Available online at www.sciencedirect.com
ScienceDirect
International Journal of Forecasting 23 (2007) 405–413
international journal of forecasting
www.elsevier.com/locate/ijforecast

The comparative accuracy of judgmental and model forecasts of American football games

ChiUng Song [a,1], Bryan L. Boulier [b,2], Herman O. Stekler [b,*]

[a] Science and Technology Policy Institute, 26F, Specialty Construction Center 395-70, Shindaebang-dong, Tongjak-ku, Seoul 156-714, South Korea
[b] Department of Economics, The George Washington University, Washington, DC 20052, United States

Abstract

This paper compares the forecasts of the outcomes of NFL games made by 31 statistical models with those of 70 experts who predicted the winners of 496 NFL games played in the 2000 and 2001 seasons. We also analyze the betting line predictions. There are nearly 18,000 expert and 12,000 statistical forecasts.

The difference in the accuracy of the experts and statistical systems in predicting game winners was not statistically significant. The variation in the success rates was higher among experts than statistical systems, but the betting line outperformed both. Moreover, having more information did not always improve the forecasting accuracy. Neither the experts nor the systems could profitably beat the betting line.

Keywords: Judgmental forecasting; Statistical forecasting; Sports forecasting; Forecast evaluation; Betting market

1. Introduction

Are predictions or forecasts derived from statistical models more accurate than those made by experts based on informed judgment? Statistical models may yield more accurate forecasts than human judgments because they employ objective criteria that guard against bias and the idiosyncratic interpretation of data. However, sometimes statistical models cannot capture non-quantitative factors. As a result, judgmental forecasting may do a better job, not only in less routine and more uncertain situations, but also in integrating qualitative factors into the forecasting process.

There have been many comparisons of the predictive abilities of judgmental and statistical forecasting methods (Bunn & Wright, 1991; Collopy, Lawrence, & Wright, 1996; Dawes, 1986; Grove & Meehl, 1996; Grove, Zald, Lebow, Snitz, & Nelson, 2000; Webby & O'Connor, 1996; Wright, Lawrence, & Collopy, 1996). These comparisons have been made in diverse fields, including medicine, college success, business decision-making, weather forecasting, legal predictions, and macroeconomics. For the most part, the evidence has favored statistical model forecasts

* Corresponding author. Tel.: +1 202 994 6150; fax: +1 202 994 6147.
E-mail addresses: cusong@stepi.re.kr (C. Song), mortile@gwu.edu (B.L. Boulier), hstekler@gwu.edu (H.O. Stekler).
[1] Tel.: +82 2 3284 1773.
[2] Tel.: +1 202 994 8088; fax: +1 202 994 6147.

0169-2070/$ - see front matter © 2007 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.ijforecast.2007.05.003

(cf Grove et al., 2000, p. 21). In addition, the "bootstrapping the expert" literature indicates that the statistical model of the expert often provides forecasts superior to those of the expert (Armstrong, 2001).

Under some circumstances, judgment may improve predictions when used as a supplement to statistical/econometric forecasting. As pointed out by Bunn and Wright (1991), judgment is important for selecting the model to be used in generating forecasts and for identifying the data and statistical methods to be used to estimate the model. In addition, forecasts from statistical models are often subject to judgmental adjustments in real-world forecasting contexts. Experts may also provide contextual information, allowing for the interpretation of data series and the identification of future events that may cause departures from past patterns (Armstrong & Collopy, 1998).

There have been a small number of studies recently that have examined the accuracy of the methods used to forecast the outcomes of sporting events. Andersson, Edman, and Ekman (2005) found that experts' forecasts of outcomes of the World Cup soccer games were not superior to those of non-experts. Dixon and Pope (2004) tested whether statistical forecasts were valuable in predicting the results of UK soccer games and whether the betting market was efficient. Neither study compared the relative accuracy of expert and statistical systems.

There are, however, a number of studies that have examined this issue. Forrest and Simmons (2000) examined the accuracy of three tipsters' predictions of the outcomes of soccer games relative to that of a model. They concluded that "while individual guidance is better than no guidance, the expertise that they can claim to offer is limited" (p. 330). Moreover, the tipsters had not incorporated information from the model into their forecasts.

The study by Forrest, Goddard, and Simmons (2005) involves a more extensive analysis of the performance of odds-makers (who can be considered experts) relative to statistical models in predicting the outcomes of UK soccer games. They show that "…a detailed benchmark statistical model proves to be far from dominant over the views of experts…" (p. 563). The experts' performance improved over time, and by the end it was better than that of the model. These results are based on nearly 10,000 observations, and must be considered significant.

Boulier and Stekler (2003) examined the predictive performances of judgmental and statistical models in forecasting American professional football games from the 1994–2000 seasons. They found that the success rate (62%) of a statistical model based on power scores published in *The New York Times* exceeded that of the newspapers' sports editors (60%), but that the forecasting success of the betting line (65%) was higher than either. Moreover, given the power scores, the sports editors' forecasts did not add any information that was useful in predicting game outcomes.

This paper analyzes and compares the forecasts of the outcomes of the National Football League (NFL) regular season games for the 2000 and 2001 seasons made by 70 experts and 32 statistical models. Our sample consists of nearly 18,000 forecasts made by experts and over 12,000 forecasts made by the statistical systems. These data enable us to determine whether there is a statistically significant difference in the predictive accuracy of the models and experts.

We describe the data in Section 2. We then compare the proportions of game winners correctly predicted for each of the two seasons, and for the two seasons combined, in Section 3. Because forecasters have less information about the relative strengths of teams early in the season than they do later in the season, we also examine whether the accuracy improves as a season progresses. We compare the accuracy of predictions against the line by experts and statistical systems in Section 4 and summarize our results in Section 5.

2. Data

Our data consist of the forecasts of the outcomes of the 496 regular season NFL games for the 2000 and 2001 seasons. These forecasts include those made by experts using judgmental techniques, forecasts generated from statistical models, and a market forecast — the betting line. All data, including the sources, can be obtained from the authors.

The forecasts of the 70 experts were collected from 14 daily newspapers, 3 weekly magazines, and the websites of two national television networks. The newspapers include *USA Today*, a national newspaper, and 13 local newspapers selected from cities that have professional football teams. The three weekly national magazines are *Pro Football Weekly*, *Sports Illustrated*, and *The Sporting News*. Two television networks, CBS and ESPN, have web-sites that present the forecasts of their staffs.

Table 1
The proportion of game winners that were correctly predicted by experts, statistical systems and the betting line: 2000 and 2001

	Experts			Statistical systems			Betting line		
	2000	2001	Both Seasons	2000	2001	Both Seasons	2000	2001	Both Seasons
	A. forecasters as the units of observation								
Mean	0.619	0.617	0.617	0.622	0.622	0.623			
Standard Deviation	0.035	0.039	0.033	0.020	0.035	0.028			
Minimum	0.524	0.524	0.524	0.583	0.535	0.535			
Maximum	0.678	0.677	.667	0.657	0.673	0.673			
Number of Forecasters	42	36	48	23	29	31			
	B. games as the units of observation								
Number of Games Predicted	9625	8351	17,976	5364	6710	12,074	244	245	489
Accuracy Ratio	0.618	0.615	0.617	0.622	0.622	0.622	0.648	0.665	0.656

Some experts predict game winners directly (also called "straight picks"), while others make predictions against the Las Vegas betting line. (The terminology that is used in this paper is explained in the Appendix.) Some experts who pick game winners also predict the margin of victory (i.e., a point spread). For those who predict a margin of victory, one can identify their implicit picks against the line by comparing their predicted margin of victory with the betting spread given by the Las Vegas betting line. Because of differences in the types of predictions that are made, we are able to use the forecasts of 48 experts in predicting game winners, and an overlapping (but not identical) set of 52 forecasters in picks against the betting line.

The initial betting line for a game is established on the Monday before the games are played on the following weekend. This opening line is based on the consensus of a small number of expert linemakers. The line may then be adjusted during the week in response to new information or the flow of bets during the week.[3] We used the Las Vegas betting line data published in *The Washington Post* on the day that the game was played.

The data from the statistical models were obtained from the Internet. The models provide point spread predictions that allow us to identify both the predicted winner of a game and the predicted winner against the betting line. The variables that are used to construct the models differ widely. Among the variables used to generate the point spread predictions are the win/loss records of the teams, offensive statistics (such as points scored or yards gained), defensive statistics (such as points or yards allowed per game), variables reflecting the strength of schedule, and the home field advantage. In many of these models, the point spread predictions are based on the power rankings of the teams. The power rankings are based on some of the aforementioned variables.

3. How accurate are the forecasts of game winners from experts, statistical systems and the betting market?

For each forecaster, we calculated accuracy ratios (i.e., the proportion of games in which the winner of the game was correctly predicted). In Section 3.1, we present a summary comparison of the accuracy of all the forecasts of game winners made by experts, systems and the betting line. In Section 3.2, we examine the distribution of the success rates of individual forecasters. In Section 3.3, we explore whether the forecasts improve as a season progresses, to see how the accumulation of information on team characteristics influences the accuracy of the predictions.

3.1. The overall accuracy of expert, system, and betting line forecasts of game winners

Table 1 provides descriptive statistics for the full sample of forecasters and the betting line, for each season and for the two seasons combined. The upper panel of the table presents data using the forecaster as the unit of observation. The number of predictions varies among forecasters, ranging from 218 to 496 for experts, and from 176 to 496 for statistical systems.

[3] Descriptions of the betting market and analyses of factors causing changes in the line during the week can be found in Gandar, Zuber, O'Brien, and Russo (1988) and Avery and Chevalier (1999).

The bottom panel shows data on the total number of forecasts by method. Table 2 summarizes the success rates for the forecasters who issued forecasts in both years.

Table 1 indicates that statistical systems and experts have similar success rates in predicting game winners for each of the two seasons and for the two seasons combined. We find that the accuracy ratios are not statistically significantly different from each other, even at the 0.35 level. The proportion of game winners correctly predicted by the betting line exceeds that of experts and statistical systems for each season. The betting line accuracy ratios for the 2000 and 2001 seasons separately are not statistically significantly different from the expert and statistical systems predictions at the 0.10 level (two-tail test). However, when the two seasons are combined, they are statistically significantly different from the experts' accuracy ratio at the 0.04 level and from the statistical systems' accuracy ratio at the 0.07 level.

We conclude that there is no difference in the overall success rates between experts and statistical systems in predicting game winners, and that the betting line is superior to both.

3.2. Distributions of forecasting success

Table 1 and Fig. 1 indicate that the success rate varies across forecasters, though no forecaster had a success rate lower than 50%. For all but one system and two experts, these success ratios were significantly different from those that could have occurred by chance.[4] At the other end of the spectrum among those issuing forecasts for both seasons, the success rate of the betting line was exceeded by only two experts and one statistical system.

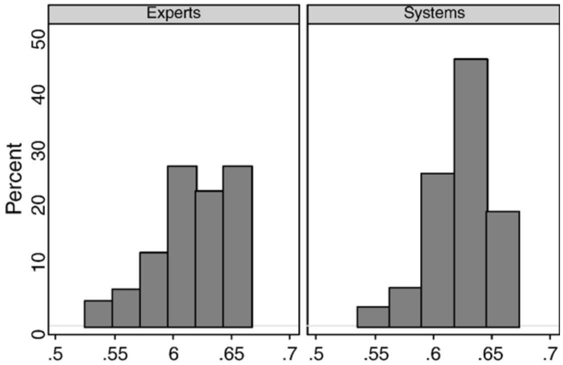

Fig. 1. Distributions of accuracy ratios, combined seasons.

While the mean accuracy ratios among experts and systems are similar, there is one important difference between their forecasts: *the dispersion of forecasting success was higher among experts than it was among systems*. Tables 1 and 2 indicate that the range and standard deviation of the accuracy ratios are higher for experts than for statistical systems for each year individually and for the combined seasons.

Moreover, within any one season, both the best and the worst records were achieved by experts. While the gap between the best expert and the best system is not very large in any one season, the worst records among statistical systems are typically much better than the worst records among experts. Among those issuing forecasts for the two seasons combined, four experts had lower accuracy ratios than the *worst* of the systems.

We also examined the forecasting performance over time. Although the detailed results are not reported here, we found that the accuracy ratios of experts for the two seasons were highly correlated (the Spearman rank correlation coefficient was 0.58). In contrast, the accuracy ratios of the statistical systems were not correlated across seasons (the Spearman rank correlation was only 0.22). Second, an *unusual* success (or *unusual* failure) in one season of forecasting was not

Table 2
The proportion of game winners that were correctly predicted by experts and statistical systems that issued predictions in both seasons: 2000 and 2001[a]

	Experts			Statistical systems		
	2000	2001	Both seasons	2000	2001	Both seasons
Mean	0.617	0.623	0.619	0.621	0.621	0.621
Standard Deviation	0.039	0.038	0.035	0.021	0.032	0.020
Minimum	0.524	0.532	0.528	0.583	0.544	0.572
Maximum	0.678	0.677	0.667	0.657	0.672	0.658
Number of Forecasters	30	30	30	21	21	21

[a] The forecaster is the unit of observation in calculating the means and standard deviations.

[4] A test based on the binomial distribution and a 5% significance level was used.

Table 3
Accuracy ratios (AR) of experts, statistical systems, and the betting line, by halves of the 2000 and 2001 seasons

Year	Half	Experts		Statistical systems		Betting Line	
		Games	AR	Games	AR	Games	AR
2000	1st	5066	0.645	2618	0.643	126	0.675
	2nd	4559	0.588 [a]	2746	0.603 [a]	118	0.619
2001	1st	4312	0.604	3384	0.582	127	0.661
	2nd	4039	0.628 [b]	3326	0.663 [a]	118	0.670
Combined Seasons	1st	9378	0.626	6002	0.608	253	0.668
	2nd	8958	0.606 [a]	6072	0.636 [a]	236	0.644

[a] Indicates that the accuracy ratios of the two halves of a season are statistically significantly different from each other at the 0.01 level.
[b] Indicates significance at the 0.05 level.

necessarily replicated in a second season, either for experts or for systems, suggesting that there is regression to the mean. This result would be consistent with chance having a lot to do with either an exceptional success or an exceptionally poor performance in a given season.

Thus, we conclude that, while the mean ratios were virtually identical, the experts had a wider variation in their performance. We also found that the relative ranking of the performance of experts is more stable over the two years of our sample than the performance of systems.

3.3. The use of new information: accuracy over a season

The information on team performance that is available to forecasters increases over the course of the season. Does this increased information improve forecast accuracy between the first and second halves of the seasons? We hypothesize that the forecast accuracy will be relatively higher in the second half of each of the two seasons. Increased information may be particularly important for statistical systems, since a larger database of games should permit better estimates to be made of the parameters of the models that are used to generate the forecasts. Unfortunately, the results are inconclusive (see Table 3).

For all methods, the relative forecast accuracy was significantly lower in the second half of the 2000 season, but significantly higher in the second half of the 2001 season. However, if the records of the two seasons are combined, the relative accuracy of the statistical systems increased over the course of the seasons, while that of the experts declined.

In conclusion, we find no evidence that the forecasts of experts improved from the first to the second halves of the seasons. However, we do find some support for the hypothesis that the second half predictions by statistical systems are superior to their predictions for the first half. The betting market outperforms experts and systems in both halves of each season, and there is no statistically significant difference in the betting line predictions across halves. One interpretation of these findings is that (a) statistical systems may systematically exploit additional information on the relative abilities of teams that accrues as the season progresses, in a way that experts do not; and (b) that the betting line includes such information even early in the season. However, the results for the 2000 season, when accuracy did not improve in the second half, suggest that some caution is required about the ability of forecasters to consistently learn from first half games to improve their subsequent performance.[5]

4. How well do experts and statistical systems predict relative to the betting line?

Experts predict against the betting line either by indicating that a specific team is expected to beat the line, or by providing a point spread for the game. Statistical systems only forecast the point spreads. If a prediction states that a team will (not) beat the spread, it is only necessary to count the number of times that this statement was correct. On the other hand, if forecasters provide point spreads, we compare the predicted point

[5] This result was confirmed when we examined forecast accuracy by week of the season. We found that there was considerable variation in performance from week to week but that there was no clear trend in forecast accuracy as the seasons progressed. (These results are available from the authors on request.)

Table 4
The proportion of games that were correctly predicted against the line by experts and statistical systems: 2000 and 2001

	Experts			Statistical systems		
	2000	2001	Both Seasons	2000	2001	Both Seasons
	A. forecasters as the units of observation					
Mean	0.504	0.486	0.496	0.516	0.496	0.507
Standard Deviation	0.031	0.030	0.027	0.026	0.033	0.026
Minimum	0.446	0.437	0.446	0.467	0.419	0.462
Maximum	0.579	0.571	0.579	0.564	0.565	0.565
Number of Forecasters	42	36	52	23	29	31
	B. games as the units of observation					
Number of Games Predicted	9,233	7,675	16,908	5,201	6,257	11,458
Accuracy Ratio	0.504	0.486	0.496	0.516	0.494	0.504

spread and the difference of the actual scores. If the point spread and the actual score difference are both greater than (or less than) the betting line, then that forecaster has made a correct prediction. We then calculate the proportion of games that are correctly predicted by experts and statistical systems.

The upper panel of Table 4 presents the basic data, with forecasters as the units of observation. The lower panel of the table uses the forecasts as the units of observation. Table 5 provides similar information for those making forecasts for both years.

Using these data, we describe the overall success rates in Section 4.1. We test two hypotheses in Section 4.2: (1) whether the individual experts or statistical systems are significantly more accurate than a naïve forecasting method of flipping a coin; and (2) whether any expert or statistical system could have yielded profits against the Las Vegas betting market. Finally, in Section 4.3 we explore whether accuracy in predicting against the line improves as the season progresses.

4.1. Overall success rates in forecasting against the betting line

The lower panel of Table 4 indicates that experts and statistical systems had similar records in predicting against the betting line. Similar results are obtained if we compare forecasts separately by year, use the individual forecaster as the unit of observation, or compare the performances of experts and statistical systems that made predictions for both seasons. Table 4 also shows that there is considerable variation in the success rates of both experts and systems in picking against the line. The range was slightly wider among experts. This result is similar to the finding about the ranges associated with picking the winning team.

We also compared success rates across seasons to see whether some forecasters were consistently better (or worse) than others. There are two findings of interest. First, we found that the accuracy ratios of statistical systems in 2000 when picking against the line were correlated with their accuracy ratios in 2001. The Spearman correlation coefficient of the ranks of statistical systems across seasons was 0.43 (statistically significant at the 0.05 level), while that of experts was only 0.03. This pattern is the opposite of that found in our examination of consistency in forecasting game winners. Second, and similar to our earlier finding with respect to forecasts of game winners, we found that experts or systems that had especially high success rates in the 2000 season did not perform as well in 2001, while those that performed especially poorly did improve on their records, again suggesting regression to the mean.

Table 5
The proportion of games that were correctly predicted against the line by experts and statistical systems that issued predictions in both seasons: 2000 and 2001[a]

	Experts			Statistical Systems		
	2000	2001	Both Seasons	2000	2001	Both Seasons
Mean	0.508	0.483	0.496	0.516	0.485	0.501
Standard Deviation	0.029	0.031	0.023	0.025	0.028	0.022
Minimum	0.468	0.437	0.465	0.467	0.419	0.462
Maximum	0.558	0.571	0.561	0.564	0.539	0.543
Number of Forecasters	26	26	26	21	21	21

[a] The forecaster is the unit of observation in calculating the means and standard deviations.

Table 6
Accuracy ratios of experts and statistical systems in predicting against the betting line by the halves of the 2000 and 2001 seasons

Year	Half	Experts		Statistical systems	
		Games	Accuracy ratio	Games	Accuracy ratio
2000	1st	4826	0.522	2547	0.533
	2nd	4407	0.484 [a]	2654	0.499 [b]
2001	1st	3909	0.479	3513	0.479
	2nd	3766	0.496	3104	0.509 [b]
Combined seasons	1st	8735	0.502	5700	0.503
	2nd	8173	0.489	5758	0.504

[a] Indicates that the accuracy ratios of the two halves of a season are statistically significantly different from each other at the 0.01 level.
[b] Indicates significance at the 0.05 level.

4.2. Comparisons of success rates against benchmarks

4.2.1. The naïve benchmark

The naïve model, i.e. flip a coin, is one benchmark for evaluating forecasts against the line. This method would be expected to pick one-half of the games correctly. In comparison with this standard, experts did slightly worse than the naïve forecast for the two seasons combined, while statistical systems fared slightly better than a naïve forecast. There are, however, considerable differences between the two seasons, with both methods beating the naïve forecast in 2000, and both failing to do so in 2001.

Among forecasters who issued predictions for both years, only about one-third of the experts and slightly under half of the statistical systems were more accurate than the naïve forecast. Moreover, only two of the experts and two of the systems had accuracy ratios that were statistically significantly greater than 0.50 at the 0.05 level. However, it should be remembered that we are conducting multiple tests of the hypothesis. Applying the Bonferroni correction, one cannot reject the hypothesis that the success rates of the two experts and the two systems were due to "luck".

4.2.2. The profitability benchmark

Another standard of comparison is whether a forecaster had an accuracy ratio greater than 52.4%. This is the success rate that is required to break-even in the betting market (Sauer, 1998). Combining the predictions of all forecasters, neither experts nor systems achieved that level of accuracy in either season. Among forecasters making predictions for both seasons, only three experts and three systems were that accurate. Only one expert had a success rate that was statistically significantly greater than 52.4% at the 0.10 level.

4.3. Do forecasts against the betting line improve over the course of a season?

Next, we inquired whether predictions against the betting line are superior in the second halves of the seasons (when forecasters have more information) than those issued in the first halves. Table 6 indicates that for the combined seasons, the forecast accuracy of the statistical systems is nearly identical in the first and second halves. However, the accuracy ratio of experts declines across halves.[6] Overall, it would appear that additional information does not improve forecasts against the betting line.

5. Overall conclusions

We have compared the accuracy ratios (success rates) of experts and statistical systems in predicting game winners and winners against the betting line for two National Football League seasons. In comparison with previous research, we have large samples of forecasts. For example, there are 48 experts and 31 statistical systems forecasting game winners, with as many as 496 predictions per forecaster. These large sample sizes allow us to test a variety of hypotheses regarding the determinants of forecasting success. There are several findings of interest. Some of these results are in conflict with the generally accepted premise that the forecasts of statistical systems are more accurate than those coming from experts.

A principal result is that there was little difference in the overall success rates between experts and statistical systems in predicting game winners, or in picking against the betting line. We have some hypotheses as to why the individuals in our sample perform as well as the statistical models. First, the individuals making predictions are "true" experts. Some had played professional football, while others had followed the sport for many years. Second, their forecasts are published and read in national magazines or newspapers, or in local newspapers with an extensive fan following. To maintain

[6] It is possible that the decreased accuracy of experts could be a result of their inability to process additional information correctly.

credibility, these experts have some incentive to provide sensible forecasts.[7] Finally, experts typically publish their forecasts late in the week, or even on game day. By that time, they have extensive information about the teams. They can consult published information, and perhaps obtain inside information from coaches or players with whom they may have developed relationships. Moreover, they can review the results of statistical models (which are generally available earlier in the week), and examine the most recently published betting line. These factors may reduce the effects of the widely-reported biases generally associated with human judgment (cf Bunn & Wright, 1991).

Our result is similar to that of Forrest et al. (2005), who found that models were not superior to the odds-makers for English soccer games. They argued that the huge financial stakes provided incentives for odds-makers to supply accurate predictions. Consistent with this argument, they found that the experts' accuracy improved relative to that of the model during a time period when there was an increase in competitive pressures on the odds-makers. To remain profitable, the odds-makers' accuracy had to increase.

We also find that, while the average success rates of experts and systems were similar, the variation in success rates was higher among experts than among statistical systems, particularly in selecting game winners. Moreover, the worst records among statistical systems were typically much better than the worst records among experts. This is potentially an important factor in deciding whether to use experts or systems in making a forecast.

Consistent with previous research in sports forecasting, the betting line was substantially superior to both experts and systems in predicting game winners. As in the case of the odds-makers in English soccer, those setting the initial betting line have a considerable financial incentive to pick the right line. The initial betting line is set by experts who use a combination of statistical models and judgment.[8] Gandar et al. (1988) found little difference between forecasts of scores using the initial and final betting lines for NFL games over the period 1980–85. We find a similar result using data gathered by Avery and Chevalier (1999) on nearly 2,500 games for the 1983 to 1994 seasons.[9] Since our results indicate that the final betting line outperformed both the models and the experts, the similarity of the initial and final betting line forecasts suggests that the combination of statistical models and the judgment of experts who have strong monetary incentives to be accurate is superior to either method alone.

Acknowledgments

We would like to thank the George Washington University Facilitating Fund for financial support in gathering the data used in this study, and Robert Goldfarb, James Lebovic, Robert Trost, Paul Goodwin and an anonymous referee for helpful comments on an earlier version of this research.

Appendix A. Characteristics of the Betting Market

Betting on American football games does not involve picking a winning team; rather, the bet involves predicting whether a particular team beats or does not beat the other team by a specific margin. This margin is called the point spread. It has also been called the betting line, or just the line. Thus, if team A is favored to beat team B by 3 points, a bettor can choose either to bet that A will beat B by more than 3 points, or that team B will win or lose by 2 points or fewer. If A wins by exactly 3 points, the bet is nullified and the money is returned to the bettor.

An additional characteristic of this market is that the bettor must bet \$11 in order to win \$10. This explains why the accuracy ratio must be 52.4% in order to break even (a detailed explanation of this betting market is presented in Sauer, 1998).

The betting line or spread is set in Las Vegas on the Monday preceding the games, which usually occur on the following Sunday or Monday. This spread is published in all of the major newspapers in cities that have professional football teams.

In our analysis of the betting market, the team that is favored to win is classified as the one that the market is

[7] The contingency forecasting model of Beach, Barnes, and Christensen-Szalanski (1986) argues that experts' forecasting strategies and their accuracy are influenced by the benefits and costs of accuracy.

[8] The process of setting the original line is explained by Schoenfeld (2003).

[9] Both the initial and final lines provided efficient forecasts of game outcomes, and the explanatory powers of the regressions of the score on the initial and final betting lines were nearly identical (R^2 = .15 vs. R^2 = .16, respectively).

forecasting will win. This is regardless of the spread that is quoted in the market. When an expert (system) indicates that a team will beat the spread, we take this to mean that the team will win by a larger margin than the one that has been established in the market.

References

Andersson, P., Edman, J., & Ekman, M. (2005). Predicting the World Cup in 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, *21*, 565−576.

Armstrong, J. S. (2001). Judgmental bootstrapping: Inferring experts' rules for forecasting. In J. S. Armstrong (Ed.), *Principles of forecasting: A handbook for researchers and practitioners* (pp. 171−192). Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S., & Collopy, F. (1998). Integration of statistical methods and judgment for time series forecasting: Principles from empirical research. In G. Wright & P. Goodwin (Eds.), *Forecasting with judgment* (pp. 269−293). New York: Wiley.

Avery, C., & Chevalier, J. (1999). Identifying investor sentiment from price paths: The case of football betting. *The Journal of Business*, *72*, 493−521.

Beach, L., Barnes, V. E., & Christensen-Szalanski, J. (1986). Beyond heuristics and biases: A contingency model of judgmental forecasting. *Journal of Forecasting*, *5*, 143−157.

Boulier, B. L., & Stekler, H. O. (2003). Predicting the outcomes of National Football League games. *International Journal of Forecasting*, *19*, 257−270.

Bunn, D., & Wright, G. (1991). Interaction of judgmental and statistical forecasting methods: Issues and analysis. *Management Science*, *37*, 501−518.

Collopy, F., Lawrence, M. J., & Wright, G. (1996). The role and validity of judgment in forecasting. *International Journal of Forecasting*, *12*, 1−8.

Dawes, R. (1986). Forecsting one's own preference. *International Journal of Forecasting*, *2*, 5−14.

Dixon, M. J., & Pope, P. (2004). The value of statistical forecasts in the UK association football betting market. *International Journal of Forecasting*, *20*, 697−711.

Forrest, D., Goddard, J., & Simmons, R. (2005). Odds-setters as forecasters: The case of English football. *International Journal of Forecasting*, *21*, 551−564.

Forrest, D., & Simmons, R. (2000). Forecasting Sport: The Behavior and Performance of Football Tipsters. *International Journal of Forecasting*, *16*, 317−331.

Gandar, J., Zuber, R., O'Brien, T., & Russo, B. (1988). Testing rationality in the point spread betting market. *Journal of Finance*, *43*, 995−1007.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, *2*, 293−323.

Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, *12*, 19−30.

Sauer, R. D. (1998). The Economics of wagering markets. *Journal of Economic Literature*, *36*, 2021−2064.

Schoenfeld, B. (2003). What's my line? *Sports Illustrated*, *99*, 76−87.

Webby, R., & O'Connor, M. (1996). Judgmental and statistical time series forecasting: A review of the literature. *International Journal of Forecasting*, *12*, 91−118.

Wright, G., Lawrence, M. J., & Collopy, F. (1996). Editorial: The role and validity of judgment in forecasting. *International Journal of Forecasting*, *12*, 1−8.