SPORTS FORECASTING

There have been an enormous number of studies involving various aspects of sports. We will concentrate only on the economic aspects. For example, Econ Lit has over 3700 entries while JSTOR, which also contains non-economic articles, contains over 40,000. The data associated with sporting events have been used to examine a number of topics and test a number of hypotheses relating to economic and financial behavior under uncertainty. These include the efficiency of betting markets, the use of information in responding to betting odds, 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.

There are many reasons for the great interest in this subject. Expenditures associated with attending sporting events are substantial. Many individuals have a great interest in sports and are devoted followers of "their" teams' performance. They may even gamble on the outcomes of sporting events. It is, thus, not surprising that a substantial number of the economic studies have been concerned with the efficiency of the betting market and have sought to determine whether there were any betting strategies that could "beat the market", i.e be profitable . (See Sauer (1998) and Vaughan Williams (1999) for surveys of the efficient market literature). These betting markets are similar to financial markets and it is possible to test behavioral hypotheses that have applicability to the broader financial markets. If there were any inefficiencies, even if they were not profitable, it would be possible to determine the biases that produced those

inefficiencies. Moreover, since a sporting event has a definite outcome at a specific point in time, it is not necessary to make assumptions about expectations of the future as is necessary in other asset markets. Even if there are no inefficiencies, the betting data that are available from these markets would permit one to extract information about the betting market process and the way it attains efficiency. (Sauer, 2005).

Finally, the huge number of observations, drawn from the real world, makes it possible to test various hypotheses and to obtain meaningful and valid results. It is not necessary to base findings on laboratory experiments with a small number of observations that may not replicate the conditions of the real world. For example, one study 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 football games.

Just as economists and financial analysts have used sports data to test hypotheses about economic and financial behavior, the forecasting profession can benefit from examining the findings derived from sports forecasts and applying them to our own specialized fields. Given the great interest in gambling on sporting events, it is not surprising that most of the empirical information about sports forecasting comes from studies that have examined these gambling markets. While most papers were not primarily concerned with forecasting, they provide crucial insights about the issues involved in making predictions.

The aim of this paper is to provide a sport-by-sport survey of (1) the type of forecast that is made, (2) whether the forecast involved picking the winner or the margin of victory, (3) the methodologies that were used in forecasting the outcomes in that sport, (4) comparisons of the results of different forecasting methodologies, and (5) the types of prediction biases that have

been observed. The concluding sections make a cross sport comparison to determine whether the results yield valid generalizations and discuss the applicability of the findings to other areas of forecasting.

I. The Sports Gambling Market

Given that the data are generally associated with and obtained from the gambling market, it is first necessary to discuss how the bets are structured. The way that gambling markets are constituted differs from sport to sport. In horse racing, baseball and soccer the market quotes odds that a particular horse or team will win. A winning bet will be paid according to those odds. In American football and in 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 than the specified margin (point spread) that is set in the market.[1]

The procedures for evaluating sports forecasts thus depends on the institutional betting arrangements. If odds are quoted and there are more than two competitors, it is not possible to determine whether forecasters correctly predicted the winners. Rather the analysis is based on a comparison of the ex ante probabilities and the ex post ratio of outcomes. The betting odds must, therefore, be converted into probabilities by the formula, $p = 1/(1+odds)$.[2] For each ex ante probability, it is possible to calculate the ex post percentage of wins for horses (teams). The ex ante probabilities and the ex post winning percentage should be calibrated, i.e. horses whose ex

---

[1]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.

[2]If the sum of the betting odds exceeds one because of the bookmakers' commissions or the parimutuel take, they must be adjusted so that the sum will equal one.

ante probability of winning was 0.30 should have won 30 percent of the time.

The evaluation procedure is different in those markets where bets are placed on the margin of victory. In those cases, it is possible to evaluate the accuracy of two types of forecasts. How accurate were the forecasts in (a) selecting the winning team and (b) in predicting whether the favored team beat the spread?

II. Types of forecasts

The forecasts that we discuss come from three sources. First, there is the market forecast itself. Experts, be they bookmakers, handicappers or sports commentators, also issue forecasts about the likely outcomes of sporting events. Finally, 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.

A. Market forecasts

For each sport, the largest number of forecasts come from the betting market. The market forecast is either the final odds that a team will win or the point spread (the expected margin by which a team is expected to win) of a particular sporting event. A sample of these forecasts can be analyzed in several different ways. How accurate was the market in predicting the winners? Are the subjective odds of the market calibrated with the objective winning percentages? Are there any biases in the forecasts? Is the market forecast more accurate than other forecasting methods?

B. Models

In order to predict the outcome of sporting events, models can be constructed at different levels. If the data are available, it is possible to begin by modeling every play. (For baseball see

4

Bukiet et al.,1997 and Sauer, 2005; for soccer see Carmichael et al., 2000). Alternatively, production functions, that explain the difference in fundamental factors such as the offensive (points or runs scored) and defensive (points or runs allowed) characteristics of teams, have been estimated.

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 differences 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 of victories of the competing teams and the time trend in this relationship. For example, the New York Times created power scores for every 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 the quality 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).

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 track handicappers should also be considered experts.

III. What has been forecast?

Since sports cover such a wide variety of activities, it is not surprising that the forecasting literature associated with this field covers a wide spectrum of topics. The outcomes of horse

races and of baseball, football, basketball, and soccer games have all been predicted. For every sport, the literature has provided forecasts of the outcomes of specific events or matches, i.e the winner of a horse race, tennis match, (baseball, football or basket ball) game, etc. In some of these cases, merely the winner is forecast; in other situations, the margin of victory is forecast. There are also forecasts about the winners of tournaments such as the NCAA basketball championships and the winners of the championship of particular leagues. Moreover, there have been theoretical analyses about tournaments within a particular sport and the probability that the best team (player) will win the tournament. In each case, we proceed by first examining the available forecasts and methodologies for each sport and discuss these ancillary topics when it is necessary.

A. Horse Racing

There have been many studies that examined the outcome of horse races. Sauer (1998) and Vaughan Williams (1999) have surveyed the major studies that analyzed these races. While the major emphasis was on the economic efficiency of the betting markets, these analyses provided insights that can have general applicability to all fields of forecasting. The observed inefficiencies provide information about the biases that exist. Moreover, the results suggest that it is even possible to model the outcome of horse races. These statistical models take into account the competition that occurs during a race.

1. Betting Market

The results indicate that the market can distinguish among horses of different quality. With a particular exception, the subjective probabilities obtained from the odds rank of the horses are well calibrated with the observed frequency of wins. (Sauer, 1998, pp. 2035 and

2044). This indicates the obvious presence of individuals who are informed forecasters, who, can predict the outcomes of horse races.

The exception to the aforementioned calibration occurs at the extremes of the odds distribution. Most studies of horse racing in the US yield a result that has been called " the favorite- long shot bias". This means that in the parimutuel market, an insufficient amount is bet upon the horses that are favored to win and an excessive amount is bet on the long shots.[3] This distorts the odds at the extremes.

This bias can be explained either by individuals' underestimates (overestimates) of favorites (long shots) or by bettors' utility functions that are locally risk seeking. (Quandt, 1986). The findings by Golec and Tamarkin (1995), however, favored the hypothesis that bettors were overconfident in their abilities to predict rather than being risk seekers. 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.[4]

Another explanation for this bias concerns the absence or presence of informed bettors and the quality of information that is available in the market. (Vaughan Williams and Paton, 1998). When more information is available publicly and the bettors are more informed, the more likely it is that the consensus forecast (represented by the market odds) will converge to the true odds. Their empirical evidence is consistent with this view, because the bias is diminished if

---

[3]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.

[4]In fact, Vaughan Williams (1999, p.8) cites one study that finds that bettors are overconfident.

either the betting pool or the number of horses in the race is increased. (Busche and Hall, 1988; Gramm and Owens, 2005).

2. Modeling

Bolton and Chapman (1986) construct a multinomial logit model of the horse race process. Their model includes characteristics of both the horse and the jockey.[5] 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 race process. The adjusted R^2 of the equation is .09 indicating that the equation explains 9% more than the null that each horse has an equal chance of winning.

Bentner (1994) and Chapman (1994) construct expanded versions of this multinomial logit model and improve the explanatory power of their equations. Most of the new variables are significant and the adjusted R^2 exceeds 12%. While the final betting odds have even more explanatory power, a combination of the model and the market odds improves upon both.[6] This finding is consistent with the results obtained from the non-sports forecasting literature which indicates that combining forecasts usually improves accuracy.

3.Experts

Figlewski (1979) examined the forecasting record of a number of individuals who

---

[5]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.

[6]However, in a real time situation an individual might not have the time and thus be in a position to use the final odds in combination with the model before placing a bet.

handicapped horse races. While the handicappers were successful 28.7% of the time in selecting the horse that would win the race, 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. In Britain, the odds in the handicapper's morning line were less accurate in predicting the probability of winning than were the final odds in the betting market. (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, in his sample, the favorite-longshot bias associated with the handicappers' morning line odds was even larger than that of the final odds of the betting market.

4. **Summary**

a. The market odds and the frequency of wins are calibrated except for th favorite-longshot bias. However, this bias is not observed universally.

b. Models can explain some of the variance of horse races. Combining models with market odds improves accuracy.

c. The odds provided by experts are better than those obtained by chance but not as accurate as the betting market odds.

d. Experts displayed even more favorite-longshot bias than the final market odds.

B. Baseball

There is so much information and so many statistics about baseball, that it is surprising

how few forecasts are available for analysis. There are a number of models that have estimated

the importance of the offensive and defensive factors that determine the outcome of a game, but

forecasts from these models have not been published in the open literature.[7]

1. The betting market

Bets in this market are made on the outcome of a game. Consequently, like the horse race

betting market, the analysis is based on odds which can then be converted into probabilities.

Unlike horse racing, 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 the underdog team would win $140,

while someone betting on the favored team would bet $150 to win $1. (In both cases the winner

would also have the bet returned). The difference is the commission. From these odds it is

possible to calculate the betting market's subjective probability that the underdog will win. The

probability is calculated at the midpoint of the line, i.e. $1/(1.45 +1) = 0.41$. This subjective

probability can then be compared with the percentage of times that the underdog won when those

odds were quoted. If the subjective probabilities are calibrated with the observed probabilities,

the forecasts would be considered rational.

Three studies examine the relationship between these subjective and observed

probabilities. 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

monotonic.[8] In order to test whether the forecasts were rational, Woodland and Woodland

(1994) regressed the objective probabilities on the subjective odds. They obtained mixed results

---

[7]Individuals may have used these models in deciding whether to place bets about the
outcome of games, but these data are not readily available.

[8]The QPS statistic also known as the Brier score could have been calculated and
decomposed to determine the degree of calibration.

that were dependent on the method of estimation. They argued that betting in baseball yielded a reverse favorite-underdog bias, with underdogs underbet. Gandar et al. made a minor correction to the Woodland-Woodland methodology and found that rationality was not rejected, and if there were any bias, it was very slight.

2. Modeling

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. This model was not used as a predictor 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.[9]

While other models provided more detail about baseball's offense and pitching, many have not yet provided forecasts that could be evaluated. Bennett and Flueck (1983) examined various characteristics of offense[10] to determine the number of runs that would be scored but did not make explicit predictions with their model. They fit the models to data from the 1969-1976 seasons and then evaluated the models by sequentially eliminating one year's data from the sample and reestimating the equations using the data of the remaining seven years. The adjusted correlation coefficients of those eight regressions did not differ significantly, but the coefficients of some of the variables did display considerable variation.

---

[9] Also see Ruggiero et al. (1997) and Horowitz (1997) for a further discussion of this subject.
    [10]These variables included the various types of hits, walks, types of outs, etc.

Similarly, 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 that have occurred since 1984. An adjusted negative binomial distribution is fit to the data and explains the number of batters that a particular pitcher faces. The number of runs that will be scored is a complex function of two distributions: this 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. None of these models produced forecasts that could be evaluated.

On the other hand, Bukiet et al. (1997) modeled each at-bat as a 25x25 transition 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. This model was used to predict the actual number of games that all of the teams in the National League would win in 1989. While the model failed to predict one of the two divisional winners and the number of runs that were scored was underestimated, the Spearman Rank Correlation between the predicted and actual number of games that each team won was .77. (Authors' calculations).

There are two other models that were used to predict National League 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. 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. 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.

3. Experts and Experts vs. Models

Despite all the predictions that are made every year by experts about the relative 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. However, the predictions of those experts were not significantly different from forecasts that were based on rankings based on models that uses payrolls as the independent variable.

4. **Summary**

a. The market odds are calibrated with the observed ratios of outcomes, but there is some debate about the possibility of a reverse favorite-longshot bias.

b. Many models that explain various aspects of a baseball game have been estimated, but they have not been used to make forecasts.

c. There is one study that examined the forecasts of experts. The experts' predictions were better than random, but not different from models that used payrolls as the predictor.

C. Football

Many studies have examined the accuracy of the forecasts about the outcomes of professional and college football games. These forecasts, like those from other sports, come

from the betting market, statistical systems, and experts.

Unlike horse racing and baseball where odds are used in establishing the payoff to a bet that involves selecting a winning horse or team, football bets do not involve selecting the winning team. Rather the bet is whether or not the favored team beats the underdog by a margin (number of points) specified in the bet. This margin is called the point spread. If an individual bets that the favored team will beat the underdog by more than this spread, the bettor only wins if the favorite is victorious by more than this number of points. Someone betting on the underdog can win if there is an upset and the favored team loses <u>or</u> if the favorite wins by less than the specified number of points. Every bettor pays $11 for a $10 payoff. The difference is the bookmaker's commission also known as the vigorish. Given this commission, a bettor must be right at least 52.4% of the time, just to break even. (Sauer, 1988, p. 211)

While the betting market is not concerned with selecting the winning team, it is possible to use the data about the spread to determine the accuracy of the market in picking the winning team. Thus, our analysis of the betting market involves two basic 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?

1. Betting Market

a. Winners

Since the focus of the previous analyses have been whether the betting market on NFL games was efficient, there have been very few studies that have considered the forecasting accuracy of the market in predicting the winners of games. Boulier and Stekler (2003) and Song et al. (2007) showed, that in every year from 1994-2001, the betting market correctly predicted

which NFL team would win at least 63% of the time. The 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.

b. Point spread

The overwhelming majority of the evidence indicates that the betting market is efficient in the sense that there is no profitable betting strategy against the spread. However, this result does not imply that the forecasts themselves do not exhibit some bias. The traditional method for determining whether a forecast is unbiased is to run the regression:

$$A = a + bF + e, \quad \text{where}$$

A is the actual value and F is the forecast. The joint null hypothesis that $a = 0$ and $b = 1$ is then tested.. If the null is rejected, the forecasts are biased.

In the football betting market, the equivalent equation is:

$$DP = a + bPS + e, \text{ where}$$

DP is the difference in the game score (actual points) and PS is the betting market point spread.[11] This equation indicates that if the forecast is unbiased, on average, the difference in the point scores will not differ significantly from the point spreads. Again, the joint null hypothesis that $a = 0$ and $b = 1$ is tested. Most studies that use this equation do not reject the null hypothesis, but the explanatory power of the equations is usually low indicating considerable unexplained variation.

On the other hand, Gray and Gray (1997) test the hypothesis that the forecasts are efficient in a different way. They estimate a probit where the dependent variable is whether the

---

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

team beat the spread or not.[12] 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.[13]

Since there is a considerable amount of unexplained variability, it is possible that there may be forecasting biases and betting strategies that may/may not be profitable. We are concerned only with the forecast biases. Vergin (2001) argued that bettors (forecasters) were subject to an overreaction bias, i.e. they overreact to the most recent 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).[14] EXPLAIN MORE All but one of the possible overreactions that Vergin tested turned out to be significant, indicating that the bettors (forecasters) over 15 seasons had displayed a bias in interpreting recent data. Similarly Gray and Gray (1997) also found that the market overreacted to the most recent information.[15]

There have been discussions in the forecasting literature about the way that individuals interpret information. Kahnemann and Tversky (1982) had argued that individuals place too much emphasis on new information. On the other hand, some experimental data suggest the opposite: people anchor on a past observations and place too little emphasis on new information. (Andreassen, 1987, 1990) The data from the football betting market seem to favor the former

---

[12] A probit places less weight on outliers than OLS does.
[13] However, there is a controversy about the appropriate way to jointly test for home team bias and underdog bias. (See Golec and Tamarkin, 1991; Dare and MacDonald, 1996; Dare and Holland, 2004).
[14]Vergin and Scriabin (1978) were also concerned with this issue.
[15]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.

view.

2.Models

There are a large number of statistical systems that are designed to predict the margin of victory of NFL or college football games.[16] However, we found only two published papers (Zuber et al., 1985; Sauer et al., 1988) that directly estimated various characteristics of a team's offense and defense that contributed to the margin of victory. Those papers were primarily concerned with testing the efficiency of the betting market: in 1983, the models correctly predicted a margin of victory that would have been profitable 59% of the time but the success rate was only 39% in 1984. The papers did not indicate the number of times that the model predicted the winners of each game, but we should note that the models explained 73-81% of the variance of the score differentials for the 1983 and 1984 NFL seasons.

Variants of power scores have also been used in forecasting both the outcomes of football games and the margins of victory. Harville (1980) derived his model from the past scores of teams and used an elaborate statistical procedure to predict both the winners of games and the probability that a given team would beat the betting spread. He found that in the 1971-77 seasons the betting market, with a 72% success rate in selecting the winner, was more accurate than the statistical procedure, which was right 70% of the time.[17] Moreover, the relative frequency of wins against the betting market was a monotonically increasing function of the conditional probability that the team would beat the spread, with the observed frequency

---

[16]Song et al. (2007) used information from 32 systems that provided information on the Internet.

[17]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.

exceeding .50 for all the probabilities.

Boulier and Stekler (2003) used the power scores published in the *New York Times* and forecast that the team with the higher score would win. These forecasts based on these scores have an accuracy ratio of 61%, less than that of the betting market which achieved 66% accuracy. Unfortunately, the forecasts based on the power scores were no more accurate than a naive forecast: 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 occurred by chance.[18] In forecasting against the betting spread, most systems, however, were not even as accurate as the naive forecast of flipping a coin.

3. Experts

Song et al. (2007) undertook the most comprehensive analysis of experts' ability to predict either the outcome of football games or the margin by which a team is expected to win. That study used 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. concluded that experts predicted the game winner approximately 62% of the time; this was the same accuracy ratio as the statistical systems

---

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

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 was 50%. On average the experts did worse than using the naive model of flipping a coin.

Less comprehensive studies report similar findings. Boulier and Stekler (2003) report that the sports editor of the New York Times selected the winner of the games during the 1994-2000 seasons 60% of the time. Even earlier, Pankoff (1968) showed that experts' accuracy in forecasting whether a team would beat the spread ranged from 48% to 56%.

4. **Summary**

a. The market correctly picks the winner of a game about 2/3 of the time. This record is better than that of the experts and systems.

b. The null that the point spread is an unbiased predictor of the margin of victory is generally not rejected.

c. Models are successful in predicting winners but they are not as accurate as the betting market.

d. Models and experts are equally good both in forecasting winners and in predicting against the spread. The predictions against the spread are not significantly better than chance.

D. Basketball

1. Betting Market

The studies that have analyzed 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).

For forecasters, the major issues of interest in the basketball betting market concern the absence or presence of the "hot hand" phenomenon and whether there are informed bettors in this market. The hot hand belief is that a team that makes a shot wins a game is more likely to win the next game. This belief 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) build a model based on teams' abilities and streak dummies in order to test this hypothesis. They conclude that the hot hand belief is embodied in the point spread and is, therefore, an important effect.[19] However, they were not able to determine whether, in fact, a hot hand phenomenon existed or whether it was a myth and bettors were misperceiving the real process and thus displaying a cognitive bias.

Using changes in the betting line, it is possible to infer the role that information and informed bettors play in the betting market. Brown and Sauer (1993) and Gandar et al (1998, 2000) examine these questions, although from different perspectives. Brown and Sauer first estimate a model based on a proxy for fundamentals: the points scored by the two teams that play against each other. This model explains 89% of the variation in the point spread and also predicts well out of sample. Brown and Sauer thus conclude that the betting market adjusts for fundamental changes in the relative team abilities that may have occurred from one season to another.

Gandar et al. examine the differences between the opening and closing point lines for

---

[19]Also see Paul and Weinbach (2005b).

NBA games.[20] They show that frequently there are large changes between the opening and closing quotations. These changes reflect betting sentiment that is different from that of the bookmaker. They test a number of hypotheses, show that the opening line is not as accurate as the closing line in forecasting the margin of victory, and conclude that informed bettors have eliminated some of the bias in the opening line.

2. Models

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. The rank of each team in terms its productive efficiency was identical to the rank based on winning percentage in the 1976 NBA season. This method has not been subsequently used for making forecasts.

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 were summed. The ranking obtained by summing the contributions of each player to team victories was remarkably close to the ranking based on the teams' won-lost records in the 1997- 1998 season. The Spearman Rank Correlation was .986 (Authors' calculations).

Other modeling approaches did not construct production functions but rather used proxy

---

[20]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. Thus it is not likely that much *new* information about the teams will have become available during the course of the day.

variables that presumably 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).

If two teams who have not played each other previously were to meet, it would be
impossible to measure the comparative strengths of those opponents. This is a particular
problem in trying to forecast the outcomes of college basketball (and football) games, because
no team plays every other team. Statistical scoring systems have been developed to overcome
this problem. 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.[21] Alternatively, in a tournament, the seedings,
which are obtained from a statistical scoring system, of the teams can be used as a predictor.
This topic will be discussed in the section on tournaments, below.

3. Experts

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.[22] (Gandar, et al., 1998). This indicates that

---

[21]The difference between two teams' Sagarin ratings is a good predictor of the margin of
victory. (Carlin, 1996).
[22] While the results are significantly different, the differences are too small to be
economically meaningful.

experts are not as accurate as the market is in forecasting the winning margins. However, this result 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 (no change). (Gandar et al., 1998, Table IV, p. 395).

4. **Summary**

a. The market does not have any observed biases; the market moves to eliminate the early biases of the bookmakers.

b. Many models of basketball games have been estimated, but except for games played in the NCAA tournaments there have been few forecasts.

c. The only evidence that we have about experts is that the bookmakers' opening line is less accurate than the final spread.

E. Soccer

Gambling in soccer is based on odds, but there is no market that can be analyzed. Rather the bookmakers fix the odds at the beginning of a week and never change them. The focus of the forecasting literature in this sport has, therefore, been on statistical procedures and the performance of these models relative to the bookmakers' odds.

1. Models

The modeling has been done at three levels. There is the production function approach where the 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 predict which team will win based on the difference in the predicted number of goals. Finally, discrete choice models based

23

on past performance are used to directly predict the probabilities of the home team winning, drawing or losing.

Carmichael et al. (2004) have been the only ones who have used the production function approach to predict the outcomes of soccer matches. They estimate the effects of specific types of plays to predict the difference in the goals scored by the two teams. Their equation was able to capture the relative performance of teams in the English Premier league, but they did not make any forecasts beyond the period of fit.

There are alternative models that predict the number of goals that teams will score, and they are not based on the production function approach. Rather, they use Poisson distributions to fit the data. (Maher, 1982). Dixon and Coles (1997) show that this distribution provides a good fit to the score data for the 1992-95 seasons. They improve upon the Maher model by adding attack and defense parameters. Moreover, they permit the parameters to vary over time to reflect changes in team strength that may have occurred over time. The probabilities obtained from their model are similar to those of the bookmakers' (as derived from the odds) but Dixon and Coles do not provide a detailed evaluation.[23] That evaluation was conducted by Dixon and Pope (2004), who showed that while the model probabilities were similar to those of the bookmakers, the model contained information that was not embodied in the odds. This finding suggests that the bookmakers' odds were inefficient.

_____

[23]Nor do they compare their predictions with a naive forecast that the home team wins 46%, draws 27%, and loses 27% of the time.

2. Experts

We have data that evaluates the forecasts of two types of experts. the first is the group of bookmakers who provide the fixed odds; the second consists of tipsters WHO WRITE FOR??.[24] Kuypers (2000, Table 2, p.1359) showed that the bookmakers' odds when converted into probabilities are closely related to the objective ratios of the outcome of the events. On the other hand, THE TIPSTERS

F. Tournaments

The first studies predicting the outcomes of tournaments were non-empirical and involved calculating the probability that the best team would win the event. Using probability theory Mosteller (1952) showed that the better team would not necessarily win the World Series if the outcome were based on a small number of games. Glenn (1960) extended this analysis by examining four different types of tournaments and the types of draws that might give the best players (teams) the highest probabilities of winning.[25] He concluded that the type of draw is more important than the type of tournament and that giving byes in the early rounds of a tournament to the best players improves their chances even more. Similarly, Moser's (1982)

---

[24]Andersson et al. (200x) evaluated the predictions of the outcome of the 200x World Cup matches made by individuals who had familiarity with soccer. They called them 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.

[25]Searls (1963) extends this analysis to other types of tournaments.

used game trees (with over 400,000 outcomes) to model the Jai Alai game and showed that post position influences the outcome.

In addition to these theoretical analyses, empirical studies have calculated the probabilities that a specified seed in the NCAA men's basketball tournament would emerge the winner of the four regional tournaments. (Schwertman et al., 1991, 1996; Carlin, 1996). These empirical probabilities were based on the rankings of the various teams, the point spread from the gambling market, or the Sagarin ratings.[26]

Since 1985, the NCAA has selected 64 college basketball teams to participate in a tournament to select a national championship. The 64 teams are divided into four groups of 16 and are ranked from 1 through 16. Using a probit based solely on the difference in ranks, Boulier and Stekler (1999) found that, between 1985 and 1995, the teams with better rankings defeated their opponents more than 73% of the time. This result was compared with the binomial distribution with a 0.50 probability of picking a winner by chance and was found to be significant.[27]

Caudill and Godwin (2002) determined that the higher seed won 74% of the time in the 1985-1998 tournaments. They used a variety of models to estimate the probability that the higher seeded team will win the game. They found that a logit that included heterogeneous skewness was the best binary choice model. This model takes into account not only the difference in seedings 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.

---

[26]Carlin evaluated the alternative models using an information theory measure.

[27] A similar result was obtained from an evaluation of the women's championship tournament for the years 1994-1995.

Kaplan and Garstka (2001) obtained a slightly lower accuracy ratio for the 1998 and 1999 tournaments, but they also included the games among Final Four that had not been examined by Boulier and Stekler. Moreover, they compared the accuracy of the forecasts from the seedings with those of the betting market and those of the Sagarin statistical scoring system. Picking the seeds was slightly more accurate than using the betting market and the Sagarin system was superior to both.[28]

Harville (2003) constructed a modified least squares ranking procedure based by placing a limit on the margin of victory. He then compared the forecasting accuracy of that method with (1) forecasting that the higher seed will win[29] and (2) the betting market. The statistical procedures had the highest accuracy ratio (71 or 72%) in forecasting the winners of the 2000 NCAA tournament. There was little difference between forecasts based on ranks (69%) and the betting market (68.5%). Kaplan and Garstka (2001) and Harville (2003) 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.[30] THERE MAY BE MORE FINDINGS HERE CHECK

IV. Results

---

[28]Kaplan and Garstka , p. 378, showed that the predictions of the three methods agreed a substantial percentage of the time, but did not determine the accuracy that would have occurred if a forecast were made only when two of the methods agreed. This would have shown whether there would have been any benefit to combining forecasts.

[29] 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.

[30] 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.

A. Cross-Sport results

So far we have considered the forecasting procedures and results only on a sport by sport basis. The results relating to the various sports are so similar that the conclusions have to be considered robust.

1. In every sport, except for horse racing, the market forecast is unbiased. 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.

2. In markets where odds are quoted, the ex ante betting probabilities and the ex post outcome ratios are calibrated.

3.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.

4. Models that explain the outcomes of games or matches have been estimated for all 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.

5. The forecasts of many models were not available. However, systems correctly predicted the winners of NFL games more than 60% of the time. This was comparable to the accuracy of experts but less than that of the market. SUMMARIZE RESULTS OF MODELS IN OTHER SPORTS

6. There is no evidence that either statistical systems or experts consistently outperform the betting market.

B. APPLICABILITY OF THESE RESULTS TO FORECASTING IN GENERAL

The analysis of the sports forecasts also provided insights about the forecasting process. 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. 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 the closing spread was closer to the margin of victory than was the opening quote. AVERY AND CHEVALIER

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.

C. CONFLICTING RESULTS

However, 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. showed that the accuracy of the two methods of forecasting was virtually identical. The accuracy of the statistical methods was, however, less variable.

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. 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. On the other hand, Sauer (1998, p. 2059) 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. This is not only agrees with the findings about economic efficiency but also with the

evidence that prediction models, in general, are the most accurate predictors in other fields. 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 Zitzwitz, 2004).

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.

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 --, 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.

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.

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

Chapman, R. G., (1994), Still searching for positive returns at the track: empirical results from 2000 Hong Kong races,--, 173-181.

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.

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

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.

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.

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, 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 --, 285-298.

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 foavourite-longshot bias?, *Applied Economics*, 34, 1309-1317.

Glenn, W. A., (1960), A comparison of the effectiveness of tournaments, *Biometrica*, 47, 253-262.

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

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.

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---, 307-???

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

Moser, L. E., (1982), A mathematical analysis of the game of Jai Lai, *The American Mathematical Monthly*, 89, 292-300.

Mosteller, F., (1952), The World Series competition, *Journal of the American Statistical Association*, 47, 355-380.

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 efficeincy 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.

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.

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.

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.

Searls, D. T., (1963), On the probability of winning with different tournament procedures, *Journal of the American Statistical Association*, 59, 1064- 1081.

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*, forthcoming.

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

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

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 and Zitzwitz,

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