Statistical Analysis of the Effectivity of A Quarterback to Increase Winning Percentage Over Time

By: Kaitlyn Jones

Advisor: Dr. Ying-Ju Chen

Objective of the Study

The preliminary study studied all 32 NFL teams with 10 variables across 4 years (2014 – 2017). In order to refine the model I have used all the same teams and variables, but across 11 years (2007 – 2017).

Over the course of this study, I have wanted to answer 3 questions:

- ❖ Does a consistent quarterback, time with the team not skill wise, increase the winning percentage of the team in the regular season? And if so how significant?
- ❖ What skills of a quarterback are significant in effecting a team's percentage?
- ❖ Is there a maximum age that a quarterback ceases to be effective?

Preliminary Study

Due to a limiting student version of EViews 9 SV, I was only able to run 4 years of data through their algorithm which uses a linear form.

Dependent Variable: YRWPCT
Method: Panel Least Squares
Date: 05/04/18 Time: 23:55
Sample: 2014 2017
Periods included: 4
Cross-sections included: 31
Total panel (balanced) observations: 124

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.084246	0.226930	-0.371243	0.7111
YRFDWNPASSPCT	-0.296135	0.265570	-1.115090	0.2671
YRQBRATING	0.000256	0.002186	0.117171	0.9069
YRQBYTTY	0.074290	0.102038	0.728058	0.4681
YRTURNO	0.012633	0.001669	7.570156	0.0000
YRYRDSPPATT	0.065509	0.030315	2.160941	0.0328
YRYRSWSAMEQB	0.006361	0.005422	1.173222	0.2431
YRQBEXP	-0.010886	0.008222	-1.324026	0.1881
YRQBAGE	0.008901	0.005414	1.644084	0.1029

These were the results of the previous model which had an R squared of 0.596421. Where only yards per pass attempt and turnover ratio are significant in raising winning percentage.

* The variables in the regression are as followed from up to down are: the intercept, first down pass percentage, quarterback rating (as given by ESPN), quarterback yards per team yards, turnover ratio, yards per pass attempt, number of years with the same quarterback, years experience as a quarterback, quarterback age. All statistics except for quarterback age, quarterback rating, years with the same quarterback, and quarterback experience were taken from the official NFL statistics.

Data Preparation

Longitudinal data is a data set with the same variable in several groups across several intervals of time.

Bayes Information Criteria (BIC): is partly based on the likelihood function that computes the likelihood of a parameter given the outcome of a random variable. It is used to compare models.

$$BIC = \ln(n)k - 2\ln(\hat{L})$$

n = number of data points in x, or the sample size
k = the number of parameters in the model
$\hat{L}$ = the maximized value of the likelihood function

A lower BIC indicates a more appropriate model. Models can be improved upon by adding more terms, but they can overfit the data if too many variables are added to them, so BIC assigns a penalized term to adjust the likelihood function. When comparing models the lower the BIC indicates the least likelihood of the model being overfit compared to the other models.

Results

Results of comparing various models:

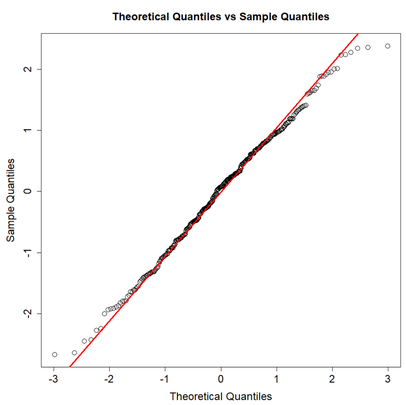
	Model A	Model B	Model C	Model D	Model E	Model F	Model G	Model H	Model I	Model J
AIC	-180.1993	-383.7514	-424.073	-437.9309	-438.3449	-435.9748	-441.201	-440.1953	-438.2453	-451.1349
BIC	-180.1993	-356.7059	-393.164	-403.1582	-407.4359	-397.3385	-398.701	-393.8317	-388.0181	-397.0440

*AIC is another information criteria similar to BIC. The only reason they would produce conflicting results would be AIC favors more variables in a model.

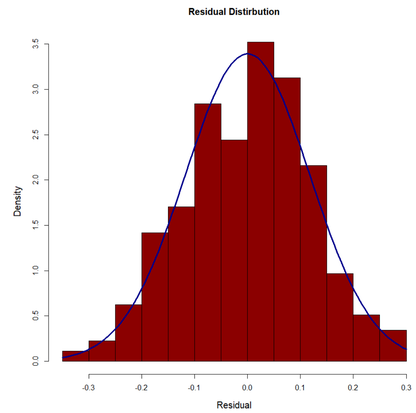
Results of the Most Appropriate Model:

	Value	Standard Error	Degrees of Freedom	t-value	p-value	Correlation to wpct
Intercept	-0.12677393	0.07529883	317	-1.683611	0.0932	-
Same Quarterback	0.04200580	0.01803871	317	2.328648	0.0205	0.205995
Turnover Ratio	0.01080909	0.00076241	317	14.177530	0.0000	0.6508583
Yards per Pass Attempt	0.08330352	0.01053868	317	7.904548	0.0000	0.5571802

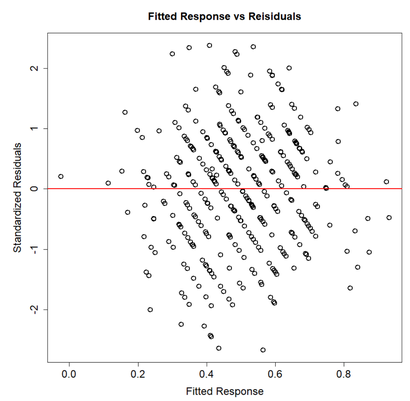
*A new variable, same quarterback, was created to show the decision of a team to keep a quarterback for a consecutive year regardless of how many years he has been present thus far.



*The sample quantiles are fairly close to the theoretical quantiles suggesting an appropriate model.



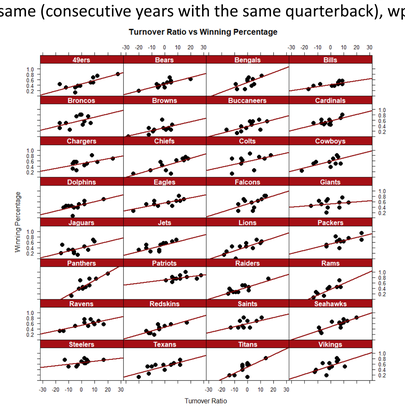
*The residuals appear to have a normal distribution



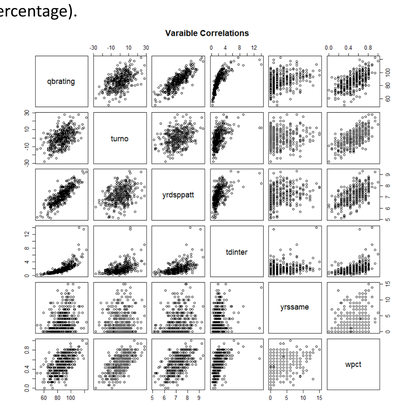
*The fitted response shows no distinct pattern with the standardized residuals, and are not farther than 3 standard deviations away from the mean.

Data Exploration

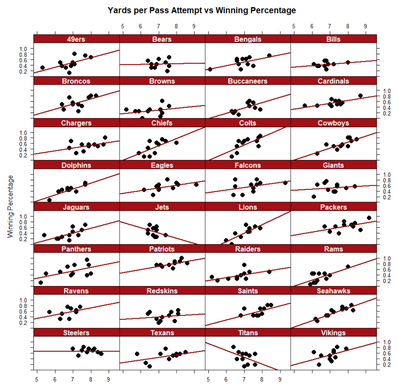
We began by plotting various variables with respect to winning percentage to see if there were any correlations. 10 variables were considered and denoted as follows: qbage (quarterback's age), qbexp (quarterback experience), qbrating (quarterback rating given by ESPN), qbypty (quarterback yard per team yard), tdinter (touchdown interception ratio), turno (turnover ratio), yrdsppatt (yards per pass attempt), yrssame (consecutive years with the same quarterback), wpct (winning percentage).



*This correlation may suggest a fixed effects model may be more due to the slight changes in slopes across teams.



*The far right column shows the relationship between the variables Of the diagonal variables with winning percentage. This is only a sample; this analysis was done to all 10 variables



*This correlation shows that a fixed effects model is more appropriate due to the wide variation in slopes, especially the negative slopes

Discussion & Conclusions

Contributions:

- ❖ There appears to be a significant increase in winning percentage if a quarterback is kept for a consecutive year.
- ❖ This model shows that turnover ratio and yards per pass attempt are the variables that significantly increase winning percentage.
- ❖ Given the appropriateness of the model, the age of the quarterback is not a significant variable to effect the winning percentage.

Suggestions for Future Research:

- ❖ I think looking at individual quarterbacks would help this model instead of looking at the quarterback who played the majority of the season.
- ❖ Being able to construct a model that shows the impact of each additional year of a consecutive quarterback could bring more understanding to this variable.

Reference

❖ Bates, Douglas M. "lme4: Mixed-effects modeling with R." (2010): 470-474.

Acknowledgement

Department of Mathematics, University of Dayton