1994

Racetrack Betting: Do Bettors Understand the Odds?

Lawrence D. Brown
University of Pennsylvania

Rebecca D'Amato

Randy Gertner

Follow this and additional works at: https://repository.upenn.edu/statistics_papers

Part of the [Applied Mathematics Commons](#), [Applied Statistics Commons](#), and the [Probability Commons](#)

Recommended Citation

Brown, L. D., D'Amato, R., & Gertner, R. (1994). Racetrack Betting: Do Bettors Understand the Odds?. *Chance*, 7 (3), 17-23. <http://dx.doi.org/10.1080/09332480.1994.10542422>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/statistics_papers/554
For more information, please contact repository@pobox.upenn.edu.

Racetrack Betting: Do Bettors Understand the Odds?

Abstract

There is considerable literature which concludes that the average person does not understand elementary probability and statistics (Tversky 1971). In one experiment of this type subjects were asked whether a family of 6 children born in the order GBGBBG was more or less likely than one in which the birth order was BGBBBB. About 80% chose the first sequence in spite of the fact that both are approximately equally likely, with the second actually being slightly more probable since male births are slightly more common.

This experiment, like most in the field, is based on questions asked of subjects in a controlled, artificial setting. A naturally occurring setting in which the subjects have a continuing stake in the outcome is a potentially better way to determine whether or not the public understands probability. Examples of such settings are racetracks and stock markets. Racetracks provide the simpler arena; the choice of actions is more restricted and one can bet only for events, not against them as is possible in short selling on the stock market. We therefore analyzed racetrack betting in an attempt to discover the patrons' betting savvy, and we looked only at the simplest type of bet - bets on a horse to win.

Our first assumption is that bettors attempt to win money. Our second one is that they have an internal perception of which horse will win a given race. This perception results in their personal, subjective probability for the outcome of the race. We want to know whether this personal probability has any basis in reality.

Disciplines

Applied Mathematics | Applied Statistics | Probability

is possible in short selling on the stock market. We therefore analyzed racetrack betting in an attempt to discover the patrons' betting savvy, and we looked only at the simplest type of bet — bets on a horse to win.

Our first assumption is that bettors attempt to win money. Our second one is that they have an internal perception of which horse will win a given race. This perception results in their personal, subjective probability for the outcome of the race. We want to know whether this personal probability has any basis in reality.

Efficient Markets

If these personal probabilities are realistic then an efficient betting market would result. In such a market the money bet on each horse is an accurate reflection of the actual odds that horse has of winning. So, the distribution of bets is determined by the bettors subjective probabilities (in a way to be explained later). Thus the *mathematical definition of an efficient betting market is that the subjective odds are equal to the objective odds.*

If a race track betting market is not efficient, there could be three explanations. First, the bettors may not have an accurate perception of which horse will win a given race, and so they do not bet on the horses optimally. Second, bettors do have an accurate perception of how likely a horse is to win, but they do not understand how to bet based on that information. Lastly, bettors could have accurate personal probabilities in mind, but they value their profits in a nonstandard way.

In order to investigate the betting market we gathered data to determine whether subjective odds are indeed equal to objective ones. Before describing what we found, let us review the way in which racetrack betting operates.

Racetrack Betting

Consider a race with just three horses for simplicity. See Table 1. Suppose \$5,000, \$2,000 and \$1,000 respectively have been bet on the three horses to win. So, \$8,000 is the total pool. If the racetrack "take" is the typical 18% then the amount available for payoff is $\$8000 \times (1 - .18) = \$8000 \times .82 = \$6560$. If the first horse wins then the amount returned on each \$1 bet is $\frac{6560}{5000} = \$1.25$, etc., and the profit on that \$1 bet is \$0.25. (Newspapers generally publish the payoff on a \$2 bet so if "Reality" won this race the published payoff would be \$2.50 to win. Also, the track usually rounds the numbers in the last column of Table 1 down to the nearest 5¢, and pockets the difference, called "breakage". Thus, if "Optimism" were to win, the payoff would be \$3.25 on \$1.00 (= \$6.50 on \$2.00) instead of \$3.28 as shown in the table).

Horse's Name	Col.1 Amount Bet (\$)	Col.2 \$ After Removing Track Take $t = .18$	Col.3 \$ Payoff on a \$1 Bet	Col.4 \$ Profit on a \$1 Bet
Reality	5000		$\frac{6560}{5000} = 1.25$	.25
Optimum	2000		$\frac{6560}{2000} = 3.28$	2.28
Fantasy	1000		$\frac{6560}{1000} = 6.56$	5.56
	<hr/> 8000	<hr/> $6560 = 8000(1 - .18)$		

Table 1: Computation of Payoffs

Now suppose the bettor's subjective odds were $\frac{3}{4}$, $\frac{3}{16}$, $\frac{1}{16}$ respectively on the three horses. Then the expected payoffs from a \$1 bet would be calculated as shown in Table 2.

	Col.1 \$ Payoff (See Table 1)	Col.2 Subjective Probability (Hypothetical)	Col.3 Expected \$ Payoff	
Reality	1.25	3/4	.94	$(= 1.25 \times \frac{3}{4})$
Optimism	3.28	3/16	.62	$(= 3.28 \times \frac{3}{16})$
Fantasy	6.56	1/16	.44	$(= 6.56 \times \frac{1}{16})$

Table 2: Calculation of Expected Payoffs

Note that the largest expected payoff in Table 2 is .94 on "Reality." At the racetrack the information in Table 1 becomes available as the betting proceeds. If our hypothetical bettor sees the information in Table 1 as he proceeds to the betting window, then he/she will naturally bet on "Reality." This will raise the amount bet on that horse. As further bettors place their bets the amounts in Column 1 in Table 1 should grow in such a way that the expected payoffs in Column 3 of Table 2 are all equal.

An Algebraic Formula for the Subjective Probabilities

All this can be expressed algebraically. Let A_i denote the amount bet on horse i . Let D denote the pool after removing the track take, as shown at the bottom of Column 2, Table 1.

Let P_i denote the payoffs on the respective horses (Col. 3, Table 1) and let S_i denote the respective subjective odds (Col. 2, Table 2). Then $P_i = D/A_i$ and at equilibrium all expected payoffs are equal, so that

$$(1) \quad P_i S_i = \text{constant for all } i.$$

Since the S_i are probabilities they add to 1, and a little algebra yields that "constant" = $\frac{D}{\sum A_i} = 1 - t$, and

$$(2) \quad S_i = \frac{1 - t}{P_i}.$$

(A minor adjustment in the above reasoning is needed to take into account the track breakage. For this purpose one can algebraically compute the actual track take, including breakage, from the values of P_i as

$$T = 1 - \frac{1}{\sum \frac{1}{P_i}}.$$

It is this value which has been used in the analysis of data reported below.)

The Data

We recorded the payoff for each horse in 5500 races, and converted these payoffs to subjective probabilities as described above. These probabilities were then grouped into intervals. The intervals were chosen so that an approximately equal number of horses fell into each category. For instance, the interval of subjective probabilities $[0.091, 0.1)$ contained 1376 horses. The average subjective probability for this category was .0954.

There were 111 winners in this category. Thus the objective probability for this category was $\frac{111}{1376} = .0807$.

(Actually, although 5500 races were typed into the computer for analysis only about 80% of these races were used. We discarded any race whose total take, T , was outside of the interval $[\.165, \.21]$ since the racetrack takes generally vary from .17 to .19. We assumed that the data on any race whose take fell outside of this interval contained a typographical error.)

These subjective and objective probabilities provide mathematical quantities which we can compare. The subjective probabilities are related to the money bet and the objective ones are a reflection of reality. As mentioned previously if these quantities correspond the market is efficient. In the interval $[0.091, 0.1)$ discussed above these two quantities are nearly equal, though not exactly so (.0954 vs .0807).

subjective probability	number of horses	average subjective probability	objective probability
[0, 0.012)	1335.	0.00865025	0.00524345
[0.012, 0.021)	2357.	0.0166076	0.0101824
[0.021, 0.025)	1067.	0.023029	0.0229991
[0.025, 0.032)	1936.	0.0284558	0.0227273
[0.032, 0.038)	1512.	0.0349155	0.0271164
[0.038, 0.044)	1404.	0.0409321	0.0306268
[0.044, 0.05)	1334.	0.0469316	0.0517241
[0, 05, 0.058)	1697.	0.0540726	0.0465527
[0.058, 0.066)	1540.	0.0619449	0.0603896
[0.066, 0.075)	1635.	0.0705189	0.0654434
[0.075, 0.083)	1401.	0.0789255	0.0735189
[0.083, 0.091)	1295.	0.0869364	0.0803089
[0.091, 0.1)	1376.	0.0954372	0.0806686
[0.1, 0.1125)	1750.	0.106015	0.0977143
[0.1125, 0.125)	1583.	0.118743	0.123816
[0.125, 0.1375)	1393.	0.13134	0.122039
[0.1375, 0.15)	1228.	0.14397	0.138436
[0.15, 0.1675)	1625.	0.158536	0.160615
[0.1675, 0.185)	1469.	0.175819	0.184479
[0.185, 0.2)	1084.	0.192217	0.200185
[0.2, 0.225)	1502.	0.211831	0.221704
[0.225, 0.25)	1323.	0.237378	0.252457
[0.25, 0.2875)	1283.	0.267602	0.268901
[0.2875, 0.35)	1378.	0.314537	0.344702
[0.35, 1.00)	1472.	0.421119	0.444293

Table 3: Betting Data

Table 3 contains the results we obtained. A good way to see whether subjective and objective probabilities correspond is to plot one versus the other. This is done

in Figure 1. The line superimposed on the plot is the 45 degree line corresponding to a perfect match between subjective and objective odds. Our first impression from this table and matching plot is that the 45° line fits the data quite well. That is: *Race track bettors do a creditable job of prediction.*

Figure 1: Betting Data, with 45° Line

(Data structurally similar to that in Table 3 and Figure 1 has been reported before. For the best summary see Asch and Quandt (1956). Those earlier results

show a somewhat larger deviation from the 45° line than do ours. It is not clear to us that the data in earlier studies was edited in order to remove the noticeable percentage of flawed result ($T \notin (.17, .21)$), nor adjusted to take account of track breakage.

Furthermore, the analyses reported in the remainder of our article require complete results, and not just summaries like Table 3, and such complete results from earlier studies were not available to us.)

Of course, the points in Figure 1 do not lie exactly on the 45° line. Their departure from a perfect relationship is emphasized in Figure 2 which plots log (objective probabilities) versus log (subjective probabilities).

Reasons have been suggested for patrons to bet less on favorites than they should and more on longshots than they should. (See the discussion of utility, below; and also Asch and Quandt (1986).) The result of such a bias for longshots over favorites would be as follows:

Where the subjective probabilities are high (favorites) the objective probabilities would be even higher (i.e., above the 45° line). Thus these subjective probabilities underestimate the corresponding objective ones. Symmetrically, where the subjective probabilities are low (longshots) the objective probabilities would be even lower (i.e., below the 45° line).

Such a pattern can be detected in Figures 1 and 2 since the objective probabilities generally lie below the 45° line at the left and above it at the right.

Figure 2: Betting Data expressed in Logs, with 45° Line

A Linear Relationship for the Longshot/Favorite Bias?

The next step in our analysis was to test the significance of this apparent longshots-over-favorites bias. Thus we performed a statistical analysis to determine whether the pattern we observed could have occurred by chance if there were

no real longshot/favorite bias among the general betting public. In this case, the deviation we observed from the 45° line would be simply a chance accident. Such a test required us to specify alternative models for the relationship between the objective and subjective probabilities. Thus we tested the null hypothesis of equality versus three increasingly complex alternative possibilities. The simplest of these alternatives was that of a linear relationship other than the null relationship of equality. Symbolically, the hypotheses were

$$H_0 : \text{obj} = \text{subj}$$

$$H_{11} : \text{obj} = a_0 + a_1(\text{subj})$$

$$H_{12} : \text{obj} = a_0 + a_1(\text{subj}) + a_2(\text{subj})^2$$

$$H_{13} : \text{obj} = a_0 + a_1(\text{subj}) + a_2(\text{subj})^2 + a_3(\text{subj})^3$$

These hypotheses were tested by a method involving the Chi-squared statistic. This method took into account the varying number of horses in each race as well as the fact that there is a relation of dependence among the objective outcomes. (For example, if Reality wins then Fantasy loses.) The test results can be summarized in a stepwise table, analogous to that encountered in a stepwise analysis of variance. The results are summarized in Table 4.

Alternative Hypothesis	Sum of Squares	Null Distribution	P-value
H_{11} (but not H_0)	20.31	χ^2_1	< .0001
H_{12} (but not H_{11})	2.38	χ^2_1	0.13
H_{13} (but not H_{12})	2.32	χ^2_1	0.13
Remainder	12.95	χ^2_{21}	0.91
Total	37.96	χ^2_{24}	0.04

Table 4: Hypothesis Tests

Note that the P -value for H_{11} but not H_0 is extremely small ($< .0001$). This means that the sample could not plausibly have occurred if only H_0 were true, and leaves us to conclude that H_{11} is a better description of the true situation. The corresponding linear relation which best fits the data is

$$(3) \quad \text{obj} = -0.0067 + 1.1165 \times \text{subj.}$$

Figure 3 shows the data of Figure 1 and also this line. Note that as conjectured this line is close to the 45° line, but is very slightly below it when the subjective probabilities are small and somewhat above it when they are large.

Figure 3: Betting Data with Linear Fit (3).

The remaining P -values in the body of Table 4 are not small. Thus a quadratic (H_{12}) or cubic (H_{13}) curve does not seem a significantly more accurate expression of reality than the above linear curve. In addition, the large (and hence, insignificant) P -value associated with the "Remainder" term is an indication that there is no other significant, simple relation between subj. and obj. (The χ^2 statistic and P -value associated with the "Total" in Table 4 reflects the effect of all possible alternatives to H_0 . This P -value is fairly small because the linear equation corresponding to H_{11} is a much better explanation of the data than is H_0 .)

In conclusion:

There appears to be a small longshot over favorite bias. It can be reasonably well described by the linear equation (3).

An Explanation from Utility Theory

Thus far in our analysis we have attempted to document and quantify the longshot/favorite bias. Now let us turn to a possible explanation for it. We will look at the third reason why a racetrack may be inefficient as expressed above; the bettors may value their winnings in a nonstandard way.

Why should bettors opt to overbet on horses that are unlikely to win. It has been conjectured that the large payoffs on these bets offer something more than monetary satisfaction; they offer excitement. (See e.g. Asch and Quandt (1986) or Thaler (1992).) A bettor would thus prefer to net \$100 from a lucky bet on a longshot than to net ten \$10 payoffs from bets on favorites. A payoff on a favorite lacks excitement. A utility function can put this excitement into mathematical terms. The utility function u (payoff) estimates the amount of satisfaction a bettor

gets from a particular payoff.

A bettor with utility function u (payoff) should bet so that at equilibrium

$$(4) \quad u(P_i)S_i = \text{constant} \quad (\text{for all } i).$$

(Compare (4) to (1) and note that they agree when $u(P_i) = P_i$.) Algebra now yields

$$(5) \quad S_i = \frac{1 - T'}{u(P_i)} \quad \text{where} \quad T' = 1 - \frac{1}{\sum \frac{1}{u(P_i)}}$$

at equilibrium as the new version of (2).

This was the starting point for deriving a likelihood ratio test that tested

$$H_0 : u(P) = P$$

$$\text{vs. } H_1 : u(P) = a + bP.$$

The appropriate χ^2 statistic has one degree of freedom, and turned out to be 18.3. This has a P -value < 0.001 , and so is highly significant. It thus appears that one explanation for the results in Table 1 is that bettors behave according to a nonstandard utility. Based on our likelihood analysis, the best estimate for this utility is the linear form

$$(5) \quad u(P) = -.58 + P.$$

Note that P is the payoff on a winning \$1 bet and so is always bigger than 1. Thus in the three horse race of Table 1 the payoffs on the three horses are in the ratio $\frac{1}{5} : \frac{1}{2} : 1 = 2 : 5 : 10$ but the utilities under (5) are in the ratio $.67 : 2.7 : 5.98 \approx 2 : 8.1 : 18$.

Which Explanation is Correct?

You may note that three different explanations have now been implicitly proposed for the deviation from 45° in Figure 1 and its associated Table 3. One is that bettors do not quite correctly evaluate horses abilities. Another is that they do but that they do not quite know how to use these evaluations to determine their bets. The third explanation is that they are behaving intelligently, but according to a utility function such as (5) rather than the more usual $u(P) = P$. It appears that some new data of a different sort would be needed to evaluate which — if any — of these explanations is correct. You might want to speculate what data would be needed, and how it could be effectively gathered.

The Last-Race Phenomenon

It has been suggested (Ziembra and Hausch (1987)) that bettors bet differently during the last race of the day. This makes sense, since by the end of the day, most people have lost money. Perhaps they feel that the best way to earn back that money is to bet on longshots, since they provide a quick high profit. Therefore, one would expect to find the longshot/favorite bias more pronounced during the last race of the day.

The results for the last race of the day are summarized in Table 5, Figure 4, and Table 6. The data does point somewhat in the direction of a longshot/favorite bias. Indeed, the underbetting on favorites appears more pronounced than in the data for all races. On the other hand there seems to be very little overbetting on longshots. None of the hypothesis tests in Table 6 are significant. (I.e., All P -values are above .05.) This lack of significance may be because there is no longshot/favorite bias

here or because the sample size here is much smaller than that in Table 3, and so is not large enough to confirm the bias which does exist.

Figure 4: Data for the Last Race of the Day, with 45° Line

subjective probability	number of horses	average subjective probability	objective probability
[0, 0.032)	1012	0.01889	0.01779
[0.032, 0.044)	390	0.03786	0.02308
[0.044, 0.058)	343	0.05091	0.04956
[0.058, 0.075)	359	0.06611	0.06407
[0.075, 0.091)	355	0.0826	0.0873
[0.091, 0.1125)	349	0.1013	0.1003
[0.1125, 0.1375)	336	0.1248	0.1250
[0.1375, 0.1675)	291	0.1515	0.1100
[0.1675, 0.2)	264	0.1827	0.1591
[0.2, 0.25)	266	0.2236	0.2481
[0.25, 0.35)	264	0.2912	0.3295
[0.35, 1.00)	95	0.4065	0.4947

Table 5: Results for the Last Race of the Day

Alternative Hypothesis	Sum of Squares	Null Distribution	P-value
H_{11} (but not H_0)	2.42	χ^2_1	0.12
H_{12} (but not H_{11})	3.58	χ^2_1	0.06
H_{13} (but not H_{12})	0.56	χ^2_1	0.46
Remainder	8.22	χ^2_8	0.42
Total	13.88	χ^2_{11}	0.25

Table 6: Hypothesis Tests for the Last Race of the Day

REFERENCES

- Asch, P. and Quandt, R. E. (1986), *Racetrack Betting: The Professors' Guide to Strategies*, Auburn House, Dover, MA.
- Thaler, R. H. (1992), *The Winner's Curse: Paradoxes and Anomalies of Economic Life*, Free Press, NY.
- Tversky, A. (1971), *Subjective probability: A judgement of representativeness*, *Cognitive Psychology* 3, 430-454.
- Ziembra, W. T. and Hausch, D. B. (1987), *Dr. Z's Beat the Racetrack*, W. Morrow, NY.