

MANAGEMENT SCIENCE
Vol. 32, No. 8, August 1986
Printed in U.S.A.

SEARCHING FOR POSITIVE RETURNS AT THE TRACK: A MULTINOMIAL LOGIT MODEL FOR HANDICAPPING HORSE RACES*

RUTH N. BOLTON AND RANDALL G. CHAPMAN

*Department of Marketing & Economic Analysis, Faculty of Business, University of Alberta,
Edmonton, Alberta, Canada T6G 2R6*

This paper investigates fundamental investment strategies to detect and exploit the public's systematic errors in horse race wager markets. A handicapping model is developed and applied to win-betting in the pari-mutuel system. A multinomial logit model of the horse racing process is posited and estimated on a data base of 200 races. A recently developed procedure for exploiting the information content of rank ordered choice sets is employed to obtain more efficient parameter estimates. The variables in this discrete choice probability model include horse and jockey characteristics, plus several race-specific features. Hold-out sampling procedures are employed to evaluate wagering strategies. A wagering strategy that involves unobtrusive bets, with a side constraint eliminating long-shot betting, appears to offer the promise of positive expected returns, even in the presence of the typically large track take encountered at Thoroughbred racing events.

(MULTINOMIAL LOGIT MODEL; HORSE RACE WAGERING; STOCHASTIC UTILITY
MODEL; RANKED CHOICE SET DATA; DISCRETE CHOICE MODELING)

Introduction

For as long as there have been horse races, bettors have searched for profitable wagering systems. The general form of any horse race wagering system involves betting against the public. If the public makes systematic and detectable errors in establishing the betting odds, it may be possible to exploit such a situation with a superior wagering strategy and make wagers with a positive expected rate of return.

Academic researchers have also searched for profitable wagering systems to evaluate the efficiency of horse race wager markets (cf., Ziemba and Hausch 1984). Such investigations have been motivated by the basic similarities of race track and stock markets, such as uncertain future returns from investments, the presence of many participants, and the availability of a variety of information concerning investments and participants. (See Copeland and Weston 1979, Fama 1970, 1975, or Rubinstein 1975 for discussions of market efficiency.)

This paper searches for a profitable wagering system to apply to win-betting in a pari-mutuel setting. Wagering systems have two components: a model of the horse race process and a wagering strategy. A model of the horse race process attempts to predict the outcome of a race. Its main output is a prediction of the probabilities of each horse winning a race. A wagering strategy then uses these probabilities as inputs to a betting algorithm which determines the amounts to wager on each horse.

Academic researchers have tended to focus primarily on devising a betting algorithm to determine the amounts to wager on each horse. These algorithms may be categorized according to whether they require knowledge of each horse's true winning probabilities. Assuming these probabilities are known with certainty, optimal wagering theorems for win-betting in a pari-mutuel setting have been developed for the expected value maximizer with infinite wealth (Isaacs 1965) and for the risk averse decision maker

* Accepted by William T. Ziemba; received August 14, 1983. This paper has been with the authors 11 months for 4 revisions.

(Rosner 1975). If the true winning probabilities are not known with certainty, wagering strategies may be fundamental (Arveson and Rosner 1971, 1973) or technical (Willis 1964; Harville 1973; Hausch, Ziemba, and Rubinstein 1981) in nature.

Previous research concerning the first component—modeling the horse race process—focused on the public's odds preferences, the existence of inside information, or handicapping ability. If the public consistently underestimates the true winning probabilities for horses in the low odds range (Fabricand 1965), then playing the favorites will improve upon a random betting strategy, but such a strategy may not yield positive returns (Snyder 1978; Asch, Malkiel, and Quandt 1982; Ali 1979). If horse race wager markets are weakly efficient (Dowie 1976), it may be possible to earn extraordinary returns by exploiting publicly available or inside information. Rosett (1965) and Ali (1979) found that horse race data are consistent with the efficient market hypothesis. However, handicapping systems based on publicly available information may yield positive returns (Vergin 1977).

In this paper, we focus on developing a statistical model of the horse race process. The multinomial logit model is used to analyze the horse race process since it explicitly recognizes that there are only a finite number of outcomes to a horse race, namely that one of the entered horses wins. It explicitly analyzes the effects of competition in modeling these outcomes. A recently developed approach to estimating the multinomial logit model's parameters is employed (Chapman and Staelin 1982): the rank order finishing data are exploited to yield improved statistical efficiency of the parameter estimates. Hold-out sampling is used to assess the model's predictions as inputs to various wagering strategies for pari-mutuel win-betting.

The Pari-Mutuel System

In a pari-mutuel system, bettors place wagers on a set of horses in a given race. These wagers form the betting pool from which δ , the track take, is deducted. Approximately 18% goes to the track and the various levels of government, although this amount varies across jurisdictions. The specific track take in any race is also influenced by "breakage," the practice of rounding payoffs down to the next lowest nickel or dime. (The analysis in this paper ignores breakage.) The remainder of the betting pool is allocated to the bettors on the winning horse in proportion to their bets. Hence, the final track probabilities are proportional to the amounts bet on the horses by all bettors. The pari-mutuel probability for horse h , π_h , can be written as follows:

$$\pi_h = w_h/W \quad (1)$$

where w_h is the total amount wagered on horse h by the public and W is the total size of the win-betting pool. These probabilities represent the public's consensus probabilities as reflected by their wagering patterns.

The π_h values cannot be determined until all the bettors have wagered. However, each bettor's wagering strategy depends on knowledge of the π_h values to place bets. Eisenberg and Gale (1959) have resolved this apparent contradiction by showing that a set of final track probabilities and individuals' wagers consistent both with the bettors' strategies and the pari-mutuel system do exist, and that the probabilities are unique. The bettors wager with reference to their subjective expectations of the final π_h values, in the absence of precise information about them.

A dollar bet on horse h will return r_h , if horse h wins the race:

$$r_h = \frac{(1 - \delta)W}{w_h} - 1. \quad (2)$$

From equation (1), it follows that $r_h = (1 - \delta - \pi_h)/\pi_h$. On the track toteboard, the odds typically appear in the form of $(1 - \delta - \pi_h)/\pi_h$ to 1.

How can the risk neutral bettor achieve positive returns at the race track? Define ρ_h to be the true *unknown* winning probability associated with horse h . Then the expected payoff of betting on horse h is given by $\rho_h(r_h + 1)$. Suppose that the public's consensus probabilities are equal to the true winning probabilities. In such a situation $r_h\pi_h = 1 - \delta - \pi_h$, and it follows that $\rho_h(r_h + 1) = 1 - \delta$. Therefore, it does not matter which horse the bettor wagers on, he will always expect to lose the track take, δ . In principle, then, it is only possible for a bettor to expect to discover a betting procedure that yields positive returns when the public misestimates the true winning probabilities (i.e., when $\pi_h \neq \rho_h$). Thus, positive returns at the track are only possible when $\rho_h(r_h + 1) > 1$.

A Statistical Model to Estimate Winning Probabilities

To operationalize any wagering strategy, a statistical model of the horse race process is required. It must estimate the true winning probabilities for each horse. Since accuracy is critically important, the estimates must possess good predictive ability. The results of Vergin (1977) suggest that it may be possible to develop such a model. Existing models of the horse race process are generally based on ad hoc filter rules ("don't bet on any horse that ran within the last 10 days and which lost ground in the stretch in its previous race") or regression analysis where the dependent variable is binary (a horse wins or not) conducted across many races. These models fail to account for the within-race competitive nature of the horse racing process. In addition, they have no theoretical foundation, and consequently may perform poorly. For example, Bratley (1973, p. 85) reports abandoning the search for a regression model using past performance information available before a race to predict its outcome due to lack of overall statistical significance.

A fundamental axiom of any model of the horse race process should be that the race is a probabilistic event. In this paper, this issue is recognized by developing a stochastic utility model to assess the worth of a horse. This model is parameterized in the form of the multinomial logit model. This model has been applied to a wide range of discrete choice problems in marketing and economics. Representative applications include college choice (Punj and Staelin 1978; Chapman 1979; Manski and Wise 1983), shopping center choice (Chapman 1980; Arnold, Oum, and Tigert 1983), and transportation model choice (Domencich and McFadden 1975; Hensher and Johnson 1981). In the horse racing context, Figlewski (1979) used the multinomial logit model to measure the information content of the published forecasts of professional handicappers and found that the track odds had already accounted for most of it.

A Stochastic Utility Model of the Horse Racing Process

A horse race may be thought of as an event in which a decision maker—"nature"—chooses the winning horse from among the available horses in a given race. In each race, "nature" is presented with a choice set which consists of the horses scheduled to run. Each horse h has a vector of K observed attributes (e.g., class, speed rating performance, etc.) associated with it, denoted $\mathbf{x}_h = [x_{h1}, x_{h2}, \dots, x_{hK}]$. In addition, each horse is ridden by a jockey characterized by a vector of M attributes, $\mathbf{y}_h = [y_{h1}, y_{h2}, \dots, y_{hM}]$.

A general specification of a statistical model of the horse racing process may be postulated as follows:

$$\rho_h = \rho(\mathbf{X}, \mathbf{Y}), \quad (3)$$

where \mathbf{X} and \mathbf{Y} are the relevant horse and jockey data, respectively, for all of the horses in a given race. A suitable parameterization of this choice model must be

chosen so that the estimated winning probabilities satisfy the standard axioms of nonnegative probabilities and probabilities which sum to unity across all of the horses in a race. The multinomial logit model, described below, intrinsically satisfies these axioms.

Let us now assume the existence of a function which measures the worth (or "utility") of a horse h with attribute vector x_h , ridden by a jockey with attribute vector y_h in a given race. The overall worth of horse h in a race can then be written as follows:

$$U_h = U(x_h, y_h). \quad (4)$$

There is typically some measurement error in the modeling process because the attribute vectors do not capture all of the factors operating in the "choice" of a winning horse, the correct functional form for the model may not be specified, and there may be idiosyncratic aspects to any single race. Thus, the overall worth of a horse is assumed to have two parts. One part is a deterministic component, denoted $V_h = V(x_h, y_h)$. The other part is a random component, $\epsilon_h = \epsilon(x_h, y_h)$, which reflects the measurement errors in the modeling process. Assuming that the stochastic error term is independent of the deterministic component, equation (4) can be decomposed as follows:

$$U_h = V_h + \epsilon_h. \quad (5)$$

The presence of the stochastic error term in equation (5) leads to this model being described as a stochastic utility model.

Suppose that horse h^* is observed to win a race. This is equivalent to observing that nature "chose" alternative h^* from the choice set. Since nature is "rational" by definition (i.e., nature "chooses" the best horse at the time the race occurs), revealed preference implies that $U_{h^*} \geq U_h$, for $h = 1, 2, \dots, H$. Since the utility function is partly stochastic, the probability of horse h^* winning the race may be written as:

$$P_{h^*} = \text{Prob}(U_{h^*} \geq U_h, h = 1, 2, \dots, H). \quad (6)$$

Further development and simplification of equation (6) requires that a joint distribution function be specified for the error terms. A natural assumption would be to invoke the normal distribution. With such an assumption, the parametric form of the model would become the multiple choice generalization of the probit model. Daganzo (1979) and Maddala (1983, pp. 62-64) may be consulted for the details of this particular discrete choice probability model. The normal distribution assumption is not without considerable cost: a formidable series of numerical integrations is required to explicitly determine the choice probabilities. Alternative error term specifications must be considered. One possible candidate is the logistic distribution. The logistic distribution assumption leads to a tractable choice probability expression, as described below. In addition to considerations of parsimony and reduced computational complexity, it may be noted that the cumulative logistic and normal distributions exhibit little numerical differences, except at the extremes. All of these considerations have led researchers to favor the logit model form over the probit model form in discrete choice modeling.

By assuming that the stochastic error terms are identically and independently distributed according to the double exponential distribution,

$$\text{Prob}(\epsilon_h \leq \epsilon) = \exp[-\exp(-\epsilon)], \quad (7)$$

then the choice probabilities assume the tractable, closed-form expression of the multinomial logit model:

$$P_{h^*} = \frac{\exp(V_{h^*})}{\sum_{h=1}^H \exp(V_h)} \quad \text{for } h^* = 1, 2, \dots, H. \quad (8)$$

To operationalize the choice probability expression in equation (8), the functional form of the deterministic component of the stochastic utility model must be specified. A linear-in-parameters specification leads to:

$$V_h = \sum_{n=1}^N \theta_n Z_{hn} \quad (9)$$

where $Z_{hn} = Z_{hn}(\mathbf{x}_h, \mathbf{y}_h)$ is the measured value of attribute n for horse h in a race. The Z functions describe either the horse (\mathbf{x}), the jockey (\mathbf{y}), or both. θ_n is the relative importance of attribute n in the determination of the winning horse. The θ values in equation (9) are the parameters of the stochastic utility model that must be estimated from a sample of races.

Estimating the Parameters of the Multinomial Logit Model

The likelihood function associated with a particular sample of races can be written in the following form for the multinomial logit model:

$$\exp(L) = \prod_{j=1}^J P_{jh^*} \quad (10)$$

where the j subscript denotes a race ($j = 1, 2, \dots, J$), h^* in equation (10) is the horse that is observed to win race j , and L refers to the log-likelihood function. Standard software packages, such as Manski (1974), are available to calculate the maximum likelihood estimates. Since maximum likelihood estimates are, in general, consistent and asymptotically normally distributed, approximate large sample confidence bounds may be constructed for parameter estimates and hypotheses may be tested in standard ways.

It is useful to describe several features of the multinomial logit model that are used extensively in this study. First, an overall goodness-of-fit measure has been proposed by McFadden (1974), which is analogous to the familiar multiple correlation coefficient in linear statistical models:

$$\bar{R}^2 = 1 - \frac{L(\theta = \hat{\theta})}{L(\theta = \mathbf{0})} \quad (11)$$

To the extent that the MLE parameters, $\hat{\theta}$, explain the horse race process completely, \bar{R}^2 will approach unity in value. If the vector of MLE parameters is essentially equal to $\mathbf{0}$ (implying an equal chance of each horse winning the race), then \bar{R}^2 will approach zero in value. Hence, \bar{R}^2 varies between zero and one depending on the "explanatory power" of $\hat{\theta}$.

Second, there is a convenient statistical test to assess whether two data subsets are characterized by the same underlying parameter vector, which would imply that the two subsets should be pooled for estimation purposes. To test the null hypothesis that $\theta^{(1)} = \theta^{(2)}$, the appropriate test statistic is:

$$-2(L(\theta = \hat{\theta}^{(1+2)}) - [L(\theta = \hat{\theta}^{(1)}) + L(\theta = \hat{\theta}^{(2)})]) \quad (12)$$

where $\hat{\theta}^{(1+2)}$ is the MLE obtained by pooling the two data subsets, and $\hat{\theta}^{(1)}$ and $\hat{\theta}^{(2)}$ are the MLEs for the two data subsets, respectively (Watson and Westin 1975). This test statistic will be asymptotically distributed χ^2 with N degrees of freedom, the number of parameters in the model (Wald 1943).

Exploiting Rank Ordered Choice Set Data

The multinomial logit model is estimated on the basis of choice set observations of the form: nature "chooses" horse h^* from among all of the competing horses in a

race. However, in addition to observing the winning horse in a race, it is also possible to conveniently observe the second finishing horse, the third finishing horse, and so on. Chapman and Staelin (1982) describe how the extra information inherent in such rank ordered choice sets may be exploited. The Chapman and Staelin "explosion process" is based on a Ranking Choice Theorem developed by Luce and Suppes (1965, pp. 354-355) for the class of models of which the stochastic utility model is a member.

To illustrate this procedure, suppose that a race results in the finishing order (from first to last) 4, 2, 1, and 3. By applying the Chapman and Staelin explosion process, it is possible to decompose these rank ordered data into the following three statistically independent choice sets: [horse 4 finished ahead of horses 2, 1, and 3], [horse 2 finished ahead of horses 1, and 3], and [horse 1 finished ahead of horse 3], where no ordering is implied among the "nonchosen" horses in each "choice" set.

These exploded choice sets are statistically independent, so they are equivalent to completely independent horse races. This leads to an increase in the number of independent choice sets available for analysis and, ultimately, to more precise parameter estimates. Extensive small-sample Monte Carlo results reported in Chapman and Staelin (1982) document the improved precision that can accrue by making use of the explosion process. This is valuable because it is costly to generate a sufficiently rich set of horse race data to estimate a multinomial logit model of reasonable complexity.

The rank order explosion process should only be used if it illuminates the choice process, and not if it just adds random noise. This is an important estimation issue because the reliability of the rank order finishing data may decrease for horses far behind the winner and the runners-up. The first three finishers typically receive a portion of the purse and are subject to considerable public scrutiny from track officials and bettors due to the existence of the win, place, and show betting pools. It seems reasonable to assume that those horses and their jockeys are trying and that their finishing position reflects well on their relative "worths." However, this may not be true for horses that finish out of the money.

An approach to resolving this rank order reliability issue is to only partially explode the data. Define E as the researcher-chosen depth of explosion. Then the number of independent choice set observations that may be generated from J races is defined as follows:

$$J(E) = \sum_{j=1}^J \min(E, d_j, H_j - 1). \quad (13)$$

$J(E)$ is only defined for nonnegative values of E and d_j represents the depth of available rank ordered choice set information for choice set (race) j .

There are three approaches to determining the appropriate depth of explosion. First, the researcher's a priori knowledge about the choice process may provide clues to the range of plausible values of E . For reasons described above, E might be as large as three in the horse race context.

Second, a heuristic approach may be utilized. This approach involves plotting values of the likelihood ratio index, \tilde{R}^2 , versus different values of E . Since this index does not depend on the number of available choice sets, the calculated \tilde{R}^2 values should remain approximately constant as E increases, if the subsequently "exploded" observations are of equivalent reliability. If the values of \tilde{R}^2 start to decrease substantially after some value of E , this would imply that "noisy" observations had been added and the explosion process should be terminated prior to that value of E .

Third, a formal approach to choosing the appropriate value of E involves grouping choice observations by depth of explosion and sequentially testing whether the observation groups may be pooled. Define the first subset of choice observations to

consist of the $J(E)$ choice sets generated by an explosion to a depth of E . The second subset then consists of the incremental $J(E + 1) - J(E)$ choice sets generated by exploding to a depth of $E + 1$. Assuming that the $J(E + 1) - J(E)$ subset is large enough, the hypothesis that $\theta^{(E)} = \theta^{(E+1)}$ can be tested using the Watson and Westin (1975) procedure. This grouping and sequential hypothesis testing procedure can be iterated for successive values of E until the hypothesis that the subset parameter vectors are equal is rejected, or the quantity $J(E + 1) - J(E)$ yields too few exploded observations to permit meaningful maximum likelihood parameter estimates to be obtained.

Estimating a Multinomial Logit Model of the Horse Race Process

The Data Base

The horse race data base was assembled from information reported in the *Daily Racing Form*. The 200 race observations are from Aqueduct (43), Pimlico (52), Garden State (42), Keystone (32), and Suffolk Downs (31).

Each of the races satisfied the following restrictions: (i) the race was run over good or fast tracks; (ii) the race distance was in the 1–1.25 mile range; (iii) each horse in the race was a separate betting entry (i.e., there were no coupled entries); and (iv) the horses were at least three years of age. The first two restrictions were applied because the logit model only permits the direct inclusion of variables which vary across the choice set alternatives. Thus, track characteristics, which are the same for every horse in a race, cannot be directly included in the model, except possibly as interaction effects with other horse and/or jockey variables. Races were selected to hold track condition and distance roughly constant, thus obviating the need for inclusion of these variables in the model. The third restriction was made to simplify the win-betting procedure. The fourth restriction ensured that adequate past data would be available on the relevant attributes of each horse. This restriction also avoids the potential problem of dealing with high variability in performance among two-year olds.

The relevant data for each race required about one hour to assemble and code. The use of the previously described explosion process thus had a significant influence on reducing data collection costs. Instead of gathering 600 races at a corresponding cost of 600 hours, it was possible to obtain about the equivalent statistical precision in the multinomial logit model estimates by using each of the 200 races exploded to a depth of three.

Specification of the Model

The specific form of the multinomial logit model employed in this study was as follows:

$$U_h = \theta_1 \text{LIFE\%WIN}_h + \theta_2 \text{AVESPRAT}_h + \theta_3 \text{W/RACE}_h + \theta_4 \text{LSPEDRAT}_h \\ + \theta_5 \text{JOCK\%WIN}_h + \theta_6 \text{JOCK\#WIN}_h + \theta_7 \text{JMISDATA}_h + \theta_8 \text{WEIGHT}_h \\ + \theta_9 \text{POSTPOS}_h + \theta_{10} \text{NEWDIST}_h + \epsilon_h. \quad (14)$$

This model specification is explained in the following text.

The quality of the competing horses is presumed to be the primary determinant of horse race outcomes. The long-term quality of a horse is reflected by two aspects of its past performance: winning potential and competitive level. Current quality/performance will also be influenced by weight, post position, whether a horse is running at a new distance, and recent workout data. The final component of the model concerns the jockey's characteristics. Each of these model components is discussed below.

A horse's quality is reflected by measures of its winning potential and competitive level. Measures of winning potential may include races won or earnings. Competitive level refers to the types of races in which a horse has previously competed. A horse which changes class is competing with horses of different quality levels. Its past performance (e.g., races won or earnings) is not directly comparable with competitors' measures. Winning potential should be adjusted by a measure of competitive level which is comparable across different past performance conditions, such as speed rating.

Overall winning potential is proxied by LIFE%WIN, the percentage of races won of those entered in the past two years. Overall competitive level is represented by AVESPRAT, an average speed rating for the last four races of each horse. (A speed rating for a horse compares its time in a race with the track record for that distance. The track record is assigned a value of 100 and a point is deducted for each one-fifth of a second that the horse's time is below that mark. The horse's raw speed rating is then adjusted by a factor to equate the track records at the various tracks used in this study, to attempt to account for differences in tracks. Thus, speed rating has been transformed to be comparable across tracks.) Recent winning potential is proxied by W/RACE, winnings per race in the current year (in \$000s). The recent competitive level component of past performance is LSPEDRAT, a track-adjusted speed rating for the previous race in which the horse ran.

The effect of weight (WEIGHT) on winning probability may be positive or negative. Since weight levels are designed to handicap better horses and result in more even competition, a higher weight should result in a decrease in winning probability, *ceteris paribus*. However, higher weight levels are assigned to higher quality horses, so a positive effect may actually be observed because weight carried is positively correlated with a horse's quality.

An inside (lower) post position theoretically improves the probability of a horse winning because a slightly shorter race distance is involved. Higher values of POSTPOS are expected to result in a decrease in winning probability.

A horse running at a new (unfamiliar) longer distance may not perform as well initially due to the different requirements of pace, stamina, and speed. Several races may be required at the new distance before a horse is fully acclimated. Thus, running at a new distance may have a negative effect on performance. An indicator variable, NEWDIST, captures this effect. NEWDIST equals one if a horse had run three or four of its last four races at distance levels of less than one mile, and zero otherwise.

Workouts could be important in assessing a horse's current condition. If a horse has changed tracks, has not raced recently, or has experienced an injury, workout data represent an important signal as to current performance capabilities. Unfortunately, such data are difficult to interpret since a trainer's objectives for a given workout may not require that a horse perform at the maximum possible level. For this reason, workout data were considered to be of dubious value especially in comparison with the other horse quality and performance variables in the model. Thus, a workout variable was not included in the model.

Jockey characteristics may be of secondary importance in determining a horse's overall "worth." A jockey, no matter how skilled, cannot consistently win with an inferior horse. However, given horses of roughly equal quality, the more accomplished jockey may be more likely to win. There will be some positive correlation between horse and jockey quality, because owners of better horses seek to employ the better jockeys and better jockeys, in turn, prefer to ride better horses since jockeys are compensated partially on a commission system based on horses' earnings.

Jockey data on percentage of winning rides, JOCK%WIN, and number of winning rides, JOCK#WIN, over the current year were included in the model. Some jockeys' records were not available in the *Daily Racing Form*. Such missing data were accounted

for by creating an indicator variable, JMISDATA, which takes on the value one when the other jockey variables are missing, and is zero otherwise. (By construction, when JMISDATA equals one, JOCK%WIN and JOCK#WIN equal zero.) Since the jockey data were missing for those who were not in the published list of leading jockeys, such missing data correspond to relatively inexperienced jockeys who are probably of lower quality than the leading jockeys. The coefficient on JMISDATA will serve as a proxy for the average values of JOCK%WIN and JOCK#WIN for such nonleading jockeys.

Results of Estimating the Base Model

The base model in equation (14) was estimated using the 200 races in the study data base. The associated empirical results are displayed in Table 1.

TABLE 1
Multinomial Logit Model Results—Base Model Estimated on 200 Races with Explosion Depths of 1, 2, and 3

Variable	Coefficient Estimates			Standard Deviation of the Variable	Standardized Coefficient Estimates
	<i>E</i> = 1	<i>E</i> = 2	<i>E</i> = 3		
LIFE%WIN	0.0143 (0.0082)	0.0076 (0.0061)	0.0066 (0.0050)	11.74	0.077
AVESPRATT	0.0789 (0.0178)	0.0615 (0.0123)	0.0546 (0.0101)	10.29	0.562
W/RACE	0.0865 (0.0506)	0.0846 (0.0398)	0.1103 (0.0356)	2.07	0.228
LSPEDRAT	0.0073 (0.0113)	0.0046 (0.0079)	0.0067 (0.0064)	14.78	0.099
JOCK%WIN	0.0205 (0.0339)	0.0297 (0.0242)	0.0236 (0.0201)	8.26	0.195
JOCK#WIN	0.0017 (0.0042)	0.0019 (0.0029)	0.0017 (0.0025)	43.41	0.076
JMISDATA	-0.0284 (0.4159)	0.1962 (0.2927)	0.1765 (0.2425)	0.48	0.086
WEIGHT	0.0227 (0.0225)	0.0076 (0.0158)	0.0030 (0.0130)	4.01	0.012
POSTPOS	-0.0804 (0.0332)	-0.0478 (0.0231)	-0.0439 (0.0189)	2.49	-0.109
NEWDIST	-0.2548 (0.2005)	-0.3902 (0.1446)	-0.3754 (0.1189)	0.46	-0.172

Summary Statistics

# of "Exploded"			
Choice Sets	200	400	600
$L(\theta = 0)$	-401.4	-773.6	-1110.5
$L(\theta = \hat{\theta})$	-364.9	-724.1	-1049.3
\bar{R}^2	0.091	0.064	0.055

Notes: ¹ Asymptotic standard errors of parameter estimates are in parentheses below each coefficient estimate.

² The standard deviations of the variables are taken from the raw data for an "explosion" depth of 1. Thus, the "unexploded" standard deviation is used as an approximation for the "exploded" data standard deviation for each of the variables in the model.

³ "Standardized Coefficient Estimates" are equal to the product of each variable's *E* = 3 coefficient estimate and its standard deviation.

The test of the null hypothesis that the parameter vector is equal to zero is rejected at the 0.005 level of significance. Thus, it may be concluded that the model is explaining a statistically significant amount of the variation in racing performance. Note that the overall goodness-of-fit index, \bar{R}^2 , has a different interpretation than the multiple correlation coefficient in linear statistical models. The standard of reference for \bar{R}^2 within the multinomial logit model is the equal probability model, $\theta = 0$ (where each horse has an equal probability of winning). The multinomial logit model then attempts to explain a significant amount of the variation in the win probabilities based on the available independent variables. Furthermore, low values of \bar{R}^2 should be expected in the horse racing context, since most races have a variety of constraints placed on the competing horses by the racing secretary. These constraints are designed to equalize, to some extent, the chances of the competing horses. Weight allowances, age restrictions, and specific past performance profiles (such as nonwinners in the last two months) are examples of explicit devices used to attempt to equalize the horses' win probabilities. For $E = 1$, \bar{R}^2 equals about 9%, indicating that the model explains 9% more variation than the null hypothesis that all the horses have equal probabilities of winning.

The choice set data were exploded to depths of two and three. Three was chosen as the maximum depth of explosion for which prior theory would suggest that reliable rank order information might be available from the finishing order information. The results displayed in Table 1 illustrate the main value of the explosion process: the standard errors of the parameter estimates decrease when the rank ordered data are exploded to yield more choice sets for analysis. In going from $E = 1$ to $E = 2$ the average decrease in the standard errors is about 28.3%; a further 16.8% average decrease in the standard errors is achieved in going from $E = 2$ to $E = 3$. This pattern is consistent with the Monte Carlo results reported in Chapman and Staelin (1982).

The Watson and Westin (1975) sequential pooling and hypothesis testing procedure was employed to determine whether $E = 3$ was appropriate. The relevant log-likelihood values and χ^2 are described in Table 2. The null hypothesis being tested in each case is whether $\theta^{(E)} = \theta^{(E+1)}$. To assess whether a move from $E = 1$ to $E = 2$ is appropriate, the relevant χ^2 test statistic (with ten degrees of freedom, the number of variables in the model) is:

$$\chi^2 = -2(-724.13 - [(-364.94) + (-353.69)]) = 11.0.$$

Comparing this calculated test statistic with the relevant critical values leads to the conclusion that the null hypothesis should not be rejected on the basis of this sample evidence. Therefore, it is feasible to pool the observations and explode the rank ordered choice set data to a depth of two. In iterating this test to determine if an explosion

TABLE 2
Relevant Log Likelihood Values for Determining the Optimal Explosion Depth

Choice Observation Group	# of Races in This Set	Log Likelihood Value	Calculated χ^2 Value*
$J(E = 1)$	200	$L(\theta = \hat{\theta}^{(1)}) = -364.94$	
$J(E = 2) - J(E = 1)$	200	$L(\theta = \hat{\theta}^{(2)}) = -353.69$	
$J(E = 3) - J(E = 2)$	200	$L(\theta = \hat{\theta}^{(3)}) = -322.06$	
$J(E = 2)$	400	$L(\theta = \hat{\theta}^{(1+2)}) = -724.13$	11.1
$J(E = 3)$	600	$L(\theta = \hat{\theta}^{(1+2+3)}) = -1049.26$	6.0

* This is the statistic calculated to test whether the choice set data can be pooled to level E . Critical values of χ^2_{10} are 18.3 and 23.2 at the 5% and 1% levels of statistical significance, respectively. The degrees of freedom in this test are 10, corresponding to the number of variables in the model.

depth of three is appropriate, the relevant test statistic value is equal to 6.1, which again is sufficiently small that the null hypothesis should not be rejected on the basis of the sample evidence. Hence, an explosion depth of three is appropriate for these horse racing choice set data.

In examining the results reported in Table 1 for $E = 3$, it may be noted that the signs of the coefficients are consistent with a priori theoretical expectations. It is useful to attempt to measure the relative importance of each variable. This is done by calculating the coefficient estimates which would have been obtained if all variables had been standardized (to unit variance) prior to estimation. This is equivalent to assessing a variable's relative importance by taking the product of its coefficient estimate and its standard deviation. The interpretation of the standardized relative importances, also displayed in Table 1, is subject to the usual difficulties in uniquely partitioning the explained variance among any set of collinear independent variables. The results in Table 1 suggest that average speed rating (AVESPRAT) accounts for the most variation in the model. Winnings per race in the current year ($W/RACE$) appears to be more important than lifetime percentage wins ($LIFE\%WIN$). This may be attributed to $W/RACE$ taking into account high but nonwinning performances and to $W/RACE$ being based on recent performance information. $WEIGHT$ does not seem to be an important determinant of finishing position, given the presence of the other variables in the model. Post position ($POSTPOS$) and new distance ($NEWDIST$) appear to exhibit nontrivial effects on winning probabilities. The jockey variables appear to have less overall importance than the horse's attributes in determining winning probabilities, although this finding may be due to collinearity among the horse and jockey variables.

This model has substantial face validity on several dimensions. First, the multinomial logit model considers the competitive nature of the horse racing process. The choice probability expression explicitly includes the characteristics of each horse in comparison with all other horses in a specific race, and not relative to all horses in the universe. Second, an intuitively appealing theoretical utility maximizing (revealed preference) framework was utilized in developing the model. Third, the empirical results indicate that the model operationalization passes the usual tests of statistical significance. The empirical findings are consistent with a priori theoretical beliefs. However it remains to be determined whether this model is sufficiently accurate to allow for the development of a superior wagering system which will earn positive returns.

Analysis of Wagering Systems: Searching for Positive Returns at the Track

In this section, the multinomial logit model of the horse race process is employed to evaluate alternative wagering strategies. Two classes of wagering strategies are considered: algorithms involving multiple bets per race and algorithms involving a single bet per race. A sequential hold-out sampling procedure was used to evaluate each wagering strategy. The model in equation (14) was estimated separately on four data subsamples drawn from the available 200 races. Each sample was a set of 150 (overlapping) choice set observations exploded to a depth of three. For each of the four estimated models, a hold-out sample of the remaining 50 races was then available to evaluate the wagering strategies. This validation approach avoids the upward bias of goodness-of-fit statistics calculated by applying a statistical model back on to the same data base from which it was originally estimated.

Strategies Involving Multiple Bets Per Race

An "optimal" set of wagers can be derived from a variety of wagering strategies based on different objective functions. For example, a wagering algorithm based on

expected value maximization might be appropriate for a risk neutral bettor. Alternatively, an algorithm that maximizes expected log returns would be consistent with risk averse behavior. In addition, wagering strategies may entail either large bets whose influence on the track odds is explicitly taken into account or unobtrusive bets which do not influence the track odds.

Isaacs' Wagering Strategy. Isaacs (1953) determined the optimal amounts to wager for a risk neutral bettor with infinite wealth who has perfect estimates of the true winning probabilities. His algorithm incorporates the impact of the expected value maximizing bettor's wagers on the track odds. In operationalizing Isaacs' strategy, it is necessary to assume that the expected value maximizing bettor is the last bettor. If not, then some subsequent bettor might place wagers which would change the track odds, and thus the optimal amount the expected value maximizing bettor should have wagered. It should be recognized that there are nontrivial logistical problems associated with being the final bettor in a race, particularly if large wagers are being placed.

Isaacs' algorithm was applied to each of the four hold-out samples of 50 races. The winning probabilities were predicted using the multinomial logit model estimated on each set of 150 remaining races. The algorithm identified an average of 3.46 bets per race with expected positive returns. The average amount wagered was \$958. (This was calculated as a weighted average across the four data subsets, where the weights were the total number of bets placed in each subset of 50 races.) The average return per race was -39.5%, while the weighted average return across our four hold-out samples of races was -27.8%.¹ This is considerably worse than a random betting strategy might be expected to yield. It is also much worse than Isaacs' wagering algorithm would perform if the true winning probabilities were known, rather than using fallible estimates. There was considerable variation in the returns across the four data subsets (of 50 races each). The individual 50-race subsamples had average returns of -2.6%, -65.9%, -35.6%, and -7.7%. Even allowing for sampling variation, these results suggest that the probability estimates are too imprecise to be useful in implementing Isaacs' optimal wagering strategy for the expected value maximizing bettor.

Why does Isaacs' algorithm perform so poorly? Modest errors in the estimates of the true winning probabilities could cause substantial deviations from the optimal returns of Isaacs' strategy. Isaacs' wagering algorithm determines the amounts of the wagers based on four factors: the true winning probabilities, the public's consensus probabilities (as reflected by their actual betting behavior), the size of the track take, and the size of the betting pool. The optimal amount to bet involves a trade-off between the attractiveness of wagering large amounts and the feedback effect of the resultant changes in the track odds. The bettor observes the discrepancy between the true winning probabilities and the track odds, and subsequently makes wagers which yield payoffs according to the *revised* track odds, where the revisions take into account the bettor's obtrusive wagers. The wagering of the expected value maximizing bettor results in the track consensus probabilities being driven closer to the true winning probabilities. When a fallible estimate is substituted for the true winning probability, the bettor observes the discrepancy between the estimated winning probability and the public's consensus winning probability, and then wagers in such a way as to drive the public's consensus winning probabilities toward the bettor's estimated winning probabilities. Since the estimates of the winning probabilities may be different from the true (unknown) values, the odds will be driven toward revised odds which may not necessarily yield the optimal payoffs. Therefore, Isaacs' strategy is unlikely to be

¹ "Return per race" is defined to be return per wager averaged across a number of races. "Return across races" is defined to be average return divided by average wager. These measures of return would, of course, be identical if wagers were of constant value.

profitable—unless the bettor has very precise estimates of the true winning probabilities—because the wagers will significantly lower the odds.² This finding suggests that wagering necessarily yield the optimal payoffs. Therefore, Isaacs' strategy is unlikely to be profitable unless the estimates of the true winning probabilities are sufficiently accurate.

Rosner's Wagering Strategy. Rosner (1975) determined the optimal amounts to wager for a risk averse bettor who has perfect estimates of the true winning probabilities. He derived a closed form solution for the optimal amount to bet under the assumption that the wagers have no effect on the odds. His algorithm maximizes expected log return. It has the desirable property of maximizing the long-run rate of asset growth, termed the Kelly criterion (Thorp 1975). Rosner's wagering strategy involves differential bets, where the size of the bet is a function of the attractiveness of the wager.

Unlike Isaacs' wagering strategy, Rosner's closed form solution does *not* take into account the effect of the wagers on the track odds. Consequently, if we take Rosner's suggested optimal wagers and do not correct the odds for our bettor's wager, then we will overestimate the returns to some extent. However, this simplification is progressively more reasonable as the public's wagers increase and/or our bettor's bankroll decreases. Another effect of this simplification is that the performance of the system should be less sensitive to errors in the estimation of the true winning probabilities.

Rosner's strategy was evaluated in the following way. The winning probabilities were predicted using the multinomial logit model estimated on each set of 150 races. For each of the four hold-out samples, wealth was assumed to be \$1000 at the start of the first race, and then updated after each race.³ The algorithm was applied to each of the four hold-out samples of 50 races. It identified an average of 3.48 bets per race. The average amount wagered per bet was \$85. The bettor's initial wealth of \$1000 decreased to \$95.63 (an average across the four data subsets). In other words, the bettor had lost most of his "stake"! It is useful to examine the average return per race and average return across 50 races for Rosner's strategy, in order to compare it to Isaacs' wagering strategy. The average return per race was -14.1%. This is a substantial improvement over Isaacs' wagering strategy. However, the average return across 50 races was -37.4%. This is worse than a random wagering strategy, and worse than Isaacs' wagering strategy.

As before, there was considerable variation in the return per race within the 50 races of each data subset. (The standard error of the mean return per race is about 27.5% for each data subset.) This variation leads to large fluctuations in wealth from race to race and creates two problems. First, the bettor's wagers in a given race can be large enough to significantly affect the track odds. Since Rosner's closed form solution for the optimal amount to wager does *not* take into account the effect of the wagers on the track odds, these large wagers will not be "optimal." Second, large fluctuations in wealth from race to race lead to considerable variability in the size of the bettor's

² One of the referees commented that Isaacs' wagering strategy does poorly because it tries to "grab all the inefficiency." Since the probabilities are estimated with error, the strategy sometimes wagers incorrectly or too much. As a result, Isaacs' wagering strategy performs poorly relative to a strategy that makes a single "small" bet on a randomly chosen horse. However, it's likely that the estimated probabilities are sufficiently accurate that Isaacs' wagering strategy would perform well (in large samples) compared to a strategy that places multiple "large" bets on randomly chosen horses.

³ The selection of \$1000 for initial wealth is arbitrary. This amount was selected for calculation/demonstration purposes. The closed form solution from Rosner's wagering strategy assumes the wagers do not affect the track odds. Hence, initial wealth of \$1000 was selected with the intention that the resultant wagers from Rosner's strategy would not materially affect the existing track odds. Note that initial wealth of \$1 (which would yield wagers which do not affect the existing track odds) would yield the same relative returns (on a percentage basis). As the discussion in the text indicates, the resultant wagers exhibited considerable variability in size, and this variability led to a modification of the strategy.

wagers per race, since the bettor's wagers on horses are calculated as a fraction of current wealth. Hence, the return across 50 races for a particular data subset will depend on the sequence of the races. For example, the return across 50 races will be lower in the situation where all the winning wagers occur in the "early" races, than in the situation where all the winning wagers occur in the "later" races. In other words, race sequence critically affects the average return across 50 races in the relatively small data subsets of races used in this study.

An ad hoc modification of Rosner's wagering strategy was utilized to make the wagers less obtrusive (i.e., to make the wagers have less impact on the track odds) and to remove the effect of race sequence on the return across races.⁴ Both difficulties were removed by eliminating the variability in the size of the bettor's current wealth. In the modified strategy, wealth is fixed to be equal to \$1000 for *each* race, and the bettor wagers some fraction of this amount.⁵ The average return across 50 races was -6.4%. As before, there was substantial variation in the average returns for the four data subsets: 18.7%, -58.6%, -32.1%, and 46.4%. This is an improvement over a random wagering strategy (and over Isaacs' wagering strategy). However, this finding still indicates negative returns across races and a decrease in initial wealth.

The results of the "fixed wealth" modification provide an estimate of the returns across races that Rosner's wagering strategy should generate in the long run (i.e., when race sequence effects would "cancel out"). However, this estimate of the returns will be inaccurate for two reasons. First, returns are somewhat overestimated because they do not take into account the effect of the wagers on the track odds. Second, returns may be underestimated because the modification to Rosner's wagering strategy is suboptimal in the sense that it no longer maximizes the long run rate of asset growth. Therefore, the results of the fixed wealth modification suggest that Rosner's wagering strategy may generate returns across races of *approximately* -6.4% in the long run.

Constrained Versions of Rosner's Wagering Strategy. Rosner's wagering strategy fails to yield positive returns across races because the true winning probabilities are estimated with error. As a result, the algorithm generates some wagers with low or negative actual returns. How can the bettor avoid these wagers? Two constrained versions of Rosner's wagering strategy are considered here.

Studies of place and show betting suggest that the bettor should wager only on horses with estimated expected returns which are substantially greater than one (Harville 1973; Hausch, Ziemba, and Rubinstein 1981). That is, the bettor should wager only if $\hat{p}_h(r_h + 1) > \alpha$, where α is some constant exceeding one. Rosner's wagering strategy and its "fixed wealth" modification were re-evaluated utilizing the α

⁴ An alternative solution to the problem of large wager effects is to extend Rosner's wagering strategy to take into account the effect of wagers on the track odds (e.g., Hausch, Ziemba and Rubenstein 1981). However, the results of applying Isaacs' wagering strategy indicate that a wagering strategy that alters the track odds requires very precise estimates of the true winning probabilities in order to be profitable. Hence, we chose to modify Rosner's wagering strategy to constrain the size of the wagers. An alternative solution to the problem of race sequence effects is to use a measure of wagering strategy performance which is not affected by race sequence. Since the effect of race sequence is reflected in average return across races, but not in final wealth, final wealth could be used to compare the performance of alternative wagering strategies. Final wealth is a relevant criterion for Rosner's strategy, but less useful for other wagering strategies. For example, Isaacs' wagering strategy assumes that the bettor has infinite wealth. These considerations led us to modifying Rosner's strategy by eliminating the variability in the size of the bettor's current wealth. The modification described in the text "solves" both difficulties (i.e., the effect of "large" wagers and race sequence), albeit in an ad hoc fashion.

⁵ The selection of \$1000 for the fixed wealth modification is arbitrary. This amount was selected for calculation/demonstration purposes. It is meant to yield wagers which are unobtrusive in the sense that they do not materially affect the existing track odds. Fixed wealth of \$1 (which would not affect the existing track odds) would yield the same relative returns (on a percentage basis). See footnote 3.

constraint, for levels of α between 1.0 and 1.8 in increments of 0.1. (About 50% of the wagers are eliminated at α equals 1.8.) The average return per race and the average return across 50 races does not improve with the addition of the α constraint.⁶

It is likely that poor estimates for long shots cause the most serious errors in the calculation of expected returns and the formulation of the optimal wagering strategy. Why? First, in relative terms, errors for long shots will be larger. To illustrate the nature of this problem, a misestimate of 0.01 on a favorite whose true winning probability is 0.20 is, in percentage terms, quite small compared to a misestimate of 0.01 on a long shot whose true winning probability is 0.04. While identical in absolute size, the former represents an error of only 5% while the latter represents an error of 25%! Second, it may be easier to predict winning probabilities for favorites since their performance is likely to be more "regular", and thus more easily represented and predicted with a statistical model. This reasoning suggests the bettor may be able to avoid wagers with low or negative actual returns by wagering only on horses with an estimated probability of winning greater than some minimum value. That is, the bettor should wager only if $\hat{p}_h > \rho_{\min}$, where ρ_{\min} is a specified minimum winning probability estimate. It should be noted that fewer wagers are made as the value of ρ_{\min} is raised. Hence, the race returns are calculated on correspondingly smaller sample sizes, and should be interpreted with caution.

Rosner's strategy and its "fixed wealth" modification were re-evaluated utilizing the ρ_{\min} constraint, for levels of ρ_{\min} between 0.00 and 0.25 in increments of 0.01. The average return per race and the average return across races are displayed in Table 3. As the prespecified value of ρ_{\min} increases, fewer wagers are made because long shots are omitted from consideration. The average returns per race associated with $\rho_{\min} > 0.17$ are positive for 6 of the 8 tabulated values. (The two negative values may be small sample results.) The ρ_{\min} constraint improves the average return across races generated by Rosner's strategy and its fixed wealth modification for the majority of the tabulated values. However, the effect of the ρ_{\min} constraint is more evident in the average return across races generated by the "fixed wealth" modification. The average return across 50 races improves for 18 of the 25 tabulated ρ_{\min} values, and the improvement is quite large. In fact, seven of these values are positive!

The above results indicate that Rosner's wagering strategy may yield long-run positive returns when a side constraint eliminates wagers on horses for which the logit model provides poor estimates of the winning probabilities and expected returns. The bettor should wager on the horses identified by Rosner's strategy, except in the case of long shots. Long shots are horses with (estimated) winning probabilities which do not meet a ρ_{\min} constraint of at least 0.07. (The constraint could be as high as 0.11, but apparently not higher.) Such a constraint eliminates about 17% of the wagers, and generates returns across races of about 1.3%. It is important that the ρ_{\min} constraint not be set too high (e.g., higher than 0.11), or too many horses will be eliminated, resulting in negative returns across races. For example, at $\rho_{\min} = 0.12$, 55% of the bets are eliminated and Rosner's differential wagers generate returns across races of -3.1%.

Multiple Unit Bets Strategy. The results obtained by applying Isaacs' and Rosner's strategies identify two features which should be incorporated in an "optimal" wagering strategy that employs fallible estimated probabilities as inputs. First, the results of

⁶ Isaacs' strategy was evaluated with an α constraint, for α equals 1.05, 1.10, 1.15, and 1.20. The race returns for Isaacs' strategy with an α constraint do not differ substantially from the results reported in this paper. This result is not unexpected because the primary reason for the poor performance of Isaacs' strategy is the feedback effect on the track odds. The wagering strategies discussed in the remainder of this paper were also evaluated with an α constraint. The race returns do not differ substantially from the results reported.

TABLE 3

*Strategies Involving Multiple Bets: Weighted Average Returns Across the Four Data Subsets****

ρ_{\min}	# of Races*	Rosner's Wagering Strategy			Unit Wagers Strategy	
		Per Race	Across Races		Per Race	Across Races
			Updated W	Fixed W^{**}		
0.00	200	-0.1411	-0.3737	-0.0638	-0.1599	-0.2178
0.01	200	-0.1411	-0.3737	-0.0638	-0.1599	-0.2178
0.02	200	-0.1411	-0.3736	-0.0635	-0.1599	-0.2138
0.03	200	-0.1402	-0.3730	-0.0617	-0.1473	-0.2050
0.04	199	-0.1324	-0.3684	-0.0516	-0.1370	-0.1818
0.05	199	-0.1265	-0.3619	-0.0386	-0.1242	-0.1479
0.06	199	-0.1197	-0.3486	-0.0169	-0.1103	-0.1040
0.07	199	-0.1019	-0.3322	0.0131	-0.0799	-0.0506
0.08	198	-0.0613	-0.3051	0.0649	-0.0282	0.0230
0.09	198	-0.1527	-0.2897	0.0442	-0.0840	0.0293
0.10	197	-0.2241	-0.2934	0.0238	-0.1671	-0.0060
0.11	191	-0.1830	-0.2495	0.0820	-0.1151	0.0833
0.12	180	-0.2277	-0.2737	-0.0307	-0.1699	0.0045
0.13	170	-0.1650	-0.3342	-0.0629	-0.1349	-0.0207
0.14	159	-0.2547	-0.4648	-0.2005	-0.2604	-0.1744
0.15	148	-0.1763	-0.4795	-0.1972	-0.1633	-0.1763
0.16	135	-0.0874	-0.4433	-0.1416	-0.0896	-0.0872
0.17	127	-0.0608	-0.4318	-0.1219	-0.0476	-0.0680
0.18	119	0.0374	-0.3889	-0.0487	0.0627	0.0393
0.19	111	0.0545	-0.3687	-0.0591	0.0757	0.0638
0.20	101	0.0780	-0.3472	-0.0401	0.0861	0.0696
0.21	93	0.0670	-0.3566	-0.0360	0.0780	0.0609
0.22	81	0.1744	-0.3074	0.0513	0.1713	0.1504
0.23	72	0.2366	-0.2888	0.0946	0.2199	0.2019
0.24	62	-0.0656	-0.3978	-0.2723	-0.0802	-0.0630
0.25	53	-0.1145	-0.4427	-0.3417	-0.1306	-0.1050

* This column describes the number of races in which at least one bet was placed for Rosner's Wagering Strategy. The number of races in which at least one bet was placed for the Unit Wagers Strategy was identical to Rosner's strategy for most values of ρ , and never differed by more than one race.

** "Updated W " denotes the pure Rosner's Wagering Strategy, in which wealth equals \$1000 at the beginning of the first race, and is updated with the results of subsequent races. "Fixed W " denotes the modification of Rosner's Wagering Strategy, in which wealth is fixed to be equal to \$1000 for each race.

*** The weighted averages reported in this table were calculated by using the number of races in which bets were placed (of the 50 races in a data subset) as the weight for the return from a data subset.

Isaacs' strategy indicate that wagers should be sufficiently unobtrusive that the track odds are not affected (in order to maximize average return across races and average return per race). Second, the results of Rosner's strategy indicate that the bettor should wager on horses with positive expected returns—i.e., wagers should be placed on horses when $\hat{\rho}_h(r_h + 1) > 1$. Since the logit model provides relatively poor estimates of expected returns for long shots, the results of Rosner's strategy imply that the bettor should wager on horses with estimated probabilities of winning which are greater than some minimum value. One possible ad hoc wagering strategy which satisfies these concerns is the following:

Wager one unit on each horse for which $\hat{\rho}_h(r_h + 1) > 1$ as long as $\hat{\rho}_h > \rho_{\min}$, where ρ_{\min} is a

specified minimum winning probability estimate and a "unit" is a dollar amount which is sufficiently small that it does not affect the track odds.

Unlike Rosner's wagering strategy, the Multiple Unit Bets strategy does not wager larger amounts on more attractive bets.

The results of applying the Multiple Unit Bets strategy to the four data subsets are displayed in Table 3. The table shows the mean returns from a \$1 bet on each horse that satisfied the two conditions stipulated above. The strategy of wagering on all horses with positive predicted expected values yields a return per race of -16.0%. The average return across 50 races is -21.8%, somewhat worse than would be expected on the basis of random betting. As before, this poor performance illustrates the impact that inaccurate winning probability estimates can have on betting outcomes. As the prespecified value of ρ_{\min} increases, fewer wagers are made because long shots are omitted from consideration. Once extreme long shots (more than 20 to 1) are removed from consideration, the picture improves tremendously. In particular, the average returns across races associated with ρ_{\min} in excess of 0.07 are positive for 10 of the 18 tabulated values.

It is interesting to compare the Multiple Unit Bets strategy with the fixed wealth modification of Rosner's wagering strategy. The Multiple Unit Bets strategy yields a higher average return *per race* for almost all values of ρ_{\min} , such that $0.04 < \rho_{\min} < 0.22$. However, the fixed wealth modification of Rosner's strategy yields a higher average return *across races* for all values of $\rho_{\min} < 0.11$. The pattern of dominance then reverses, and the Multiple Unit Bets strategy dominates Rosner's strategy for all values of $\rho_{\min} \geq 0.11$.

This comparison indicates that the estimates of the winning probabilities are sufficiently accurate to justify employing a differential betting strategy (such as Rosner's) to maximize returns across races, rather than a unit betting strategy. Differential betting dominates unit betting for strategies involving unconstrained multiple bets. This finding can be explained by the observation that Rosner's strategy tends to bet very lightly on long-odds horses, for which the winning probabilities seem to be poorly estimated. Differential betting also yields positive returns, as well as dominates unit betting, if a small number of misestimated long shots are eliminated from consideration. However, if too many wagers are eliminated by the ρ_{\min} constraint (i.e., at $\rho_{\min} \geq 0.11$), unit betting frequently yields positive returns, as well as dominating differential betting. This finding cannot be completely explained by the fact that the (misestimated) long shots—for which differential wagers are better than unit wagers—have been eliminated.

Strategies Involving a Single Bet Per Race

In the popular literature, the bettor is often advised to bet only on the horse with the highest winning probability. Most handicapping systems are based on attempting to identify the "best" horse, where "best" means most likely to win. This is, of course, suboptimal since such a betting strategy does not take the expected return (the public's wagers) into account. Thus, a superior approach for the bettor desiring to wager only on one horse would be to bet on the horse with the maximum expected return in the race, as long as that expected return exceeds one. The bettor could wager the same amount in each race, or wager differential amounts across races. Strategies involving a single bet per race could be constrained to eliminate long shots, in the same way that strategies involving multiple bets per race were constrained.

Single Unit Bet Strategy. This strategy can be formally described as follows:

Wager one unit on the horse for which the expected return, $\hat{\rho}_h(r_h + 1)$, is a maximum, as long as the expected return exceeds one and $\hat{\rho}_h \geq \rho_{\min}$.

The results of applying this strategy to the four data subsets are reported in Table 4. Note that the average return per race and the average return across 50 races must,

TABLE 4
*Single Bet Strategies: Weighted Average Returns
 Across the Four Data Subsets***

ρ_{\min}	# of Races	Mean Return Per Race*	Return Across Races With Differential Wagers
0.00	200	0.0310	0.0359
0.01	200	0.0310	0.0359
0.02	200	0.0310	0.0359
0.03	199	0.0362	0.0407
0.04	195	0.0574	0.0465
0.05	188	0.0968	0.0673
0.06	181	0.1392	0.0712
0.07	170	0.2129	0.0968
0.08	151	0.3656	0.1756
0.09	135	0.1919	0.0287
0.10	119	0.2009	0.0416
0.11	103	0.3874	0.1688
0.12	94	-0.1000	-0.2004
0.13	84	-0.1726	-0.2331
0.14	75	-0.0733	-0.1883
0.15	64	-0.1406	-0.2274
0.16	54	0.0185	-0.1317
0.17	44	0.2500	-0.0060
0.18	41	0.3415	0.0565
0.19	37	0.1622	-0.0282
0.20	33	0.3030	0.0506
0.21	33	0.3030	0.0506
0.22	28	0.3143	0.1608
0.23	27	0.1889	0.1756
0.24	21	-0.7333	-0.8077
0.25	20	-0.7200	-0.8079

* This is the average return per race from any strategy involving a single bet per race (i.e., a unit or differential wagering strategy). In addition, it is the average return across races for a Single Unit Bet strategy.

** The weighted averages reported in this table were calculated by using the number of races in which a bet was placed as the weight for the return from a data subset.

by definition, be identical for this wagering strategy. The results indicate that wagering on the horse with the maximum expected return yields an average return of 3.1%.

This strategy wagers on the horse with the highest expected percentage return and does not wager additional dollars on horses with lower returns. Hence, it would be expected that the average return *per race* from the Single Unit Bet strategy would dominate the average return *per race* from any of the strategies involving multiple bets per race. (This statement refers to percentage returns, not absolute returns.) In fact, the average return per race from the Single Unit Bet strategy does dominate the majority of the values in Table 3. In addition, the average return *across* 50 races is typically much higher for the Single Unit Bet strategy. In fact, 19 of the 26 tabulated values are positive! These positive values are quite large, ranging from 3.1% to 38.7%!

Single Differential Bet Strategy. A method to increase the average return across 50 races generated by the Single Unit Bet strategy is to wager different amounts of money in different races, wagering larger amounts when the betting opportunity is particularly

attractive. For example, Rosner's strategy wagers smaller amounts on long shots. An ad hoc single differential bet strategy is the following:

Wager on the horse for which the expected return, $\hat{p}_h(r_h + 1)$, is a maximum, as long as the expected return exceeds one and $\hat{p}_h \geq \rho_{\min}$. The amount of the wager should be the amount that Rosner's wagering strategy recommends for that horse, assuming a current wealth of \$1000.

The results of applying this strategy to the four data subsets are reported in Table 4. The average return per race is the same for both single bet wagering strategies. The average return across 50 races for the Single Differential Bet strategy (without a ρ_{\min} constraint) is 3.6%. This is a slight improvement over the Single Unit Bet strategy. However, the Single Unit Bet strategy dominates for all values of ρ_{\min} greater than 0.03. Apparently, a differential betting strategy cannot improve returns when long shots have already been eliminated.

Concluding Remarks

A trade-off exists between the "optimality" of a wagering strategy and the accuracy of the statistical model of the horse race process. When the true (unknown) winning probabilities are fallibly estimated, it is necessary to recognize the existence of such estimation errors within the wagering strategy. The major consequence of using estimates of the winning probabilities within a wagering strategy is that the size of the wagers and the number of wagering opportunities must be constrained.

The size of the wagers must be constrained so that the wagering strategy does not affect the track odds. Since Isaacs' wagering strategy includes feedback effects on the track odds, errors in estimating the true winning probabilities (particularly for horses which are long shots) result in an average return of -27.8% across 50 races. In contrast, Rosner's wagering strategy—which involves unobtrusive bets—improves upon a random wagering strategy. It yields an average return of -6.4% across 50 races, after adjusting for the race sequence effects that arise when this strategy is applied to small samples.

The number of wagering opportunities must be constrained so that the wagering strategy involves bets on horses which are more predictable. The multinomial logit model provides relatively poor estimates of long shots' winning probabilities and, consequently, their expected returns. Hence, the returns of most wagering strategies improve when a side constraint eliminates wagers on long shots. For example, a modification of Rosner's wagering strategy may generate positive returns once horses with predicted winning probabilities of less than about 0.07 are eliminated from consideration.

Two simple wagering strategies equal or surpass the results of Rosner's wagering strategy. The bettor can achieve comparable returns by betting identical amounts on all the favorites (winning probabilities greater than 0.19) that have positive expected returns. Or, the bettor can generate returns of 3.1% (or more) by betting a fixed amount on the horse with the highest expected return. Both these strategies involve unit bets on a limited number of horses. A strategy involving differential bets on multiple horses, such as Rosner's wagering strategy, has the potential to yield higher *long-run* returns than any strategy involving a limited number of unit bets. This study did not have a large enough data base to adequately assess the long-run rates of return across races for the various wagering strategies.

Can a horse race wagering system involving win betting yield positive returns? Given this paper's results, there appears to be room for some optimism. While this study represents a pioneering effort in statistical modeling of the determinants of horse race outcomes, a variety of avenues for followup research exist. The variability in the results across the four data subsets suggests the need for additional empirical analyses of larger

samples of races to confirm or refute these findings. It is also possible that future multinomial logit modeling efforts might lead to reduced estimation errors. For example, a separate multinomial logit model could be estimated for each track, instead of pooling across tracks as was the case in this study. It would be useful to attempt to devise a complete wagering system for win, place, and show betting. Such a research effort could combine a fundamental wagering strategy similar to the one developed in this paper with a technical approach similar to Ziemba and Hausch (1984). These and other related questions will no doubt draw the future attention of researchers interested in race track wager markets.⁷

⁷ The helpful comments of the Departmental Editor, an Associate Editor and the referees are gratefully acknowledged.

References

- ALI, M. M., "Probability Estimates for Racetrack Betting," *J. Political Economy*, 85 (4) (1977), 803-815.
- , "Some Evidence on the Efficiency of a Speculative Market," *Econometrica*, 47 (1979), 387-392.
- ARNOLD, S. J., T. H. OUM, AND D. J. TIGERT, "Determinant Attributes in Retail Patronage: Seasonal, Temporal, Regional, and International Comparisons," *J. Marketing Res.*, 20 (May 1983), 149-157.
- ARVESON, J. AND B. ROSNER, "Optimal Pari-Mutuel Wagering," in S. S. Gupta and J. Yackel (Eds.), *Statistical Decision Theory and Related Topics*, Academic Press, Inc., New York, 1971.
- AND ———, "Multivariate Analysis of Horse Race Betting," *Proc. Business and Economics Section of the American Statistical Association*, 1973, 238-240.
- ASCH, P., B. MALKIEL, AND R. E. QUANDT, "Racetrack Betting and Informed Behavior," *J. Financial Economics*, 10 (2) (1982), 187-194.
- BRATLEY, P., "A Day at the Races," *Infor.*, 11 (2) (1973), 81-92.
- CHAPMAN, R. G., "Pricing Policy and the College Choice Process," *Research in Higher Education*, 10 (1) (1979), 37-57.
- , "Retail Trade Area Analysis: Analytics and Statistics," in R. A. Leone (Ed.), *Proc.: Marketing Measurement and Analysis*, TIMS College on Marketing and The Institute of Management Sciences, Providence, R.I., 1980, 40-49.
- AND R. STAELIN, "Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model," *J. Marketing Res.*, 19 (August 1982), 288-301.
- COPELAND, T. E. AND J. F. WESTON, *Financial Theory and Corporate Policy*, Addison-Wesley Publishing Company, Reading, Mass., 1979.
- DAGANZO, C., *Multinomial Probit: The Theory and Its Application to Demand Forecasting*, Academic Press, Inc., New York, 1979.
- DOMENCICH, T. AND D. MCFADDEN, *Urban Travel Demand: A Behavioral Analysis*, North-Holland Publishing Company, Amsterdam, 1975.
- DOWIE, J., "On the Efficiency and Equity of Betting Markets," *Economica*, 43 (May 1976), 139-150.
- EISENBERG, E. AND D. GALE, "Consensus of Subjective Probabilities: The Pari-Mutuel Method," *Ann. Math. Statist.*, 30 (1959), 165-168.
- FABRICAND, B. P., *Horse Sense*, David McKay Company Inc., New York, 1965.
- FAMA, E. F., "Efficient Capital Markets: A Review of Theory and Empirical Work," *J. Finance*, 25 (May 1970), 383-417.
- , *Foundations of Finance*, Basic Books, New York, 1976.
- FIGLEWSKI, S., "Subjective Information and Market Efficiency in a Betting Market," *J. Political Economy*, 87 (1) (1979), 75-88.
- HARVILLE, D. A., "Assigning Probabilities to the Outcome in Multi-Entry Competitions," *J. Amer. Statist. Assoc.*, 68 (June 1973), 312-316.
- HAUSCH, D. B., W. T. ZIEMBA, AND M. RUBINSTEIN, "Efficiency in the Market for Race Track Betting," *Management Sci.*, 27 (12) (1981), 1435-1452.
- HENSHER, D. A. AND L. W. JOHNSON, *Applied Discrete Choice Modeling*, Croom Helm Ltd., London, 1981.
- ISAACS, R., "Optimal Horse Race Bets," *Amer. Math. Monthly*, 60 (May 1953), 310-315.
- LUCE, R. D. AND P. SUPPES, "Preference, Utility, and Subjective Probability," in R. D. Luce, R. B. Bush, and E. Galanter (Eds.), *Handbook of Mathematical Psychology. Volume 3*, John Wiley & Sons, New York, 1965, 249-410.
- MADDALA, G. S., *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge, 1983.

- MANSKI, C. F., "The Conditional/Polytomous Logit Program: Instructions For Use," School of Urban and Public Affairs, Carnegie-Mellon University, 1974.
- AND D. A. WISE, *College Choice in America*, Harvard University Press, Cambridge, Mass., 1983.
- McFADDEN, D., "Conditional Logit Analysis of Qualitative Choice Behavior," in P. Zarembka (Ed.), *Frontiers in Econometrics*, Academic Press, Inc., New York, 1974, 303-328.
- PUNJ, G. AND R. STAELIN, "The Choice Process For Graduate Business Schools," *J. Marketing Res.*, 15 (November 1978), 588-598.
- ROSETT, R. N., "Gambling and Rationality," *J. Political Economy*, 73 (1965), 595-607.
- ROSNER, B., "Optimal Allocation of Resources in a Pari-Mutuel Setting," *Management Sci.*, 21 (9) (1975), 997-1006.
- RUBINSTEIN, M., "Securities Market Efficiency in an Arrow-Debreu Market," *Amer. Economic Rev.*, 65 (5) (1975), 812-824.
- SNYDER, W. W., "Horse Racing: Testing the Efficient Markets Model," *J. Finance*, 33 (4) (1978), 1109-1118.
- THORP, E. O., "The Portfolio Choice and the Kelly Criterion," in W. T. Ziemba and R. G. Vickson (Eds.), *Stochastic Optimization Models in Finance*, Academic Press, Inc., New York, 1975, 599-619.
- VERGIN, R. C., "An Investigation of Decision Rules for Thoroughbred Race Horse Wagering," *Interfaces*, 8 (1) (1977), 34-45.
- WALD, A., "Tests of Statistical Hypotheses Concerning Several Parameters, When the Number of Observations Is Large," *Trans. Amer. Math. Soc.*, 54 (1943), 426-482.
- WATSON, P. L. AND R. B. WESTIN, "Transferability of Disaggregate Mode Choice Models," *Regional Sci. and Urban Economics*, 5 (2) (1975), 227-249.
- WILLIS, K. E., "Optimum No Risk Strategy for Win-Place Pari-Mutuel Betting," *Management Sci.*, 10 (3) (1964), 574-577.
- ZIEMBA, W. T. AND D. B. HAUSCH, *Beat the Racetrack*, Harcourt Brace Jovanovich Publishers, San Diego, Cal., 1984.