ISSN 1750-9823 (print)
International Journal of Sports Science and Engineering
Vol. 06 (2012) No. 03, pp. 147-152

Examining the Outcome Effects of the Turnover Margin in Professional Football

Rhonda C. Magel+ and Geoffrey Childress

Department of Statistics North Dakota State University

(*Received October 24, 2011, accepted December 9, 2011*)

Abstract. This paper investigates a belief within professional football that whichever team commits fewer turnovers ultimately will win the game. Data was collected for seven seasons of NFL regular season games and was analyzed using logistic regression techniques. A variable was included to control for whether or not the team was playing at home or as a visitor. The team's final turnover margin for a game was found to be a significant predictor. The model was then used to predict games for the 2008 NFL regular season and its accuracy was noted.

Key Words: logistic regression; estimating winning probabilities

1. Introduction

In professional football, there is a widely held belief that, for a single game, whichever team wins the turnover battle ultimately wins the game. This belief has proved to be a pivotal component in the way teams prepare for a game, in addition to the way sports analysts interpret a team's ability to perform well, or even win a game.

Coaches place a special importance on attempting to be the dominant team in the turnover statistic. The Defensive Coordinator for the 2006 National Football League Super Bowl Champion Pittsburgh Steelers, Dick LeBeau, noted, "When you look at this game in hindsight, the team that wins the turnover battle will probably win the game" [1]

Sports writers are also strong proponents of the turnover margin statistic. Sporting News writer Vinnie Iyer claimed, "In the modern NFL game, where several teams are evenly matched, the disparity between takeaways and giveaways is crucial" [2].

Even the National Football League, which is the premier professional football league in the United States, has noted the special importance of turnovers relating to victory in the Super Bowl. One article stated, "… expect turnovers to give one team a distinct advantage in Super Bowl XLIII" [3].

With coaches, writers, and even the NFL all proclaiming the importance of the turnover statistic, the need becomes quite clear to provide evidence for this belief. A proper understanding of the terminology regarding turnovers is necessary to truly understand how the statistics are collected. The important definitions to note are giveaways, takeaways, and the turnover margin.

For a given team, a giveaway is defined to be either an interception thrown or a fumble lost by a member of the team's offense. A takeaway is defined to be either an interception or a fumble recovered by a member of the team's defense.

A given team's turnover margin (denoted TOM) in a single game is defined to be the number of takeaways caused by the team's defense minus the number of giveaways caused by the team's offense. A team's turnover margin can be either positive, zero, or negative, but will always take on an integer value.

These three statistics are all mathematically related since they are counted only for a single game. The result is that one team's number of takeaways is equal to the opponent's number of giveaways. In addition, one team's turnover margin will always be the negation of the other team's turnover margin.

Some past researchers have developed models to predict whether or not a team wins a particular football game. Harville [4] considered NFL teams and used a linear model consisting of two independent variables

+ Corresponding author. Tel: 701-231-7177, E-mail address: *Rhonda.magel@ndsu.edu*

to predict the home team's final score minus the visiting team's final score for an individual game. Harville's model used the results of previous games to predict future games. The model proposed by Harville did not use any individual sports statistics such as turnovers, but Harville did note that the home team had a clear advantange. Harville reported that his model had a 70.3% accuracy when used in situations for which it was created.

Willoughby [5] proposed models estimating the probability of winning a game for three specific teams in the Canadian football league. Each team had their own model which used individual sports statistics as predictors through the use of a logistic regression approach. The individual statistics used in the models were marginal statistics, which calculated the difference between the prediction team's statistics and those of the opposing team. The statistics consisted of rushing yards, passing yards, interceptions, fumbles, and quarterback sacks.

The results reported in Willoughby indicated that the two statistics that make up a turnover margin (interceptions and fumbles) were not consistently significant parameters for predicting victory for all three teams. In the case of one team, differences in number of interceptions yielded significant results at the 1% level, while differences in number of fumbles were not significant. For another team, both turnover statistics were significant at the 1% level. The third team yielded significant results for differences in the number of interceptions at the 1% level, and differences in the number of fumbles at the 5% level. Willoughby reported an accuracy of the three models as 85.9%, 90.2%, and 78.8%. This was based on using the model to predict the outcomes of the games used in the creation of the model. However, since there was an individual model for each team, the results are only good for the three teams upon which the analysis was based.

This current research study will illustrate the statistical importance of the turnover margin statistic in predicting a team's probability of winning. In addition, it will show how a team's probability of winning a game changes with additional turnovers. A model will be developed for the use of any NFL team in trying to estimate the probability of winning a game based on turnover margin and whether or not they are playing at home.

Section 2 details the collection of data and methodology used in analyzing the effect of turnovers. Section 3 reports the results of the methods used in Section 2, while Section 4 discusses an application of the results of the analysis to the 2008 season of the NFL. Section 5 reports the conclusions of this paper.

2. Methodology

The overall goal of this study was to determine if the turnover margin for an individual game was a reliable predictor of eventual victory. To test the hypothesis, a model was fit to estimate the probability of winning a game based on the turnover margin. The accuracy of the model was then calculated.

Data were collected from the National Football League website for all regular season games in the 2001 to 2007 seasons [6]. Pre-season and post-season games were not included, due to the different circumstances under which those games are played. Only games where a victory was able to be recorded were used; tied games were ignored. In total, 1,783 games were observed and recorded.

One problem that was faced in collecting the data was in determining a coding scheme to collect the independent and dependent variables. The goal of the study is to determine a team's likelihood of winning based off that team's turnover margin within the game. Therefore, the prediction model will be predicting a single team's probability of victory. However, the data source is for a single game, consisting of two teams. For each game, the data must be coded relative to a single team.

To select a single team for the coding scheme, the option chosen was to randomly select one of the two teams for each game. The method used was to flip a coin, which was assumed to have been fair, with equal probabilities for heads and tails. Heads represented that the data would be coded relative to the home team, while tails meant the data would be recorded relative to the visiting team. This was done once for every game utilized in the analysis.

A binary variable was recorded for whether the home team or the visiting team was used for the collection of the independent and dependent variables. This variable was coded as a '1' if the home team was used, and a '0' if the visiting team was used.

Next, the turnover margin was collected as an independent variable. This turnover margin, as defined in Section 1, was recorded relative to the team selected randomly. For instance, assume the results were selected to be coded for the visiting team, and the visiting team had five fewer takeaways than the home team. The turnover margin would then be recorded as '-5'.

Additionally, two independent variables were recorded to track any potential difference in results over time. The season in which the game took place was recorded as the numerical year of the first week of the season. The week the game took place was a number between 1 and 17, since each NFL season consists of 17 weeks of play. These values were the same, regardless of which team was selected in the coding scheme.

A single dependent variable measured victory or loss for the team randomly selected. A victory was coded as a '1' while a defeat was coded as a '0'.

Since the dependent variable is a binary response, being either a victory or a defeat for the team chosen, the ideal prediction will be a proportion, estimating the probability that a team wins, conditioned on the independent variables. This dichotomous nature of the dependent variable suggests that logistic regression is the most appropriate regression model [7].

The collected data was fit to a logistic regression model using the SAS v9.2 logistic procedure [8]. This method assumes that the observations are independent from one another. The claim of independence between individual NFL games is one that must be defended in order to justify the techniques used in this paper. Between seasons, the rosters of each NFL team change significantly with players being drafted, retiring, or signing with new teams. Due to this yearly change, it is assumed that games are independent from season to season. Within a single regular season, no two teams will play each other on the same field twice. This fact, along with the changes in injury status and random weekly weather effects, allows the assumption to be made that each game within a season is independent from all others. Any potential effects of season, week, and home-field are accounted for by the inclusion of such factors as independent variables in the model, as well as any potential interaction between the season and week variables with the TOM and home variables.

The logistic model was fit estimating the probability of winning the game using four sets of independent variables: season, week, home/visitor, and TOM (turnover margin). The season variable had 7 levels, resulting in the need for 6 indicator variables in order to perform the analysis. The week variable had 17 levels, resulting in 16 indicator variables. The home/visitor variable was already a binary response and therefore only one indicator variable was needed. The variable TOM was a quantitative variable. Interaction terms for season*TOM, week*TOM, season*home/visitor, and week*home/visitor were initially added to the model. Once the model was fit for all sets of independent variables and interaction terms, Wald Chi-Square Tests were performed [7].

3. Results

The model was fit for all four sets of independent variables and four sets of interaction terms, and the Wald Chi-Square test results were observed. The logistic model was tested for having a good fit through the Hosmer-Lemeshow Goodness-of-Fit Test [6]. The null hypothesis of this test is that the model fit is adequate. This test reported a test statistic of 6.1534 with 8 degrees of freedom. The p-value for this test was 0.6301, which indicates an adequate fit. This implies that the logistic model is appropriate to use in this situation.

The interaction terms in the model were tested for significance (at alpha equal to 0.05) and found not to be significant. Therefore, they were taken out of the model..

Once the interaction variables were removed, both the season and week sets of variables were tested and found to be nonsignificant. Only the TOM and home/visitor independent variables remained in the model.

The stepwise logistic regression procedure was also used in SAS to help derive a model. The model obtained using this procedure was the same as the previous model. Namely, only the TOM, and home/visitor variables were entered and kept in the model.

Table 1 gives the p-values of the test results from the model without the season and week variables, and none of the interaction terms. This reduced model found both the TOM and home/visitor independent variables to be significant, as well as the intercept parameter. The actual parameter estimates were found to be -0.3575 for the intercept, 0.7796 for the TOM variable, and 0.6958 for the home/visitor variable. The final estimated model was as follows:

where $\pi(TOM,|ome)$ is the estimated probability of winning based on the turnover margin TOM and whether the team is playing at home (home=1) or away (home=0).

The standard error for the intercept parameter was found to be .0835, the standard error for the TOM parameter reported as .0394, and the standard error for the home/visitor parameter was .1173.

Table 1. Reduced Model Wald Chi-Square Test Results.

Variable	DF	$\hat{\beta}_i$	SE($\hat{\beta}_i$)	Chi-Square	p-value
intercept	1	-0.3575	0.0835	18.3383	<.0001
TOM	1	0.7796	0.0394	391.4709	<.0001
home/visitor	1	0.6958	0.1173	35.2038	<.0001

DF = degrees of freedom

$\hat{\beta}_i$ = parameter estimate

SE($\hat{\beta}_i$) = standard error of parameter estimate

Table 2 gives estimated probabilities of winning, $\pi(x)$, from the reduced model for a given set of independent variables. For instance, a team with a -2 turnover margin for a single game played at home, the probability of winning is estimated to be 0.22778.

The estimated parameters for the model also provide an opportunity to analyze the odds ratios for comparing two fixed sets of independent variables. The parameter estimate for TOM was reported to be .7796. This would result in an odds ratio of 2.181 for an additional turnover. A Wald 95% confidence limit for this odds ratio was found to be (2.019, 2.356). An interpretation of this odds ratio is that for each additional takeaway a team causes, the odds of winning for that team increase by a factor of 2.181, assuming the number of giveaways remains constant. The parameter estimate for the home/visitor variable was found to be .6958, resulting in an odds ratio of 2.005. The Wald 95% confidence limits of this estimate are (1.594, 2.524). An interpretation of this odds ratio is that, for a fixed turnover margin, a team's odds of winning will be increased by a factor of 2.005 if they are playing at home as opposed to playing as a visitor.

Table 2. Estimated Probabilities of Winning Based on Reduced Model.

TOM	Home $\pi(x)$	Visitor $\pi(x)$
-6	0.01288	0.00646
-5	0.02766	0.01399
-4	0.05841	0.03001
-3	0.11915	0.06319
-2	0.22778	0.12823
-1	0.39143	0.24285
0	0.58378	0.41156
1	0.75360	0.60399
2	0.86961	0.76883
3	0.93566	0.87882
4	0.96943	0.94053
5	0.98575	0.97182
6	0.99341	0.98688

4. Model Application

In an attempt to measure the model's ability to perform well in an actual setting, data from the 2008 NFL regular season was collected [9]. For every game that took place during the 2008 season, the turnover margin of the winning team was noted. Additionally, it was observed whether the winning team was the home team or the visiting team. If the winning team was at home, a '1' was recorded, while a '0' was recorded for a visiting team.

For each observation the predicted probability of victory was estimated using the model obtained in Section 3. If the model's predicted probability of victory for the given home/visiting and TOM variables was greater than .50 it was determined that the model predicted a 'victory'. If the predicted probability was less than .50 the model predicted a 'loss'. A frequency table of the prediction results is given in Table 3.

Table 3. Model Predictions for the 2008 NFL Regular Season

	Winning Team		
Model Prediction	Visiting	Home	**Total**
Loss	41	30	71
Victory	68	116	184
Model Accuracy	**62.39%**	**79.45%**	**72.16%**

Table 3, illustrates that the model boasted an overall 72.16% accuracy when applied to the 2008 NFL regular season. This was based only on the turnover margin of a team and whether or not it was playing at home. Among games where the home team was the victor, the model accurately predicted 79.45% of the contests correctly. The prediction accuracy rate dropped to 62.39% among games where the visiting team was the victor. This difference seems to indicate that when a team plays a game on the road, a positive turnover margin is less reliable as a predictor of victory than if a team has a positive turnover margin while at home.

5. Conclusions

When the model was fitted using all four sets of independent variables and four sets of interaction terms, it was found that the week and season sets of variables, along with interactions involving them, were not significant. This seems to indicate that the predictions based on turnover margin are homogeneous with respect to these two sets of variables. That is to say, there is no statistical difference between results among different weeks or years. With no difference in results between different seasons or weeks, using it on future games is more strongly justified. This is an important result in terms of the model's usability.

The results of the Hosmer-Lemeshow Goodness-of-Fit test indicate that the logistic model is a good model to use. In addition, when applied to the 2008 NFL regular season games, the model boasted 72.16% accuracy. The fact that one can reasonably predict whether or not a team will win based only on the team's location and turnover margin is a very strong result. The results of this paper tend to support the claim that winning the "turnover battle" in a game is a reliable predictor of eventual victory.

In the reduced model, even after controlling for the effect of turnovers, the home/visitor variable was still found to be significant. An interpretation of this result is that there is still a measurable "home-field advantage" in terms of winning a game, beyond an effect that the home playing field will have on a team's turnover margin. This effect could potentially be quantified more explicitly in future research.

An additional option for further research would be to analyze the effect of time of possession on victory. Within the football industry, the idea of "ball control" is suggested to be correlated to victory, much in the same way that the turnover margin is. A similarly-designed study using a time-of-possession statistic as an independent variable could be analyzed and compared to the results provided in this paper.

6. References

[1] Zilgitt, J. *Turnovers a big part of winning and losing*. Retrieved Febraury 10, 2009, from USATODAY: http://www.usatoday.com/sports/football/nfl/2006-02-04-turnovers_x.htm. 2006, February 4.

[2] Iyer, V. *Positive turnover margin a solid measure for postseason success*. Retrieved February 10, 2009, from SportingNews: http://www.sportingnews.com/yourturn/viewtopic.php?t=450396. 2008, August 27.

[3] NFL.com. *Turnover differential often determines Super Bowl outcome*. Retrieved February 10, 2009, from NFL.com: http://www.nfl.com/superbowl/story?id=09000d5d80e5d3d9&template=with-video&confirm=true. 2009, January 26.

[4] Harville, D. Predictions for National Football League Games Via Linear-Model Methodology. *Journal of the American Statistical Association.* 1980, **75** (371): 516-524.

[5] Willoughby, K. A. Winning Games in Canadian Football: A Logistic Regression Analysis. *The College*

Mathematics Journal . 2002, **33** (3): 215-220.
[6] *NFL Scores*. Retrieved July 17, 2008, from NFL.com: http://www.nfl.com/scores. 2008.
[7] Hosmer, D. W., & Lemeshow, S. *Applied Logistic Regression* (2nd ed.). New York: John Wiley & Sons, Inc. 2000.
[8] SAS Institute, Inc. LOGISTIC Procedure. *SAS 9.2 Help and Documentation* Cary, NC, USA. 2007.
[9] *NFL Scores*.. Retrieved February 2, 2009, from NFL.com: http://www.nfl.com/scores , 2009.