HOW DO MARKETS FUNCTION?
AN EMPIRICAL ANALYSIS OF GAMBLING
ON THE NATIONAL FOOTBALL LEAGUE

Steven D. Levitt

How Do Markets Function? An Empirical Analysis of
Gambling on the National Football League
Steven D. Levitt
NBER Working Paper No. 9422
December 2002
JEL No. G12, G14, D81

ABSTRACT

   The market for sports gambling is structured very differently than the typical financial market. In sports betting, bookmakers announce a price, after which adjustments are small and infrequent. As a consequence, bookmakers do not play the traditional role of market makers whose primary function is to match buyers and sellers, but rather, take large positions with respect to the outcome of game. Using a unique data set that includes both prices and quantities of bets placed over the course of an NFL season, I demonstrate that this peculiar price-setting mechanism allows bookmakers to achieve substantially higher profits than would be possible if they played the role of the typical market maker. Bookmakers are more skilled at predicting the outcomes of games than bettors and are able to systematically exploit bettor biases by choosing prices that deviate from the market clearing price. While this strategy exposes the bookmaker to risk on any particular game, in aggregate the risk borne is minimal. Bookmaker profit maximization provides a simple explanation for heretofore puzzling deviations from market efficiency that were observed in past empirical work. I find little evidence that there exist bettors who are systematically able to beat the bookmaker, even given the distorted prices that bookmakers set. The results concerning whether aggregating across bettor preferences improves the ability to forecast outcomes are inconclusive.

Steven D. Levitt
Department of Economics
1126 East 59th Street
University of Chicago
Chicago, IL 60637
and NBER
slevitt@midway.uchicago.edu

There are many parallels between trading in financial markets and sports wagering. First, in both settings, investors with heterogeneous beliefs and information seek to profit through trading as uncertainty is resolved over time. Second, sports betting, like trading in financial derivatives, is a zero-sum game with one trader on each side of the transaction. Finally, large amounts of money are potentially at stake. Americans are estimated to wager as much as $380 billion annually on sporting events (National Gambling Impact Study Commission 1999).

In light of these similarities, it is surprising that these two types of markets are organized so differently. In most financial markets, prices change frequently. The prevailing price is that which equilibrates supply and demand. The primary role of market makers is to match buyers with sellers. With sports wagering, however, market makers (casinos and bookmakers) simply announce a "price" (which takes the form of a point spread, e.g., the home team to win a football game by at least 3.5 points), after which adjustments are typically small and infrequent.[2] If that price is not the market clearing price, then the bookmakers may be exposed to substantial risk. If bettors are able to recognize and exploit mispricing on the part of the bookmaker, the bookmaker can sustain large losses. The risk borne by bookmakers on sports betting is categorically different than the casino's risk on other games of chance such as roulette, keno, or slot machines. In those games of chance, the odds are stacked in favor of the casino and the law of large numbers dictates profits for the house. In contrast, however, if the bookmaker sets the wrong line on sporting events, it can lose money in the long run. The presence of even a small number

---

[2]   In the five days preceding a game, the posted price changes an average of 1.4 times per game. When the price does change, in 85 percent of the cases the line moves by the minimum increment of one-half of a point. Thus, the posted spread on Tuesday is within one point of the posted spread at kickoff on Sunday in 90 percent of all games. These calculations are based on information on changes in the casino lines reported at www.wagerline.com.

of bettors whose skills allow them to achieve positive expected profits could prove financially disastrous to the bookmaker. Such bettors could either amass large bankrolls, or in the presence of credit constraints, sell their information to others.

Although the mechanism used for price-setting in sports betting seems peculiar, there are at least three scenarios in which the bookmaker can sustain profits implementing it.   In the first scenario, bookmakers are extremely good at determining in advance the price which equalizes the quantity of money wagered on each side of the bet. If this occurs, the bookmaker makes money regardless of who wins the game since the bookmaker charges a commission (known as the "vig") on bets.[3] Following this strategy, bookmakers do not have to have any particular skill in picking the actual outcome of sporting events, they simply need to be good at forecasting how bettors behave. Popular depictions of bookmaker behavior have stressed this explanation.[4]

An alternative scenario under which this price-setting mechanism could persist is if bookmakers are systematically better than gamblers in predicting the outcomes of games. If that were the case, the bookmaker could set the "correct" price (i.e. the one which equalizes the

---

[3] Typically, bettors must pay the casino 110 units if a bet loses, but are paid only 100 units if the bet wins.

[4] For instance, a website devoted to educating novice gamblers (www.nfl-betting.org) writes, "A sports bettor needs to realize that the point spread on a game is NOT a prediction by an oddsmaker on the outcome of a game. Rather, the odds are designed so that equal money is bet on both sides of the game. If more money is bet on one of the teams, the sports book runs the risk of losing money if that team were to win. Bookmakers are not gamblers--they want to make money on every bet regardless of the outcome of the game." Similarly, Lee and Smith (forthcoming) write, "Bookies do not want their profits to depend on the outcome of the game. Their objective is to set the pointspread to equalize the number of dollars wagered on each team and to set the total line to equalize the number of dollars wagered over and under. If they achieve this objective, then the losers pay the winners $10 and pay the bookmaker $1, no matter how the game turns out. This $1 profit (the "vigorish") presumably compensates bookmakers for making a market and for the risk they bear that the pointspread or total line may be set incorrectly.

probability that a bet placed on either side of a wager is a winner). Although the money bet on any individual game would not be equalized, on average the bookmaker will earn the amount of the commission charged to the bettors. Unlike the first scenario above, however, if prices are set in this manner, the bookmaker will lose if gamblers are actually more skilled in determining the outcome of games than is the bookmaker.

The third possible scenario combines elements of the two situations described in the preceding paragraphs. If bookmakers are not only better at predicting game outcomes, but also proficient at predicting bettors preferences, they can do even better in expectation than to simply collect the commission. By systematically setting the "wrong" prices in a manner that takes advantage of bettor preferences, bookmakers can increase profits. For instance, if bookmakers know that local bettors prefer local teams, they can skew the odds against the local team. There are constraints on the magnitude of this distortion, however, since bettors who know the "correct" price can generate positive returns if the posted price deviates too much from the true odds.[5]

In this paper, I attempt to better understand the structure of the market for sports gambling by exploiting a data set of approximately 20,000 wagers on professional football placed placed by 285 bettors at an online sports book as part of a high-stakes handicapping contest. Two aspects of this data set are unique. First, in contrast to previous studies of betting that only had information on prices, I observe both prices *and* quantities of bets placed. That information

---

[5] This assumes that bookmakers are unable to offer different prices to different bettors. Indeed, there is evidence that local bookmakers who deal repeatedly with the same clients are able to exercise some degree of price discrimination. See Strumpf (2002) for empirical evidence that bookmakers both shade the odds against the home team and offer different odds to bettors with different past betting histories.

allows me to determine whether the bookmaker appears to be equalizing the amount of bets on each side of a wager. Second, I am able to track the behavior of individual bettors over time, which provides a means of determining whether some bettors are more skillful than others. Although my data are from one bookmaker, the patterns observed here are likely to generalize since all bookmakers offer nearly identical spreads on a given game.

A number of results emerge from the analysis. First, I demonstrate that the bookmaker does not appear to be trying to set prices to equalize the amount of money bet on either side of a wager. In almost one-half of all games, at least two-thirds of the bets fall on one side of the gamble. Moreover, the spread chosen systematically fails to incorporate readily available information (e.g. which team is the home team) that would help in equalizing the money bet on either side of a wager.[6] For instance, in games where the home team is an underdog, *on average* two-thirds of the wagers are on the visiting team. These findings argue strongly against the first scenario presented above and the popular depictions of bookmaker behavior. A rationale for this failure to equalize the money emerges in the paper's second finding: bookmakers appear to be strategically setting prices in order to exploit bettors' biases. Bettors exhibit a systematic bias toward favorites, and to a lesser extent, towards visiting teams.[7] Consequently, the bookmakers

---

[6] Of course, it is not the number of bets on either side of a wager that the bookmaker wants to equalize, but rather, the total dollars bet on either side. In my data, however, all wagers are constrained to have the same dollar value, so the two are equivalent. If there are large bankroll bettors outside my sample who systematically bet against the prevailing sentiment of other bettors, conclusions based on my sub-sample may be erroneous.

[7] This finding is not to be confused with the bias towards longshots that has been observed in parimutuel horse-race betting (e.g. Ali 1977, Golec and Tamarkin 1998, Jullien and Salanie 2000, Thaler and Ziemba 1988). In football, the odds are set to make the chance of each team covering the spread about fifty percent, making this consideration irrelevant.

are able to set odds such that favorites and home teams win less than fifty percent of the time, yet

attract more than half of the betting action. By choosing these prices, it appears bookmakers

increase their gross profit margins by 20-30 percent over a price-setting policy that attempts to

balance the amount of money on either side of the wager. On the dimension of favorites versus

underdogs, bookmakers appear to have distorted prices as much as possible without allowing a

simple strategy of always betting on underdogs to become profitable. The fact that home teams

and underdogs cover the spread a disproportionate share of the time has been well established in

the literature (e.g. Golec and Tamarkin 1991, Gray and Gray 1997). My findings provide an

explanation for that empirical regularity: it is profit maximizing for the bookmaker who sets the

spread. Third, there is little evidence that there are individual bettors who are able to

systematically beat the bookmaker.[8] The distribution of outcomes across bettors is consistent

with data randomly generated from independent tosses of a 50-50 coin. Moreover, how well a

bettor has done up to a certain point in time has no predictive value for future performance.

Finally, the evidence is mixed as to whether aggregating bettor preferences has any predictive

value in helping to beat the spread. In my sample, there is some weak and ultimately statistically

insignificant, evidence that bettors are more likely to predict correctly when there is agreement

among them as to which team looks attractive. Altogether, the results are consistent with the

conclusion that the bookmakers are at least as good at predicting the outcomes of games than are

even the most skilled gamblers in the sample, and the bookmakers exploit their advantage by

---

[8] While only tangentially related to the issues addressed in this paper, it is worth noting
that there is an extensive academic literature devoted to the question of testing for market
efficiency in wagering markets (e.g. Asch, Malkiel, and Quandt 1981, Sauer et al. 1988,
Woodland and Woodland (1994), Zuber, Gandar, and Bowers 1985).

strategically setting prices to achieve profits that are likely higher than would be possible if they simply acted as market makers letting supply and demand equilibrate prices.

The remainder of the paper is organized as follows. Section II provides some background on wagering on professional football in the United States. Section III describes the data set used in the paper. Section IV presents the empirical findings. Section V concludes.

Section II: Background on football wagering

Football is the most popular sport for wagering in the United States, generating 40 percent of sports-betting revenue for legal bookmakers in Nevada (Nevada Gaming Control Board 2002). *USA Today* reports that half of all Americans have a wager on the outcome of the Super Bowl. The most common type of bet in pro football involves picking the winner of a game against a point spread (a so-called "straight bet").[9] For instance, if the casino posts a betting line with the home team favored by 3 points, a bettor can choose either (1) the home team to win by more than that amount, or (2) the visiting team to either lose by less than three points or to win outright. In the event the game ends exactly on the point spread, all bets are refunded. Regardless of which team is chosen, the bettor typically pays the casino 110 units if they lose the bet and collects 100 units when victorious. The difference in the amount paid on a loss versus

---

[9] There are many other types of bets available. For instance, one can bet on whether the total number of points scored in a game is above or below a certain level. One can also bet on which team will win the game (not against the spread), with the payouts appropriately adjusted to reflect the probability of these outcomes. It is also possible to parlay bets on a series of games such that the bettor receives a large payout if correctly picking all the games and receives zero otherwise. Because my data only covers straight bets, I do not focus on these other bet types. Woodland and Woodland (1991) argue that the use of point spreads as opposed to odds that depend only on which teams wins or loses is a bookmaker profit-maximizing response to risk aversion on the part of bettors.

the amount won for a victory is the casino's commission, known as the "vig." Because the bettor

can take either side of the wager at the same payout rate, the casino needs to pick a betting line

that roughly equalizes the probability of the two events occurring (in this case home team

winning by more than three or more points, or failing to do so). The bettor receives the spread in

force at the time a wager is placed, regardless of later adjustments made by the bookmaker.

Define terms as follows: p is the probability that the favorite wins a particular game, f is

the fraction of the total dollars bet on the game that go to the favorite, and v is the vig or

commission charged by the bookmaker, which is paid only on losing bets.[10] The bookmaker's

expected gross profit per unit bet[11] is given by

(1) $E[\text{Bookmaker profit}] = [(1\text{-}p)f)+p(1\text{-}f)](1+v)\text{- }[(1\text{-}p)(1\text{-}f)+pf]$

The terms inside the left set of brackets is the fraction of dollars bet in which the bookmaker

wins. That amount is multiplied by $1+v$ to reflect the bookmakers vig. The terms in the right set

of brackets are the cases in which the bookmaker loses and has to payout to the bettor.

Rearranging terms, equation (1) simplifies to

(2) $E[\text{Bookmaker profit}] = (2+v)(f+p\text{-}2pf)\text{-}1$

---

[10] I define the model in terms of favorites and underdogs simply because this is the most salient dimension empirically. Game outcomes could be characterized along any relevant set of dimensions.

[11] For simplicity, I treat the number of wagers placed as fixed in the analysis. To the extent that changing the odds affects the total volume of bets, the bookmaker's overall gross profit would be a function of both the number of bets and the gross profit per unit bet.

If either the probability the two teams win is equal (p=.5) or the money bet on both teams is equal (f=.5), the bookmakers gross profit simplifies to $v/2$. In either of these instances, the bookmaker is indifferent about the outcome of the game and earns a profit proportional to the size of the commission charged. As noted in the introduction, therefore, the bookmaker does not need to be able to predict the outcome of the games more accurately than the bettors to ensure a profit. The bookmaker just needs to be able to predict better preferences so as to balance out the money on each side of the wager.

Equations (1) and (2) take f and p as given. Of course, the fraction of money bet on the favorite will be a function of the probability the favorite actually wins, i.e. $f=f(p)$, with $\partial f/\partial p>0$.[12] Taking the derivative of (2) with respect to p, an optimizing bookmaker will set p such that

(3) $[1-2f(p)] + (1-2p)\partial f/\partial p = 0$

The term in square brackets is the benefit the bookmaker would achieve from distorting the odds if gamblers did not respond to changes in prices. The remaining term captures the impact on profits of the behavioral response of bettors who switch towards the team with bettor odds. Note that if bettors preferences are unbiased in the sense that $f(p=.5)=.5$, then the bookmaker's optimum is to choose p=.5, which implies f=.5 as well.

---

[12] Throughout this analysis, I treat v (the commission charged by the bookmaker) as a parameter rather than a decision variable for the bookmaker. Commissions are virtually always 10 percent. Gaining a better understanding of the reasons for the uniformity of commissions across bookmakers and over time presents an interesting puzzle for future research.

If, on the other hand, bettors preferences are biased so that $f(p=.5)>.5$, as is true

empirically in the data set, then the bookmaker can increase profits by reducing p below .5

(Kuypers 2000 makes this same point).[13] Intuitively, if bettors prefer favorites at fair odds, the

bookmaker can offer odds slightly worse than fair on favorites and still attract more than half of

the wagers on the favorite, yielding profits that are strictly higher than is the case at $p=.5$.

Mathematically, at $p=.5$, the term in square brackets in equation (3) is negative, but the other

term on the left-hand-side is equal to zero, demonstrating that the bookmaker is not at an

optimum. The bookmaker will not want to push p too far away from .5 for two reasons,

however. First, as p diverges from .5, it becomes increasingly costly to the bookmaker when a

bettor switches from the favorite to the underdog. This is because the bet on the favorite is at

(increasingly) unfair odds, whereas the bet on the underdog is at (increasingly) better than fair

odds. Note that when p is lowered to the point where $f(p)=.5$, gross profits are back to the level

attained when $p=.5$. Thus, the bookmaker would never want to distort prices to that point. The

second reason that the bookmaker cannot distort prices too much is that if some subset of bettors

do not have biased preferences, those bettors can exploit the distorted prices. With the standard

vig, a bettor must win 52.4 percent of bets to make profit.[14] One could imagine that the volume

of capital available to bettors with positive expected profits could be enormous, both because

their bankrolls would grow over time and because the availability of such profits would attract

new investors. Thus, it would be surprising to observe price distortions so large that simple

---

[13] Although I use the term "bias" to describe bettor preferences, I do not necessarily
imply irrationality on their part. If there is more consumption value associated with betting on
favorites, then bettor preferences for favorites could be completely rational.

[14] A bettor breaks even when $p-(1+v)(1-p)=0$. The solution to that expression is $p=.5238$.

strategies (e.g. always bet the underdog) could yield a positive profit.

III. The data set

The data used in this paper are wagers placed by bettors as part of a handicapping contest offered at an online sportsbook during the 2001-2002 NFL season. In the contest, bettors were required to pick five games per week against the spread for each of the seventeen weeks of the NFL regular season (a total of 85 games). Bettors could choose those five games from any of the 13-15 games being played in a given week. One point was given for each correct pick, and one-half point if a game ended exactly on the spread. The entry fee was $250 per person, and there were 285 entrants. All of the entry fees were returned as prize money, so participants were competing for a total pool of $71,250.[15] Sixty percent (or $42,750) went to the bettor with the most correct picks. Second through fifth place finishers received declining shares of the pool. The last-place finisher (conditional on making 85 picks) received five percent of the pool. In the data, I observe the ID number of each bettor, all wagers placed as part of this contest, the spread at which the bet was placed, and the outcome of the game.

There are a number of potential shortcomings with these data. First, these are not wagers in the traditional sense. The bettor does not receive a direct payoff from winning any particular game; the payoff is only based on the cumulative number of wins. Nonetheless, the presence of large monetary rewards to the winners provides strong incentives to the participants. Presumably, the picks made by bettors in the contest closely parallel the actual betting wagers

---

[15] The apparent purpose of the contest was to ensure that the bettors had a reason to visit the website each week. The fact that the worst-place finisher (conditional on having bet each week) received a payoff, reinforces this point.

they were making; anecdotally that is true among the contest participants known to the author.[16]

Second, the nature of the data make it impossible to ascertain the intensity of preferences across games since all selections receive equal weighting. Bettors may have much stronger preferences for their most favored pick of the week than would be the case for the fifth-favorite pick. Third, there is substantial attrition in the sample. Of the 285 bettors who entered the contest, 100 (a little more than one-third) made their entries all 17 weeks of the season. More than 60 percent participated in at least 15 of the 17 weeks of the season. Less than 10 percent of the contestants recorded data for fewer than eight weeks. For bettors in the middle of the pack as the season progresses, the incentive to continue participating decreases substantially. Bettors who miss a week receive zero points, greatly reducing their likelihood of winning the contest, and disqualifying themselves from eligibility for the last-place prize. It should be noted, however, that attrition in this context adversely affects my ability to test only one of the hypotheses: what the overall distribution of bettor success rates looks like. Because attrition is non-random, the set of bettors who continue to the end will be skewed. Fourth, the spreads used in the handicapping contest are fixed on the Tuesday preceding the game and do not fluctuate with the actual spread, even though a bettor's contest picks are not due until the Friday before the game. As a consequence, in some games, the actual spread and the contest spread differ at the time an entry is made. All of the results of the paper, however, are robust to dropping games in which there are substantial fluctuations in the spread between Tuesday and Friday. Finally, these data are not

---

[16] One important way in which the contest wagers might be expected to differ from actual wagers placed is that there is an incentive in the contest to pick outcomes that the bettor believes will be unpopular with other gamblers. That is because of the tournament structure of the contest in which rewards are great in the extreme right-hand tail, but no differentiation is made elsewhere in the distribution.

the universe of bets placed at the sportsbook (although they are the universe of bets in this contest), much less at sportsbooks in general. Nor, as discussed below, are the bettors who participated in the contest likely to be a random subset of all bettors.

On the other hand, these data do offer enormous advantages over that which is typically available. Virtually all previous analyses of sports wagering have focused exclusively on prices, but have not had access to any information about the quantity of bets on each side of the wager (see, for example, Avery and Chevalier 1999, Golec and Tamarkin 1991, Gray and Gray 1997, Kuypers 2000).[17] In my sample, there are a total of 19,770 bets in the data set covering 242 different games. An average of 80.5 different bettors make a selection on a game, with the minimum and maximum number of bets on a game ranging from 28 to 146. Although the bettors included in my sample are not a random selection of all gamblers, they represent a particularly interesting subset. These bettors are likely to be relatively sophisticated, serious bettors. Because they are betting at an online sportsbook, they are likely to be geographically quite diverse. In signing up for the contest, they are indicating an expectation that they plan to visit (and presumably bet at) the internet sports book every week of the season. In addition, to the extent there are differences in skill across bettors, this contest should attract the most skillful players because it rewards exceptional long-run performance.

Section IV: Empirical Results

---

[17] In his innovative work, Strumpf (2002) does have actual bets based on seized bookmaker records. Although his data are extremely informative on a number of questions, they are less than ideal for the questions posed in this paper both because they cover a short time period and are geographically localized.

I begin the analysis by addressing the issue of whether the spread is set so as to equalize the wagers on either side, as well as testing the predictions of a model of profit-maximizing bookmakers who exploit biases on the part of bettors. The analysis then turns to the question of whether bettors differ in their skill at picking winners. Finally, I examine whether aggregating information across bettors provides any valuable information.

How do bookmakers set prices?

The first question addressed is whether bookmakers set prices so as to equalize the amount of money on either side of the wager. Figure I presents a histogram of the fraction of the total wagers placed on the team that bettors most prefer. By definition, this fraction must lie between .5 and 1. If the bookmaker balances bets, these values will be concentrated near .50. In the data, however, this is clearly not the case. In only twenty percent of the games are 50-55 percent of the wagers placed on the preferred team. In the median game, almost two-thirds of the bets fall on one side. In almost 10 percent of the games, more than 80 percent of the bets go one direction.

The dispersion in Figure I is not simply the result of sampling error due to the fact that I observe only 80 bets per game on average. If bettor choices were independent and each bettor had a fifty percent chance of picking either team, than one would expect the preferred team to garner between 50 and 55 percent of the wagers in nearly two-thirds of the games, compared to only twenty percent in the data. Furthermore, if the sample is divided in half based on the number of bets placed, the dispersion of the fraction of wagers placed on the preferred team is

actually greater in the games with more total bets.[18]

Two alternative hypotheses could explain the failure of the bookmaker to equalize wagers on the two sides of the spread. The first possibility is that the bookmaker would like to balance the bets, but is unable to do so because it is difficult to accurately predict what team bettors will prefer. The second hypothesis is that balancing the wagers is not the objective of the bookmaker. Indeed, as demonstrated earlier, if bettors exhibit systematic biases, a profit maximizing bookmaker does not want to equalize the money bet on both sides. Rather, the bookmaker intentionally skews the odds such that the preferred team attracts more wagers, but wins less than half of the time.

If bookmakers are attempting to balance the money bet on each side of the wager, one would expect that observable characteristics of a team or game would have no power in predicting the fraction of bettors preferring that team. Otherwise, the bookmaker could have used that information to set a spread that would have better equalized the distribution of bets. If the bookmaker is attempting to exploit bettor biases by setting skewed odds, however, the opposite is true. Dimensions along which bettors exhibit bias should be systematically positively related to bet shares (and as demonstrated below, will also be systematically negatively related to win percentages).

Figures II and III provide some initial evidence on the question of whether observable characteristics are correlated with the distribution of wagers on a game. Figure II presents a

---

[18] Further confirmation of the patterns in Figure I come from data available at www.wagerline.com. At that website, visitors make hypothetical wagers on game outcomes as part of small-payoff contests. The breakdown of bets on each team is available. The patterns in that data are strikingly similar.

histogram of the fraction of bets placed on the home team for the subset of games in which the

home team is the favorite.[19] Since the identity of the home team is readily observable, the

histogram in Figure II should be centered around 50 percent if the bookmaker is attempting to

equalize bets on either side of the wager. The distribution is clearly skewed to the right,

implying that home teams systematically attract more than half of the bets in games in which they

are favored. In almost three-quarters of the games, more bets are placed on home favorites than

on their opponents. In the median game, roughly 58 percent of the bets go to the home favorite.

Figure III is identical to Figure II, except that the sample is games in which the visiting team is

favored, and the values in the figure are the fraction of bets placed on the visiting team. In these

games, the distribution of bets is even more skewed. In more than 90 percent of games with a

visiting favorite, more bets are placed on the visitor than the home team. In the median game in

Figure III, two-thirds of the money is wagered on the visitor. Thus, Figures II and III

demonstrate quite definitively that the spreads are set such that substantially more than half of the

bets are placed on favorites.

   Table I further explores the issue of whether observable characteristics are correlated with

betting patterns. The dependent variable in Table I is the percent of bettors who choose the

favorite. The unit of observation is a game. The method of estimation is weighted least squares,

with the weights proportional to the total number of bets placed on the game. The first column

_____

[19] Note that despite the similarity in the word "favorite" and the phrase "bettors'
preferred team," these are two completely different concepts. " Favorite" refers to the team
judged most likely to win the game by the bookmaker. The "bettors' preferred team," on the
other hand, is the team that the bettors think is most likely to cover the spread. In other words,
the bettors are making their choices conditional on the bookmaker already setting a spread that
ostensibly equalizes the chance of the favorite and the underdog covering.

of the table demonstrates that, not controlling for anything else, 60.6 percent of the bets accrue to the favorites. The standard error on that point estimate is .009, so the null hypothesis that 50 percent of the bets are placed on favorites is strongly rejected. Column II adds five indicator variables corresponding to which team is favored and by how much (the omitted category is games in which the visitor is favored by more than 6 points). Consistent with the figures presented earlier, home favorites do not attract as high a fraction of the bets as do visiting favorites. These five variables capturing the spread are jointly highly statistically significant, as reported in the bottom of the table. The third column adds dummy variables for each week of the season. These week dummies are also jointly statistically significant at the .01 level. Although not shown individually, the dummies suggest that a greater fraction of the bets are placed on favorites early in the season. Finally, the last column adds thirty-one indicator variables corresponding to each team in the league. These variables take the value of one if a team is favored and -1 if the team is an underdog. Once again, the team variables are highly statistically significant.[20] The R^2 in the final column of the table is .483, implying that these observable characteristics explain a great deal of the variation in the fraction of money bet on the favorite.

The results in Table I uniformly argue against the hypothesis that bookmakers are doing the best they can to even out the bets on each game, suggesting instead that the imbalance is intentional. If that is true, then the model presented earlier makes strong predictions about the expected pattern of winning percentages for bets of different kinds. In particular, teams with attributes that attract a disproportionate share of the money from bettors (e.g., being the favorite)

---

[20] Based on these estimates, the five teams that attracted the most bettors, controlling for other factors, were Green Bay, Kansas City, New Orleans, Oakland, and San Francisco. The least popular teams were Baltimore, Chicago, Jacksonville, Pittsburgh, and Saint Louis.

should cover the spread less than fifty percent of the time. But, the deviation from 50-50 cannot be too large (more than a few percentage points), or bettors who do not suffer from biases can profitably exploit the price distortion.

Table II presents evidence consistent with those predictions. The unit of observation in the table is an individual bet. The first three columns captures the fraction of the bets placed on a team; columns 4-6 are the percentage of those wagers in which the bettor correctly picks the winners. Bets are categorized according to whether they are placed on games in which the home team is favored (top row) or the visiting team is favored (second row), and whether the bet is for the favorite or the underdog (the columns in table). Mirroring the results presented earlier, visiting favorites attract a disproportionate share of the bets placed in such games: 68.2 percent of the total.[21] To a lesser degree, home favorites also attract excess bets (56.1 percent). Shifting focus to columns 4 and 5, the model predicts that a high fraction of bets in columns 1 and 2 will be associated with low winning percentages in columns 4 and 5. Indeed, the results confirm the prediction. Across the four categories considered, the rank-order correlation between the percentage of bets placed and the win percentage is -1. Bets placed on visiting favorites win only 47.8 percent of the time. Bets on home favorites are successful in 49.1 percent of cases. Bets on underdogs, which are under-represented in the data, win more than half of the time. Notably, bets on home underdogs have win rates of 57.7 percent, well above the threshold required for a

---

[21] By definition, the sum of the fraction of bets on visiting favorites (row 2, column 1) and home underdogs (row 1, column 2) must add up to 100 percent. The same holds for home favorites and visiting underdogs.

bettor to break even.[22]

One might be tempted to discount the observed relationship between a high fraction of bets made and low winning percentages since the results are based on the outcomes of only 236 games of a single season. Although it is impossible to obtain the data on quantities for earlier years, outcomes of games relative to the spread are readily available. Assuming that the strong tendencies towards betting on favorites are persistent across years, one would expect the patterns in winning percentages to also be persistent. Indeed, the winning percentage patterns have previously been documented by Golec and Tamarkin (1991) and Gray and Gray (1997). They are also present in Table III, which includes results on game outcomes for twenty-one seasons of data (1980-2001), covering almost 5,000 games. The first column of the table presents the percentage of bets made on teams of a particular type in the 2001-02 data; the second column is the win percentage over the past twenty-one years. The results are quite consistent with those of Table II. Overall, favorites win less than half their games. The null hypothesis of a 50 percent win rate for favorites is rejected at roughly the .01 level. As predicted, visiting favorites (who attract a greater share of the bets) do especially poorly, winning only 46.7 percent, again rejecting the null of a .50 win percent at approximately the .01 level. Given this win percentage, a naive strategy of always betting against visiting favorites would actually have yielded positive profits over these two decades (as was also true in 2001-02).[23] In light of the consistent lack of success of

---

[22] The win percentages on visiting favorites and home underdogs need not sum to 100 percent because the fraction of the bets on the favorite and the underdog varies across games.

[23] Although not shown in tabular form, the year-by-year patterns in the data confirm the overall findings. In only 4 of the last 21 years have favorites covered the spread in as many of 50 percent of the games. The likelihood of that occurring if the true win likelihood is .50 is less than 1in 300. Only 3 times in 21 years have visiting favorites won against the spread in 50

favorites, especially visiting favorites, it is remarkable that the strong bettor bias towards such

bets persists. The bettor bias is not concentrated among a small fraction of bettors, either.

Three-fourths of the contestants chose favorites more often than underdogs. Only two percent of

bettors chose visiting underdogs for at least half of their picks.

Just how much do bookmakers increase their profits by exploiting bettor biases?

Assuming that (1) the total distribution of bets in this sample is representative of overall betting,

and (2) there is no information in aggregate bettor preferences (the evidence presented below

cannot reject this), it is straightforward using the values in Table II to calculate that the way

spreads are currently set, bettors should win 49.45 percent of their bets.[24] Given the standard

"vig" of bettors risking 110 units to win 100, a bookmaker who wins half his bets has a gross

profit rate of 5.0 percent. If bettors win only 49.45 percent of their bets, the expected gross profit

rate jumps to 6.16 percent (.5055*110-.4945*100). Thus, in expectation, this seemingly minor

distortion of the win rate increases gross profits by 23 percent. It is true, of course, that the

bookmaker must bear some risk when the bets are not balanced on both sides of the wager.

Because game outcomes are likely to be independent, however, the risk is minimized as the

number of games played increases. For instance, in the case where 63 percent of the money is

one side of each wager and that team wins 48 percent of the time, over the course of the NFL

season (roughly 250 games) the bookmaker's expected gross profit rate is 6.1 percent with a

standard deviation of 2.5 percent. Thus, the bookmaker would be expected to make negative

percent of games.

[24] The number 49.45 is obtained by multiplying the probability that the home favorite
wins a game times the percent of total bets on home favorites plus the probability that the visiting
underdog wins times the percent of overall bets on visiting underdogs, etc.

gross profits less than once every one hundred seasons. If one looks over a five-year time frame, the standard deviation drops to 1.1 percent. So the probability of a bookmaker losing over any given five-year period in this scenario is less than one in 10,000. Relative to a 23 percent increase in gross profit, the costs associated with bearing this level of risk appear minimal.[25]

The last two panels of Table III explore the relationship between other factors and betting on favorites. In the 2001-02 data, a higher fraction of bets were placed on favorites in the first half of the season than in the second half (63.4 percent versus 56.3 percent), and the week of the season was highly statistically significant in predicting bet shares in Table I. Whether this is simply an idiosyncracy of the 2001-02 season is uncertain. Consistent with the theory, the win percentages for favorites over twenty-years are higher in the second half of the season (48.5 versus 47.7), although the differences is not statistically significant. Finally, the bottom panel of the table demonstrates that the size of the spread has little impact on the distribution of bets on the favorite; correspondingly, there is little apparent difference in win percentages across these games in column 2.

Is there evidence that some bettors are especially skillful in picking winners?

In order for the current system of price-setting (in which the bookmakers set a price and do not adjust that price to equilibrate supply and demand) to survive, there cannot exist a sufficient number of bettors with an ability to pick winners that exceeds that of the bookmaker. This is particularly true when the bookmaker distorts prices to exploit the subset of bettors with

---

[25] A major puzzle in this industry is the rarity of price competition, i.e. the vig is almost universally 10 percent. It is possible that the bearing of risk somehow supports this equilibrium. One website, www.tradesports.com, acts as a traditional financial market-maker, matching buyers and sellers, but taking no positions on game outcomes. The commission charged for this match-making service is less than 1 percent of the bet – far smaller than the traditional vig.

biases. In that case, a sophisticated bettor only needs to be slightly better than the bookmaker in determining the true odds to turn a profit.[26] Indeed, Strumpf (2002) argues that much of the internal structure of bookmaker organizations is designed to protect the bookmaker against adverse selection by these talented bettors.

Testing for bettor skill is complicated in my data set by the fact that there is a great deal of attrition over the course of the sample, and the attrition is not random. Bettors who have performed poorly up to that point in time are much more likely to leave the sample since the chances of receiving a prize are very low for these contestants.

I consider two possible approaches for testing for heterogeneity in skill across bettors. The first approach is to look at the overall distribution of games won over the course of the season and to test whether that distribution is consistent with that which would have been generated by homogeneous bettors.[27] Because of attrition, however, these data are incomplete. Approximately 82 percent of possible bets were actually placed. Under the assumption that the outcomes of bets on the missing games can be modeled as being generated by independent coin tosses with probability .5, it is possible to simulate what the distribution of wins would have been without attrition. This approach has the obvious drawback that the simulated portion of the data is generated by the process that I have defined as the null hypothesis against which to test. Thus,

---

[26] And, as demonstrated above, even a naive strategy of betting against all visiting favorites has been marginally profitable.

[27] Because the worst-place finisher gets a payoff, those near the bottom have an incentive to try to intentionally pick losers. If they have some ability to do this, that will exaggerate the bottom tail, exacerbating deviations from normality.

such a test is biased against rejecting the null of no heterogeneity.[28] Figure IV presents a

representative histogram of the distribution of these simulated final win totals. Superimposed on

the histogram is the corresponding normal distribution which the data would be expected to

approximate if generated by i.i.d. coin tosses with a win probability of .50. Visually, the

observed distribution closely mirrors the normal distribution. P-values for the three generally

applied tests of normality (skew test, Shapiro-Francis, and Shapiro-Wilk) are well within the

acceptable range. Thus, with the caveat that the test is biased against rejection due to the

simulated data, there is no evidence to reject the null hypothesis of no differences in skill across

bettors in the sample.

The second approach to testing whether there is heterogeneity across bettors in ability to

pick winners is to look for persistence in win rates. A priori, it is not known who the skillful

bettors are. Success early in the season, however, is likely to be a (possibly noisy) signal of

talent. Thus, in the presence of heterogeneity in skill, one would expect those who do better at

the beginning of the season to also outperform later in the season. Like the test of homogeneity

of skill presented above, this approach is not robust to particular sources of attrition. If high-

skilled bettors who have been unlucky in the early part of the season are less likely to quit than

similarly placed low-skilled bettors, than this approach will be biased against finding

heterogeneity in skill across bettors. The poor-performing bettors who persist will be

disproportionately drawn from the high-skilled group and thus will be expected to perform well

on average later in the season.

---

[28] In defense of the manner in which the missing data are generated, other results
presented below suggest that there is no evidence of serial correlation across weeks in a given
bettors ability to pick winners.

Bearing in mind this caveat, I estimate equations of the form

$$(4) \quad \text{WIN}_{bwg} = \alpha + \beta_1 H_{bw} + \beta_2 (H_{bw})^2 + \beta_3 (H_{bw})^{3+} \beta_4 (H_{bw})^4 + \gamma X_{bw} \, \epsilon_{bwg}$$

where *b*, *w*, and *g* denote bettors, weeks of the season, and specific games respectively. *WIN* is a variable equal to one if the bettor picks the game and covers the spread, .5 if the game is a push, and zero otherwise. The variable *H* represents the bettors cumulative historical winning percentage across all games played thus far in the season. The vector *X* captures other predictors of whether or not the bet is won, for example if the team chosen is a visiting favorite. The quartic in the cumulative winning percentage is designed to non-parametrically capture the serial correlation across betting performances. The first week of the season is omitted from the regression because there is no bettor history. The equation is estimated using weighted least squares, with weights determined by the number of games making up the history. The reported standard errors have been corrected through clustering to account for the fact that *H* is correlated across games for a given contestant.

Table IV reports results of the estimation. In the first column, the cumulative betting success rate is constrained to enter linearly. Although not statistically significant, the point estimate implies that bettors who have been more successful up until that point in the season are predicted to do slightly *worse* in the current week. This argues against heterogeneity in skill across bettors, which would lead to a positive coefficient. The bottom panel of the table reports the predicted success rate for bettors with varying win percentages up to this point in the season. Bettors across the entire range are predicted to get slightly more than fifty percent of the bets correct. The second equation adds the quartic in betting history. Although the history variables

23

are jointly statistically significant, the R^2 is very low (.0004). Bettors averaging 40 or 50 percent correct up to this point are predicted to perform right around 50 percent, bettors who have won 60 percent of games prior to this week are projected to win only 48.7 percent of games. Once again, these results argue against persistent differences in skill. Adding covariates in column 3 has little impact on the conclusions. Columns 4 and 5 restrict the sample to exclude the first five and ten weeks respectively, on the rationale that cumulative win percentages early in the season may not be very informative. The results provide no evidence that strong past performance predicts wins today.

In summary, there is little in the data to suggest that, at least in this particular sample, there is heterogeneity in skill across bettors. This result may be due partially to the relative sophistication of bettors in the sample – perhaps the most naive bettors are unlikely to frequent internet bookmakers. [29]

Does pooling information across bettor preferences help in predicting the outcome of games?

In other contexts, it has been argued that aggregating information across agents provides valuable information in predicting future outcomes. For example, Clemen and Winkler (1986) and Fomby and Samant (1991) find that the consensus estimate of future GNP growth is a better predictor than any one individual's estimate. One might also expect such a pattern to be present in sports betting, especially because price is set unilaterally by the bookmaker. To the extent that the bookmaker sometimes makes mistakes, one would expect that many bettors will simultaneously recognize the presence of the mistake and disproportionately pick one team.

---

[29] Strumpf (2002), for instance, reports the existence of a fraction of New York bettors who always bet on the Yankees, even though the bookmakers, knowing their preferences, systematically offer these bettors substantially worse odds than other clients.

There is one simple result in my data which suggests that aggregating opinions across

bettors may carry valuable information: despite the fact that more than half the money is bet on

favorites and the bookmaker set the odds so that favorites win less than half the games, the

overall winning percentage for bets placed is 50.1 percent. As noted earlier, based on the odds

offered by the bookmaker and the distribution of money bet, one would expect 49.45 percent of

all bets to win if there was no correlation between the percentage of bettors choosing a game and

the game's outcome. The difference between 50.1 percent and 49.45 percent implies that games

in which a greater fraction of bettors choose the favorite (or alternatively the underdog) are more

likely to be won by the favorite (underdog). Thus, in principle one might believe that knowledge

of aggregate bettor preferences might be useful in prediction, making access to quantity data

(which is in general very difficult to obtain, but is available prior to the start of the games

through this contest) valuable.[30]

Table V provides additional suggestive evidence on this issue. Bets are divided into four

categories: home favorites, visiting favorites, home underdogs, and visiting underdogs. Within

each category, bets are divided into quartiles according to how popular that team was with

bettors. For instance, the top quartile of bets on home favorites would capture those bets in

which an especially high share of the people making selections in that game chose the favorite.

Results presented in the table show the total for all games (the first column) and then broken

down by quartile (columns 2-5). If aggregating bettor preferences is valuable in predicting

_____

[30] Interestingly, the win percentage for mock bets placed at www.wagerline.com were
49.5 percent in 2001 – precisely what would be expected if aggregate bettor preferences
contained no information. One possible explanation for the difference in win rates between the
bettors in my sample and those at wagerline is that my sample contains a more sophisticated set
of gamblers who have much more money at risk.

performance against the spread, then one would expect to observe a monotonic, declining win percentage moving from the top quartile to the bottom quartile of bets by popularity. In practice, bets in the second quartile yield the highest winning percentage (52.9 percent). The top and third quartiles also win more than 50 percent of bets. The worst performance by far (46.6 percent) is for those selections that are least popular with bettors. Thus, while the evidence is mixed, there is at least a suggestion in the data of the possibility that popular bets fare better, at least relative to the least popular choices in the bottom quartile.

Table VI tackles this question more rigorously in a regression framework that estimates the impact of a bet's popularity non-parametrically using a fourth-order polynomial and controlling for other factors such as week of the season dummies and team-fixed effects. In addition, Table VI reports standard errors corrected for the fact that although there are over 19,000 bets made, there are only a few hundred game outcomes; for a given game, the bet outcomes will be perfectly correlated, so effectively there are only a few hundred degrees of freedom. The results of the table provide suggestive, but ultimately quite weak, evidence in support of the value of aggregating preferences. In columns 1 and 2, a bet's popularity among bettors enters linearly. The coefficient is .11, implying that a ten percentage point increase in the number of bettors favoring a particular gamble is associated with a 1.1 percentage point increase in the likelihood that team wins. While substantively large, the estimate is not statistically significant. When allowing the coefficient on a bet's popularity to vary in a non-linear fashion, the individual coefficients are no longer readily interpretable, so I report average win percents by quartile in the bottom panel of the table. The patterns observed in columns 3-5 are similar to those in the raw data, with the bets in the top half of the distribution winning more than fifty

percent and bets in the bottom quartile doing especially poorly. It is important to note, however, that the coefficients on bet popularity are in all cases jointly insignificant (as reflected in the P-values reported near the bottom of the table). Thus, great caution must be exercised in drawing any conclusions from these regressions.

Section V: Conclusions

This paper utilizes a unique data set that includes information on both the prices and quantities of bets placed to analyze the workings of the market for gambling on NFL football. The results provide an answer to the apparent puzzle of why these markets do not look much like markets at all, characterized by small and infrequent price changes. The answer is that the bookmakers are better at predicting game outcomes than the typical bettor. As a consequence, the bookmakers are able to set prices in order to exploit their greater talent, and apparently yielding greater profits than could be obtained if the bookmakers acted like traditional market makers and attempted to equilibrate supply and demand, avoiding taking large stakes in the outcomes of games. In the presence of some bettors who are as skillful in picking games as are the bookmakers, there are limits to how much prices can be distorted without creating profit opportunities. It appears that the bookmakers have distorted prices to the point where they bump up against this limit.

In order to support this mode of price-setting, it cannot be the case that a significant fraction of bettors (or more precisely dollars bet) have better information than the bookmaker. In that case, it would be the bookmaker who would be exploited by this system. At least within my sample of bettors, I see little evidence of heterogeneity in ability to pick winners, suggesting that

27

the bookmaker might be as good or better at picking game outcomes than any of the bettors.

Given the incentives for the bookmaker to get the spread right, it is hardly surprising that the

most talented individuals would be employed as the odds makers.

Perhaps, then, a fundamental difference between gambling and financial markets is that it

is possible in gambling to find and hire a small set of individuals (the odds makers) who can

systematically do better in predicting game outcomes than can bettors overall. In financial

markets, on the other hand, the flow of inside information or the inherent complexity in valuing

companies may make it impossible for one individual to do better than the market, meaning that

a market maker who acted like a bookmaker would do worse than one who simply equilibrated

supply and demand and took advantage of the bid-ask spread.[31] Notably, betting on horse racing

and dog racing, which is characterized by extreme insider information on the part of those who

own and train the animals, is typically (though not always) done in a parimutuel framework that

ensures the market maker a fixed profit.

---

[31] Certainly, the weight of the evidence regarding the inability of fund managers to systematically beat the market indexes is consistent with this idea.

References

Ali, Mukhtar, 1977, "Probability and Utility Estimates for Racetrack Bettors," *Journal of Political Economy* 85(August): 803-815.

Asch, Peter, Burton Malkiel, and Richard Quandt, 1984, "Market Efficiency in Racetrack Betting," *Journal of Business* 57:165-175.

Avery, Christopher, and Judith Chevalier, 1999, "Identifying Investor Sentiment from Price Paths: The Case of Football Betting," *Journal of Business* 72 (October): 493-521.

Clemen, Robert, and Robert Winkler, 1986, "Combining Economic Forecasts," *Journal of Business and Economic Statistics* 4:(January): 369-91.

Fomby, Thomas, and Subarna Samanta, 1991, "Application of Stein Rules to Combination Forecasting," *Journal of Business and Economic Statistics* 9(October): 391-407.

Golec, Joseph, and Maurry Tamarkin, 1991, "The Degree of Price Inefficiency in the Footbll Betting Markets," *Journal of Financial Economics* 30:311-323.

Golec, Joseph, and Maurry Tamarkin, 1998, "Bettors Love Skewness, Not Risk, at the Horse Track," *Journal of Political Economy* 106 (February): 205-225.

Gray, Philip, and Stephen Gray, 1997, "Testing Market Efficiency: Evidence from the NFL Sports Betting Market," *Journal of Finance* 52 (September): 1725-1737.

Jullien, Bruno, and Bernard Salanie, 2000, "Estimating Preferences under Risk: The Case of Racetrack Bettors," *Journal of Political Economy* 108(June): 503-530.

Kuypers, Tim, 2000, "Information and Efficiency: An Empirical Study of a Fixed Odds Betting Market," *Applied Economics* 32: 1353-1363.

Lee, Marcus, and Gary Smith, forthcoming, "Regression to the Mean and Football Wagers," *Journal of Behavioral Decisionmaking.*

National Gambling Impact Study Commission, 1999, <u>Final Report</u>, Washington, D.C.

Nevada Gaming Control Board, 2002, "Nevada Gaming Revenues Calendar Year 2001
Analysis," Press release, February 12.

Sauer, Raymond, Vic Brajer, Stephen Ferris, and M. Wayne Marr, 1988, "Hold Your Bets:
Another Look at the Efficiency of the Gambling Market for National Football League
Games," *Journal of Political Economy* 96 (February): 206-213.

Strumpf, Koleman, 2002, "Illegal Sports Bookmakers," Mimeo, University of North Carolina
Department of Economics.

Thaler, Richard, and William Ziemba, 1988, "Parimutuel Betting Markets: Racetracks and
Lotteries," *Journal of Economic Perspectives* 2:161-174.

Woodland, Bill, and Linda Woodland, 1991, "The Effects of Risk Aversion on Wagering: Point
Spreads versus Odds," *Journal of Political Economy* 99(June): 638-653.

Woodland, Bill, and Linda Woodland, 1994, "Market Efficiency and the Favorite-Longshot Bias:
The Baseball Betting Market," *Journal of Finance* 49 (March): 269-279.

Zuber, Richard, John Gandar, and Benny Bowers, 1985, "Beating the Spread: Testing the
Efficiency of the Gambling Market for National Football League Games," *Journal of
Political Economy* 93:(August) 800-806.

Figure I: Share of Bets on the Team
that Bettors Prefer
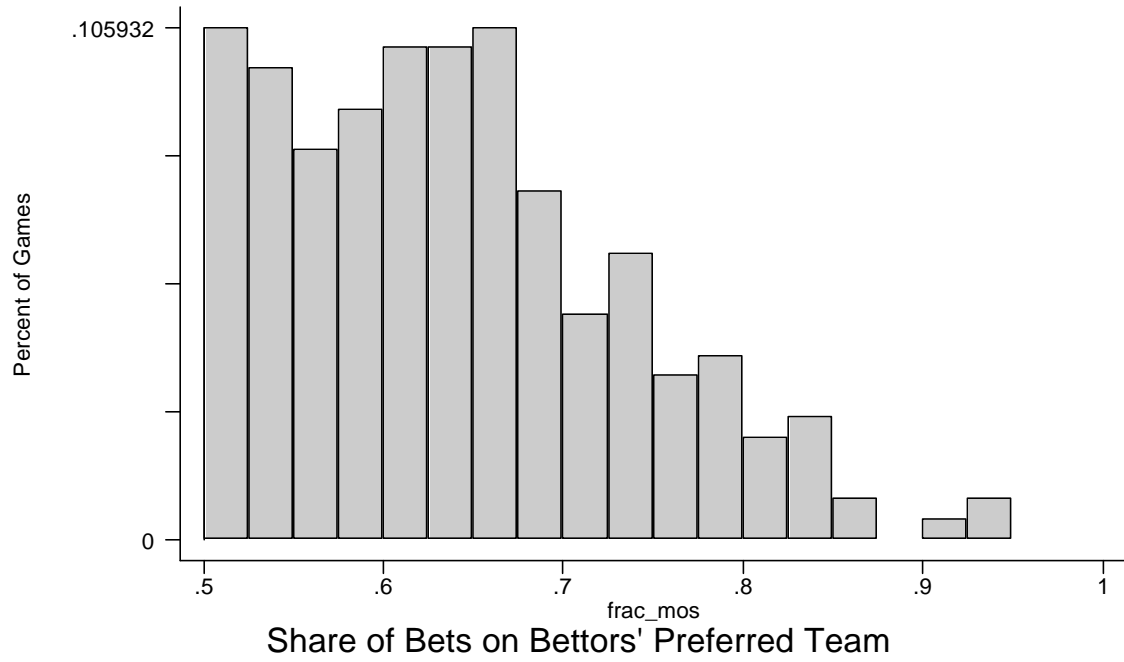
Share of Bets on Bettors' Preferred Team

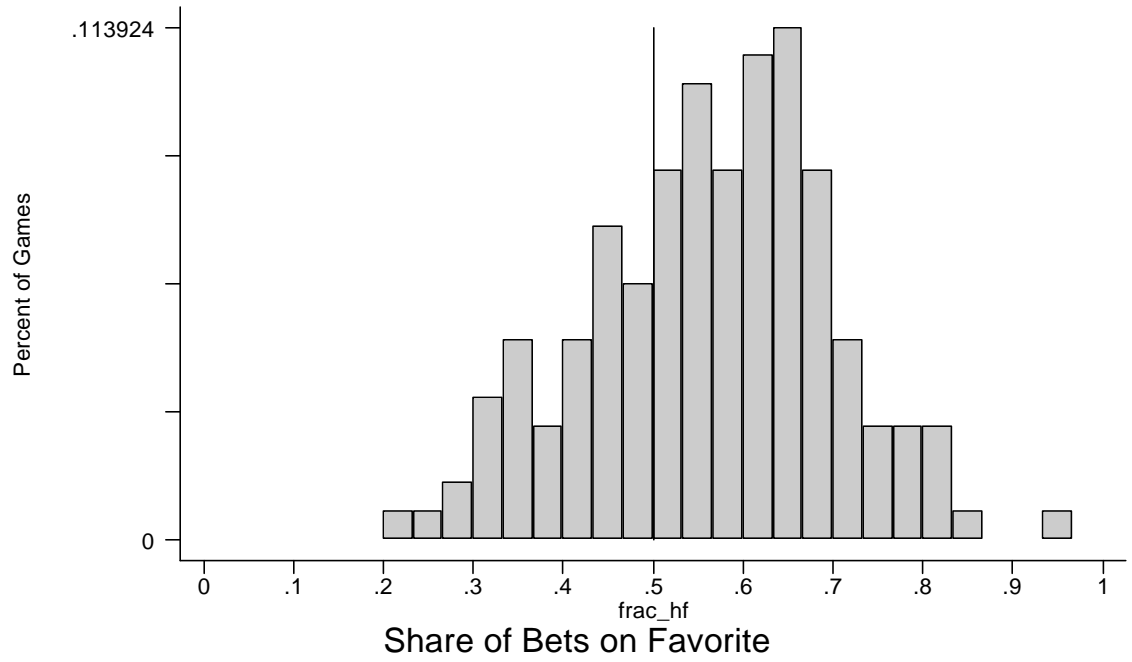Figure II: Share of Bets on the Favorite
when the Home Team is the Favorite

Figure III: Share of Bets on the Favorite
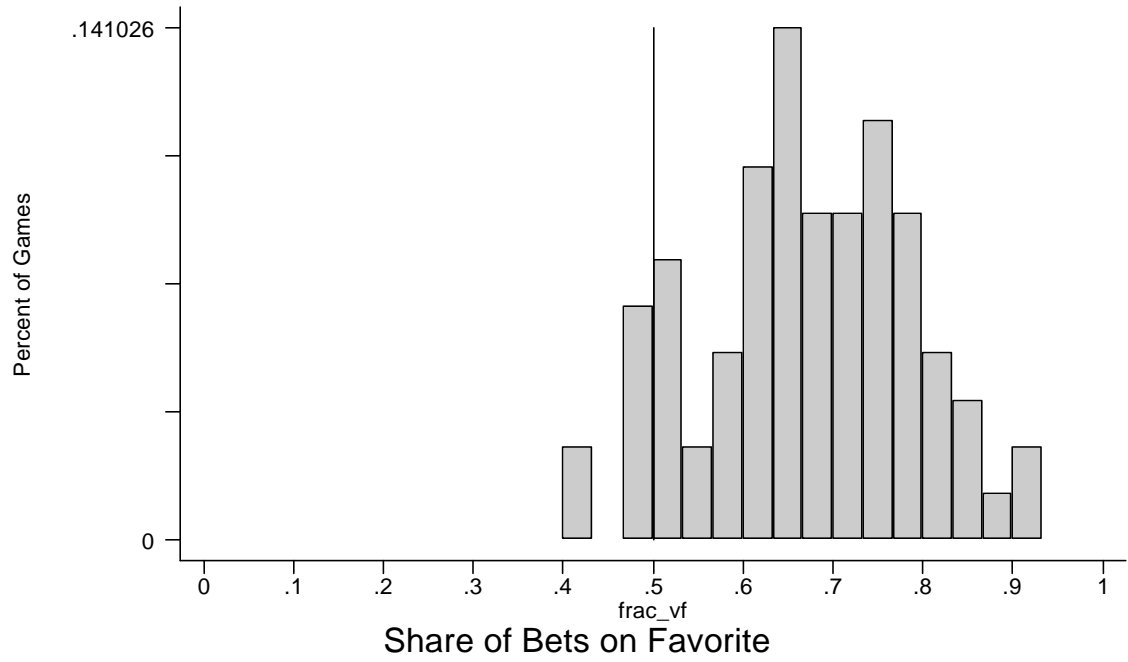when the Visiting Team is the Favorite

Figure IV:Simulated Distribution of Total Wins
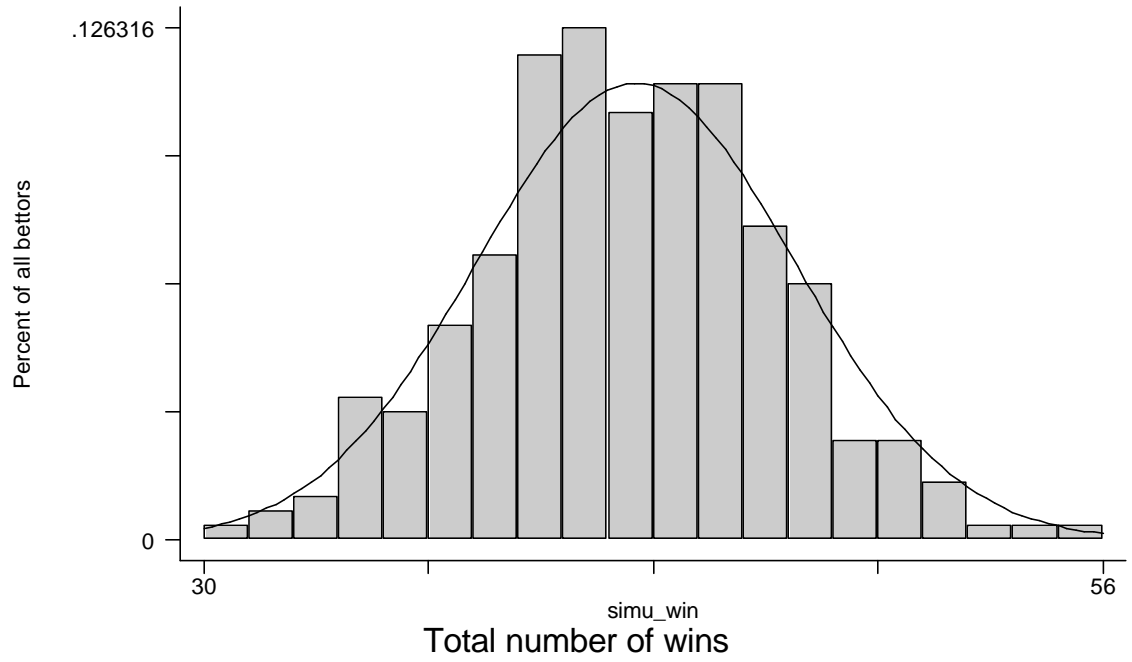Relative to Null Hypothesis of Random Coin Flips

Table I: Predicting the Fraction of Bets Placed on the Favorite

Variable	Dependent variable: Percent of bettors placing bets on the team that is favored			
	(1)	(2)	(3)	(4)
Constant	.606 (.009)	.689 (.025)	-----	-----
Home team favored by more than 6 points	-----	-.129 (.031)	-.131 (.031)	-.144 (.031)
Home team favored by 3.5 to six points	-----	-.127 (.033)	-.123 (.032)	-.136 (.037)
Home team favored by 3 or fewer points	-----	-.126 (.031)	-.126 (.031)	-.123 (.043)
Visiting team favored by 3 or fewer points		-.005 (.030)	-.026 (.030)	-.057 (.033)
Visiting team favored by 3.5 to 6 points		-.016 (.035)	-.002 (.034)	-.002 (.034)
Week of season dummies included?	No	No	Yes	Yes
Team dummies included?	No	No	No	Yes
R-squared	-----	.165	.299	.484
P-value of test of joint significance of:				
Spread variables	-----	<.01	<.01	<.01
Week dummies	-----	-----	<.01	<.01
Team dummies	-----	-----	-----	<.01

Notes: Omitted category for the spread variables are games in which the visiting team is favored by ten or more points. The unit of observation is a game. The number of observations is equal to 242 in all columns. Standard errors are in parentheses. The method of estimation is weighted least squares, with the weights proportional to the total number of bets placed on the game.

Table II: Bets Placed and Won on Favorites and Underdogs

Which team is favored in the game?	Percent of total bets on the game that are placed on:			Percent of bets placed that win (i.e. cover the spread) when a team is:		
	(1)	(2)	(3)	(4)	(5)	(6)
	Favorite	Underdog	Total, favorite and underdog	Favorite	Underdog	Total, favorite and underdog
Home team	56.1 [N=12,011]	31.8 [N=7,190]	47.0 [N=19,201]	49.1 [N=6,741]	57.7 [N=2,286]	51.2 [N=9,027]
Visiting team	68.2 [N=7,190]	43.9 [N=12,011]	53.0 [N=19,201]	47.8 [N=4,904]	50.4 [N=5,270]	49.1 [N=10,174]
Total, home and visiting team	60.6 [N=19,201]	39.4 [N=19,201]	50.0 [N=19,201]	48.5 [N=11,645]	52.6 [N=7,556]	50.1 [N=19,201]

Notes: The values reported in the first three columns of the table are the percentage of total bets placed on the named team (e.g. home favorite in row 1, column 1). The values reported in the last three columns of the table are the fraction of bets placed that win. The unit of analysis is a bet. The number in square brackets is the total number of bets placed in each cell. The results in this table exclude the six games where the spread was equal to zero, i.e. neither team was favored.

Table III: A Comparison of Betting Frequencies in my Sample and Game Outcomes against the Spread over the Last 21 Years

Type of game	Percent of bets in my sample that are placed on the favorite	Percent of games in which favorite wins over last 21 NFL seasons
All games	60.6	48.2 [N=4,793]
Games in which home team is favored	56.1	48.8 [N=3,310]
Games in which visiting team is favored	68.2	46.7 [N=1,483]
Games played in first half of season	63.4	47.7 [N=2,209]
Games played in second half of season	56.3	48.5 [N=2,584]
Games in which the point spread is greater than six points	60.4	48.5 [N=1,759]
Games in which the point spread is 3.5 to six points	59.5	48.1 [N=1,475]
Games in which the point spread is less than or equal to three points	61.5	47.8 [N=1,559]

Notes: Values in the first column are based on bets placed for the sample of 285 bettors for the 2001 season used throughout the paper. Values in the second column are game outcomes for the 21 NFL seasons covering the period 1980-2001.

Table IV: Bettor Performance as a Function of Cumulative Win Percentage up to that Point in the Season

	Dependent variable=bettor wins this bet				
Variable	(1)	(2)	(3)	(4)	(5)
Bettor's cumulative ranking on win percentage (0=last place, 1=first place)	-.012 (.015)	-.06 (.32)	.03 (.34)	.10 (.36)	.07 (.47)
Cumulative rank^2	-----	.54 (1.31)	.25 (1.36)	-.07 (1.47)	.21 (1.91)
Cumulative rank^3	-----	-1.04 (1.97)	-.69 (2.04)	-.18 (2.20)	-.68 (2.86)
Cumulative rank^4	-----	.56 (.98)	.41 (1.01)	.15 (1.09)	.41 (1.41)
Bettor bets on home favorite	-----	-----	-.048 (.010)	-.064 (.014)	-.037 (.016)
Bettor bets on home underdog	-----	-----	.046 (.012)	.026 (.014)	-.010 (.018)
Bettor bets on visiting favorite	-----	-----	-.005 (.010)	-.002 (.013)	.121 (.016)
Constant	.506 (.009)	.499 (.024)	.497 (.026)	.534 (.030)	.598 (.036)
R-squared	.0001	.0002	.0041	.0048	.0131
Number of observations	18,345	18,345	18,345	12, 985	6,785
Sample of games used in estimation	Weeks 2-17	Weeks 2-17	Weeks 2-17	Weeks 6-17	Weeks 11-17
P-value of test of joint significance of cumulative rank variables	-----	.64	.51	.73	.72
Predicted win percentage if cumulative ranking on win percentage falls in:					
Top quartile	49.4	49.0	49.0	49.7	49.0
Second quartile	49.8	50.2	50.1	50.4	50.5
Third quartile	50.1	50.7	50.8	51.0	51.2
Bottom quartile	50.4	49.9	49.9	50.1	49.4

Notes: The dependent variable in all regressions is equal to one if the bettor wins the bet, zero otherwise. The unit of observation is a bet. A bettor's cumulative ranking on win percentage is calculated by taking an unweighted average of that bettor's win percentage on all bets up to that point in time, and then computing the bettor's rank among all bettors. The method of estimation is weighted least squares, with the weights proportional to the total number of games making up the year-to-date history. Standard errors (in parentheses) are corrected to take into account correlation in the cumulative rank variable across bets for a given bettor.

Category of games:	Win percentage as a function of the degree to which bettors prefer this selection (relative to other teams in the same category):				
	All games, regardless of degree preferred by bettors	Top quartile (i.e. selections most popular with bettors)	Second quartile	Third quartile	Bottom quartile (i.e. selections least popular with bettors)
Home favorites	49.1 [N=6,741]	46.0 [N=1,666]	44.6 [N=1,697]	57.4 [N=1,696]	48.2 [N=1,682]
Visiting favorites	47.8 [N=4,904]	52.0 [N=1,240]	57.4 [N=1,210]	43.9 [N=1,206]	39.2 [N=1,205]
Home underdogs	57.5 [N=2,286]	71.0 [N=568]	52.6 [N=579]	59.9 [N=564]	47.5 [N=575]
Visiting underdogs	50.4 [N=5,270]	45.5 [N=1,325]	59.6 [N=1,288]	45.6 [N=1,318]	51.0 [N=1,339]
All favorites	48.6 [N=11,645]	48.6 [N=2,906]	49.9 [N=2,907]	51.8 [N=2,902]	44.4 [N=2,887]
All underdogs	52.6 [N=7,556]	53.2 [N=1,893]	57.4 [N=1,867]	49.9 [N=1,882]	49.9 [N=1,914]
All bets (favorites and underdogs)	50.1 [N=19,201]	50.4 [N=4,799]	52.9 [N=4,744]	51.0 [N=4,784]	46.6 [N=4,801]

Notes: The values reported in the table are fraction of bets won. The unit of observation is a bet. Games are categorized into quartiles within bet type (e.g. home favorites, visiting favorites, home underdogs, visiting underdogs) according to the fraction of bets on the team in the named category. Top quartile bets correspond to those cases where the greatest fraction of bettors selected the team in the named category.

Table VI: Does Pooling Information Across Bettors Help in Picking Winners?

Variable	Dependent variable=bettor wins game				
	(1)	(2)	(3)	(4)	(5)
Rank of this bet's popularity (relative to other games of this type) (0= least popular; 1= most popular)	.11 (.11)	.11 (.11)	-1.02 (1.82)	-1.16 (1.84)	-1.04 (1.85)
Rank of bet's popularity^2	-----	-----	4.81 (7.61)	5.48 (7.63)	5.18 (7.55)
Rank of bet's popularity^3	-----	-----	-6.77 (11.20)	-7.92 (11.23)	-7.57 (10.94)
Rank of bet's popularity^4	-----	-----	3.01 (5.41)	3.64 (5.44)	3.46 (5.21)
Bet on home favorite	-----	-.013 (.084)	-.013 (.083)	-.011 (.083)	-.013 (.082)
Bet on home underdog	-----	.073 (.072)	.073 (.072)	.080 (.074)	.055 (.078)
Bet on visiting favorite	-----	-.026 (.073)	-.026 (.073)	-.022 (.073)	-.054 (.070)
Constant	.447 (.048)	.450 (.062)	.499 (.119)	-----	-----
R-squared	.004	.008	.010	.022	.048
Include week of season dummies?	No	No	No	Yes	Yes
Include team-fixed effects?	No	No	No	No	Yes
P-value of joint significance of bettor preference variables	-----	-----	.81	.80	.78
Predicted win percentage by quartile of popularity of this bet:					
Top quartile (i.e. selections most popular with bettors)	46.1	46.1	44.5	44.6	44.6
Second quartile	48.8	48.8	49.4	49.5	49.8
Third quartile	51.5	51.4	54.2	54.1	54.0
Bottom quartile	54.2	54.1	52.3	52.3	52.1

Notes: The dependent variable in all regressions is equal to one if the bettor wins the bet, zero otherwise. The unit of observation is a bet. The rank of a bet's popularity is calculated relative to other games of this type (e.g. home underdog). The method of estimation is weighted least squares, with the weights proportional to the total number of bets placed on the game. Standard errors (in parentheses) are clustered to take into account correlation in the rank of bet popularity variables for different bets placed on the same game.