Pick Me a Winner
An Examination of the Accuracy of the Point-Spread in Predicting the Winner of an NFL Game

Richard McGowan
Boston College

John Mahon
University of Maine
Every week in the fall, millions of dollars are bet on the outcome of National Football League games (NFL) using a gambling tool called the point-spread. For most football fans, the point-spread is taken at face value and they do not question its accuracy. But why is the point-spread accepted as the most accurate way to predict the winner of an NFL game? This paper tests the accuracy and significance of the point-spread and finds that although it is significant, and able to predict the outcome of a game 66.9% of the time, it is not the most accurate predictor. Through the creation of a new NFL statistic, the “offensive-defensive ranking differential”, the accuracy of predicting the outcome of an NFL game was increased by 1.6% to 68.5%. This new finding begs further discussion on profitable betting strategies within the NFL, as well as the ultimate efficiency of the NFL betting market.
Introduction – Key Terms

• **Point-Spread:** Merriam-Webster defines as the number of points by which a team is expected to win a particular game
  • Often expressed in terms of a handicap for the home team
• **11 for 10 rule:** “Requires that bettors must pick winners in 52.4% of bets to break even. This marginal dollar is commission, or vigorish, enabling the betting house to earn a profit” (Gray 1727)
  • This rule guarantees the betting house a profit of $1 for every $22 bet. “Therefore, the game-time spread should reflect the aggregate information in the market” (Gray 1727)
  • If the market is efficient, the spread will be an unbiased predictor of the result of the game
Introduction – Why is the Topic Interesting?

Popularity

- Adweek/Harris Poll – 2,374 U.S. Adults over 4 days in September 2011
  - 64% of U.S. adults
  - 73% of U.S. men
  - 55% of U.S. women

- ESPN – *TMQ* by Gregg Easterbrook (2011)
  - 9 of 10 highest-rated TV broadcasts in 2011 were NFL (U.S. only)
    - 8 bested the Academy Awards
  - Of the 20 most watched global TV broadcasts, all 20 were Super Bowls
Introduction – Why is the Topic Interesting?

Volume

- WSJ – Hannah Karp (2009)
  - $81.5 million wagered in Nevada for 2009 Super Bowl
    - Sum has approached $100 million in previous years
  - $1.1 billion wagered (legally) on NFL & NCAA – LTM 4/30/09
    - Analysts estimate tens of billions more offshore and illegally
Introduction – What Has Been Written?

- Surprisingly little has been written on the accuracy of the point-spread
  - Andy Fodor (2013) is the only person to write on the topic
  - Prior playoff teams only

- Most NFL betting literature covers market efficiency (weak form EMH)
  - Accuracy “against the spread” rather than predicting winner
  - Profitable betting strategies
  - In-sample versus out-of-sample

- Efficiency has been debated since the 1960’s
  - Implications of efficiency effect accuracy of point-spread
  - Weak Form efficiency = point-spread most accurate predictor of winner
Introduction – Planned Accomplishments?

- Hope to show that the point-spread is an accurate way to predict the winner of an NFL game

- Are there other metrics that add predictive power?

- Successful test is inherent confirmation of inefficient market
Almost all NFL betting literature focuses on market efficiency.

Researchers have split into efficient vs. inefficient:
- Efficient – profitable trends very rarely exist out-of-sample, and those that do correct back to equilibrium
- Inefficient – Most inefficiencies have dissipated but some are still profitable

3 sources for efficiency (2 original articles & 1 rebuttal)

2 Sources for inefficiency (Holdover Bias & Bettor Overreaction)
Market Efficiency and Football Betting (1968) – Lyn Pankoff

- “The placing of bets on football games through bookmakers seems to be an efficient economic market analogous to the stock market” (213)
- OLS Regression of actual point-spread on the predicted spread
  - Used quadratic loss function for all regular season games 1956 – 1965
  - Large mean error or variance error would suggest an exploitable pattern
- Found no exploitable market patterns
  - Tried hindsight regression correction for mean/variance errors
  - Added more independent variables
- Conclusion: Some bettors are better than others, but market is efficient
Literature Review – Efficient Market


- Shows efficiency through a rebuttal of Zuber et al. (Zuber) which said market was inefficient
  - Zuber did a weak test of efficiency and was “unable to reject the hypothesis that the scores of NFL games are unrelated to gambling line predictions” (805)
  - Tested using weekly samples instead of individual game basis which increased sample variance – makes it more likely that hypothesis can’t be rejected
- Sauer et al. adjusts Zuber’s OLS regression for the sample size and can now reject the hypothesis
- **Conclusion:** Regression was set up to fail, market is efficient, and even if the inefficiency existed, it is gone now
Literature Review – Efficient Market


- Argues that all profitable strategies dissipate over time – a staple of an efficient market
- Takes issue with the OLS regression that others have run to test efficiency because of sensitivity to extreme outliers make results difficult to interpret
- Reexamined the NFL betting market using a discrete-choice probit model
  - Developed the model using data from 1976 – 1992
  - Found a number of profitable strategies both in and out-of-sample
  - “No monotonic relationship between the probit filter and out-of-sample success rates” (1735)
- **Conclusion:** Possible inefficiencies dissipate over time, thus market is efficient
Literature Review – Possible Inefficiencies

*Inefficient Pricing from Holdover Bias in NFL Point Spread Markets* (2013) – Andy Fodor

- Found “unprecedented profitability based on taking a contrarian strategy on wagering against prior playoff teams in the following seasons opening week” (1417)
  - Line errors of the games indicate a statistically inefficient market
  - Bookmakers and bettors seem to over-estimate the potential performance of teams that make the playoffs the year before, so bet against them
  - Average return of 22.6% per game if implemented from 2004 – 2012
  - Dissipates after Week 1 – but it is a long term abnormal return

- **Conclusion:** the market it inefficient because there exists an opportunity to have long term abnormal returns
Literature Review – Possible Inefficiencies

Inefficient Pricing from Holdover Bias in NFL Point Spread Markets (2013) – Andy Fodor

• Also examined the accuracy of the spread in predicting the winner
  • Only for prior playoff teams

• Prior playoff teams are favored in Week 1 85% of the time
  • Only win 51.7% of the time
  • Trend does NOT disappear after Week 1
  • Prior playoff teams continue to be favored more than they actually win in 16 of the 17 weeks

• Suggests that the point-spread is not an accurate predictor
Literature Review – Possible Inefficiencies


- Findings are remarkably similar to those that Fodor finds 12 years later
- Tested 2,300 games from 1981 to 1995
  - “Bettors overweigh outstanding positive performance when measured over the previous game, over the previous 2 to 5 games, or over the previous season” (497)
- Tested 22 strategies for betting using overreaction to positive performance
  - 21 had WR over 50%
  - 17 had WR over 52.4%
  - 8 significant at 10% and 4 significant at 5%

**Conclusion:** Market is inefficient because this overreaction bias existed for 14 years and could still exist today
Research Question – How is This Topic New?

- Most research focuses on the efficiency of the market
  - Are there profitable strategies that can be used to exploit the book maker’s line

- This paper is not going to focus on profitable strategies
  - Does the favorite win or lose

- Is the point-spread the most accurate predictor, or is there another metric?
Research Question – Narrowed Scope

- Scope of thesis narrowed in 3 ways:

1. Data will be restricted to NFL regular season games
2. Data will exclude all tie games
3. Data will be from 2002 through 2013
The point-spread is the most accurate way to predict the winner of an NFL game
Preliminary Sources of Data

Historical Closing Spreads from 2002 through 2013
- Will be gathered from various online databases such as:
  - Sunshineforecast.com
  - Goldsheet.com

Historical NFL Performance Metrics from 2002 through 2013
- Will be gathered from various online databases such as:
  - Profootballreference.com

- 16 games for 32 teams through 11 seasons results in 5,632 samples
Methodology

- Find the accuracy of the point-spread in predicting winners
  - Logistic Regression with dummy variables (1 or 0)
  - Test significance with Null Hypothesis of “the accuracy of the point-spread is not statistically significant”

- Add other metrics to the regression as independent variables
  - Test for individual significance
  - Test for joint significance
Results

Table I: Logistic Regression Classification Table – Test 1

<table>
<thead>
<tr>
<th>Observed</th>
<th>Predicted</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winner .00</td>
<td>2037</td>
<td>12</td>
</tr>
<tr>
<td>Winner 1.00</td>
<td>1004</td>
<td>15</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. The cut value is .500

Table II: Independent Variable Significance – Test 1

<table>
<thead>
<tr>
<th>Variables in the Equation</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Step 1²</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spread</td>
<td>-.141</td>
<td>.013</td>
<td>116.488</td>
<td>1</td>
<td>.000</td>
<td>.868</td>
</tr>
<tr>
<td>Constant</td>
<td>.032</td>
<td>.075</td>
<td>.188</td>
<td>1</td>
<td>.665</td>
<td>1.033</td>
</tr>
</tbody>
</table>

a. Variable(s) entered on step 1: spread.
The point spread predicts the winner of a NFL game 66.9% of the time.

Recall that according Phillip and Gray, a bettor can break even by picking correctly 52.38% of the time!

Hence it appears that the point spread is quite accurate.

It appears that the NFL betting market is inefficient.

The next question: Is the most accurate way?
Second Metric
“Offensive-defensive Rank Differential”

• Another metric that was developed will be labeled “offensive-defensive rank differential”
• This statistic ranks the offensive points scored per game and the defensive points allowed per game for each team for week of the season.
• It can be used to see if the differential shows a bias towards one side of the ball in predicting the outcome of a NFL game. The team with the higher differential was treated as the projected winner.
• In the current NFL, it is held that offense is more important than defense.
Results of Second Metric

Table III: Logistic Regression Classification Table – Test 2

<table>
<thead>
<tr>
<th>Classification Table&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Observed</th>
<th>Predicted</th>
<th>Percentage Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pred</td>
<td>.00</td>
<td>1488</td>
<td>77.4</td>
</tr>
<tr>
<td>Pred</td>
<td>1.00</td>
<td>470</td>
<td>50.7</td>
</tr>
<tr>
<td>Overall Percentage</td>
<td></td>
<td></td>
<td>68.5</td>
</tr>
</tbody>
</table>

<sup>a</sup> The cut value is .500

Table IV: Independent Variable Significance – Test 2

<table>
<thead>
<tr>
<th>Variables in the Equation</th>
<th>B</th>
<th>S.E.</th>
<th>Wald</th>
<th>df</th>
<th>Sig.</th>
<th>Exp(B)</th>
</tr>
</thead>
<tbody>
<tr>
<td>spread</td>
<td>-.093</td>
<td>.014</td>
<td>42.922</td>
<td>1</td>
<td>.000</td>
<td>.911</td>
</tr>
<tr>
<td>after8</td>
<td>-.034</td>
<td>.087</td>
<td>.151</td>
<td>1</td>
<td>.698</td>
<td>.967</td>
</tr>
</tbody>
</table>

Domenico Nicolia - Thesis Proposal
Results of Second Logistic Regression

- Using the point differential independent variable improves the predictive power to 68.5%.

- Hence, the point spread alone is not the most accurate predictor of a winner of a NFL game. If you used in conjunction with the point spread differential the accuracy of predicting the outcome of a NFL game is improved.
Final Conclusions

- It appears that this fairly simple strategy (betting the point spread and point differential) would produce profits.
- But this strategy does imply that NFL betting market is inefficient. So why haven’t bettors utilized it?
- First and foremost, bettor do allow personal bias interfere with their bets. They tend to bet their favorite even when the point spread or point differential is known. Now one might think this is irrational. Since the Efficient Market Hypothesis does not hold.
- Finally the point Spread Differential does seem to uphold the common belief that offense is more important than defense in the NFL.
Works Cited

Works Cited - Continued


- The Harris Poll © Oct. 14, 2011, “America’s Sport – A Majority of Americans Watch NFL Football”. Harris Interactive Inc. All rights reserved.
Works Cited - Continued