Research Program on Forecasting

ISSUES IN SPORTS FORECASTING

H. O. STEKLER
DAVID SENDOR
RICHARD VERLANDER
Department of Economics
George Washington University
Washington DC 20052

RPF Working Paper No. 2009-002
http://www.gwu.edu/~forcpgm/2009-002.pdf

December 10, 2009

FORTHCOMING IN THE SPECIAL ISSUE ON SPORTS FORECASTING OF THE *INTERNATIONAL JOURNAL OF FORECASTING*

RESEARCH PROGRAM ON FORECASTING
Center of Economic Research
Department of Economics
The George Washington University
Washington, DC 20052
http://www.gwu.edu/~forcpgm

ISSUES IN SPORTS FORECASTING

H. O. STEKLER*
DAVID SENDOR
RICHARD VERLANDER
Department of Economics
George Washington University
Washington DC 20052

*Corresponding author: hstekler@gwu.edu. We would like to thank Bryan Boulier, David Forrest, Ira Horowitz and Leighton Vaughan Williams for their comments on an earlier draft of this paper. This paper has also been presented at the International Symposium on Forecasting in New York in June 2007 and the Forecasting Workshop of the Department of Economics, George Washington University.

ISSUES IN SPORTS FORECASTING

ABSTRACT

A great amount of effort is spent in forecasting the outcome of sporting events, but few papers have focused exclusively on the characteristics of sports forecasts. Rather, many papers have been written about the efficiency of sports betting markets. As it turns out, it is possible to derive considerable information about the forecasts and the forecasting process from the studies that tested the markets for economic efficiency. Moreover, the huge number of observations provided by betting markets makes it possible to obtain robust tests of various forecasting hypotheses. This paper is concerned with a number of forecasting topics in horse racing and several team sports. The first topic involves the type of forecast that is made: picking a winner or predicting whether a particular team beats the point spread. Different evaluation procedures will be examined and alternative forecasting methods (models, experts, and the market) will be compared. The paper also examines the evidence about the existence of biases in the forecasts and concludes with the applicability of these results to forecasting in general.

Key words: sports forecasting, betting markets, efficiency, bias, sports models

ISSUES IN SPORTS FORECASTING

A great amount of effort is spent in forecasting the outcome of sporting events. Moreover, there are a large amount of data regarding the outcomes of sporting events and the factors that are assumed to contribute to those outcomes. Yet despite the 40,000 entries in JSTOR and 3,700 in Econ Lit that refer to sports, few papers have focused exclusively on the characteristics of sports forecasts. Rather, many papers have been written about the efficiency of sports betting markets.[1] (See Sauer (1998, 2005) and Vaughan Williams (1999, 2005) for surveys of the efficient-market literature.)

Since, most previous betting-markets studies were concerned with economic efficiency, they did not evaluate the actual (or implied) forecasts associated with those markets, i.e. the studies' emphases have been on profitability rather than predictability. As it turns out, it is possible to derive considerable information about the forecasts and the forecasting process from the studies that tested the markets for economic efficiency. Moreover, the huge number of observations provided by betting markets makes it possible to obtain robust tests of various forecasting hypotheses. It is not necessary to base findings on laboratory experiments with a small number of observations that may not replicate real-world conditions. For example, one

[1] Sports data have also been used to investigate topics such as: the strategies (minmax, risk taking) that competitive players and teams employ, the benefits of stadia and teams to cities, the business and management of professional team sports including the trading of players, the market structure and competitiveness of professional leagues with free agency and payroll caps, labor relations and the effects of strikes, the determinants of attendance at sporting events, and even the problems associated with adverse selection in the sale of thoroughbred race horses.

study (Song et al. 2007) that compared the predictive accuracy of judgmental forecasters with statistical systems was based on 31,000 observations of real-time predictions of the outcomes of American professional football games. [2] Although we make extensive use of data from these betting markets, we do not examine the economic efficiency of these markets.

This paper is concerned with a number of forecasting topics in horse racing and several team sports. The first topic involves the type of forecast that is made. In some sports the forecast is intended to determine the winner of an event. In other betting markets, individuals predict whether/or not the favored team wins by x points (called the point spread). It should be noted that the information (summary forecasts) that is obtained from these markets is related to the underlying characteristics and scoring systems of each sport. In sports where there is no binary outcome (horse racing, soccer, etc.), the market provides odds (probabilities) about the likelihood of each outcome .Even when there is a binary outcome, such as in baseball, the market quotes odds against each of the outcomes. However, in other binary outcome sports, such as basketball and American football, odds (probabilities) are not quoted. Rather the summary statistic is the point spread. This number is the median value of the probability distribution for the difference in scores.[3] This statistic thus provides information that distinguishes between teams that are closely matched and those where there is a clear favorite.

A second topic involves the procedures that are used to evaluate the forecasts. This paper will show that the evaluation procedures depend on the type of betting market that is associated with each sport. A third topic involves a comparison of alternative forecasting methods. For

[2] Moreover, since a sporting event has a definite outcome at a specific point in time, it is not necessary to make assumptions about expectations as to the future as is necessary in other asset markets.

[3] We wish to thank the referee for pointing this out.

every sport, forecasts have been made by models (systems) and experts. In some sports, it is also possible to analyze a forecast that is made by the market. The final topic is to determine whether the forecasts are biased and, if they are, the sources of the biases.

These topics will not be discussed on a sport-by-sport basis. Rather we will integrate the results across sports and in the process provide extensive citations referring to each sport. We then determine whether the results for individual team sports and horse racing yield valid generalizations about sports forecasting. Finally, we compare the findings from sports forecasts with the profession's generally accepted beliefs about forecasting knowledge and see whether they are consistent.

I. Types of Forecasts

The forecasts that we examine come from three sources. First, there is the betting market forecast itself. Second, forecasts can be derived from statistical models that are based on the fundamentals of the sports or are based on variables that are proxies for these characteristics. Finally, experts, be they bookmakers, handicappers or sports commentators, also issue forecasts about the likely outcomes of sporting events.

A. Betting Market Forecasts

Given that the forecasts are generally associated with and obtained from the gambling market, it is first necessary to discuss how the markets are structured and the type of forecast that can be analyzed. The gambling markets are constituted differently from sport to sport. In horse racing, baseball, and soccer the bet (forecast) involves picking a winner, and the market quotes odds that a particular horse or team will win.[4] As an example, if there were only two teams and there were no commissions, the odds on Team A might be 2 to1 and the odds on Team B would then be 1 to 2. This forecast implies that Team B will beat Team A 2/3 of the time. This type of forecast can be evaluated in two ways: (1) was the market correct and did Team B win more frequently or (2) in repeated trials, did teams that were favored 2-1 win 2/3 of the time?, i.e. were the probability forecasts calibrated?

In markets where odds are quoted and there are more than two competitors (horse racing) or more than two outcomes (soccer: win, draw lose), it is not possible to determine whether the market forecasts correctly predicted the winners. Rather the forecasting evaluation must be based on a comparison of the market odds (ex ante probabilities) and the ex post relative frequencies of the actual outcomes. The betting odds must, therefore, be converted into probabilities by the formula, $p = 1/(1+\text{odds})$.[5] For each ex ante probability, p_i , the ex post proportion of winning horses that went off at those odds, f_i, should be equal to p_i. Using statistical terminology, the ex ante probabilities and the ex post winning percentage should be calibrated, and horses whose ex ante probability of winning was 0.30 should have won 30 percent of the time.[6]

The betting market that involves baseball has a unique way of presenting the odds. While bets in this market are made on the outcome of a game, the odds are not quoted directly. The bookmaker quotes, a line, +140, -150 for example. This means that the winner of a $100 bet on

[4] The betting market in golf and tennis, two markets that are not analyzed here, function similarly.

[5] The sum of the betting odds exceeds one because of the bookmakers' commissions or the parimutuel take; they must be adjusted so that they sum to one.

[6] The quadratic probability score (QPS) as also known as the Brier Score can be used to evaluate probability forecasts when the relative frequencies of the outcomes are known.

the underdog team would win $140, while someone betting on the favored team would bet $150 to win $100. The difference is the commission. From these odds it is possible to calculate the betting market's forecast that the underdog will win. The probability is calculated at the midpoint of the line, i.e. 1/(1.45 +1) = 0.41 . This probability can then be compared with the percentage of times that the underdog won when those odds were quoted to determine whether the ex ante probabilities are calibrated with the observed relative frequencies.

The forecasts and the evaluation procedures are different in those markets where bets are placed on the margin of victory. In American football and basketball, bets are not made on which team will win nor are odds quoted in the market. Rather there is a bet on whether or not the favored team will win by more or less than the specified margin (point spread) that is set in the market.[7] While this type of betting market is not concerned with selecting the winning team, it is possible to use the data about the spread to determine whether the market accurately predicts who will win. Thus, our analysis of the forecasts from this type of betting market examines two questions: (1) How frequently does the team that is favored to win actually win? and (2) Are there any observed biases in the spreads that were published just before the game was played? It is also possible to determine if the market forecasts are more accurate than those of other forecasting methods, such as models and the opinions of experts.

B. Models

In order to predict the outcome of sporting events, many different types of models have been constructed, but, unfortunately, many of these models have never been used in forecasting

[7] Although odds are not quoted, the bet is not even money because the bettor must commit $11 to win 10. Since bookmakers set the point spread in an attempt to receive an equal amount of money on both sides of the bet, the deviation from the even-money bet represents the bookmakers' commission.

beyond the period of fit. Only the major characteristics of these models are explained in the text. Some models are disaggregated; others are based on production functions or power scores. There are even models that use payrolls as predictors of success within a sport. (See Smyth and Smyth, 1994 for baseball; Szymanski, 2003, p. 1154 and Forrest et al. (2005a) for soccer). The form of each of the models and the variables that are included in the equations are presented on a sport-by-sport basis in the Appendix.

At the most disaggregated level, it is possible to predict the outcome of a game by modeling the effects of every play. (For baseball see Bukiet[8] et al. (1997) and Sauer (2005); for soccer see Carmichael et al. (2000)). At a more aggregative level, production functions are used. These functions focus on the fundamental factors that determine the outcome of a game. On offense it is the *factors* that determine the number of points, runs, or goals scored; on defense the *factors* that determine the number of points, runs, or goals allowed. The model that is used to forecast the outcome of a game is then based on the differences in the fundamental characteristics of the two teams.

An alternative statistical procedure is to construct a power score or index that is a <u>proxy</u> for these fundamental characteristics or the latent skills and strengths of the teams. Such a model uses the difference in runs (points, goals) scored as a predictor. Then there are models that use power scores based on relative performance as the independent variables. The focus of these models is exclusively on the relative number (and margins) of victories of the competing teams and the time trend in this relationship. For example, the New York Times created power scores

[8] Bukiet et al. (1997) modeled each at-bat as a 25x25 transition probability matrix that explained all of the alternatives that might occur. Markov chains were then used to predict team performance based on the characteristics of each batter.

for every National Football League (NFL) team that summarized each team's relative performance in previous games. It was based on the winning percentage of each team, its margin of victory, and a quality measure for each of its opponents. Similar measures, that include the strength of schedule, have been constructed for other sports. These power scores can be transformed further into ordinal rankings (Boulier and Stekler, 2003).

Statistical scoring systems are variants of power scores. As an example, Sagarin has developed a system that can be used to predict the expected scoring by any two teams. This system is based on the number of victories of each team, the strength of the teams that were defeated, the margin of victory adjusted for blowouts, and an adjustment for the home court advantage.[9]

C. Experts

Finally, we have predictions made by individuals (experts) who may or may not reveal their methods. Some of these experts are sports writers, editors of newspapers or sports magazines, or sports commentators on the major television networks; others are tipsters. The odds makers in the betting markets and the race-track handicappers should also be considered experts.

II. Results- Betting Market

In presenting our results it should be remembered that the procedures for evaluating forecasts will differ from sport to sport depending upon the institutional structure of the betting market that is associated with that sport.

A. Horse racing

[9]The difference between two teams' Sagarin ratings is a good predictor of the margin of victory (Carlin, 1996).

The evidence is that the betting market yields accurate forecasts. The horse racing results indicate that the market can distinguish among horses of different quality. With a particular exception, the probabilities obtained from the odds rank of the horses are well calibrated with the observed frequency of wins. (Sauer, 1998, pp. 2035 and 2044). The exception to the aforementioned calibration occurs at the extremes of the odds distribution. This result found in most studies of horse racing in the US yields what has been called "the favorite-long shot bias." This means that an insufficient amount is bet upon the horses that are favored to win and an excessive amount is bet on the long shots, thus distorting the odds at the extremes.[10]

B. Baseball

In the baseball market, Woodland and Woodland (1994, p.275; 1999, p.339) and Gandar et al. (2002, p.1313) all indicate that the odds are related to the observed outcomes, but the relationship is not strictly monotonic.[11] Woodland and Woodland (1994) argued that betting in baseball yielded a reverse favorite-underdog bias, with underdogs underbet. Gandar et al. (2002) made a minor correction to the Woodland-Woodland methodology and found that if there were any bias, it was very slight.

C. American Football

We must distinguish between picking winners and betting against the spread when the forecasting performance of the American football betting market is evaluated. Boulier and Stekler (2003) and Song et al. (2007) showed, that in every year from 1994-2001, the football betting market correctly predicted the winner of NFL games at least 63% of the time. The

[10] This bias does not exist in race-horse betting in Hong Kong and Japan, for example. It also does not exist in most other sports.

[11] The QPS statistic, also known as the Brier score, was not calculated and decomposed to determine the degree of calibration.

average over this time period was about 65%. In fact, in selecting the winners of games, the betting market was the most accurate forecasting method in every year. Moreover, there was a positive relationship between the point spread and the ex post winning percentage of the home team, but the increase is not monotonic.

In analyzing the market's performance relative to the spread, the main question is: Are there any observed biases? The overwhelming majority of the evidence indicates that the betting market is efficient in the sense that, on average, there is no profitable betting strategy against the spread. The traditional method for determining whether a forecast is unbiased is to run the regression:

$$A = a + bF + e, \qquad (1)$$

where A is the actual value and F is the forecast. If the joint null hypothesis that a = 0 and b =1 is rejected, the forecasts are biased.   In the football betting market, the equivalent equation is:

$$DP = a + bPS + e, \qquad (2)$$

where DP is the difference in the game score (actual points) and PS is the betting-market point spread.[12] If the forecast is unbiased, on average, the difference in the point scores will not differ significantly from the point spreads.   Most studies do not reject the null hypothesis that a = 0 and b = 1, but the explanatory power of the equation is usually low, indicating that there is considerable unexplained variation.[13]

D. Basketball

[12]Both the scores and the point spreads are usually constructed on a home team minus away team basis.

[13] Gray and Gray (1997) use a different method to test the hypothesis that the forecasts are efficient. They estimate a probit model where the dependent variable is whether the team beat the spread or not. They find that two variables, whether a team plays at home and whether it is a favorite, are jointly significant. If the market had provided an efficient forecast, neither variable should have had any explanatory power.

The studies that have examined the basketball betting market have not found any *significant* biases. There is a slight but insignificant underestimate of the home-court advantage, (Brown and Sauer 1993a) and large favorites may be over bet (Paul and Weinbach, 2005a, 2005b). Gandar et al. (1998, 2000) examine the differences between the opening and closing point lines for NBA games[14] and show that frequently there are large changes between the opening and closing quotations. Since they show that the opening line is not as accurate as the closing line in forecasting the margin of victory, they conclude that informed bettors have eliminated some of the bias in the opening line.

F. Soccer

Gambling in soccer is based on odds, but this betting market is different from those that have been analyzed above. The bookmakers set the odds at the beginning of a week and do not change them during the betting period. In one of the few studies that searched for biases, Cain et al. (2000) showed that there was a favorite-longshot bias, similar to that found in horse racing, in the soccer betting market. Many studies have found that models contained information that was not embodied in the odds (Pope and Peel, 1989; Kuypers, 2000; Goddard and Asimakopoulos, 2004; Dixon and Pope 2004). These findings suggest that the forecasts embedded in the bookmakers' odds were inefficient.

G. Summary and Discussion of Biases

---

[14]Gandar et al. (1998) examine the winning margin (the difference in the scores of the two teams) while Gandar et al. (2000) analyze the totals betting market (the sum of the scores of the teams). The opening line is set early in the day that the game is played and the closing line is established just before the game begins. It is unlikely that much *new* information about the teams will have become available during the course of the day.

There are mixed results about the existence of biases in the market forecasts.[15] The bias in horse racing occurs at the two extremes: favorites are underbet and long-shots are overbet, but these results do not hold in all countries. Similarly, in soccer Cain et al. (2000) and Deschamps (2007) found biases. On the other hand, in markets where odds are quoted, the ex ante betting probabilities and the ex post relative frequencies are calibrated. The betting spread is an unbiased predictor of the winning margins in American football and basketball. Moreover, the betting market correctly predicted the winner of NFL games about 2/3 of the time.

However, the presence of a forecast bias in this type of financial market, such as the horse race and sports event betting markets, is an anomaly, and explanations have been based on forecasting characteristics or bettor preferences (See Sauer, 1998; Vaughan Williams, 1999, 2005; Vergin, 2001; Forrest and McHale, 2007). There are several explanations for this bias. One concerns bettors' preferences; another involves issues in forecasting, namely the role of information and individual's ability to interpret the information. We are only concerned with the forecast issues.

When the odds are wide and less information is available, a longshot bias is more likely (Vaughan Williams and Paton, 1998; Forrest and McHale, 2007). When more information is available publicly and the bettors are better informed, the more likely it is that the consensus forecast (represented by the market odds) will converge to the true odds. Empirical evidence involving horse racing is consistent with this view. The bias is diminished if either the betting pool or the number of horses in the race is increased. (Busche and Hall, 1988; Gramm and Owens, 2005).

[15] Biases were also found in the betting market for two sports that are not considered in this paper, e.g golf Shmanske (2005) and tennis Forrest and McHale (2007).

On the other hand, there is the possibility that new information, especially if it is positive, might introduce a bias. Vergin (2001) argued that bettors (forecasters) in the NFL market were subject to an overreaction bias. They overreact to the most recent *positive* information and undervalue other data: "For example, if a team won a game by a very large margin in a given week, the betting public would tend to overrate the team in the following week" (Vergin, 2001, p. 499).[16] Vergin found that, in most cases, over 15 seasons, the bettors had displayed a bias in interpreting recent positive data. Gray and Gray (1997) also found that the market overreacted to the most recent information.[17]

The non-sports forecasting literature has also analyzed the way that individuals interpret information. Kahnemann and Tversky (1982) had argued that individuals place too much emphasis on new information, but some experimental data suggested the opposite: people anchor on past observations and place too little emphasis on new information. The data from the football betting market seem to favor the former view about the role that new information plays in generating a forecast.

The manner in which information is interpreted has also been discussed in the context of the "hot hand" belief in the basketball betting market. This is a belief that a team that wins a game is more likely to win the next game and indicates that forecasters believe that these events are not independent but rather are positively autocorrelated. Camerer (1989) argues that the hot hand is a myth and that bettors have a misunderstanding of random processes, especially with small samples. Brown and Sauer (1993b) conclude that the hot hand belief is embodied in the

[16]Vergin and Scriabin (1978) were also concerned with this issue.

[17]In addition, Gray and Gray observed that the market had a slight overconfidence in the favorite's ability to cover the spread. In the 1976-94 seasons, that team won by slightly less than the market had expected.

point spread and is, therefore, an important effect.[18] They were unable, however, to determine whether this was a real phenomenon or whether bettors misperceived the real process and thus displayed a cognitive bias.

An alternative, and perhaps supplementary, explanation involves the forecasting abilities of the individuals who place bets. Golec and Tamarkin (1995) argued that bettors were overconfident in their abilities to predict. This result is consistent with the findings from some laboratory experiments indicating that individuals generally underestimate the probability of likely events and overestimate the probability that an unlikely event will occur.[19]

We conclude that, on average, the betting market in all sports generates unbiased forecasts. A continuing debate on the existence of a longshot bias should provide a greater understanding of the role of information and individuals' forecasting abilities.

III. **Results- Models**

We again must note that the models have frequently not been used to make ex ante forecasts. Consequently, we only have a limited amount of information about the forecasting records of these models.

A. Horse Racing

A multinomial logit model of horse racing that included characteristics of both the horse and the jockey had an adjusted R^2 of .09 (Bolton and Chapman, 1986). The equation explains only 9% more than the null that each horse has an equal chance of winning. Expanded versions of this model somewhat improved the explanatory power of the equation, yielding an adjusted R^2 exceeding 12% (Bentner, 1994;Chapman, 1994). While the final betting odds have even

[18] Also see Paul and Weinbach (2005b).
[19] In fact, Vaughan Williams (1999, p.8) cites one study that finds that bettors are overconfident.

more explanatory power, a combination of the model and the market odds improves upon both. This finding is consistent with the results obtained from the non-sports forecasting literature which indicates that combining forecasts usually improves accuracy.

B. Baseball

There is so much information about baseball that it is surprising how few forecasts are available for analysis. The Bukiet et al. model (1997), based on modeling the effects of every play, was used to predict the actual number of games that each team in the National League would win in 1989. The results were mixed: the model failed to predict one of the two divisional winners and the number of runs that were scored was underestimated. Nevertheless, the Spearman Rank Correlation between the predicted and actual number of games that each team won was .77 (Authors' calculations).

Two other models have been used to predict divisional winners. Barry and Hartigan (1993) used a binary-choice model to calculate the probability that in 1991 a National League team would win its division. Using simulations, the model successfully showed that Atlanta's probability of winning its division was increasing as the season progressed. Finally, Smyth and Smyth (1994) based their predictions of division winners and relative standings on the payrolls of each of the teams in each division and league. They found that the rankings within a division were correlated with the teams' payrolls.[20]

C. Football

The models that forecast the outcome of football games are frequently evaluated on their

[20]Szymanski (2003, p. 1154) presented similar results for soccer. He showed that, within each league, the winning performance of a team was associated with the relative size of the teams' payrolls. Forrest et al. (2005a) use payrolls as a measure of team quality.

ability to predict the margin of victory rather than the outright winner. The Zuber et al., (1985) and Sauer et al., (1988) models, based on the fundamental characteristics of the teams, correctly predicted a margin of victory that would have been profitable 59% of the time for games played in 1983, but the success rate was only 39% in 1984.[21] The Dana and Knetter (1994) model, using proxy variables, was accurate less than 50% of the time. On the other hand, the predictions of the Glickman and Stern (1998) model for the last eight weeks of the 1993 season were comparable to the betting line and would have been profitable.

Variants of power scores have been used in forecasting both the outcomes of football games and the margins of victory. Harville (1980) found that in the 1971-77 seasons the betting market, with a 72% success rate in selecting the winner, was more accurate than his statistical procedure, which was right 70% of the time.[22] Boulier and Stekler (2003) used the power scores published in the *New York Times* and analyzed the forecast: the team with the higher power score would win. These forecasts had an accuracy ratio of 61%, less than that of the betting market which had an accuracy ratio of approximately 65% and only comparable to a naive forecast that the home team will win.

An intensive evaluation of the forecasting record of statistical systems indicated that they had a 62% average accuracy ratio in picking the winners of the games played in the 2000 and 2001 NFL seasons. (Song et al., 2007). This ratio was comparable to the record of experts but less than the 66% accuracy of the betting market. Every system had a success rate of at least 50% and the ratios for all but one system were significantly different from those that could have

[21]The papers did not indicate the number of times that the model predicted the winners of each game, but the models explained 73-81% of the variance of the score differentials for the two NFL seasons.

[22]These success rates are higher than the accuracy that has been observed in more recent seasons. One explanation is that more ties occurred in the earlier seasons and Harville counted a tie as ½ of a successful forecast. Harville did not report the methods' record in betting against the spread.

occurred by chance.[23] In forecasting against the betting spread, most systems, however, were not even as accurate as the naive forecast of flipping a coin.

D. Basketball

Zak et al. (1979) developed a production function that represented the defensive and offensive elements of a basketball game, yielding a team's productive efficiency. The rank of each team in terms of its productive efficiency was identical to the rank based on winning percentage in the 1976 NBA season. Berri (1999) used a similar model that was designed to measure the contribution of individual players to a team's wins. The ranking obtained by summing the contributions of each player to team victories was remarkably close to the ranking based on the teams' actual won-lost records in the 1997- 1998 season. The Spearman Rank Correlation was .986 (Authors' calculations).

Alternatively, in a tournament, the seedings of the teams, which are obtained from a statistical scoring system, can be used as a predictor. Since 1985, the NCAA has selected 64 college basketball teams to participate in a tournament to select a field to compete for the men's national championship. The 64 teams are divided into four regional tournaments of 16 teams that are ranked from 1 through 16.[24] Boulier and Stekler (1999), Caudill and Godwin (2002), Kaplan and Garstka (2001), Caudill (2003), and Harville (2003) all found that the difference in ranks predicted the winner around 70% of the time.[25]

The accuracy of forecasts based on ranks has been compared with that of other methods.

---

[23] The test was based on the binomial distribution and a 5% level of significance.

[24] The seeds are determined from a statistical scoring system, the RPI, called the ratings percentage index. It gives weights of .25, .50, and .25 to the team's winning percentage, the winning percentage of its opponents, and winning percentage of the opponents' opponents, respectively.

[25] Caudill (2003) pointed out that probit models do not maximize the number of correct predictions. He uses a maximum score estimator and achieves a slight increase in predictive accuracy.

Kaplan and Garstka (2001) found that forecasts based on picking the higher seeds was slightly more accurate than using the betting market and that forecasts based on the Sagarin system were superior to both. Harville (2003) then compared the forecasting accuracy of his statistical method with (1) forecasting that the higher seed will win and (2) the betting market. The statistical procedures were the most accurate in forecasting the winners of the 2000 NCAA tournament, but there was little difference between forecasts based on ranks and the betting market. Kaplan and Garstka and Harville have been the only authors who have found that forecasts obtained from the market were not more accurate than those obtained from either experts or statistical systems.[26]

E. Soccer

The Poisson distribution is a model used to predict the number of goals that teams will score. Dixon and Coles (1997) show that this distribution provides a good fit to the score data for the 1992-95 seasons. Dixon and Pope (2004) then showed that the probabilities obtained from the Dixon-Coles (1997) model are similar to those of the bookmakers (as derived from the odds).[27] The model of Cain et al. (2000) used the win-lose odds prices quoted by the bookmakers rather than attack-defense proxies as independent variables to predict the total number of goals each team scored. The model generated probability forecasts that approximated the observed distribution of particular scores. Goddard (2005) showed that models based on goals and those derived from scores gave similar results.

Since the abilities and performance of teams can change over time, some models have

[26] Harville also found that there was no significant difference between the market and statistical systems in the football bowl games played after the 2001 regular season.

[27] Dixon and Coles do not provide a detailed evaluation, nor do they compare their predictions with a naive forecast that the home team wins 46%, draws 27%, and loses 27% of the time.

become dynamic to capture these effects. Dixon and Coles were among the first to incorporate dynamic factors into a model. Crowder et al. (2002) derived an approximation to the Dixon-Coles model and show that the two models yield similar results. The success ratio, associated with the prediction that the home team will win, is, however, only around 50%. The Bayesian dynamic model of Rue and Salveson (2000) yielded model likelihood measures that were very similar to the bookmakers' odds. Moreover, they used retrospective analysis to predict the posterior final rankings of the teams in the English Premier League. The relationship between the actual and predicted rankings in the 1997-1998 season was not perfect. The model forecast that Manchester United had a 43% chance of being the highest-ranked team; it finished second to an Arsenal team that had been given a 25% chance. Nevertheless, the model correctly selected the top four teams in the League.

The discrete choice models were based on ordered probits that included a variety of explanatory variables. Kuypers (2000) and Goddard and Asimakopoulos (2004) both indicated that there was little difference between their models' and bookmakers' probabilities.

F. Summary

Models that explain the outcomes of games or matches have been estimated for many sports. Sometimes the models were derived from the fundamental characteristics of the sport. In other instances, variables that were proxies for these fundamental characteristics were used as explanatory variables or discrete-choice models were used. The forecasts of many models were not available. Betting systems, however, correctly predicted the winners of NFL games more than 60% of the time which was comparable to the accuracy of experts but less than that of the market. The soccer models were comparable in accuracy to bookmaker odds.

IV. Results-Experts

A. Horse Racing

Figlewski (1979) examined the forecasting record of a number of horse-racing handicappers. While the handicappers were successful 28.7% of the time in selecting the winning horse, the favorite, as measured by the betting odds, won 29.4% of the time. Both the track-odds and the handicappers improved over the null that all horses had an equal probability of winning, but combining the handicappers' selections with the market odds did not significantly improve forecasting accuracy.[28] In Britain, the odds in the handicapper's morning line were also less accurate in predicting the probability of winning than were the final market odds (Crafts, 1985).

The experts also displayed the favorite-longshot bias. Snyder (1978) found that the favorite-long shot bias of official race-track handicappers and newspaper forecasters was greater than that of the general public. Lo (1994) showed that the favorite-longshot bias associated with the handicappers' morning-line odds was even larger than that of the final odds of the betting market.

B. Baseball

Despite all the predictions that are made every year by experts about the relative expected performance of the Major League teams, we found only one study that examined the quality of those forecasts. Smyth and Smyth (1994) found that the experts' forecasts were better than random guesses. The predictions of those experts, however, were not significantly different from those based on the rankings of teams' payrolls.

[28] Bird and McCrae (1987) found that the odds of Australian bookmakers, who could be considered experts, were fully incorporated into the racetrack odds.

C. Football

We found information about two types of experts: bookmakers and sports commentators (analysts). [29] In both NFL and college football games the bookmakers set an opening line that is biased. In both cases, the opening line contains sentimental beliefs of some bettors. The final line is more accurate indicating that some of the inefficiencies are eliminated (Avery and Chevalier, 1999; Dare et al., 2005).

Song et al. (2007) undertook a comprehensive analysis of the other group of experts, the sports commentators, and examined their ability to predict either the outcome of NFL football games or the margin by which a team was expected to win. The Song et al. study was based on the forecasts of 48 experts who predicted which team would win and an overlapping (but not identical) set of 52 forecasters who made selections against the betting line. All told, the forecasts of 70 experts were analyzed. Based on this sample of nearly 18,000 forecasts for the 2000 and 2001 seasons, Song et al. (2007) concluded that experts predicted the game winner approximately 62% of the time; this was the same accuracy ratio as the statistical systems achieved, but was less than the betting market's 66%. Similarly, the accuracy ratios of both the experts and systems in forecasting against the betting line were 50%. On average, the experts did worse than if they had used the naive model of flipping a coin. Avery and Chevalier (1999) reported a similar result.

Less comprehensive studies yield similar findings. Boulier and Stekler (2003) report that 60% of the time the sports-editor of the New York Times selected the winner of the games played during the 1994-2000 seasons. Even earlier, Pankoff (1968) showed that experts'

[29] For information about the process of setting the opening line (spread) on football games, see Schoenfeld, (2003).

accuracy in forecasting whether a team would beat the spread ranged from 48% to 56%.

D. Basketball

While there are no studies that have directly examined the forecasts of experts in predicting the outcomes of basketball games, there is one piece of indirect evidence. The bookmakers who set the opening line or point spread can be considered experts. The evidence is that the opening line that is established by the bookmakers is somewhat less accurate than the closing line established by the betting market (Gandar, et al., 1998). [30] This indicates that experts are not as accurate as the market in forecasting the winning margins. This result, however, does not imply that the experts exhibit a bias, because the changes between the opening and closing lines seem to be normally distributed around zero, the point of no change (Gandar et al., 1998, Table IV, p. 395).

E. Soccer

We have data that evaluates the forecasts of two types of experts. The first is the group of tipsters who write for newspapers;[31] the second consists of the bookmakers who provide the fixed odds. The evidence is that the tipsters' forecasts have little value and that they do not process public information properly (Pope and Peel, 1989; Forrest and Simmons, 2000).

In contrast, Forrest et al. (2005) demonstrate that there is virtually no difference between the accuracy of the forecasts of the odds makers and those obtained from a complex statistical model. This result is consistent with previous results because Kuypers (2000, Table 2, p.1359) had shown that the bookmakers' odds, when converted into probabilities are closely related to

---

[30] While the results are significantly different, the differences are too small to be economically meaningful.

[31] Andersson et al. (2005) evaluated the predictions of the outcome of the 2002 World Cup matches made by individuals who had familiarity with soccer. The participants were called experts but in reality they were not "real" experts. In any event, their predictions were no better than those that could have occurred by chance.

the relative frequencies of the outcome of the events. (Also see Goddard and Asimakopoulos, 2004).

F. Summary

There are many types of experts and the extent of their knowledge differs among the various groups. Experts who have a financial stake in the forecast are likely to be more knowledgeable. There is no evidence that experts consistently outperform the betting market and in football their accuracy is about the same as that of statistical systems. It should, however, be noted that models that tried to explain the behavior of bookmakers, who can be considered to be experts, showed that these individuals had not omitted important information in setting odds. (See Graham and Stott, 2008).

V. **Applicability of These Results to Forecasting in General**

The results relating to the various sports are so similar that the conclusions have to be considered robust. Some of these results are in accord with the generally accepted views of the forecasting profession; others are in conflict with those beliefs or require further research.

A. The findings that agree with our a priori views:

1. Forecasters correctly used information to reduce the biases that they observed. In horse racing, more information reduced the favorite-longshot bias; the final odds in horse racing were less biased than those of the racetrack's handicapper; in basketball and NFL games the closing spread was closer to the margin of victory than was the opening quote.

2. Forecasters are overconfident in their ability to predict.

3. Many forecasters have a misunderstanding of random processes as evidenced by their belief in the hot hand.

4. Combining forecasts does improve accuracy.

B. Findings that conflict with our generally accepted views:

1. Our analysis of these sports forecasts seriously conflicts with the widely held belief that the predictions derived from statistical methods are more accurate than those of experts. The analysis of 31000 NFL forecasts by Song et al. (2007) showed that the accuracy of the two methods of forecasting was virtually identical. The accuracy of the statistical methods was, however, less variable.

2. Similarly in soccer, there was no difference between the accuracy of the models' and bookmakers' forecasts.

C. Further Research Required

An area that requires further research concerns the relative weight that forecasters place on new and old information. There is a gambler's fallacy that the next outcome, even though it is independent of previous events, depends on events that have previously occurred. This fallacy has been observed in horse-racing studies and the hot hand belief in basketball. This is akin to placing too much weight on new information (Vaughan Williams, 1999, pp. 15-16). The majority of the evidence indicates that forecasters overreact to new information rather than anchor on the old forecast and adjust it in the face of the new data. Sauer (1998, p. 2059), however, reports on situations where recent information is given too little weight relative to what is optimal.

D. MOST IMPORTANT RESULT

There is no evidence that either statistical systems or experts *consistently* outperform the market. Kaplan and Garstka and Harville have been the only authors who have found that forecasts obtained from the market (basketball in this case) were not more accurate than those obtained from either experts or statistical systems.[32] This finding not only agrees with the theory about economic efficiency but also with the evidence that the market price is the best predictor of the event because the market aggregates all the information that is relevant to the event (Wolfers and Zitzewitz, 2004).

Biograhical Information

H.O. Stekler is a Research Professor in the Economics Department of George Washington University. His research interests involve all aspects of forecasting with an emphasis on forecast evaluations.

David Sendor graduated from George Washington University with a B.S. in economics. He has been doing research work at the consulting firm, BatesWhite in Washington DC.

Lincoln Verlander graduated from GWU with a dual B.S. degree in Economics and Statistics. He then worked for two years at the Federal Reserve Board of Governors as a Senior Research Assistant in the Capital Markets section. He currently works for Digital Sandbox Inc. as a Risk Engineer.

---

[32] Harville also found that there was no significant difference between the market and statistical systems in the football bowl games played after the 2001 regular season.

Bibliography

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

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

Barry, D. and J. A. Hartigan, (1993), Choice models for predicting divisional winners in Major League Baseball, *Journal of the American Statistical Association*, 88, 766-774.

Bennett, J. M. and J. A. Flueck, (1983), An evaluation of Major League Baseball Offensive performance models, *The American Statistician*, 37, 76-82.

Bentner, W., (1994), Computer based horse race handicapping and wagering systems: A report, in D. B. Hausch, V. S. Y. and W. T. Ziemba, eds. Efficiency of Racetrack Betting Markets, (Academic Press, New York, 1994), 183-198.

Berri, D. J., (1999), Who is 'most valuable'? Measuring the player's production of wins in the National Basketball Association, *Managerial and Decision Economics*, 20, 411-427.

Bird, R. and M. McCrae, (1987), Tests of the efficiency of racetrack betting using bookmaker odds, *Management Science*, 33, 1552-1562.

Bolton, R. N. and R. G. Chapman, (1986), Searching for positive returns at the track: A multinomial logit model for handicapping horse races, *Management Science*, 32, 1040-1060.

Boulier, B. L., and H. O. Stekler, (1999), Are sports seedings good predictors?: An evaluation, *International Journal of Forecasting*, 15, 83-91.

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

Brown W. O. and R. D. Sauer, (1993a), Fundamentals or noise? Evidence from the professional basketball betting market, *Journal of Finance*, 48, 1193-1209.

Brown W. O. and R. D. Sauer, (1993b), Does the basketball market believe in the hot hand? Comment, *American Economic Review*, 83, 1377-1386.

Bukiet, B., E. R. Harold, and J. L. Palacios, (1997), A Markov chain apporach to baseball, *Operations Research*, 45, 14-23.

Busche, K. and C. D. Hall, (1988), An exception to the risk preference anomaly, *Journal of Business*, 61, 337-346.

Cain, M., D. Law, and D. Peel, (2000), The favourite-longshot bias and market efficiency in UK football betting, *Scottish Journal of Political Economy*, 47, 25-36.

Camerer, C. F., (1989), Does the basketball market believe in the hot hand? *American Economic Review*, 79, 1257-1261.

Carlin, B. P., (1996), Improved NCAA basketball tournament modeling via point spread and team strength information, *The American Statistician*, 50, 39-43.

Carmichael, F., D. Thomas and R. Ward, (2000), Team performance: The case of English Premiership football, *Managerial and Decision Economics*, 21, 31-45.

Caudill, S. B., (2003), Predicting discrete outcomes with the maximum score estimator: The case of the NCAA men's basketball tournament, *International Journal of Forecasting*, 19, 313-317.

Caudill, S. B. and N. H. Godwin, (2002), Heterogeneous skewness in binary choice models: Predicting outcomes in the men' NCAA basketball tournament, *Journal of Applied Statistics*, 29, 991-1001.

Chapman, R. G., (1994), Still searching for positive returns at the track: empirical results from 2000 Hong Kong races, in D. B. Hausch, V. S. Y. and W. T. Ziemba, eds. Efficiency of Racetrack Betting Markets, (Academic Press, New York, 1994), 173-181.

Crafts, N. F. R., (1985), Some evidence of insider knowledge in horse race betting in Britain, *Economica*, 52, 295-304.

Crowder, M., M. Dixon, A. Ledford, and M. Robinso, (2002), Dynamic modelling and prediction of English Football League matches for betting, *The Statistician*, 51, 157-168.

Dana J. D. and M.M. Knetter, (1994), Learning and efficiency in a gambling market, *Management Science*, 40, 1317-1328.

Dare, W. H. and S. S. MacDonald, (1996), A generalized model for testing the home and favorite team advantage in point spread markets, *Journal of Financial Economics*, 40, 295-318.

Dare, W. H. and A. S. Holland, (2004), Efficiency in the NFL betting market: Modifying and consolidating research methods, *Applied Economics*, 36, 9-15.

Dare, W. H., J. M. Gandar, R.A. Zuber and R. M. Pavlik, (2005), In search of the source of informed trader information in the college football betting market, *Applied Financial Economics*, 15, 143-152.

Deschamps, B. and O. Gergaud, (2007), Efficiency in Betting Markets, Evidence from English football, *Journal of Prediction Markets*, 1, 61-73.

Dixon, M. J. and S.G. Coles, (1997), Modelling association football scores and inefficiencies in the football betting market, *Applied Statistics*, 46, 265-280.

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

Figlewski, S., (1979), Subjective information and market efficiency in a betting market, *Journal of Political Economy* , 87, 75-88, reprinted in D. B. Hausch, V. S. Y. and W. T. Ziemba, eds. Efficiency of Racetrack Betting Markets, (Academic Press, New York, 1994), 285-298.

Forrest D. K. and R. Simmons, (2000), Forecasting Sport: The behaviour and performance of football tipsters, *International Journal of Forecasting*, 16, 337-331.

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

Forrest, D., R. Simmons, and B. Buraimo, (2005a), Outcome uncertainty and the couch potato audience, *Scottish Journal of Political Economy*, 52, 641-661.

Forrest, D. and I. McHale, (2007), Anyone for tennis (betting)?, *European Journal of Finance*, 13, 751-768.

Gandar, J. M., W. H. Dare, C. R. Brown, and R. A. Zuber, (1998), Informed traders and price variations in the betting market for professional basketball games, *Journal of Finance*, 53, 385-401.

Gandar, J. M., R. A. Zuber, and W. H. Dare, (2000), The search for informed traders in the totals betting market for National Basketball Association games, *Journal of Sports Economics*, 1, 177-186.

Gandar, J. M., R. A. Zuber, R. S. Johnson, and W. Dare, (2002), Re-examining the betting market on Major League Baseball games: Is there a reverse favourite-longshot bias?, *Applied Economics*, 34, 1309-1317.

Glickman, M. E. and H. S. Stern, (1998), A state space model for National Football League games, *Journal of the American Statistical Association*, 93, 25-35.

Goddard, J., (2005), Regression models for forecasting goals and match results in association football, *International Journal of Forecasting*, 21, 331-340.

Goddard, J. and I. Asimakopoulos, (2004), Forecasting football results and the efficiency of fixed-odds betting, *Journal of Forecasting*, 23, 51-66.

Golec, J. and M. Tamarkin, (1991), The degree of inefficiency in the football betting market, *Journal of Financial Economics*, 30, 311-323.

Graham, I. and H. Stott, (2008), Predicting bookmaker odds and efficiency for UK football, *Applied Economics*, 40, 99-109.

Gramm, M. and D. H. Owens, (2005), Determinants of betting market efficiency, *Applied Economics Letters*, 12, 181-185.

Gray, P. K. and S. F. Gray, (1997), Testing market efficiency: Evidence from the NFL sports betting market, *Journal of Finance*, 52, 1725-1737.

Hadley, L., M. Poitras, J. Ruggiero, and S. Knowles, (2000), Performance evaluation of National Football League, teams, *Managerial and Decision Economics*, 21, 63-70.

Harville, D., (1980), Predictions for National Football League games via linear-model methodology, *Journal of the American Statistical Association*, 75, 516-524.

Harville, D., (2003), The selection or seeding of college basketball or football teams for postseason competition, *Journal of the American Statistical Association*, 98, 17-27.

Harville, D. and M. H. Smith, (1994), The home-court advantage: How large is it and does it vary from team to team? *The American Statistician*, 48, 22-28.

Horowitz, I., (1994), Pythagoras, Tommy Lasorda, and me: On evaluating baseball managers, *Social Science Quarterly*, 75, 187-194.

Horowitz, I., (1997), Pythagoras's petulant persecutors, *Managerial and Decision Economics*, 18, 343-344.

Kaplan, E. H. and S. J. Garstka, (2001), March madness and the office pool, *Management Science*, 47, 369-382.

Kahnemann, D. and A. Tversky, (1982), Intuitive prediction: biases and corrective procedures, in Judgment under uncertainty: Heuristics and biases, (Eds.) D. Kahnemann, P. Slovic, and A. Tversky, Cambridge University Press, London.

Kuypers, T., (2000), Information and efficiency: An empirical study of a fixed odds betting market, *Applied Economics*, 32, 1353-1363.

Lo, V. S. Y., (1994), Application of logit models in racetrack data, in D. B. Hausch, V. S. Y. and W. T. Ziemba, eds. Efficiency of Racetrack Betting Markets, (Academic Press, New York, 1994), 307-???

Maher, M. J., (1982), Modelling association football scores, *Statist. Neerland*, 36-109-118.

Malios, W. S., (2000), *The analysis of sports forecasting: Modeling parallels between sports gambling and financial markets*, (Kluwer Academic Publishers, Boston)

Oorlog, D. R., (1995), Serial correlation in the wagering market for professional basketball, *Quarterly Journal of Business and Economics*, 34, 96- 109.

Pankoff, L. D., (1968), Market efficiency and football betting, *Journal of Business*, 41, 203-214.

Paul, R. and A. Weinbach, (2005a), Market efficiency and NCAA college basketball gambling, *Journal of Economics and Finance*, 29, 403-408.

Paul, R. and A. Weinbach, (2005b), Bettor misperceptions in the NBA: The overbetting of large favorites and the "hot hand", *Journal of Sports Economics*, 6, 390-400.

Pope, P. F. and D. A. Peel, (1989), Information. Prices and efficiency in a fixed-odds betting market, *Economica*, 56, 323-341.

Porter, P. K. and G. W. Scully, (1982), Measuring managerial efficiency: The case of baseball, *Southern Economic Journal*, 48, 642-650,

Quandt,R. E., (1986), Betting and equilibrium, *Quarterly Journal of Economics*,101, 201-207.

Rosner, B., F. Mosteller, and C. Youtz, (1996), Modeling pitcher performance and the distribution of runs per inning in Major League Baseball, *The American Statistician*, 50, 352-360.

Rue, H. and O. Salvesen, (2000), Prediction and retrospective analysis of soccer matches in a league, *The Statistician*, 49, 399-418.

Ruggiero, J., L. Hadley, G. Ruggiero, and S. Knowles, (1997), A note on the Pythagorean theorem of baseball production, *Managerial and Decision Economics*, 18, 335-342.

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

Sauer, R. D., (2005), The state of research on markets for sports betting and suggested future directions, *Journal of Economics and Finance*, 29, 416-426.

Sauer, R. D., V. Brajer, S. P Ferris, and M. W. Marr, (1988), Hold your bets: Another look at the efficiency of the market for National Football League games, *Journal of Political Economy*, 96, 206-213.

Shmanske, S., (2005), Odds-setting efficiency in gambling markets: Evidence from the PGA Tour, *Journal of Economics and Finance*, 29, 391-402.

Schoenfeld, B., (2003), What's my line?, *Sports Illustrated*, December 8, 76- 89.

Schwertman, N. C., T. A. McCready and L. Howard, (1991), Probability models for the NCAA regional basketball tournaments, *The American Statistician*, 45, 35-38.

Schwertman, N. C., K. L. Schenk and B. C. Holbrook, (1996), More probability models for the NCAA regional basketball tournaments, *The American Statistician*, 50, 34-38.

Smyth, D. J. and S. J. Smyth, (1994), Major League Baseball division standings, sports journalists' predictions and player salaries, *Managerial and Decision Economics*, 15, 421-429.

Snyder, W. W., (1978), Horse racing: Testing the efficient markets model, *Journal of Finance*, 33, 1009-1118.

Song, C. U., B. L. Boulier, and H. O. Stekler, (2007), Comparative accuracy of judgmental and model forecasts of American football games, *International Journal of Forecasting*, 23, 405-413.

Szymanski, S., (2003), The economic design of sporting events, *Journal of Economic Literature*, 41, 1137-1187.

Turocy, T. L., Offensive performance, omitted variables, and the value of speed in baseball, *Economics Letters*, 89, 283-286.

Vaughan Williams, L., (1999), Information efficiency in betting markets: A survey, *Bulletin of Economic Research*, 51, 1-30.

Vaughan Williams, L., (2005), *Information efficiency in financial and betting markets*, (Cambridge: Cambridge University Press).

Vaughan Williams, L. and D. Paton, (1998), Why are some favourite- longshot biases positive and others negative?, *Applied Economics*, 30, 1505-1510.

Vergin, R. C., (2001), Overreaction in the NFL point spread market, *Applied Financial Economics*, 11, 497-509.

Vergin, R. C. and M. Scriabin, (1978), Winning strategies for wagering on National Football League games, *Management Science*, 24, 809-818.

Wodland, B. M. and L. M. Woodland, (1999), Expected utility, skewness, and the baseball betting market, *Applied Economics*, 31, 337-345.

Wodland, L. M. and B. M. Woodland, (1994), Market efficiency and the favorite-longshot bias: The baseball betting market, *Journal of Finance*, 49, 269-279.

Wolfers, J. and E. Zitzewitz, (2004), Prediction markets, *Journal of Economic Perspectives*, 18, 107-126.

Zak, T. A., C. J. Huang, and J. J. Siegfried, (1979), Production efficiency: The case of professional basketball, *Journal of Business*, 52, 379-392.

Zuber, R.A., J. M. Gandar, B. D. Bowers, (1985), Beating the spread: Testing the efficiency of the gambling market for National Football League games, *Journal of Political Economy*, 93, 800-806.

Appendix- Characteristics of the Models

A. Horse Racing

Bolton and Chapman (1986) construct a multinomial logit model of horse racing that includes characteristics of both the horse and the jockey.[33] While the final equation includes many variables that are not statistically significant, one characteristic, the speed of the horse contributes the most to explaining the variance of the horse races. Expanded versions of this multinomial logit model improve the explanatory power of the equations (Bentner, 1994; Chapman, 1994).

B. Baseball

Many of the basic models of a baseball game consider either the characteristics of the offense to determine the number of runs scored or the qualities of the pitching staff in permitting runs to be scored. Thus Porter and Scully (1982) estimate a production function based on a team's slugging average and its strike-out-to-walk pitching ratio.[34] Other models provided more detail about baseball's offense and pitching; for example Bennett and Flueck (1983) examined various characteristics of offense to determine the number of runs that would be scored but did not make explicit predictions with their model.[35]

---

[33]The characteristics of the horse include the percentage of races won, winnings per race, an index measuring speed, weight and post position. The jockey characteristics include the number of races won and his winning percentage.

[34]This model was not used to make predictions but rather was employed to measure the relative performance of baseball managers. Horowitz (1994) uses a power score variant of this production function (runs scored/ runs allowed) in a similar analysis of managerial performance. See Ruggiero et al. (1997) and Horowitz (1997) for a further discussion of this subject. Hadley et al. (2000) used a similar approach to evaluate football coaches.

[35]These variables included the various types of hits, walks, types of outs, etc.

Rosner et al. (1996) estimated relationships that measured pitcher performance and determined the number of runs that would be scored in each inning. They were able to do this because play-by-play data have been available for all Major League Baseball games since 1984. An adjusted negative binomial distribution was fit to the data to explain the number of batters that a particular pitcher would face. The number of runs that will be scored is a complex function of two distributions: the negative binomial distribution and a conditional distribution of the number of runs that score given that the pitcher has faced a specified number of batters. (Rosner et al. 1996, p.352.) Other studies include Malios (2000) who listed the factors that determined offensive and pitching performance, and Turocy (2005) who added a speed variable to the conventional production function.

Barry and Hartigan (1993) used a binary-choice model to calculate the probability that in 1991 a National League team would win its division. The model was based on the strength of the teams as the season progressed with greater weight placed on the most recent sequence of games as well as the teams' home-field advantage.

C. Football

The many statistical systems that are designed to predict the margin of victory of NFL or college football games are estimated in different ways. Zuber et al., (1985) and Sauer et al., (1988) derived models using the fundamental characteristics of a team's offense and defense to explain the margin of victory. Dana and Knetter (1994) developed a dynamic model using point score indices as a measure of the abilities of the teams while Glickman and Stern (1998), developed a state space model based on Bayesian techniques for predicting NFL scores.

Other models are based on power scores that summarize the latent abilities of the football teams. Variants of these power scores have been used in forecasting both the outcomes of football games and the margins of victory. For example, see Harville (1980) and Boulier and Stekler (2003).

D. Basketball

Zak et al. (1979) developed a production function that represented the defensive and offensive elements of a basketball game. The model was designed to measure the relative contribution of each of those elements to the winning margin. Each team's productive efficiency was then calculated. Berri (1999) used a similar model that was designed to measure the contribution of individual players to a team's wins. Rather than directly predicting a team's wins, each player's contribution towards his team's wins was summed.

Other modeling approaches did not construct production functions but rather used proxy variables that measured the latent skills or strengths of each team. The margin of victory in a contest between two teams was considered a measure of the comparative strengths of the two (Brown and Sauer, 1993; Harville and Smith, 1994; Oorlog, 1995; Kaplan and Garstka, 2001; Harville, 2003). A variant of this approach is a statistical scoring system. As an example, Sagarin's system can be used to predict the expected scoring by any two teams. This system is based on the number of victories of each team, the strength of the teams that were defeated, the margin of victory adjusted for blowouts, and an adjustment for the home court advantage.

Alternatively, in a tournament, the seedings of the teams, which are obtained from a statistical scoring system, can be used as a predictor. Since 1985, the NCAA has selected 64 college basketball teams to participate in a tournament to select a field to compete for the men's national championship. The 64 teams are divided into four regional tournaments of 16 teams that are ranked from 1 through 16.[36] Boulier and Stekler (1999), Caudill and Godwin (2002), Kaplan and Garstka (2001), Caudill (2003), and Harville (2003) all base their analyses on the differences in the ranks of the teams, but their statistical models differed.

Boulier and Stekler (1999) used a probit model based solely on the difference in ranks. Caudill and Godwin (2002) developed a heterogeneous skewness model that takes into account not only the difference in ranks but also the level of the seed. Thus the probability that a Number 1 seed beats a Number 5 seed is greater than the probability that a Number 5 seed beats a Number 9-ranked team. Finally, Harville (2003) constructed a modified-least-squares ranking procedure based by placing a limit on the margin of victory. He then compared the forecasts of the model solely with the difference in ranks (Also see Schwertman et al., 1991, 1996; and Carlin, 1996).

E. Soccer

The modeling has been done at three levels. In the production function approach variables that are associated with attack and defense are embodied in the model. A second approach is to model each team's goal scoring abilities and then predict which team will win based on the difference in the predicted number of goals. Finally, discrete-choice models based on past performance are used to directly predict the probabilities of the home team winning, drawing, or losing.

In the production function approach, Carmichael et al. (2004) estimated the effects of

[36] The seeds are determined from a statistical scoring system, the RPI, called the ratings percentage index. It gives weights of .25, .50, and .25 to the team's winning percentage, the winning percentage of its opponents, and winning percentage of the opponents' opponents, respectively.

specific types of plays on the difference in the number of goals scored by the two teams. The Poisson distribution is an alternative model for predicting the number of goals that teams will score. Dixon and Coles (1997) show that this distribution provides a good fit to the score data for the 1992-95 seasons, but they also add attack and defense parameters to the basic model of Maher (1982). Moreover, they permit the parameters to vary to reflect changes in team strength that may have occurred over time. The model of Cain et al. (2000) differed from that of Dixon-Coles for two reasons: it used the negative binomial distribution to model the number of goals scored, and the independent variables were the win-lose odds prices quoted by the bookmakers rather than proxy attack-defense variables.[37]

Since the abilities and performance of teams can change over time, some models have become dynamic to capture these effects. Dixon and Coles were among the first to incorporate dynamic factors into a model. Crowder et al. (2002) derive an approximation to the Dixon-Coles model and show that the two models yield similar results. The Bayesian dynamic model of Rue and Salveson (2000) yielded likelihood measures.

The discrete choice models were based on ordered probits that included a variety of explanatory variables. Kuypers (2000) included the bookmakers' odds as well as some performance variables in his model. The win ratios of the two teams playing the match were included as independent variables in the model of Goddard and Asimakopoulos (2004).

[37]The Poisson distribution is a limiting case of the negative binomial.

H. O. Stekler is a research professor at George Washington University and an associate editor of the *International Journal of Forecasting*. He has written extensively on topics related to macroeconomic forecasting, with an emphasis on forecast evaluation.

David Sendor graduated from George Washington University as an economics major. He is currently working at Bates White in Washington DC.

Richard Verlander graduated from George Washington University as an economics major. He is currently working at the Board of Governors of the Federal Reserve System in Washington DC.