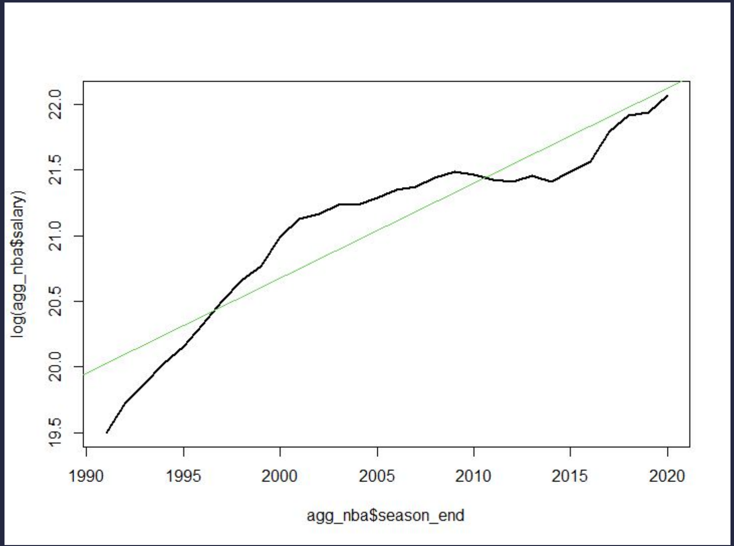
We manipulated the data in order to deal with a more linear scenario

Log transformation of the salaries

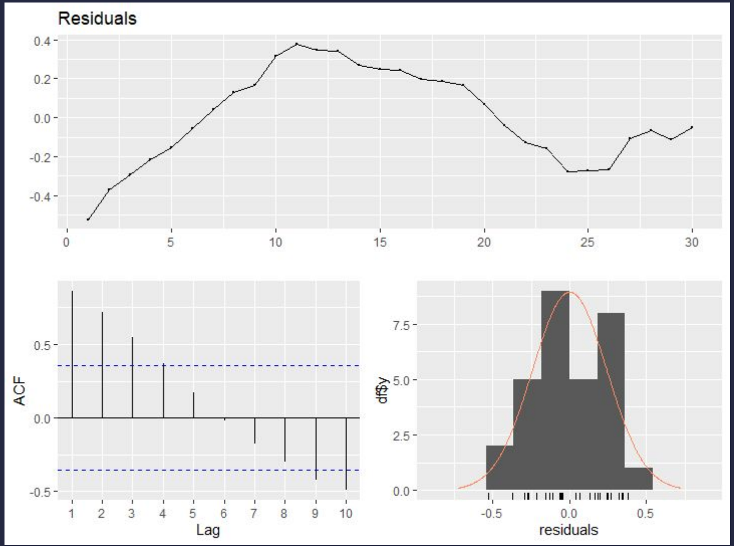


LINEAR MODEL

Decent fit capturing the trend
Adj R-sq of 0.8696



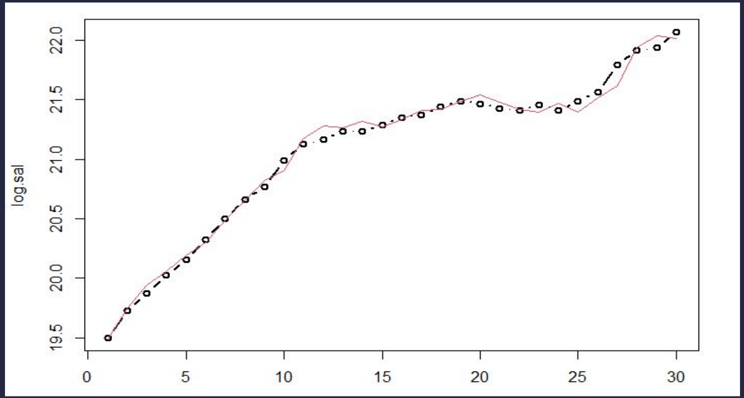
DW-Test of 1.07, some positive autocorrelations in the residuals



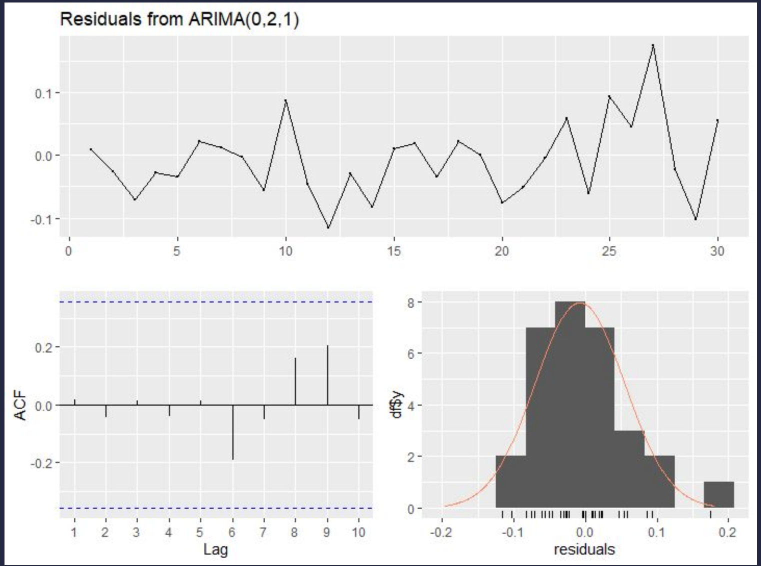
ARIMA MODEL(0,2,1)

Good Fit and the best AIC score

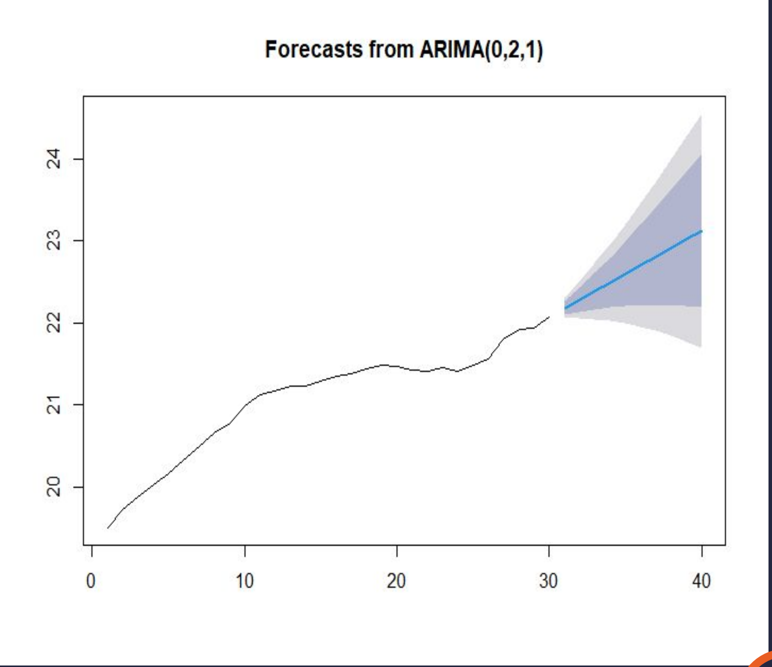
	(0,1,0)	(1,1,0)	(0,2,1)
AIC	-39,84	-67,42	-70,16



White noise residuals with no evident autocorrelation between them



ARIMA MODEL(0,2,1)



2021 and 2022 predictions

LogValue	Conversion	TrueValue
22.17	4,25e+09	4,20e+09
22.28	4,74e+09	4.95e+09

2023 coming soon

...

WORLDWIDE PHENOMENON?

LET's TAKE A LOOK INSIDE

In 1992

93%

Of the Nba players was USA born

In 2018

21.8%

Of the players are classed as international players

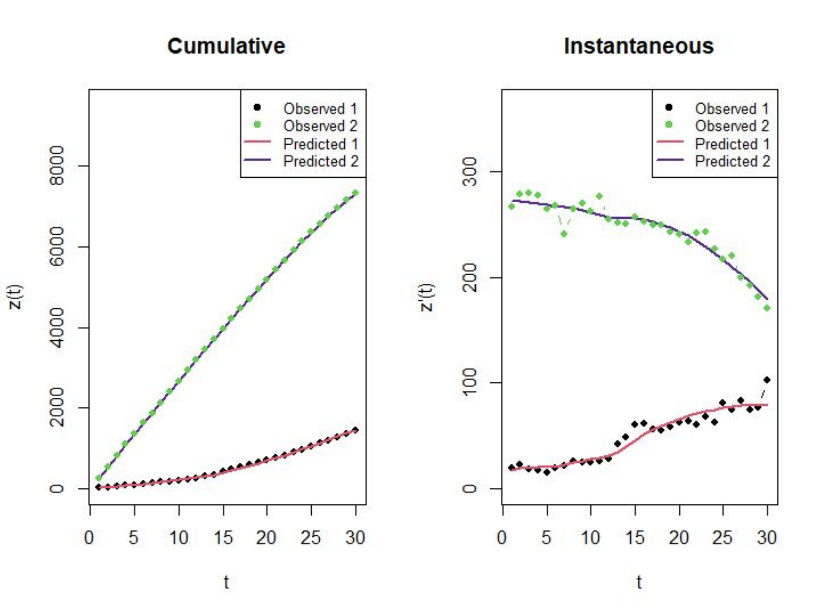
SO HOW DO THEY RELATE?

The UCRCD model shows that USA players compete with foreign players, but the latter collaborate with the former.

Coefficients:

	Estimate	Std.Error	Lower	Upper	p-value	
mc	1.208020e+04	4.585191e+02	1.118152e+04	1.297888e+04	4.00e-32	***
p1c	1.356544e-03	4.462677e-04	4.818757e-04	2.231213e-03	3.67e-03	**
p2	2.246560e-02	1.054755e-03	2.039831e-02	2.453288e-02	1.20e-27	***
q1c	-1.200337e-02	4.921878e-03	-2.165008e-02	-2.356670e-03	1.81e-02	*
q2	9.961600e-03	5.702106e-03	-1.214322e-03	2.113752e-02	8.64e-02	.
delta	2.589305e-01	4.137010e-02	1.778466e-01	3.400144e-01	6.99e-08	***
gamma	-2.040296e-01	6.611383e-02	-3.336104e-01	-7.444890e-02	3.22e-03	**

Nba global expansion

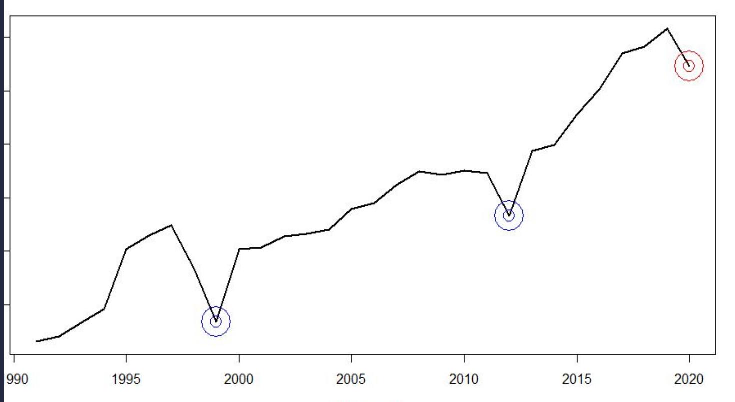


EVOLUTION OF THE GAME

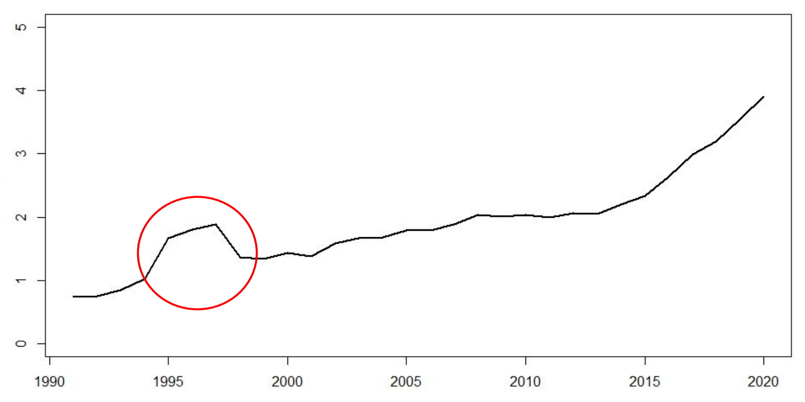
1979/80: 3 Pointers

3PT SHOTS THROUGH TIME

Number of triples attempted during the season

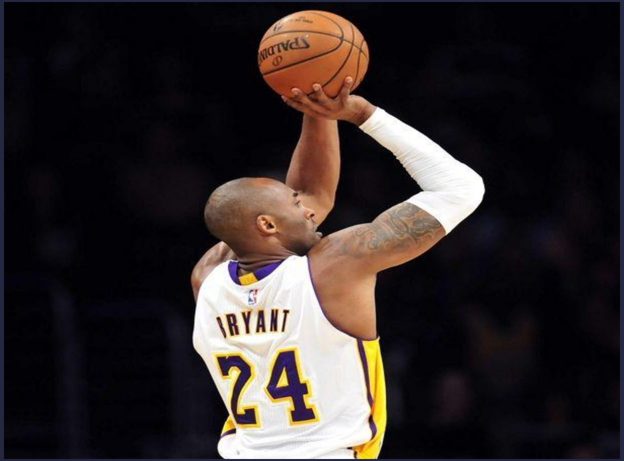


Average triples attempted by player per game

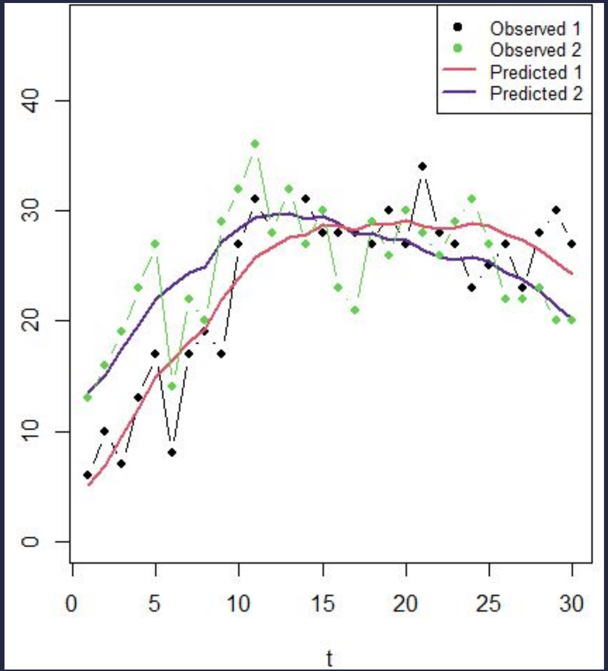


TOP PLAYERS EVOLUTION

Big Men vs Guards



TOP PLAYERS EVOLUTION



Counting top players for each year (salary > 18% of salary cap).

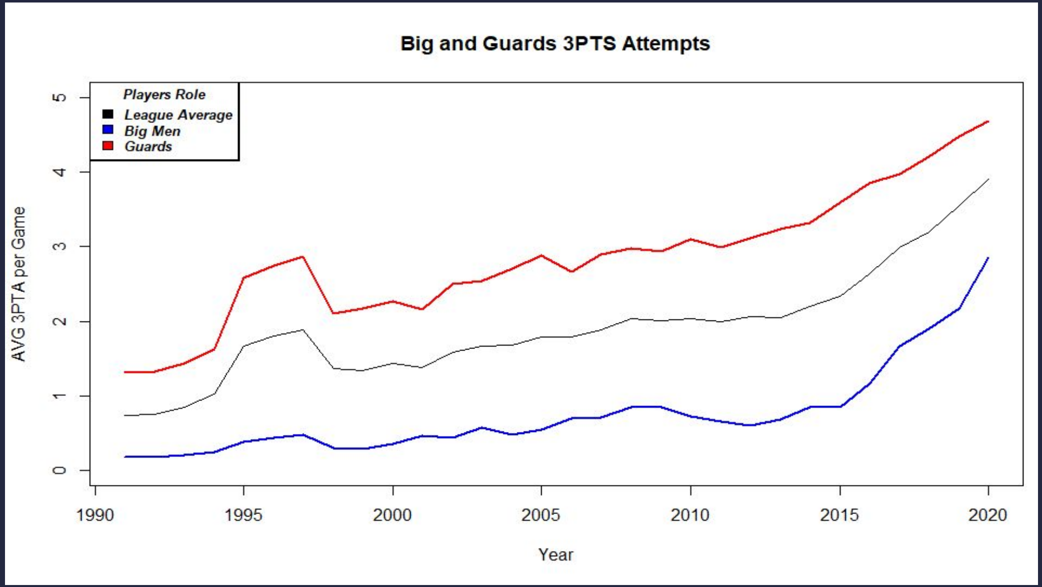
UCRCD shows that guards compete with big men

Coefficients:

	Estimate	Std.Error	Lower	Upper	p-value	
mc	2.321884e+03	2.747241e+02	1.783434e+03	2.860333e+03	2.14e-11	***
p1c	1.425297e-03	1.112674e-03	-7.555035e-04	3.606097e-03	2.06e-01	
p2	5.154554e-03	1.293462e-03	2.619416e-03	7.689693e-03	2.07e-04	***
q1c	1.932371e-01	5.877932e-02	7.803176e-02	3.084425e-01	1.80e-03	**
q2	2.003779e-01	5.416839e-02	9.420982e-02	3.065460e-01	5.15e-04	***
delta	-3.111397e-01	1.291790e-01	-5.643258e-01	-5.795354e-02	1.95e-02	*
gamma	3.533030e-01	1.173993e-01	1.232047e-01	5.834013e-01	4.00e-03	**

Evolution of the game promoted guards way to play

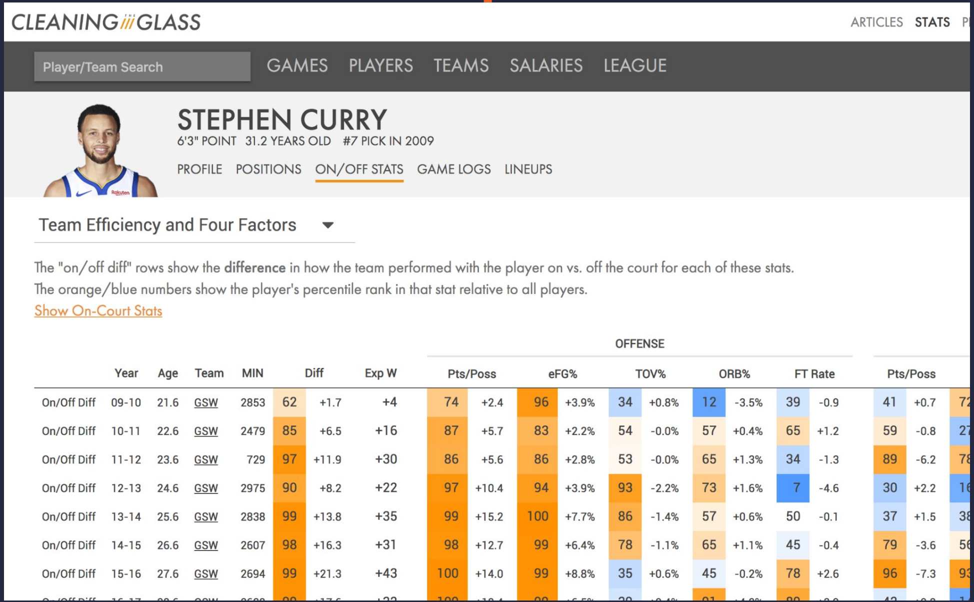
ADAPT TO SURVIVE



From 2015 the increasing trend of three points shots started to be a trait also on the Big Men line.
Showing an even faster growth

BACK TO THE WHOLE LEAGUE

What are the most significant statistics that influence the players salaries?



CLEANING GLASS

ARTICLES STATS

Player/Team Search

GAMES PLAYERS TEAMS SALARIES LEAGUE

STEPHEN CURRY

6'3" POINT 31.2 YEARS OLD #7 PICK IN 2009

PROFILE POSITIONS ON/OFF STATS GAME LOGS LINEUPS

Team Efficiency and Four Factors

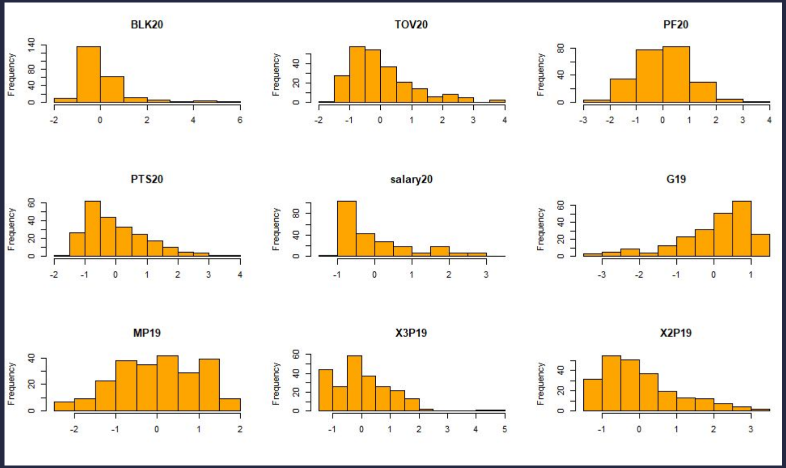
The "on/off diff" rows show the **difference** in how the team performed with the player on vs. off the court for each of these stats.
The orange/blue numbers show the player's percentile rank in that stat relative to all players.

Show On-Court Stats

								OFFENSE											
	Year	Age	Team	MIN	Diff		Exp W	Pts/Poss		eFG%		TOV%		ORB%		FT Rate		Pts/Poss	
On/Off Diff	09-10	21.6	GSW	2853	62	+1.7	+4	74	+2.4	96	+3.9%	34	+0.8%	12	-3.5%	39	-0.9	41	+0.7
On/Off Diff	10-11	22.6	GSW	2479	85	+6.5	+16	87	+5.7	83	+2.2%	54	-0.0%	57	+0.4%	65	+1.2	59	-0.8
On/Off Diff	11-12	23.6	GSW	729	97	+11.9	+30	86	+5.6	86	+2.8%	53	-0.0%	65	+1.3%	34	-1.3	89	-6.2
On/Off Diff	12-13	24.6	GSW	2975	90	+8.2	+22	97	+10.4	94	+3.9%	93	-2.2%	73	+1.6%	7	-4.6	30	+2.2
On/Off Diff	13-14	25.6	GSW	2838	99	+13.8	+35	99	+15.2	100	+7.7%	86	-1.4%	57	+0.6%	50	-0.1	37	+1.5
On/Off Diff	14-15	26.6	GSW	2607	98	+16.3	+31	98	+12.7	99	+6.4%	78	-1.1%	65	+1.1%	45	-0.4	79	-3.6
On/Off Diff	15-16	27.6	GSW	2694	99	+21.3	+43	100	+14.0	99	+8.8%	35	+0.6%	45	-0.2%	78	+2.6	96	-7.3

STATS AND SALARY

Stats distribution



Salary correlations

Age20	0.2081655
G20	0.1647131
MP20	0.5768627
X3P20	0.3361096
X2P20	0.5808521
FT20	0.6154182
TRB20	0.4655156
AST20	0.5520381
STL20	0.4160207
BLK20	0.2536754
TOV20	0.6023417
PF20	0.3410836
PTS20	0.6311433

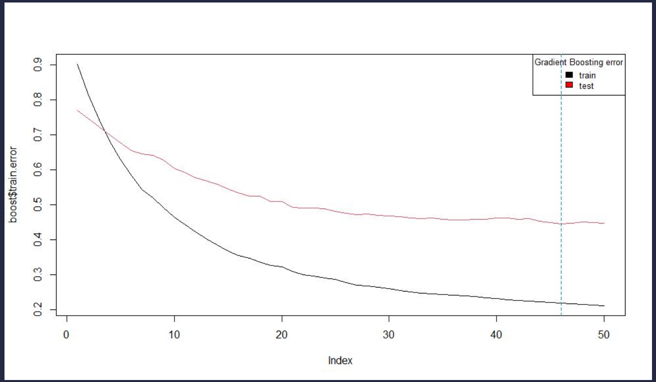
TOV19	0.6265392
PF19	0.3044048
PTS19	0.6842217
G18	0.1419813
MP18	0.5405299
X3P18	0.3753790
X2P18	0.5762544
FT18	0.6554624
TRB18	0.4530440
AST18	0.5153997
STL18	0.4268380
BLK18	0.2815855
TOV18	0.5584488
PF18	0.3446330
PTS18	0.6740730

GBM RELEVANT VARIABLES

Points made from previous seasons are the most significant variables to predict current salary

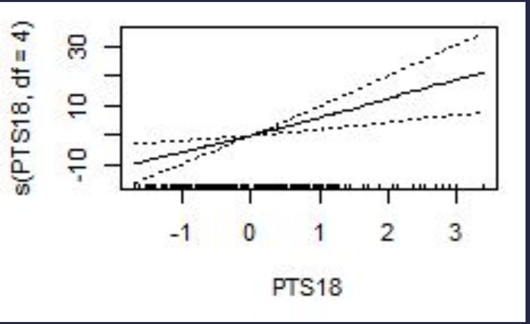
var	rel.inf
PTS18	14.7961503
PTS19	13.4803639
Age20	9.6089299
MP19	9.3173514
FT18	5.5464770
TRB20	5.5274050
STL19	4.6290067
MP20	3.9002723
FT20	3.7140912
TRB18	3.6694478
AST18	3.2767505
G19	2.2788476
X2P20	2.2641125
MP18	1.9512670
BLK20	1.8920508

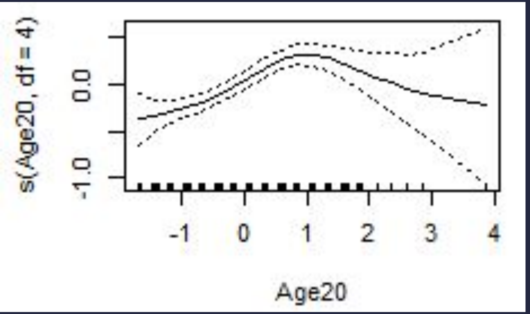
Train and test error plot



Hard to predict due to few samples and “bad deals”

GAM RELEVANT VARIABLES





“Points made” have an evident positive linear relation with the salary

The partial function of the age has a bell shape. Being in at the prime has a positive influence on salaries

CONCLUSIONS

- GGM has decent fit and gives reasonable interpretations, however it is not reliable because it analyzes the early phase of the process

- Arima captures very well the shape of the phenomenon and all the assumptions of the model are satisfied

- UCRCD application suggest a partial competition on US vs Rest of World and the class of top players analysis.

- Prediction task is hard, but both GB and GAM method agree that the most influent statistics, are related to previous seasons.