Peter Asch
Rutgers University

Burton G. Malkiel
Yale University

Richard E. Quandt
Princeton University

Market Efficiency in Racetrack Betting

Do market inefficiencies in racetrack betting permit profitable strategies to be devised on the basis of observed betting patterns? Logit analysis of market data indicates that profits cannot be earned in win betting, although it is possible to outperform the average bettor substantially. Surprisingly, profits net of transactions costs on place and show betting appear possible, although there are reasons to suspect that such opportunities may not be exploitable on a substantial scale. The findings do not suggest irrational behavior by bettors.

I. Introduction

Betting at racetracks and investing in the stock market have several characteristics in common. In neither situation are future earnings known with certainty. Both situations are characterized by a large number of participants and by the availability of extensive information and professional advice. Horse racing data, therefore, permit interesting tests of the efficient market theory.

Numerous authors have analyzed horse racing data and various conclusions have been confirmed: (1) There is a systematic tendency to overbet long-shots and underbet favorites (Rosett 1965, 1971; Snyder 1978; and Ali 1979). (2) The tendency to overbet long-shots is strongest in late races (McGlothlin 1956; Asch, Malkiel, and Quandt 1982). (3) The utility function of bettors is convex, indicating risk love (Weitzman 1965; Asch et al. 1982). Figlewsky (1979) tested hypotheses concerning whether the betting "market" is efficient with respect to track odds and handicappers' picks of winning horses. By fitting a logit model with horses' win-loss records representing the dependent variable and track odds and handicappers' picks the independent variables, it is shown by likelihood ratio

(*Journal of Business*, 1984, vol. 57, no. 2)

0021-9398/84/5702-0003$01.50

tests that handicappers' picks are substantially accounted for in the track odds. Finally, Hausch, Ziemba, and Rubinstein (1981) analyzed betting horses to place (come in first or second) or to show (first, second, or third) and found that inefficiencies exist in place and show betting.

The present paper analyzes a new sample of racetrack results from Atlantic City, New Jersey. We investigate how well the track odds and the morning line odds (rather than handicappers' picks) predict the victors of races. But there is an important further question that requires attention. If track odds and other publicly available information do very well in predicting winners, one might want to use such information in betting; yet this use of information will tend to lower the odds (and payoffs) of the horses predicted to win.

The pertinent question, then, concerns the efficiency of the racetrack betting market. If efficiency is perfect, the rate of return to every betting strategy based on available information should approximate closely the track "take." For the sample of races analyzed here, the take is about 18.5%; thus investors who pursue any strategy should earn a negative return close to 18.5%. It is of particular interest to ask, however, not whether the racetrack betting market is efficient in this very stringent sense, but whether there exist departures from efficiency sufficiently large to permit profitable betting strategies. Can one, in other words, devise strategies based on observable betting patterns that imply positive rates of return? Only if such unexploited arbitrage opportunities were being consistently neglected could we conclude that the racetrack betting market was inefficient.[1]

A second question is, If one were able to predict winners by using publicly available information, would such a finding be inconsistent with the assumption of rational behavior in the racetrack market? Rosett (1965) has shown that rationality imposes requirements on the relationship between return and the probability of winning by exploiting the possibility of combination strategies, such as parlays. He found specifically that the return to low-win-probability horses appeared lower than predicted by the rationality assumption. We comment on this issue in the concluding section.

1. An analogy may be made to empirical research in the stock market. Departures from perfect efficiency have often surfaced in empirical tests, such as some findings of serial correlation and evidence of inferrable information from "insider" trading that is not immediately reflected in market prices. Nevertheless, the conclusion remains that, in general, any departures from efficiency are not large enough to permit an investor to devise an active strategy that can consistently beat a passive buy-and-hold strategy. In this sense, the empirical evidence generally supports a finding of efficiency for the stock market.

II. Data and Definitions

The Data Set

Observations are based on the entire 1978 thoroughbred racing season at the Atlantic City Race Course, which includes 712 races and 5,714 horses. The data are

1. The "morning" line odds for each horse in each race, determined by the track's professional handicapper and printed in the daily racing program. These are the handicapper's estimates of the winning probabilities for each horse that confront bettors before the start of each betting period.

2. Parimutuel odds for each horse in each race at various points ("cycles") during the betting period. Twenty-four cycles are typically recorded, showing the minute-by-minute course of the actual betting for each race. Only some of the available cycles were employed in the analysis.

3. Final parimutuel odds for each horse, given at the final betting cycle. The odds are determined by the ratio of the betting pool available for distribution to the amount bet on each horse. The racetrack subtracts a percentage or "take" from the total betting pool to cover taxes, expenses, and profits. In addition, the track enjoys "breakage," the gain from being allowed to round payoffs downward (to the nearest 10 cents or 20 cents on a \$2.00 bet), or to round odds downward (to the nearest tenth). The total betting pool minus the take, including breakage, is paid on the winning horse.

4. The outcome of each race.

Some Definitions

The following definitions are employed: B = the total amount bet on the race; b_h = the amount bet on horse h, $h = 1, 2, \ldots, H$, where H is the number of horses in the race. It follows that

$$B = \sum_{h=1}^{H} b_h.$$

Let t = the total take of the track including breakage, all expressed as a percentage of the total bet. The odds of horse h may then be stated as

$$O_h = \frac{B(1 - t) - b_h}{b_h} = \frac{B(1 - t)}{b_h} - 1.$$

We define the (bettors' subjective) probability that horse h will win the race, P_h. It follows then that

$$P_h = \frac{b_h}{B} = (1 - t)/(1 + O_h)$$

One can also define implicit or marginal odds derived from bets made late in the betting period. The betting period is approximately 25 minutes, the time between races. Define cycle 1 as the betting cycle encompassing the first 17 minutes (the first ⅔) of the betting period, and cycle 2 as the betting in the last 8 minutes (the last ⅓) of the betting period. The marginal odds on horse h in betting cycle 1, $O_{h,1} = [B_1(1 - t)/b_{h,1}] - 1$, where B_1 is the total amount bet up to the end of cycle 1, and $b_{h,1}$ is the amount bet on horse h up to the end of cycle 1. The final odds on horse 1 (which are also the odds on the horse derived from all betting up to and including cycle 2), are

$$O_{h,2} = \frac{B_2\,(1 - t)}{b_{h,2}} - 1,$$

where B_2 is the total amount bet through both cycles (1 and 2) and $b_{h,2}$ is the total amount bet on horse h through both cycles (1 and 2).

Define the marginal odds as the odds derived from the betting during a particular cycle, but not including previous cycles. We seek the marginal odds created by bettors in cycle 2, namely, the bets of the "late money," or what is sometimes hypothesized to be the "smart money." The argument is that bettors with true inside information will prefer not to signal that information to the public until late in the betting cycle to minimize "following" behavior on the part of other track bettors.

The marginal odds on horse h in cycle 1 are simply the regular odds, that is:

$$O^m_{h,1} = O_{h,1}.$$

The marginal odds on horse h in cycle 2 can be derived from the odds at the end of cycles 1 and 2, as follows:

$$O^m_{h,2} = \frac{O_{h,2}b_{h,2} - O_{h,1}b_{h,1}}{b_{h,2} - b_{h,1}}.$$

Finally, define a payoff W_h from horse h as unity plus the final odds if h wins and zero otherwise. A horse that goes off at odds of three to one pays \$4.00 for each dollar bet if the horse wins. A rate of return R from betting one dollar on one horse in each of N races is $R = (\Sigma W_h - N)/N$, where ΣW_h is the sum of payoffs in the N races.

III. Models and Results

Define W_{ij} as the win function of the ith horse in the jth race, by

$$W_{ij} = \boldsymbol{\beta}'\mathbf{x}_{ij} + u_{ij},$$

where $\boldsymbol{\beta}'$ is a vector of coefficients, $\mathbf{x}_{ij}$ a vector of observable variables, and u_{ij} an error term. W_{ij} may be thought to measure the "winningness" of a horse and to depend on a systematic part ($\boldsymbol{\beta}'\mathbf{x}_{ij}$) and a stochastic part. The ith horse wins the jth race if $W_{ij} > W_{lj}, \ldots, W_{ij} > W_{n_j j}$, where n_j is the number of horses entered in the jth race. If the u_{ij} are assumed to have independent extreme value (Weibull) distributions, this leads to the well-known logit model, giving the probability that the ith horse wins the jth race as

$$P_{ij} = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ij})}{\sum_{k=1}^{n_j} \exp(\boldsymbol{\beta}'\mathbf{x}_{kj})}.$$

Estimates of the $\boldsymbol{\beta}$'s are obtained by maximizing the likelihood function. It is well known that the logit model possesses the independence of irrelevant alternatives property: the relative odds for horses i and k (P_{ij}/P_{kj}) depend only on the characteristics of these horses. As a first approximation this appears to be reasonable for horse racing.

Table 1 displays some basic results. All independent variables were odds converted to probabilities. Thus *ML* represents the win probabilities implied by the morning line odds, *F* the probability based on the final odds, and *M*2 the probability based on the marginal odds for cycle 2 (the last ⅓ of the betting period). The variable *M*2 is included in some of the computations to account for the possibility that the late bettors (smart money) have better information than the aggregate of all bettors. All coefficients are highly significant and, comparing runs 1, 2, and 3 with 4 and 5, respectively, the likelihood ratio test emphatically rejects the null hypothesis that the morning line odds or the final odds or the marginal odds alone provide a satisfactory explanation. In particular, we reject the null hypothesis that the morning line odds contain no information not already accounted for by the track odds. The probability of winning is positively and significantly affected by the morning line and final track probabilities as well as by the marginal probabilities. Including *F* and *ML* as well as *M*2 causes the standard errors to

TABLE 1 **Logit Results**

	Logit Run				
Variable	1	2	3	4	5
ML	9.99	...	...	4.15	4.74
	(.54)	...	...	(1.02)	(.96)
F	...	6.42	...	4.26	...
	...	(.33)	...	(.62)	...
*M*2	...	...	5.74	...	3.54
	...	...	(.30)	...	(.54)
Log likelihood	−1,280.0	−1,265.0	−1,270.0	−1,257.0	−1,258.0

NOTE.—Asymptotic standard errors in parentheses.

increase sharply, and no further significant improvement occurs in the log likelihood when both these variables are included.[2]

Simulations Using Only Win Bets

Using the results of the logit estimation one can perform the following simulations. For each race, using the coefficients in run 4 or 5 one may compute the win probability $\hat{P}_{ij}$ for each horse. We may then consider the following strategies:

Strategy 1: In each race $j(j = 1, \ldots, N)$ bet \$2.00 on the horse with the highest $\hat{P}_{ij}$.

Strategy 2: In each race bet \$2.00 on the horse with the highest $\hat{P}_{ij}$ if $\hat{P}_{ij} \geqq \gamma$, where γ is a "filter" coefficient.

Strategy 3: In each race bet \$2.00 on the horse with the highest $\hat{P}_{ij}$, if the highest $\hat{P}_{ij}$ is at least 3.5 times greater than the second highest $\hat{P}_{ij}$. Strategy 3 differs from strategy 2 in its concern, not for a high $\hat{P}_{ij}$ per se, but rather for selecting horses to bet only if the second choice by the criterion function is sufficiently inferior to the first choice.

Running such simulations on the sample from which the coefficients were obtained would bias the results. We therefore divided the sample into two halves and reestimated runs 4 and 5 from each of the two halves.[3] We then used the coefficients from each half sample for simulations on the other half sample.

For each simulation one can compute the rate of return for each of the strategies. These are displayed in table 2. The results show that all but one of the rates of return are better than the rate of return implied by the track take (-0.185) if the market is perfectly efficient, as defined above. Averaging the rates over the two half samples shows that marginally better results are obtained with run 5, that is, when the marginal odds pertaining to the last 8 minutes of betting time are used instead of the final odds. For one of the strategies the average rates of return indicate that betting can almost be a fair game. The results demonstrate clear departures from perfect efficiency but do not suggest the possibility of betting strategies that are profitable after accounting for the track take.

Simulations Using Place Bets and Show Bets

An alternative to placing win bets is to bet the selected horse to place or to show. The selection of the horse in each race was based on the same strategies discussed before but in the present simulations place or show bets were made. The rates of return are shown in table 3. All

2. The standard errors increase because of collinearity between F and M_2.

3. The coefficients from the half samples and the results of the likelihood ratio tests comparing the half-sample runs 4 and 5 with half-sample runs 1, 2, and 3 are similar to those reported in table 1.

TABLE 2 **Rates of Return in Simulations Using "Win Bets"**

	Estimation: First Half Simulation: Second Half		Estimation: Second Half Simulation: First Half		
	N	*R*	*N*	*R*	*A*
Run 4:					
Strategy 1	356	−.160	356	−.168	−.164
Strategy 2:					
$\gamma = .4$	132	−.111	100	−.101	−.106
$\gamma = .5$	62	−.077	57	.002	−.038
$\gamma = .6$	31	−.145	22	−.141	−.143
Strategy 3	52	−.150	51	−.047	−.098
Run 5:					
Strategy 1	356	−.186	356	−.116	−.151
Strategy 2:					
$\gamma = .4$	127	−.091	107	−.028	−.068
$\gamma = .5$	61	−.070	58	−.052	−.061
$\gamma = .6$	28	−.114	21	−.090	−.102
Strategy 3	47	−.183	54	−.007	−.095

NOTE.—N = number of horses; R = rate of return, A = average of rates of return over the two half samples.

TABLE 3 **Rates of Return in Simulations Using Place Bets and Show Bets**

	Run 4					Run 5				
	Est.: First Half Sim.: Second Half		Est.: Second Half Sim.: First Half			Est.: First Half Sim.: Second Half		Est.: Second Half Sim.: First Half		
	N	*R*	*N*	*R*	*A*	*N*	*R*	*N*	*R*	*A*
A. Place bets:										
Strategy 1	356	.160	356	.127	0.143	356	.187	356	.165	.176
Strategy 2:										
$\gamma = .4$	132	.241	100	.183	0.212	127	.258	107	.269	.263
$\gamma = .5$	62	.328	57	.214	0.271	61	.301	58	.198	.249
$\gamma = .6$	31	.179	22	.527	0.353	28	.130	21	.400	.265
Strategy 3	52	.253	51	.190	0.221	47	.224	54	.287	.255
B. Show bets:										
Strategy 1	356	.090	356	.046	0.068	356	.094	356	.052	.073
Strategy 2:										
$\gamma = .4$	132	.133	100	.054	0.093	127	.140	107	.085	.112
$\gamma = .5$	62	.202	57	.016	0.109	61	.149	58	.006	.087
$\gamma = .6$	31	.100	22	.189	0.144	28	.064	21	.126	.095
Strategy 3	52	.108	51	.041	0.074	47	.102	54	.135	.118

NOTE.—See table 2 and n.

returns are positive. Betting the horse with the highest win probability to place yields a higher return than betting it to show. The number of horses on which bets are made under strategy 2 with $\gamma = .5$ or $.6$ and under strategy 3 is sufficiently small so that one cannot place great confidence in the resulting return figures. But for strategy 1, for example, involving 356 bets in each subsample, the return to place betting ranges from 0.143 to 0.176. Finally, one may note that for those strategies which allow a large number of horses to be bet on (strategy 1 and strategy 2 with $\gamma = .4$), the results of run 5, which employ the morning line and marginal odds, are slightly better.

Even simpler strategies can yield positive returns. One could, for example, predict the winner on the basis of logit runs 1, 2, or 3, employing a single explanatory variable.[4] The results are broadly comparable but somewhat less favorable. Thus, for example, strategy 1 yields an average return from place bets (over the two subsamples) of 0.081 when the morning line alone is used to predict the winner, 0.076 when the final odds are used, and 0.101 when the marginal odds are employed. This still compares favorably with placing win bets on the favorites, from which the return is -0.173.

We were not able to improve on the results of table 3 by employing more sophisticated compound strategies. The winners and runners-up can, for example, be predicted from a generalization of the logit model that explicitly takes into account not only which horse won, but which horses were second and third.[5] These predictions have been employed for combination strategies involving win, place, and show bets simultaneously.[6] The most successful of these combination strategies yielded average returns of 0.133–0.140, depending on which explanatory variables were used in the generalized logit estimation.

IV. Conclusions

The results from win betting strategies show that strategies can be developed that exceed the return from simply betting the favorite to win. The results from using place and show bets are surprising in the sense that we now find that net profits can be made. Several conjectures may be offered to explain this finding. It is possible, for example, that the tendency to overbet long shots is accentuated in place and

4. For strategy 1 this is equivalent to betting on the favorite.

5. The generalization follows from Harville (1973).

6. The most successful strategy requires predicting for each race three highest win probabilities denoted by $P_1 \geqq P_2 \geqq P_3$, where 1, 2, and 3 are the indices of the corresponding horses. According to this strategy, if (i) $P_1 - P_2 > .25$ and $P_2 - P_3 > .15$, bet horse 1 to win, 2 to place, 3 to show; (ii) if $P_1 - P_2 \leqq .25$ and $P_2 - P_3 > .15$, bet horse 1 to place, 2 to show; (iii) if $P_1 - P_2 > .25$, $P_2 - P_3 \leqq .15$, bet 1 to win and 1 to place; (iv) if $P_1 - P_2 \leqq .125$, $P_2 - P_3 \leqq .15$, bet 1 to show.

show betting. It is also possible that all the information contained in the win pool may not be efficiently impounded in the place and show pool. This finding confirms the broad results of Hausch et al. (1981).

Are these results incompatible with the assumption of rationality? We think not. First, we may ask whether the relationship between return and the probability of winning predicted by Rosett (1965) would still hold today. The reason is that the opportunities for low-probability parlay-type bets at racetracks have greatly increased in the last decade and there is now a plethora of exactas, trifectas, "pick fours," and so on, that are more profitable than parlays in the sense that they involve only one track take rather than n. More important, however, it is quite possible, as Rosett suggests, that even today the market fails to provide enough low-probability betting opportunities, given a substantial "taste" by the betting public for such bets. This clearly will have the effect of lowering the return on low-probability bets and is compatible with our results. Our findings that positive returns can be earned do not suggest irrational behavior; rather they suggest that the taste for certain nonpecuniary benefits associated with certain types of bets make positive profits possible for some bettors.

In other words, it is not a paradox that bets on different horses have different expected returns (and variance of returns); this is simply a consequence of bettors' having different utility functions and selecting different points among the available opportunities (characterized by different objective and subjective winning probabilities).

It should be noted, in conclusion, that these results cannot be interpreted as firm evidence of unexploited arbitrage opportunities for at least two reasons. First, for any strategy that employs more than just morning line odds, there is a serious difficulty in implementation, since final odds are typically not posted until thirty or more seconds after the betting windows close. Second, the place and show pools are usually much smaller than the win pools. Any attempt to employ the implied arbitrage strategies may prove infeasible since a relatively small volume of betting could quickly wipe out the apparent profit opportunity. Finally, it goes without saying that our simulations need to be repeated over alternative and large samples before we can have considerable confidence in the results.

References

Ali, Muktar M. 1979. Some evidence of the efficiency of a speculative market. *Econometrica* 47:387–92.

Asch, P.; Malkiel, Burton G.; and Quandt, Richard E. 1982. Racetrack betting and informed behavior. *Journal of Financial Economics* 10:187–94.

Figlewsky, Stephen. 1979. Subjective information and market efficiency in a betting market. *Journal of Political Economy* 87:75–88.

Harville, D. A. 1973. Assigning probabilities to outcomes of multi-entry competition. *Journal of the American Statistical Association* 68:312–16.

Hausch, D. B.; Ziemba, W. T.; and Rubinstein, M. 1981. Efficiency of the market for racetrack betting. *Management Science* 27:1435–52.
McGlothlin, W. H. 1956. Stability of choices among uncertain alternatives. *American Journal of Psychology* 69:604–15.
Rosett, R. N. 1965. Gambling and rationality. *Journal of Political Economy* 73:595–607.
Rosett, R. N. 1971. Weak experimental verification of the expected utility hypothesis. *Review of Economic Studies* 38:481–92.
Snyder, Wayne N. 1978. Horse racing: Testing the efficiency markets model. *Journal of Finance* 22:1109–18.
Weitzman, M. 1965. Utility analysis and group behavior: An empirical study. *Journal of Political Economy* 73:18–26.