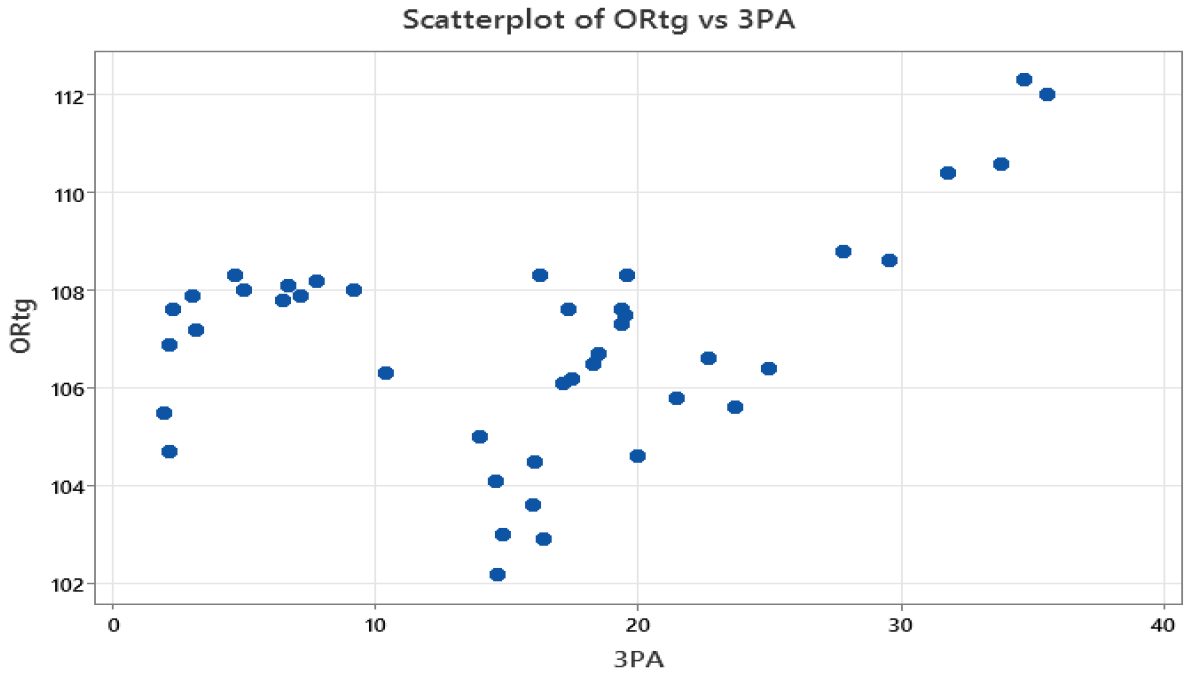
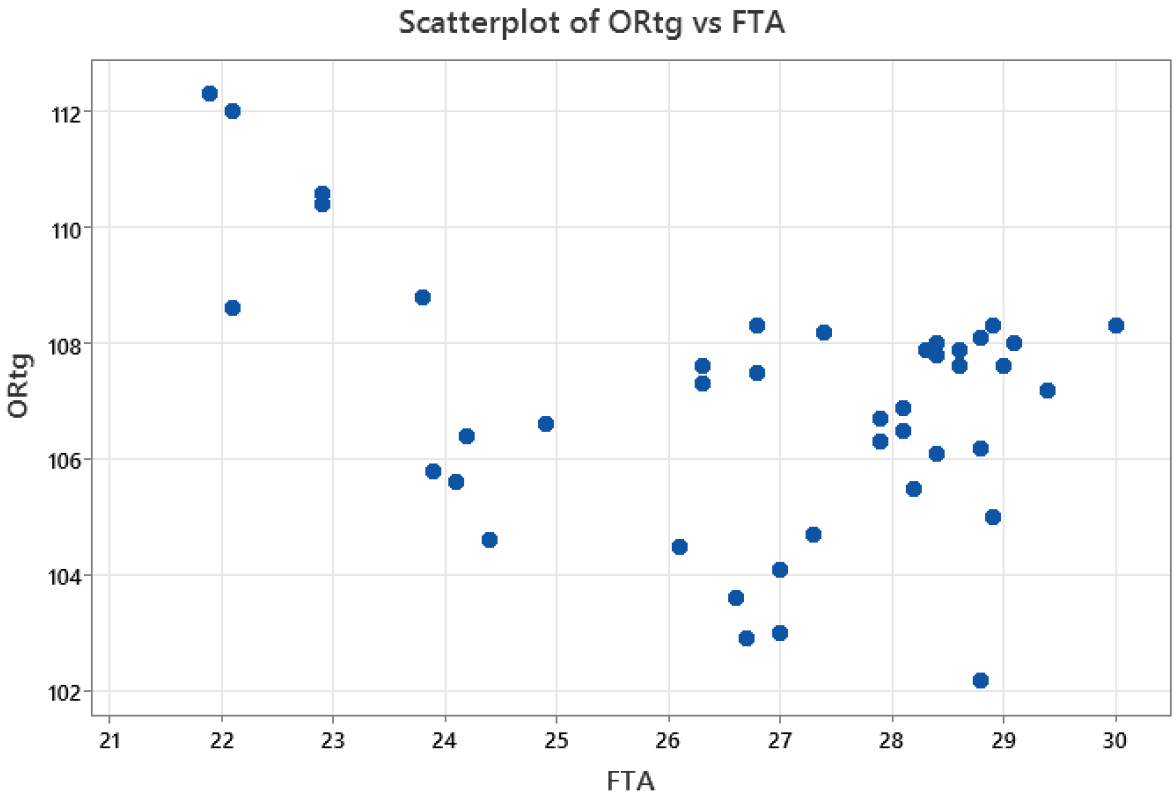
Mihir Paranjape
11/17/2022

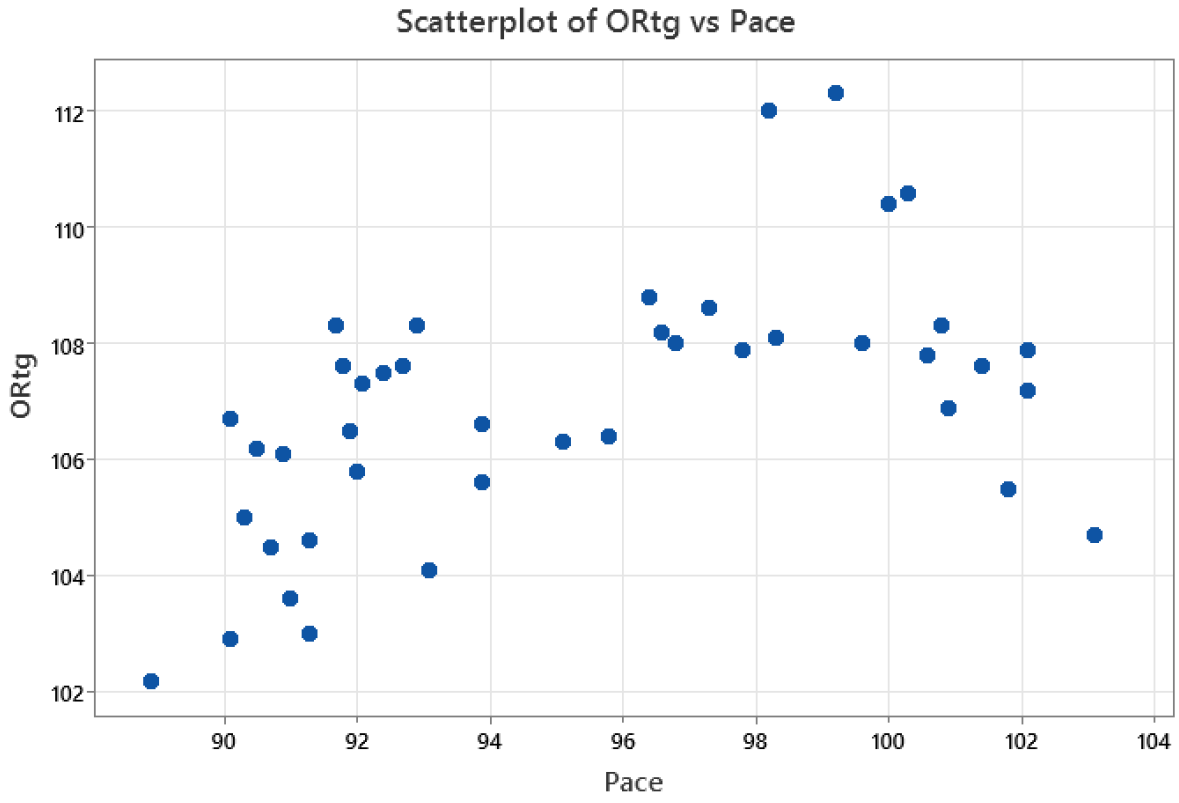
NBA Historical Offense Analysis

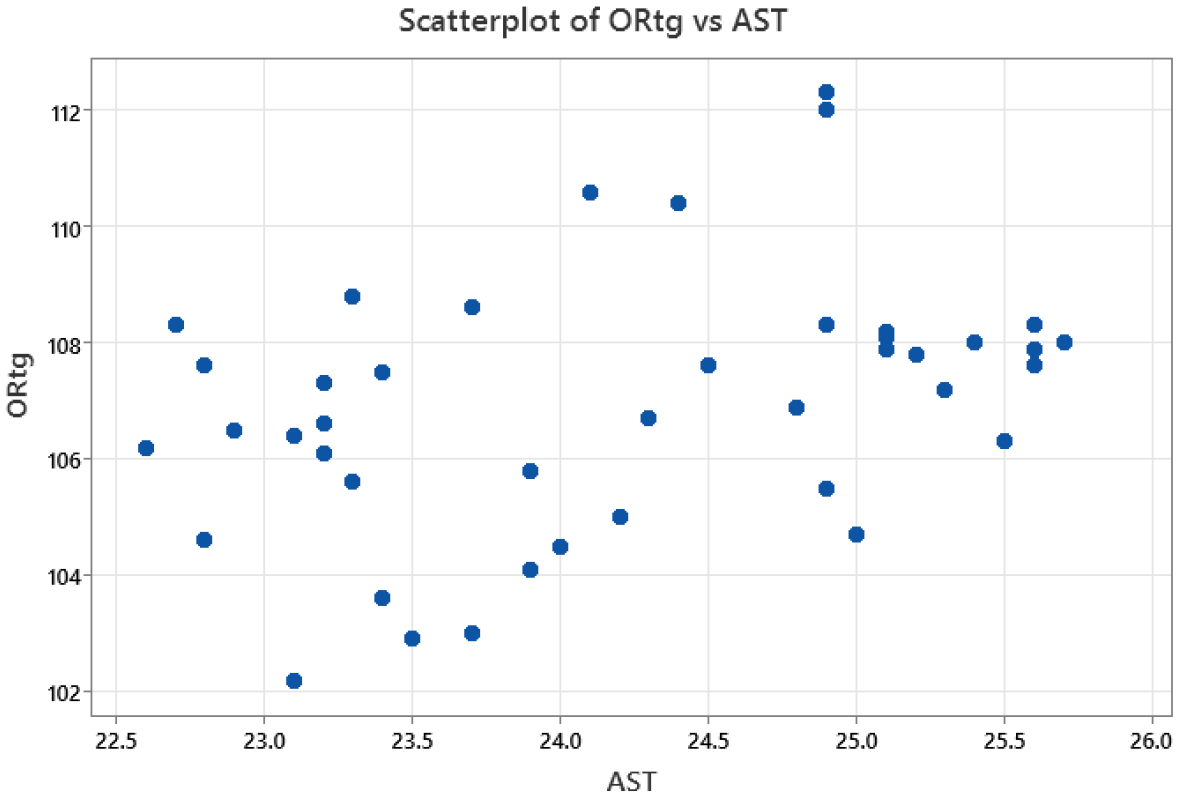
It is safe to say that the NBA has gone through a stark transformation over the last couple of decades. Everything from the way the game has been officiated to the skills that most players possess has seen drastic changes in one way or another. One thing which particularly interests me is how offenses have changed league-wide over the past 40 years or so. Obviously there has been a league-wide shift towards 3 pointers in teams' offensive diet, but I would be interested to see what other changes there might be and how it has affected the league's average offensive efficiency. To do this, I will run a regression with 42 observations initially, starting from league averages in the 1980-1981 season going to league averages in the 2021-2022 season. I will be using the average offensive rating of the league as my response variable. This measures how many points a team scores per 100 possessions, meaning it adjusts for the pace of that year. For the predictors, I will look at areas that I think are notable in basketball offenses: passing, 3-point shooting volume, turnovers, pace, and free throw attempts. The predictors I will be using will be per 100 possessions where applicable. I will be looking at league-wide 3 pointers attempted, turnover percentage, assists, pace, and free throws attempts. Turnover percentage is an estimate of the percentage of turnovers committed per 100 plays, and pace is the number of possessions per game. All of these numbers are league-wide averages for a particular season. Obtaining this data was fairly simple. I simply went to Basketball reference's page for league averages in history and selected per 100 possession numbers. I exported this to excel, ordered it by time and decided to go as far back as 1980, and then I exported this excel data into minitab where it was ready for analysis. I also added a second time column numbered 1-42 instead of using the year.

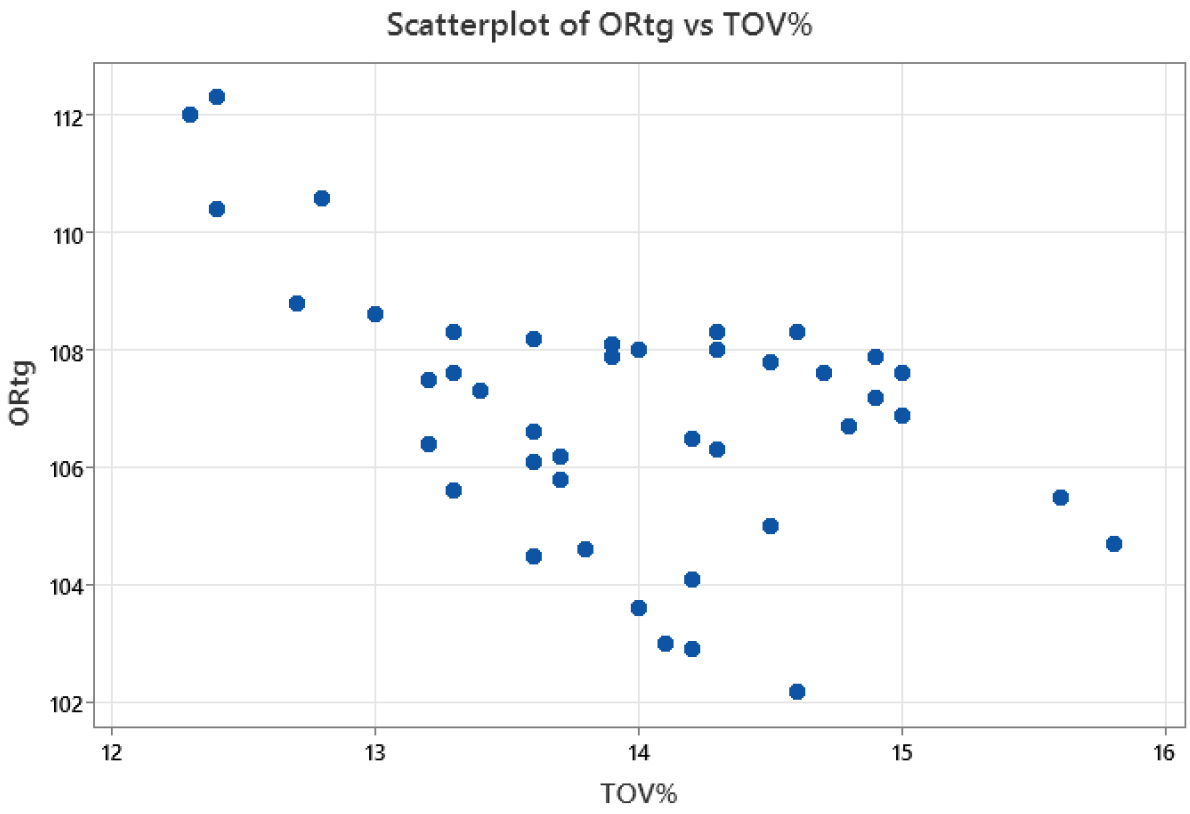
I started off my analysis by looking at scatterplots of the five predictor variables I chose versus the response variable, offensive rating.


Scatterplot of ORtg vs 3PA
ORtg
112
110
108
106
104
102
0
10
20
30
40
3PA

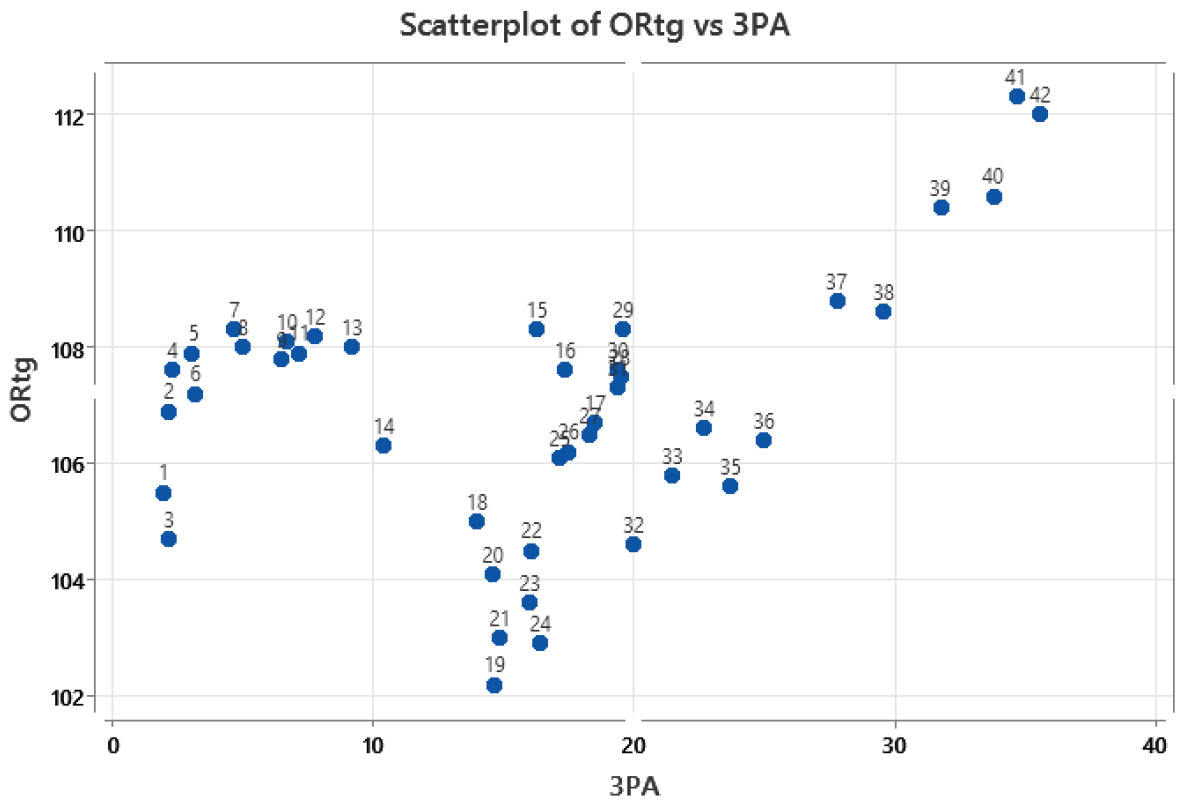

Scatterplot of ORtg vs FTA
ORtg
112
110
108
106
104
102
21
22
23
24
25
26
27
28
29
30
FTA

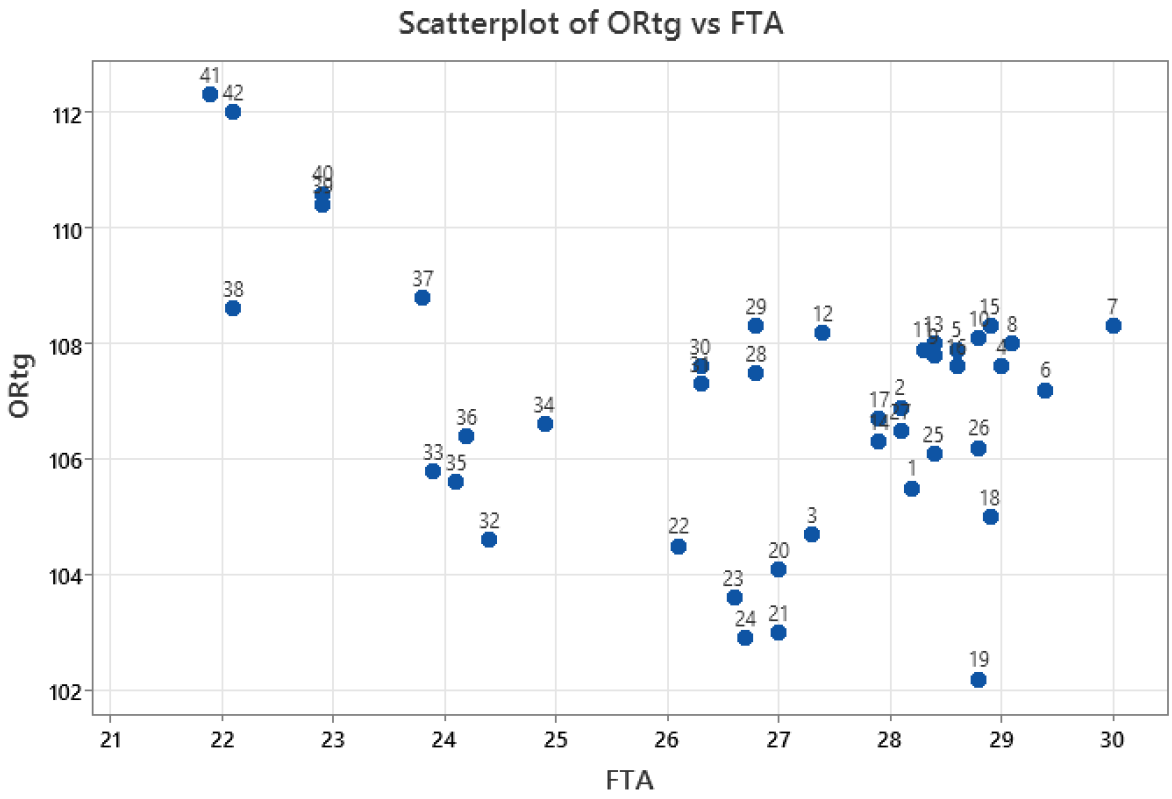

Scatterplot of ORtg vs Pace
ORtg
Pace
112
110
108
106
104
102
90
92
94
96
98
100
102
104


Scatterplot of ORtg vs AST
ORtg
AST
112
110
108
106
104
102
22.5
23.0
23.5
24.0
24.5
25.0
25.5
26.0



From the results above, it seems that most of the data is centered around an offensive rating between around 108 and 104, with a couple of seasons eclipsing an offensive rating of 110. Looking a little further, we can see that the last 6 seasons have had the 6 highest offensive ratings in the last 42 years. This is quite the anomaly and shows the improvements in offensive efficiency over the past several years. The only variable with a seemingly strong relationship is turnover percentage and maybe 3 pointers attempted, however there seems to be a segmenting of the data with 3 pointers attempted roughly around 10 3 pointers attempted. Here are the observation numbers of the scatterplot:


Scatterplot of ORtg vs 3PA
ORtg
3PA
112
110
108
106
104
102
0
10
20
30
40



We can see that there was a considerable jump from observation 14(1993) to observation 15(1994) in terms of 3 pointers attempted. This is largely due to the shortening of the 3 point line in the 1994-1995 that led to teams attempting more 3 pointers. Something like this could potentially skew the regression in some way due to subgrouping. In addition, there seems to be another partition in the data with the free throws attempted scatterplot. After observation 31(2010), the league average number of free throws attempted per 100 possessions dropped considerably. Perhaps some of this due to an increase in pace, and therefore a greater divisor for free throws attempted when looking at free throws attempted per 100 possessions. Other than this, I do not have an explanation for why this might be the case. There were no major rule changes to my knowledge at this point in time, and the three point revolution had not quite begun yet back in 2011. However, I will still see the regression residual results to determine whether these may be violations of assumptions. Here are the results with all the predictors in the model:

Regression Equation

ORtg = 51.00 + 0.2268 3PA + 0.745 FTA + 0.741 AST + 0.3586 Pace - 1.417 TOV%

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	51.00	9.91	5.14	0.000	
3PA	0.2268	0.0468	4.85	0.000	8.38
FTA	0.745	0.161	4.62	0.000	5.76
AST	0.741	0.263	2.81	0.008	2.81
Pace	0.3586	0.0631	5.68	0.000	3.14
TOV%	-1.417	0.355	-3.99	0.000	3.70

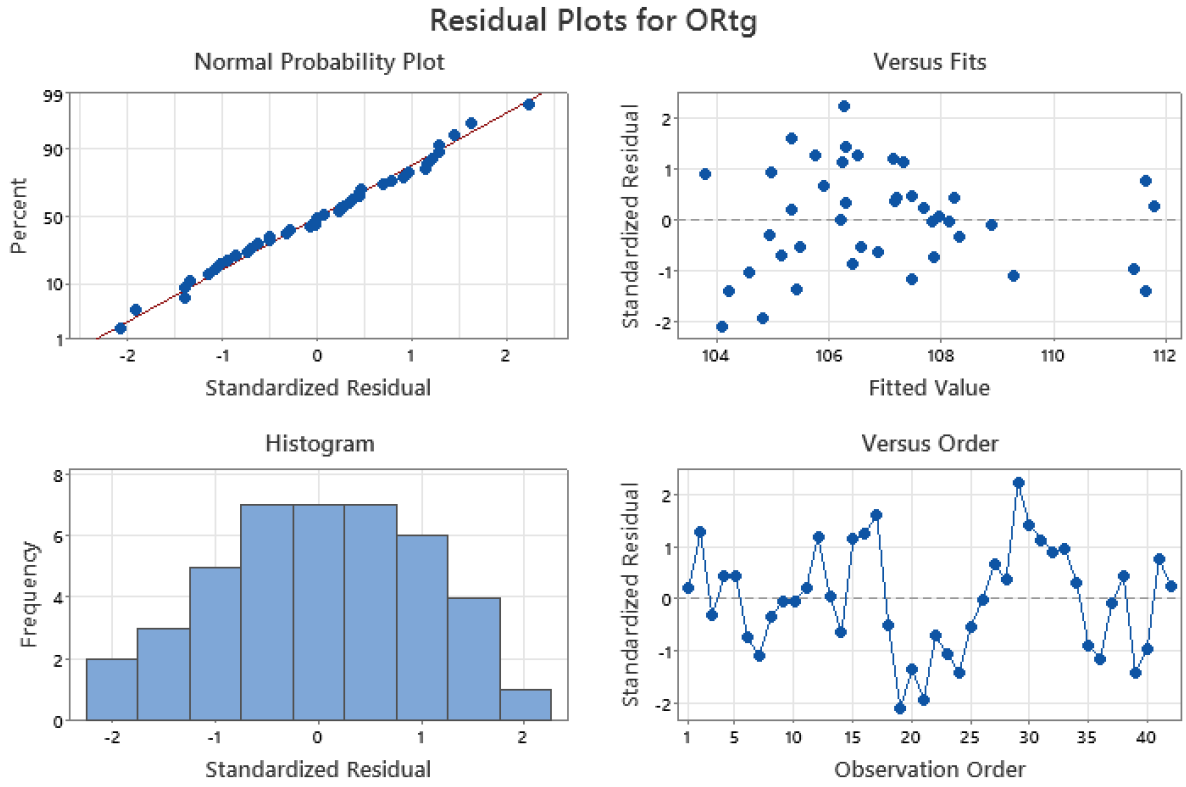
Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.980126	82.90%	80.53%	76.97%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	167.698	33.5396	34.91	0.000
3PA	1	22.561	22.5609	23.49	0.000
FTA	1	20.515	20.5154	21.36	0.000
AST	1	7.607	7.6071	7.92	0.008
Pace	1	31.019	31.0192	32.29	0.000
TOV%	1	15.258	15.2577	15.88	0.000
Error	36	34.583	0.9606		
Total	41	202.281			

The R^2 of this regression is at 82.9%, which means that 82.9% of the variability in offensive rating is explained by the variability in the predictors. This seems to be a pretty strong regression. None of the variance inflation factors are above 10, which means that multicollinearity is not much of a concern. The F value for the regression is statistically significant with a p-value of below 0.001, which means there is extremely strong evidence of an effect on the response variable, offensive rating. The intercept has no meaningful interpretation in this case. For the coefficients, assuming all other predictors are held constant, an increase in 3 point attempts by one per 100 possessions is associated with a 0.2268 increase in average offensive rating. A single unit increase in free throw attempts per 100 possessions is associated with a 0.745 increase in offensive rating. An assist increase in assists per 100 possessions is associated with a 0.741 increase in offensive rating. An increase of 1 possession per game is associated with a 0.3586 increase in offensive rating. An increase of 1 percentage point in turnover percentage is associated with a -1.417 decrease in offensive rating. The signs for these coefficients are more or less what I would expect, although it is surprising that as the game has gotten faster (more possessions per game) there has been an associated increase in scoring per possession. The standard error of the estimate of 0.98 says that we can predict within plus or minus 1.96 for the response variable 95% of the time. I will use k-fold cross validation to determine the standard error of the estimate for the final model.

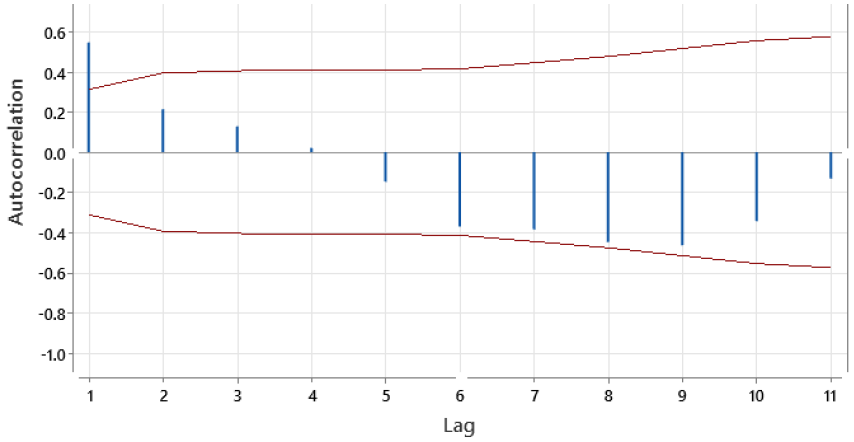


Taking a look at the residual plots, we can see that the errors seem to be normally distributed according to the histogram and normal probability plot. However, we can see that there are four points with fitted

values far above the rest of the fitted values. These happen to be the last four NBA seasons. The residuals in the fitted values plot besides those 4 points seem to show an almost quadratic pattern, first starting low, then increasing, and then ending low. Another interesting point would be the time series plot with residuals. It seems that there is a sharp decrease in the standardized residuals near the middle of the observations (around 20-25) that is worth noting.

Durbin-Watson Statistic

Durbin-Watson Statistic = 0.880495



Runs Test: SRES

Descriptive Statistics

		Number of Observations	
N	K	≤ K	> K
42	0	21	21

Test

Null hypothesis	H_0: The order of the data is random
Alternative hypothesis	H_1: The order of the data is not random

Number of Runs		
Observed	Expected	P-Value
13	22.00	0.005

The durbin watson statistic shows a value of 0.8804, which is well below the upper limit of 1.583 for the parameters here (p=5, n= ~40). If we also look at the ACF plot, we can see that for lag-1, there is statistically significant positive autocorrelation, although the correlations do not seem to have the strongest evidence of first order autoregressive decaying as is assumed by the Durbin-Watson test. The runs test also seems to indicate that there might be autocorrelation since the number of runs observed is significantly different from the number of runs expected. This also indicates positive autocorrelation.

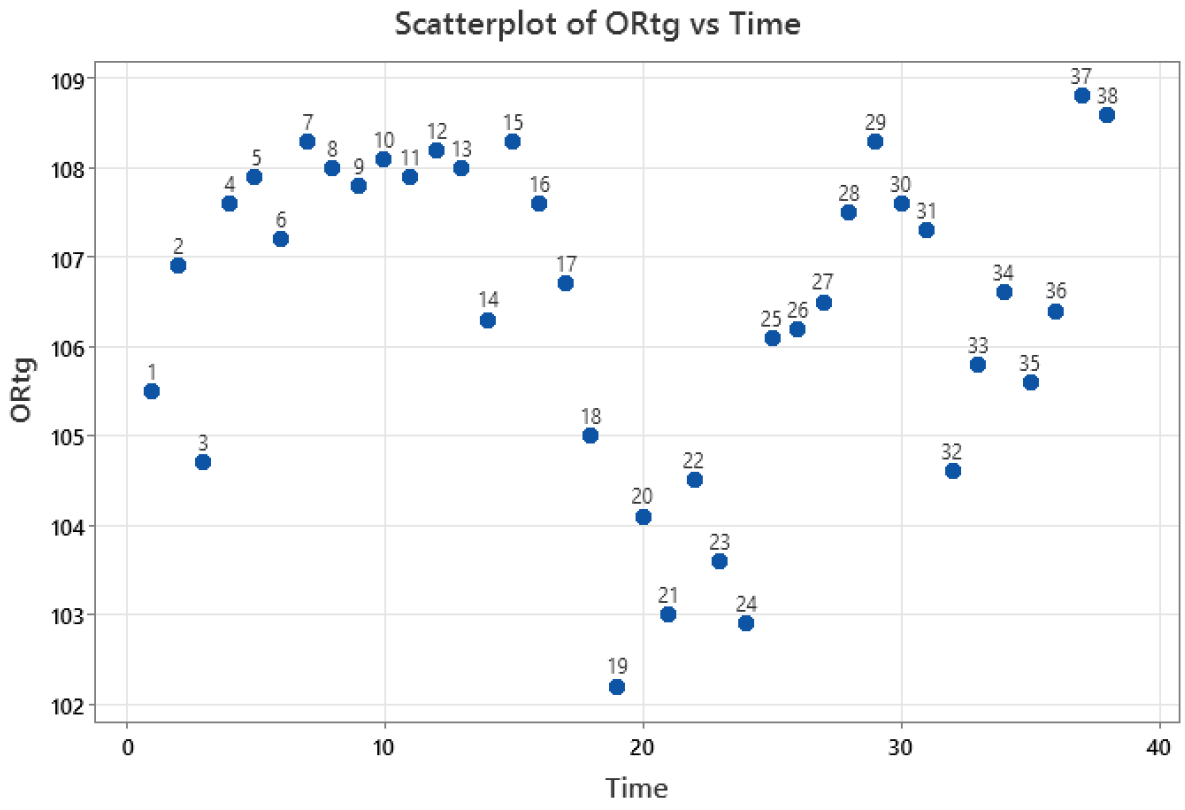
Tally					
SRES	Count	HI	Count	COOK	Count
-2.07147	1	0.041155	1	0.000002	1
-1.91057	1	0.063782	1	0.000012	1
-1.39937	1	0.064328	1	0.000014	1
-1.39324	1	0.064447	1	0.000097	1
-1.34961	1	0.072128	1	0.00012	1
-1.13993	1	0.072814	1	0.001108	1
-1.07513	1	0.077154	1	0.001249	1
-1.03237	1	0.083103	1	0.002065	1
-0.95099	1	0.090719	1	0.002241	1
-0.86798	1	0.097389	1	0.002279	1
-0.73323	1	0.101282	1	0.003522	1
-0.69977	1	0.10467	1	0.003912	1
-0.63115	1	0.10496	1	0.004112	1
-0.51162	1	0.109235	1	0.005103	1
-0.50712	1	0.109983	1	0.005657	1
-0.32994	1	0.111015	1	0.006055	1
-0.29994	1	0.111588	1	0.00737	1
-0.07623	1	0.116595	1	0.010498	1

-0.03027	1	0.119617	1	0.010713	1
-0.02486	1	0.121879	1	0.01303	1
-1	1	0.12378	1	0.013468	1
0.07036	1	0.130662	1	0.013808	1
0.22065	1	0.131041	1	0.014598	1
0.23282	1	0.13498	1	0.014758	1
0.26826	1	0.138939	1	0.015505	1
0.3301	1	0.140642	1	0.023368	1
0.36974	1	0.156802	1	0.028788	1
0.43253	1	0.15965	1	0.030024	1
0.43304	1	0.164461	1	0.032562	1
0.4594	1	0.172994	1	0.037295	1
0.68631	1	0.176989	1	0.03792	1
0.7733	1	0.19846	1	0.041909	1
0.9087	1	0.205285	1	0.042511	1
0.95066	1	0.205942	1	0.042651	1
1.14005	1	0.206269	1	0.048323	1
1.14707	1	0.212317	1	0.056988	1
1.20724	1	0.21643	1	0.062745	1
1.26873	1	0.231509	1	0.069574	1
1.28061	1	0.236735	1	0.069579	1
1.42916	1	0.242767	1	0.107848	1
1.61091	1	0.245955	1	0.115921	1
2.2189	1	0.329544	1	0.134147	1
N=	42	N=	42	N=	42

Here is the table for standardized residuals, leverage values, and cook's distance. There seems to be no residual close to greater than an absolute value 2.5. For leverage values, the value we are looking to flag

is 2.5*((5+1)/42) = 0.357. There are no values near this, as the closest value is 0.329. There is no value for Cook's distance greater than 1, which means there are no uniquely influential values.

Looking at all these statistics about the residuals as well as the original scatterplots, it seems that there might be subgroups in the data, as well as autocorrelation. This means that at least two of the assumptions about errors are potentially violated: Constant variance and lack of correlation. First I will take a look at the response variable versus time.



Here we see something peculiar which explains some of the issues I was seeing with the residuals. The years from 1997 to 2003 have a noticeably lower offensive rating as compared to the data points around that period. The reason for this is that the league moved the 3 point line back to its original length in the 1997 season, which caused the league wide offensive rating to plummet for the next several years. I think the best way to account for this shift would be to include an indicator variable for the years between 1997 and 2003. I also will add an indicator variable for seasons before 1994, when the three point line was shortened. Here is the regression with the new indicator variables:

Regression Equation

ORtg = 62.96 + 0.497 FTA + 0.2376 Pace - 1.539 TOV% + 0.0930 3PA - 1.91 Before 1994 + 1.199 AST - 2.514 Between 1997 and 2003

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	62.96	7.12	8.84	0.000	
FTA	0.497	0.118	4.23	0.000	6.74
Pace	0.2376	0.0542	4.38	0.000	5.09
TOV%	-1.539	0.297	-5.19	0.000	5.64
3PA	0.0930	0.0615	1.51	0.140	31.70
Before 1994	-1.91	1.07	-1.79	0.082	24.20
AST	1.199	0.232	5.17	0.000	4.78
Between 1997 and 2003	-2.514	0.385	-6.53	0.000	1.98

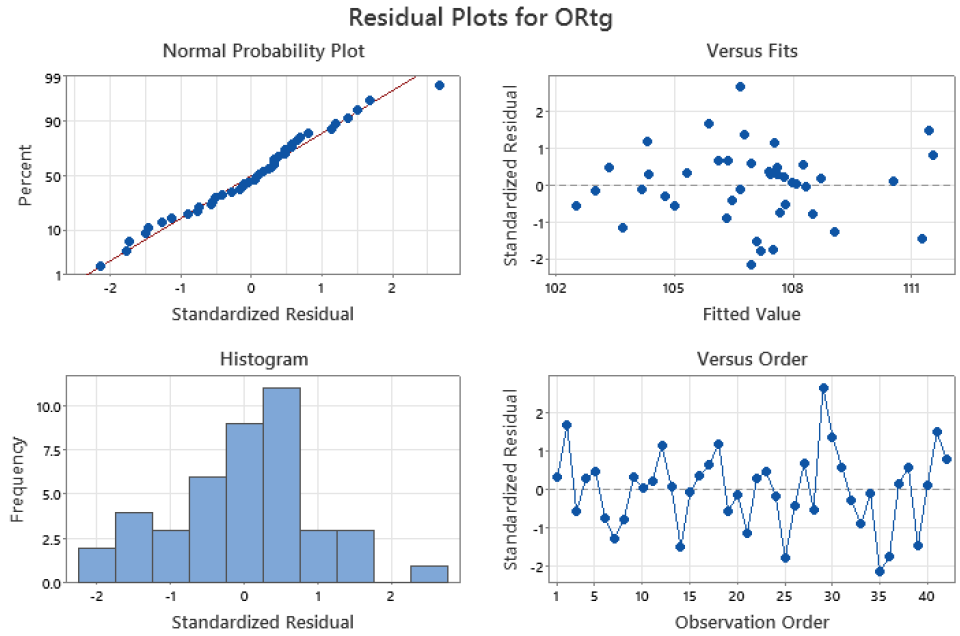
Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
0.661847	92.64%	91.12%	89.21%

Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	187.388	26.7697	61.11	0.000
FTA	1	7.821	7.8214	17.86	0.000
Pace	1	8.407	8.4071	19.19	0.000
TOV%	1	11.790	11.7902	26.92	0.000
3PA	1	1.003	1.0028	2.29	0.140
Before 1994	1	1.403	1.4035	3.20	0.082
AST	1	11.727	11.7273	26.77	0.000
Between 1997 and 2003	1	18.660	18.6603	42.60	0.000
Error	34	14.893	0.4380		
Total	41	202.281			

The R-Squared has increased to 92.64% and the F-statistic is still quite statistically significant, signaling strong evidence of a strong effect. The coefficient for the indicator variable for before 1994 suggests that seasons before 1994 have an associated 1.91 decrease in offensive rating, holding all other predictors constant. The indicator variable for between 1997 and 2003 (inclusive) has a coefficient of -2.514, which means that seasons within that time period have an associated -2.514 decrease in offensive rating, holding all other predictors constant. However, it is the residual plots that are more important. Here they are:

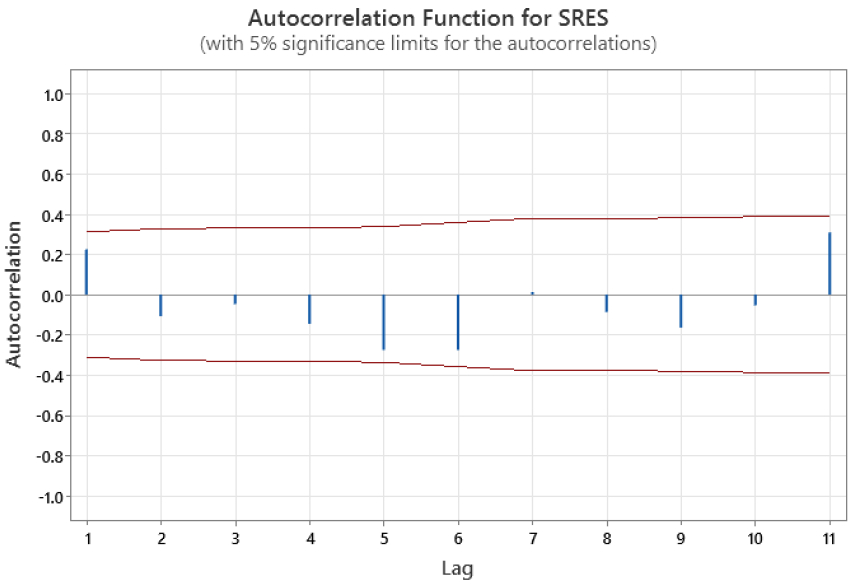


The Residuals seem normally distributed and the versus fits scatterplot looks random and without pattern as compared to the plot from the last regression. The time series plot with residuals looks more random and without autocorrelation. But we can verify this using the three tests we have.

Durbin-Watson Statistic

Durbin-Watson Statistic = 1.51255

The Durbin-Watson statistic is now 1.51, which is still below the upper bound critical value of 1.997 for lambda =8(7 predictors) and n= 42. Let us take a look at the ACF plot and runs test.



Descriptive Statistics

	Number of Observations		
N	K	≤ K	> K
42	0	19	23

Test

Null hypothesis	H_0: The order of the data is random
Alternative hypothesis	H_1: The order of the data is not random

Number of Runs		
Observed	Expected	P-Value
17	21.81	0.129

The lagged correlations do not seem to follow a first-order autoregressive pattern, which means that we should question the Durbin-Watson statistic in its validity. Both the ACF plot and the runs test do not show strong evidence of autocorrelation at the .05 level of significance. Finally, let's take a look at the standardized residuals, leverage values, and cook's distance.

SRES	Count	HI	Count	COOK	Count
-2.13215	1	0.075764	1	0.000032	1

-1.7745	1	0.086557	1	0.000119	1
-1.7212	1	0.091373	1	0.000141	1
-1.49628	1	0.099044	1	0.000152	1
-1.45287	1	0.100297	1	0.000422	1
-1.27203	1	0.112795	1	0.000455	1
-1.12594	1	0.120816	1	0.000661	1
-0.88976	1	0.12786	1	0.000692	1
-0.77103	1	0.127894	1	0.001031	1
-0.73523	1	0.129089	1	0.001182	1
-0.56717	1	0.130615	1	0.001811	1
-0.55678	1	0.132444	1	0.002934	1
-0.51166	1	0.136641	1	0.003536	1
-0.41812	1	0.136862	1	0.003694	1
-0.28115	1	0.137644	1	0.004591	1
-0.16131	1	0.139329	1	0.004642	1
-0.1202	1	0.144675	1	0.005137	1
-0.1076	1	0.147334	1	0.005223	1
-0.04644	1	0.153013	1	0.005995	1
0.04146	1	0.154986	1	0.006523	1
0.07137	1	0.168834	1	0.006968	1
0.1131	1	0.177839	1	0.010212	1
0.16916	1	0.186107	1	0.010939	1
0.23711	1	0.189271	1	0.013323	1
0.28795	1	0.192689	1	0.018471	1
0.3105	1	0.196804	1	0.019817	1
0.31586	1	0.20883	1	0.023453	1
0.31643	1	0.220285	1	0.027348	1

0.37859	1	0.227075	1	0.028628	1
0.47334	1	0.228926	1	0.034301	1
0.48126	1	0.229874	1	0.04071	1
0.57408	1	0.235643	1	0.041842	1
0.57716	1	0.250729	1	0.043442	1
0.65324	1	0.250731	1	0.04375	1
0.68272	1	0.262744	1	0.047519	1
0.80858	1	0.297171	1	0.055144	1
1.13753	1	0.302011	1	0.057145	1
1.19095	1	0.305404	1	0.068012	1
1.36812	1	0.322798	1	0.069645	1
1.49084	1	0.33013	1	0.085651	1
1.66956	1	0.343463	1	0.134384	1
2.6537	1	0.38761	1	0.177133	1
N=	42	N=	42	N=	42

Descriptive Statistics: 3PA, FTA, AST, Pace, TOV%

Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
3PA	42	0	15.92	1.46	9.46	2.00	7.08	16.35	20.38	35.60
FTA	42	0	26.764	0.352	2.281	21.900	24.775	27.350	28.600	30.000
AST	42	0	24.186	0.150	0.975	22.600	23.275	24.150	25.100	25.700
Pace	42	0	95.421	0.663	4.298	88.900	91.600	94.500	99.700	103.100
TOV%	42	0	13.933	0.128	0.828	12.300	13.300	13.950	14.525	15.800

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
	Season	3PA	FTA	AST	Pace	TOV%	ORtg	Time	Before 1994	Between 1997 and 2003
26	2005	17.5	28.8	22.6	90.5	13.7	106.2	26	0	0
27	2006	18.3	28.1	22.9	91.9	14.2	106.5	27	0	0
28	2007	19.5	26.8	23.4	92.4	13.2	107.5	28	0	0
29	2008	19.6	26.8	22.7	91.7	13.3	108.3	29	0	0

There is one observation with a residual of 2.6537(2008). As for Cook's distance, there are no points above or near 1. There are no leverage values above or near [2.5((7+1)/42)= 0.4762] 0.4762. The 2008 season had an unusually high offensive rating league-wide considering the pace was within the first quartile and the assists per 100 possessions was also within the first quartile. However, other than these factors, I am not completely sure why else this observation would be an outlier. To deal with this outlier I will use an imputed value based on the average of the time period before and after. Instead of 108.3, I will use 107.55(average of 107.5 and 107.6) as the offensive rating for 2008. With this, I will then move onto a best subsets regression with all the variables, forcing the indicator variables into the model.

Response is ORtg

Total Vars	R-Sq	R-Sq (adj)	PRESS	R-Sq (pred)	Mallows Cp	S	AICc	BIC
3	74.0	71.9	64.1	68.1	103.4	1.1727	140.035	147.056
3	70.7	68.4	72.1	64.1	120.5	1.2435	144.962	151.983
4	87.9	86.6	29.8	85.2	32.0	0.81103	110.673	118.699
4	82.8	80.9	42.7	78.7	58.8	0.96608	125.369	133.395
5	89.5	88.0	27.8	86.1	25.5	0.76533	107.545	116.414
5	88.4	86.8	30.0	85.0	31.0	0.80291	111.572	120.442
6	93.0	91.8	19.8	90.2	8.8	0.63203	93.356	102.894
6	89.8	88.1	28.3	85.9	25.7	0.76407	109.293	118.831
7	93.6	92.2	19.2	90.4	8.0	0.61662	93.327	103.341

Looking at the best subsets model, the Mallows Cp value, the standard error of the estimate, and the AICc are all minimized with the 7 predictor model, in addition to the predictive R-squared being the highest. Therefore, this is the model I will use with the imputed value.

Here are the regression results:

Regression Equation

ORtg = 61.74 + 0.487 FTA + 0.2359 Pace - 1.499 TOV% + 0.0954 3PA - 1.906 Before 1994 + 1.243 AST - 2.478 Between 1997 and 2003

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	61.74	6.64	9.30	0.000	
FTA	0.487	0.110	4.44	0.000	6.74
Pace	0.2359	0.0505	4.67	0.000	5.09
TOV%	-1.499	0.276	-5.43	0.000	5.64
3PA	0.0954	0.0573	1.66	0.105	31.70
Before 1994	-1.906	0.993	-1.92	0.063	24.20
AST	1.243	0.216	5.75	0.000	4.78
Between 1997 and 2003	-2.478	0.359	-6.91	0.000	1.98

S	R-sq	R-sq(adj)	R-sq(pred)	42-fold S	42-fold R-sq
0.616622	93.56%	92.24%	90.43%	0.676468	90.43%

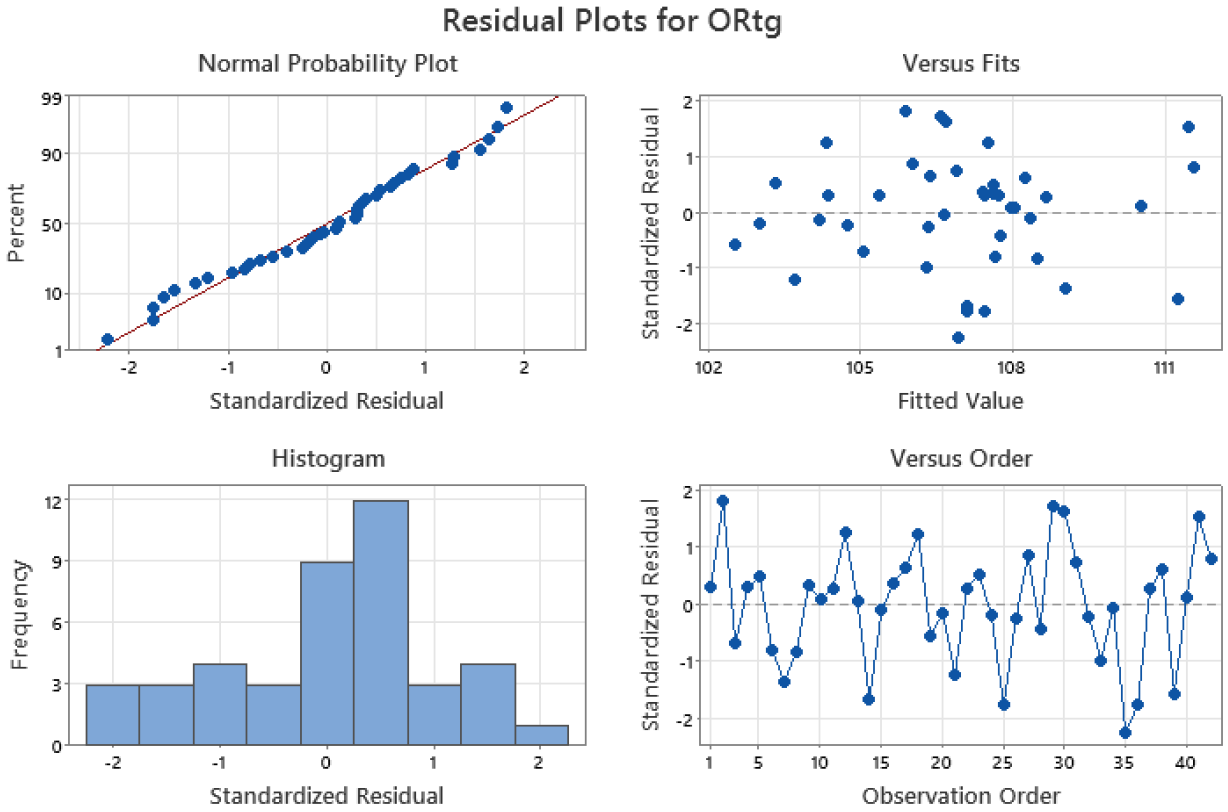
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	7	187.863	26.8376	70.58	0.000
FTA	1	7.506	7.5057	19.74	0.000
Pace	1	8.289	8.2886	21.80	0.000
TOV%	1	11.194	11.1939	29.44	0.000
3PA	1	1.054	1.0537	2.77	0.105
Before 1994	1	1.401	1.4007	3.68	0.063
AST	1	12.590	12.5899	33.11	0.000
Between 1997 and 2003	1	18.130	18.1303	47.68	0.000
Error	34	12.928	0.3802		

The R-Squared for this regression model is 93.56%. Using k-fold cross validation, a more accurate standard error of the estimate would be 0.676, meaning that this model can predict to within plus or minus 1.352 units of offensive rating about 95% of the time. The coefficient for free throw attempts means that holding all else in the model fixed, an increase of one free throw attempt per 100 possessions league wide

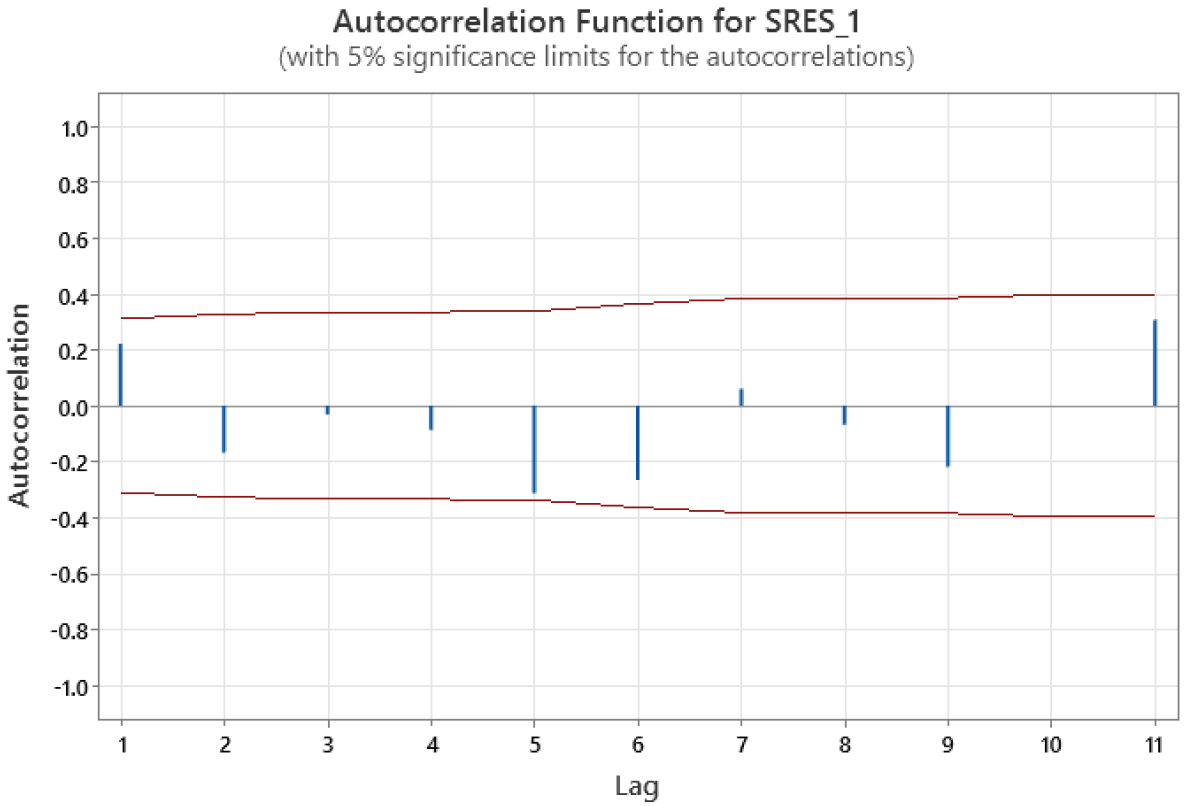
is associated with a 0.487 increase in offensive rating league-wide. An extra possession per game league wide is associated with a 0.2359 increase in offensive rating. A percentage point increase in turnover percentage is associated with a decrease of 1.499 in offensive rating. A one 3-pointer attempt increase league-wide per 100 possessions is associated with a 0.0954 increase in offensive rating. An increase of one assist per 100 possessions is associated with a 1.243 increase in offensive rating league-wide. Seasons before 1994 were associated with a 1.906 decrease in offensive rating and the seasons between 1997 and 2003 were associated with a 2.478 decrease in offensive rating. The multicollinearity for the 3 pointers attempted variable is very much explained by its association with the "Before 1994" indicator variable. The only reason I included this indicator variable in the first place was to account for the shortening of the 3 point line within the model. All t-tests seem to have highly significant values except for the three pointers attempted variable and the "Before 1994" indicator variable.

Here are the standardized residual plots:



Durbin-Watson Statistic

Durbin-Watson Statistic = 1.51573



Descriptive Statistics

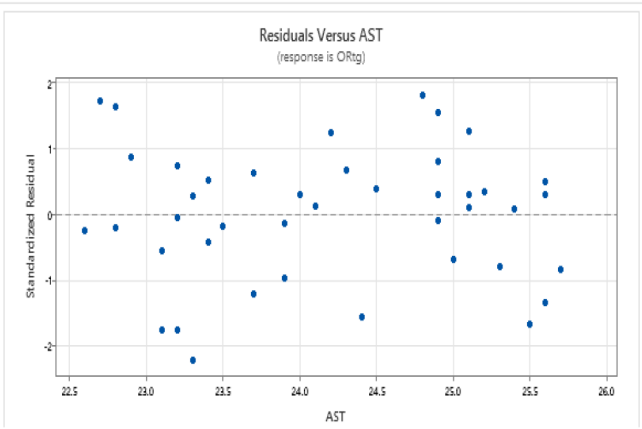
N	K	Number of Observations ≤ K	Number of Observations > K
42	0	19	23

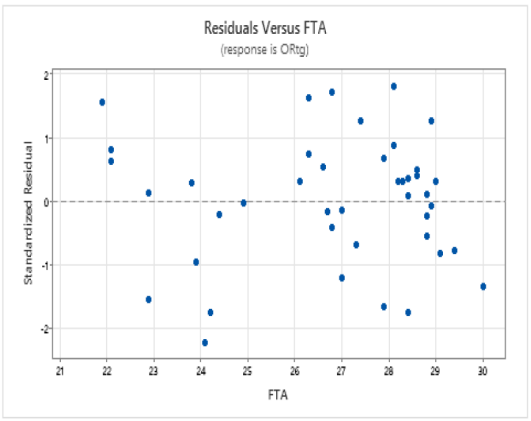
Test

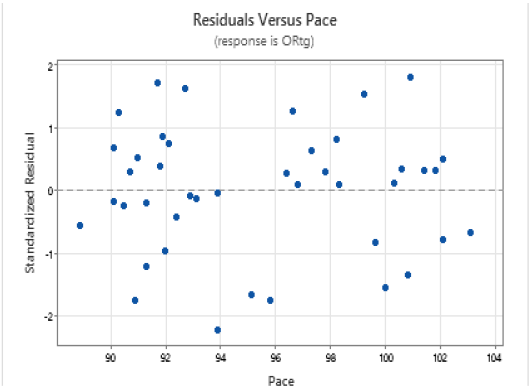
Null hypothesis	H_0: The order of the data is random
Alternative hypothesis	H_1: The order of the data is not random

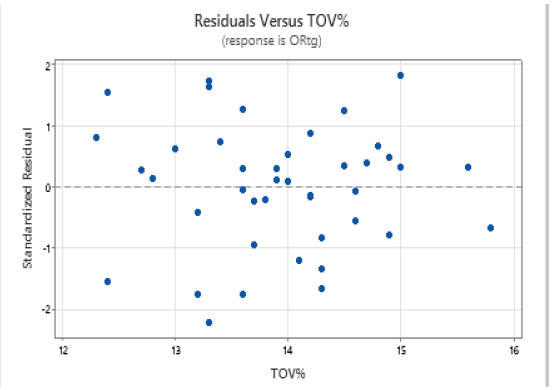
Number of Runs		
Observed	Expected	P-Value
17	21.81	0.129

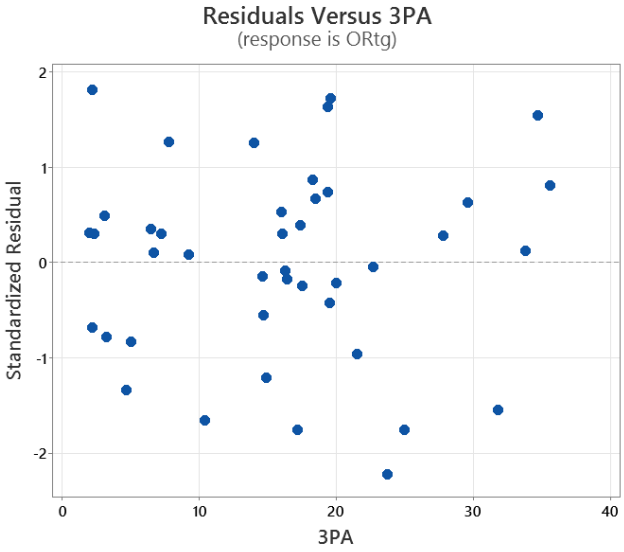
The residuals seem to be normally distributed and there seems to be mostly constant variance based on the versus fits plot and the accounting for subgroups within the data. The time series residual plot also seems to indicate lack of autocorrelation. The Durbin Watson statistic is again below the critical value. However, the ACF plot indicates that there is not a first order autoregressive pattern to the data, and that there are no statistically significant autocorrelations with any lag amount given the regression. The runs test also seems to indicate that there is not enough evidence to suggest that there is autocorrelation. Here are the residual plots versus each predictor:











These plots all generally seem to not include any particular pattern with the residuals. Moving on to the standardized residuals, leverage values, and cook's distance, here is a table of this data.

SRES_1	Count	HI_1	Count	COOK_1	Count
-2.22033	1	0.075764	1	0.000019	1
-1.75541	1	0.086557	1	0.000181	1
-1.75303	1	0.091373	1	0.000191	1
-1.65776	1	0.099044	1	0.000443	1
-1.54654	1	0.100297	1	0.000601	1
-1.33871	1	0.112795	1	0.000801	1
-1.20704	1	0.120816	1	0.000809	1
-0.95844	1	0.12786	1	0.001218	1
-0.82983	1	0.127894	1	0.00141	1
-0.78246	1	0.129089	1	0.001612	1
-0.67996	1	0.130615	1	0.001647	1
-0.55461	1	0.132444	1	0.001747	1
-0.41933	1	0.136641	1	0.002213	1
-0.24353	1	0.136862	1	0.003387	1
-0.21063	1	0.137644	1	0.003508	1
-0.17846	1	0.139329	1	0.00398	1
-0.1435	1	0.144675	1	0.004697	1
-0.08228	1	0.147334	1	0.006268	1
-0.04289	1	0.153013	1	0.006364	1
0.08007	1	0.154986	1	0.007607	1
0.09884	1	0.168834	1	0.008315	1
0.12167	1	0.177839	1	0.011829	1
0.27681	1	0.186107	1	0.012389	1
0.29651	1	0.189271	1	0.018327	1

0.29889	1	0.192689	1	0.021227	1
0.30502	1	0.196804	1	0.024378	1
0.3097	1	0.20883	1	0.027474	1
0.34507	1	0.220285	1	0.028482	1
0.38711	1	0.227075	1	0.032901	1
0.4868	1	0.228926	1	0.042344	1
0.52685	1	0.229874	1	0.047682	1
0.62712	1	0.235643	1	0.048457	1
0.66599	1	0.250729	1	0.048551	1
0.73883	1	0.250731	1	0.048591	1
0.81044	1	0.262744	1	0.056156	1
0.86178	1	0.297171	1	0.058696	1
1.24772	1	0.302011	1	0.061969	1
1.26425	1	0.305404	1	0.064609	1
1.53926	1	0.322798	1	0.066377	1
1.62836	1	0.33013	1	0.078914	1
1.71544	1	0.343463	1	0.091304	1
1.80716	1	0.38761	1	0.21743	1
N=	42	N=	42	N=	42

There are no standardized residuals over +- 2.5, no leverage values over 0.4762, and no values of Cook's distance over 1. This final regression seems to have best addressed assumptions and has the least chance of providing misleading information or insights.

I would say that a lot of what I speculated about was proven correct after looking at this model, but I also learned many different things I did not consider prior to this. For example, I did not realize how much the movement of the three point line affected NBA offenses, both before and after it was shortened. I was also surprised at how many of the variables had high significance t-tests. This seems to suggest that there are common measures that affect NBA offenses no matter the era of basketball. Many of these statistics were not directly related to scoring the basketball. In fact, only "Assists per 100 possessions"

was directly related to scoring. It seems that, in any era, taking care of the basketball, getting trips to the free throw line, playing team basketball, and playing at a faster pace can lead to greater offensive efficiency. I was especially surprised at how pace was a statistically significant factor in the overall regression and ended up in the final model. It is not obvious at all higher paced offenses are necessarily more efficient per possession, but a potential reason for this might be that higher paced offenses tend to create more easy transition opportunities, which are generally very efficient. Either way, transition offense do not comprise a large majority of a team's offense, at least in today's NBA. Looking at 3 pointers attempted, I thought that this would have a larger impact on offensive efficiency, considering the most efficient years league-wide have come in the past decade, where three point shooting volume has rapidly increased. Instead, 3 pointers attempted was the least significant predictor in terms of the t-test. This analysis shed light on things I underlooked, but also confirmed many things I believed to be true. If I were to do another analysis of historical NBA data, I would perhaps look at the prevalence of the mid-range shot and how that has affected league-wide offensive efficiency.