Journal of Applied Statistics, Vol. 25, No. 2, 1998, 221–229

Probability models on horse-race outcomes

MUKHTAR M. ALI, *Department of Economics, University of Kentucky, USA*

SUMMARY *A number of models have been examined for modelling probability based on rankings. Most prominent among these are the gamma and normal probability models. The accuracy of these models in predicting the outcomes of horse races is investigated in this paper. The parameters of these models are estimated by the maximum likelihood method, using the information on win pool fractions. These models are used to estimate the probabilities that race entrants finish second or third in a race. These probabilities are then compared with the corresponding objective probabilities estimated from actual race outcomes. The data are obtained from over 15 000 races. it is found that all the models tend to overestimate the probability of a horse finishing second or third when the horse has a high probability of such a result, but underestimate the probability of a horse finishing second or third when this probability is low.*

1 Introduction

In many respects, the pari-mutuel horse-race wagering market is similar to the stock market. In both markets, returns from investments are uncertain, there are many participants and there is a variety of information concerning investments and participants. This has generated considerable interest in studying the efficiency of the wagering market (see, for example, Dowie, 1987; Ali, 1979; Figlewski, 1979; Hausch *et al.*, 1981; Asch *et al.*, 1982). The common method of attacking this problem has been to devise a profitable betting strategy. If such a strategy exists, then the market is inefficient. A prerequisite for developing a profitable betting strategy is to have accurate prediction of the probability of the outcomes of a horse race. Thus, probability models which assign accurately the probability of the outcome of a horse race would be of utmost interest to academic researchers who want to study the efficiency of the wagering market.

Harville (1973) examines one such probability model. His analysis of 335

Correspondence: M. M. Ali, Department of Economics, University of Kentucky, Lexington, KY 40506, USA. Tel: 606 257 3626.

0266-4763/98/020221-09 $7.00

thoroughbred races suggests that his model overestimates the probability of finishing second or third for horses that have a high probability of such a result, and underestimates the probability of finishing second or third for other horses. Stern (1990) examines a class of models that includes Harville's model, and applies two models from this class to analyze 47 races. Analysis seems to corroborate the findings of Harville (1973). Unfortunately, both these studies are limited in scope, in terms of the number of models and the number of races being analyzed. Henery (1981) proposes an alternative model. Bacon-Shone *et al.* (1992) propose logistic models based on probability obtained from Harville (1973), Henery (1981) and a number of Stern (1990) models. They fit these logistic models to data on races held at racetracks in Hong Kong and Meadowlands, NJ. Using a likelihood criterion, they found that logistic models based on probability obtained from the Henery (1981) model fit the data best. Their conclusion was further confirmed by Lo and Bacon-Shone (1994), who fit logistic models based on probability obtained from Henery (1981) and Harville (1973) probability models. These studies suggest that the Henery (1981) model is likely to provide accurate estimates of the ranking probability of horse-race outcomes.

Unfortunately, these studies did not examine the accuracy of such estimates. A major purpose of this paper is to investigate the accuracy of a number of commonly advocated probability models. The analysis will be based on more than 15 000 races. The models are described in Section 2. Also described in Section 2 is the maximum likelihood estimate (MLE) of the model parameters. Details of the data analysis and findings are reported in Section 3. Some concluding remarks are given in Section 4.

2 Probability models and their estimates

2.1 *Probability models*

Assigning the probability of the outcomes of horse-races in which k horses are competing is the same as assigning the probability for the permutations of the first k integers. The k integers can be interpreted as ranks of k objects. A number of probability models for such ranking (permutations) have been proposed in the statistical and psychological literature (see Critchlow *et al.*, 1991). Among these models is a class of models which assign to each ranking the probability of the corresponding ordering of independent, not necessarily identically distributed random variables. More specifically, let $X_1, X_2, \ldots, X_k$ be k independent random variables with probability distribution functions $F(x; \alpha_i)(i = 1, 2, \ldots, k)$, and let $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ represent a permutation of k objects in which object π_j has rank j $(j = 1, \ldots, k)$. Then, these models assign the probability to the permutation π as

$$\Pr(\pi) = \Pr(X_{\pi_1} < X_{\pi_2} < \ldots < X_{\pi_k})$$

The models are known to be ranking models. Two well-studied cases of the ranking models are the model of Thurstone (1927), Daniels (1950) and Mosteller (1951), also known as the normal ranking model, where the random variables are normally distributed with mean $\alpha_i (i = 1, 2, \ldots, k)$ and variance $= 1$; and the Luce (1959) model, where the distribution of the random variables is Gumbel. The Luce model is also the first-order model in the Plackett (1975) system of logistic models. Henery (1981) proposes the normal ranking model for horse-race outcomes. Henery (1983) and Stern (1990) investigate a ranking model known as the gamma ranking model, where the random variables X_i ($i = 1, 2, \ldots, k$) have gamma distributions with scale parameter α_i and a common shape parameter r. The probability density of X_i is given by

$$f(x; \alpha_i, r) = \left[\frac{\alpha_i^r}{\Gamma(r)}\right] x^{r-1} \exp(-\alpha_i x), \quad x \geqslant 0$$

With shape parameter $r = 1$, the random variables have exponential distributions and the model becomes the Luce model. Harville (1973) applies the Luce model and Stern (1990) applies the gamma ranking models with shape parameter $r = 1, 2$ to horse racing. Bacon-Shone *et al.* (1992) and Lo and Bacon-Shone (1994) fit logistic models based on the probability obtained from both normal and gamma ranking models. In this paper, we examine the normal ranking model of Thurstone (1927), Daniels (1950) and Mosteller (1951), and the gamma ranking model of Henery (1983) and Stern (1990). The gamma ranking model is a class of models where a variety of models are obtained by considering different values of the shape parameter r. A number of these models will be examined in this paper.

2.2 *Parameter estimates*

It is a simple matter to obtain MLEs of the parameters if we have the complete set of data. This is because, in the complete data set, a random sample of permutations is observed and the empirical distribution $p_n(\pi)$ of permutation π is known. Consequently, the log likelihood of the ranking model can be obtained as

$$\ln[L(p_n(\pi), \alpha)] = \sum_\pi np_n(\pi) \ln[p(\pi)] + C$$

where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_k)$ are the parameters $np_n(\pi)$ is the number of times that the permutation π is observed, and $p(\pi)$ is the probability of the permutation π under the model (expressed as a function of the unknown parameters α). The likelihood estimate of α maximizes the term ln L. Unfortunately, we do not have a complete data set. In fact, we observe only one outcome from a race and the data from different races cannot be combined, because each race is different, in the sense that the model parameters will vary from race to race.

Fortunately, however, not only is the outcome of a race observed but the empirical probability $p_n(i)$ that horse i wins can also be accurately estimated from the published win odds O_i. The win odds are determined by the amount of money bet on each horse[1] to win[2] the race, the track take and breakage[3]. If W_i is the amount bet on horse i to win and β is the take-out rate and breakage, then the total win pool is $W = \Sigma W_i$ and the odds O_i are given by

$$1 + O_i = (1 - \beta)W/W_i \tag{1}$$

It has been observed in several studies (see, for example, Fabricand, 1965; Weitzman, 1965; Ali, 1977; Snyder, 1978) that the empirical probability that horse i wins can be accurately estimated by taking it to be proportional to $(W_i/W)^\delta$, for some parameter δ. From equation (1), the win pool fraction W_i/W is given by

$$W_i/W = [1/(1 + O_i)]/\Sigma[1/(1 + O_i)] \tag{2}$$

so that it can be obtained from the published win odds O_i. Knowing the empirical probability $p_n(i)$, it is easy to write the likelihood function and, as shown by Stern (1990), the MLEs satisfy the system of equations

$$p(i) = p_n(i) \tag{3}$$

where $p(i)$ is the model-based probability that horse i wins the race:

$$p(i) = \int f(x; \alpha_i) \prod_{j \neq i=1}^{k} [1 - F(x; \alpha_j)]\, dx \tag{4}$$

$p(i)$ is a function of the parameters α.

It may be noted that $p(i)$ is unchanged if a constant is added to each α_i term in the case of the normal ranking model, and if each α_i term is multiplied by a positive constant in the case of the gamma ranking model. Therefore, without loss of generality, we have chosen α_i so that $\Sigma\,\alpha_i = 0$ for the normal ranking model and $\Sigma\,\alpha_i = 1$ for the gamma ranking model. Hence, MLEs of α_i are obtained by solving equation (3) subject to the condition that $\Sigma\,\alpha_i = 0$ for the normal ranking model and $\Sigma\,\alpha_i = 1$ for the gamma ranking model. The solutions are obtained by minimizing

$$S = \Sigma [p(i) - p_n(i)]^2 \tag{5}$$

$p(i)$, as defined in equation (4), is obtained by numerical integration. The Gauss–Newton method as modified by Marquardt (1963) is used to minimize the function S numerically. To start off the iterative solution procedure, initial estimates of α_i are required. These are obtained as follows.

For the normal ranking model, following Henery (1981), we have

$$\alpha_i = \frac{(k-1)\phi(z_0)(z_i - z_0)}{\Phi^{-1}[(i - 3/8)/(k + 3/4)]} \tag{6}$$

where $z_i = \Phi^{-1}[p_n(i)]$, $z_0 = \Phi^{-1}(1/k)$, $p_n(i)$ is the empirical win probability, and $\Phi(.)$ and $\phi(.)$ are, respectively, the distribution and density functions of the standard normal variable. The initial estimates of α_i are deviations, $\alpha_i - \Sigma\,\alpha_i/k$, where the α_i terms are given in equation (6).

For the gamma ranking model, the probability $p(i)$ depends on the shape parameter r. To make this dependency explicit, let us write $p(i, r)$ for $p(i)$. Some elementary computations suggest that $p(i, r)$ is approximately proportional to $p(i, r = 1)^{-g(r)}$, where $g(r)$ is a positive and monotonically decreasing function of r. Moreover, it can be shown that $p(i, r = 1) = \alpha_i$. Therefore, from the likelihood equation (3), $\alpha_i^{-g(r)}$ is approximately proportional to $p_n(i)$ and estimates of α_i can be taken to be proportional to $p_n(i)^{g(r)}$. Thus, some preliminary estimates of α_i can be obtained once the function $g(r)$ is known.

To determine $g(r)$, we note that, besides being positive and decreasing in r, it must be equal to 1 when $r = 1$. That $g(r = 1) = 1$ follows from the fact that, when $r = 1$, because $p(i, r = 1) = \alpha_i$, the MLEs estimate of α_i, which solve equation (3), are $p_n(i)$. With some trial and error, we found that reasonable estimates of α_i can be obtained by setting $g(r)$ to $2/(1 + r)$. This is what is used in our subsequent analysis.

3 Data analysis

To check the accuracy of the normal and gamma ranking models in assigning the probability of horse-race outcomes, data on 20 247 harness horse races were

TABLE 1. Data description

Racetrack and year	No. of racing dates	No. races	Average daily attendance	Average bet per race (US$)[a]	Average bet per person in a race (US$)
Saratoga					
1970	193	1 909	3 784	23 701	6.26
1971	182	1 779	3 873	25 500	6.58
1972	187	1829	3 393	23 706	6.98
1973	189	1 834	3 486	25 350	7.27
1974	172	1 721	3 541	25 694	7.26
Roosevelt					
1970–71[b]	192	1 698	20 426	228 884	11.21
1972	154	1 355	17 014	213 082	12.52
1973	153	1 347	17 148	214 110	12.49
1974	159	1 406	15 789	216 435	13.71
Yonkers					
1971	155	1 381	18 025	224 289	12.44
1972	145	1 268	17 258	234 837	13.61
1973	160	1 407	15 871	225 639	14.22
1974	148	1 313	15 988	227 853	14.25

Note: All tracks are in the state of New York.
[a] Includes all possible betting opportunities.
[b] October–December 1970 and March–October 1971. In the analysis, the time period is treated as the year 1971.

collected. The data are described in Table 1. Previously, these data have been reported and analyzed by Ali (1977). Data from races that involve a 'dead heat' at any finishing position are not considered. For the sake of comparability, only races with the same number of betting interest are analyzed. Specifically, our analysis is limited to races with eight entrants. In our data, eight entrants competed in 15 402 out of the 20 247 races.

Besides the 'win' bet, there are at least two regular betting opportunities in a race, known as 'place' and 'show' bets. Betting on a horse to place is successful if the horse finishes first or second, and betting on a horse to show is successful if the horse finishes first, second or third. The probabilities that horse i finishes second or third are of direct interest to those making the place and show bets. To check the model accuracy, the model-based probability of a horse finishing second (or third) is compared with its corresponding objective probability. The objective probability of a horse finishing second (third) is defined to be the proportion of times that the horse finishes second (third) when the race is repeated an infinitely large number of times.

Both model-based and objective probabilities are different for horses in a race, and they also differ in different races. The model-based probabilities can be obtained from the estimated model but, because we have only one observation, a reliable estimate of the objective probabilities cannot be obtained. In our study, the horses are grouped and their average model-based probability is compared with an unbiased estimate of the corresponding average objective probability. The average objective probability is estimated by the relative frequency of the horses in a group finishing in a specified position (second or third). If the estimate is ϕ, then its standard error is estimated as the square root of $\phi\ (1-\phi)/n$, where n is the number of races from which the estimate is obtained.

Following Ali (1977), horses are grouped by 'favorites'. The horse with the lowest odds O_i to win a race is known as the first favorite; the horse with the second lowest odds O_i to win the race is known as the second favorite, and so on. On the basis of information of win odds on the horses in a race, the likelihood estimates of the parameters α_i of each model are obtained. Following the analysis of this data by Ali (1977) and Lo (1992), the parameter δ is set to 1.16 for the races at Saratoga, and to 1.13 for the races at both Roosevelt and Yonkers. We then use the estimated model to estimate the probability that favorite i finishes second ($p(i, 2)$) or third ($p(i, 3)$). These are model-based probabilities. These probabilities are obtained by numerical integration, utilizing the following formulae:

$$p(i, 2) = \int f(x; \alpha_i) \sum_{j \neq i}^{k} F(x; \alpha_j) \prod_{l \neq i,j}^{k} [1 - F(x; \alpha_l)]\, \mathrm{d}x \tag{7}$$

$$p(i, 3) = \int f(x; \alpha_i) \sum_{\substack{j_1 j_2 \neq i \\ j_1 < j_2}}^{k} F(x; \alpha_{j_1})\, F(x; \alpha_{j_2}) \prod_{l \neq i, j_1, j_2}^{k} [1 - F(x; \alpha_1)]\, \mathrm{d}x \tag{8}$$

where $f(x; .)$ and $F(x; .)$ are, respectively, the appropriate density and distribution functions. Average model-based probabilities are simple averages of $p(i, 2)$ and $p(i, 3)$ over the number of races being analyzed.

The results from comparisons of the probabilities of finishing second and finishing third are given in Tables 2 and 3 respectively. For each model, the results are reported in three rows. The probabilities estimated from the model are reported in the first row, while the observed relative frequencies (objective probabilities) are in the second row. The deviations of the model-based probabilities from the corresponding relative frequencies are expressed as ratios to the respective standard errors and are reported in the third row. In general, irrespective of the ranking model, the probabilities of finishing second and third are overestimated for those horses which have high probabilities of finishing second or third (or of winning the race), and are underestimated for those horses which have low probabilities of finishing second or third (or of winning the race).

This agrees with the findings of Harville (1973), who examined the gamma ranking model with shape parameter 1; it also agrees with the findings of Stern (1990), who analyzed 47 races based on the gamma ranking model with shape parameter $r = 1$ and 2. However, the standardized deviations of the model-based probabilities from the corresponding average relative frequencies (deviations divided by their standard errors) are almost always smaller when the probabilities are estimated from the normal ranking model than when the probabilities are estimated from any of the gamma ranking models. For example, for favorite 1, in estimating the probability of finishing second, the standardized deviation is 5.737 for the normal ranking model, which is the smallest standardized deviation among all the models. The next smallest deviation resulted for the gamma model with the shape parameter $r = 20$. This is expected, because the gamma model with shape parameter $r = 20$ is closer to the normal model than is a gamma model with shape parameter less than 20. Thus, it seems that the normal ranking model best fits the data. This is consistent with the findings of Bacon-Shone *et al.* (1992) and with those of Lo and Bacon-Shone (1994). However, contrary to the conclusions drawn by Bacon-Shone *et al.* (1992) and Lo and Bacon-Shone (1994), this best-fitting model significantly overestimates the probability of finishing second or third for

TABLE 2. Probability of finishing second: model-based vs objective probability

Model $[n]^a$	Probability	Favorite horse 1	2	3	4	5	6	7	8
$N(0, 1)^b$	M^c	0.210	0.197	0.165	0.134	0.107	0.083	0.061	0.042
[15402]	O^d	0.192	0.193	0.168	0.139	0.111	0.089	0.068	0.042
	$(M-O)/\mathrm{SE}^e$	5.737	1.409	− 0.880	− 1.490	− 1.558	− 2.522	− 3.135	− 0.162
$G(0.5)$	M	0.269	0.224	0.166	0.121	0.088	0.063	0.043	0.026
[14059]	O	0.193	0.189	0.164	0.137	0.112	0.091	0.070	0.044
	$(M-O)/\mathrm{SE}$	23.018	10.484	0.569	− 5.333	− 0.156	− 11.503	− 12.670	− 10.187
$G(0.75)$	M	0.263	0.223	0.167	0.123	0.090	0.064	0.043	0.027
[15400]	O	0.192	0.193	0.168	0.139	0.111	0.089	0.068	0.042
	$(M-O)/\mathrm{SE}$	22.231	9.457	− 0.251	− 5.448	− 8.283	− 10.669	− 11.957	− 9.391
$G(1.0)$	M	0.256	0.219	0.167	0.125	0.092	0.066	0.045	0.028
[15402]	O	0.192	0.193	0.168	0.139	0.111	0.089	0.068	0.042
	$(M-O)/\mathrm{SE}$	20.204	8.385	− 0.163	− 4.781	− 7.338	− 9.759	− 11.114	− 8.577
$G(2.0)$	M	0.242	0.212	0.167	0.128	0.097	0.971	0.050	0.032
[15400]	O	0.192	0.193	0.168	0.139	0.111	0.089	0.068	0.042
	$(M-O)/SE$	15.833	6.173	− 0.237	− 3.642	− 5.478	− 7.616	− 8.975	− 6.287
$G(5.0)$	M	0.230	0.206	0.166	0.131	0.101	0.076	0.054	0.035
[15402]	O	0.192	0.193	0.168	0.139	0.111	0.089	0.068	0.042
	$(M-O)/SE$	11.906	4.277	− 0.399	− 2.781	− 3.920	− 5.628	− 6.704	− 4.058
$G(10.0)$	M	0.224	0.204	0.166	0.132	0.103	0.078	0.056	0.037
[15402]	O	0.192	0.193	0.168	0.139	0.111	0.089	0.068	0.042
	$(M-O)/\mathrm{SE}$	10.037	3.406	− 0.543	− 2.361	− 3.224	− 4.692	− 5.606	− 2.935
$G(20.0)$	M	0.220	0.202	0.166	0.133	0.104	0.080	0.058	0.039
[15402]	O	0.192	0.193	0.168	0.139	0.111	0.089	0.068	0.042
	$(M-O)/\mathrm{SE}$	8.750	2.777	− 0.599	− 2.110	− 2.701	− 4.012	− 4.955	− 2.080

[a] n is number of races in the sample.
[b] $N(0,1)$ is the normal ranking model and $G(r)$ is the gamma ranking model with shape parameter r.
[c] M is the average model-based probability.
[d] O is the average objective probability.
[e] $\mathrm{SE} = [O(1-O)/n]^{1/2}$ is the standard error of O, where n is the number of races in the sample.

horses which have a high probability of winning, and overestimates the probability of finishing second or third for horses that have a low probability of winning.

4 Concluding remarks

We have examined the normal ranking model and a number of gamma ranking models for their accuracy in estimating ranking probabilities in horse races. The analysis is based on over 15 000 races. It is found that the normal ranking model best fits the data, but all models suffer from favorite–longshot bias, i.e. the probability of finishing second or third is overestimated for horses which have a high probability of winning and is underestimated for horses which have a low probability of winning. The data were further subdivided by year and racetrack and an analysis of these data led to the same conclusion about favorite–longshot bias. This shows that none of the ranking models may be used to construct profitable betting strategy.

Hausch *et al.* (1981) reported positive profits, by applying a betting strategy in

TABLE 3. Probability of finishing third: model-based vs objective probability

		Favorite horse							
Model $[n]^a$	Probability	1	2	3	4	5	6	7	8
$N(0, 1)^b$ [15402]	M^c	0.147	0.165	0.161	0.146	0.127	0.106	0.085	0.063
	O^d	0.129	0.153	0.152	0.146	0.136	0.118	0.095	0.071
	$(M-O)/\mathrm{SE}^e$	6.793	3.858	2.942	0.153	− 3.077	− 4.599	− 4.288	− 3.779
$G(0.5)$ [14059]	M	0.191	0.210	0.188	0.144	0.106	0.076	0.052	0.032
	O	0.131	0.151	0.152	0.143	0.135	-.118	0.097	0.073
	$(M-O)/\mathrm{SE}$	21.061	19.730	11.852	0.454	− 9.939	− 15.411	− 18.098	− 18.680
$G(0.75)$ [15400]	M	0.182	0.205	0.186	0.147	0.111	0.080	0.055	0.034
	O	0.129	0.153	0.152	0.146	0.136	0.119	0.095	0.071
	$(M-O)/\mathrm{SE}$	19.600	17.637	11.764	0.462	− 9.122	− 14.744	− 17.039	− 17.684
$G(1.0)$ [15402]	M	0.178	0.199	0.183	0.148	0.113	0.084	0.059	0.037
	O	0.129	0.153	0.152	0.146	0.136	0.118	0.095	0.071
	$(M-O)/\mathrm{SE}$	18.089	15.718	10.586	0.655	− 8.060	− 13.177	− 15.388	− 16.328
$G(2.0)$ [15400]	M	0.169	0.188	0.176	0.148	0.119	0.091	0.066	0.044
	O	0.129	0.153	0.152	0.146	0.136	0.118	0.095	0.071
	$(M-O)/\mathrm{SE}$	14.692	11.825	8.064	0.791	− 6.224	− 10.428	− 12.125	− 12.977
$G(5.0)$ [15402]	M	0.160	0.178	0.169	0.148	0.122	0.098	0.074	0.051
	O	0.129	0.153	0.152	0.146	0.136	0.118	0.095	0.071
	$(M-O)/\mathrm{SE}$	11.658	8.575	5.917	0.648	− 4.848	− 8.012	− 9.007	− 9.559
G(10.0) [15402]	M	0.156	0.174	0.166	0.147	0.124	0.100	0.077	0.054
	O	0.129	0.153	0.152	0.146	0.136	0.118	0.095	0.071
	$(M-O)/\mathrm{SE}$	10.149	7.079	4.933	0.439	− 4.288	− 6.887	− 7.628	− 7.841
$G(20.0)$ [15402]	M	0.153	0.171	0.165	0.147	0.125	0.102	0.080	0.057
	O	0.129	0.153	0.152	0.146	0.136	0.119	0.095	0.071
	$(M-O)/\mathrm{SE}$	9.026	6.070	4.334	0.383	− 3.919	− 6.298	− 6.642	− 6.564

[a] n is the number of races in the sample.
[b] $N(0, 1)$ is the normal ranking model and $G(r)$ is the gamma ranking model with shape parameter r.
[c] M is the average model-based probability.
[d] O is the average objective probability.
[e] $\mathrm{SE} = [O(1-O)/n]^{1/2}$ is the standard error of O, where n is the number of races.

which the probabilities were estimated by the gamma ranking model with shape parameter $r = 1$ to 627 races held during the 1973–74 winter season at Santa Anita Racetrack in Arcadia, CA and to 1065 races held during the 1978 summer season at Exhibition Park, Vancouver, BC. The achievement of such positive profits may have been as a result of the peculiarity of the samples of races that were analyzed. Lo *et al.* (1994) applied the betting system of Hausch *et al.* (1981) and its variants to 705 races held during 1984 at Meadowlands, NJ, as well as applying it to 905 races held during 1981–82 in Hong Kong and to 983 races held during 1990–91 in Japan. In each case, the system of Hausch *et al.* produced negative profits.

Notes

1. In some cases, several horses are grouped—known as an 'entry' or 'field'—for a single betting interest, and the group is assigned a single number. A bet on this number is successful if one of the horses in the group is successful. Thus, if it is a 'win' bet, then the bet is successful if one of the horses in the group finishes first. Without loss of generality, this group is taken as a single horse.
2. Races with a 'dead heat' at any finishing position are not considered.

3. A fixed proportion of the amount bet in a race is taken out by the track before it distributes the rest to the successful betters. This proportion is known as the 'take-out rate'. The breakage arises because of the following two restrictions: (a) odds cannot be below a certain minimum; (b) odds have to be rounded downward, except when restriction (a) is in effect, in which case, it is rounded upward. For the races that are analyzed, all the odds are rounded to 10 cents and the minimum odds are also 10 cents.

REFERENCES

ALI, M. M. (1977) Probability and utility estimates for racetrack betters, *Journal of Political Economy*, 84, pp. 803–815.

ALI, M. M. (1979) Some evidence on the efficiency of a speculative market, *Econometrica*, 47, pp. 387–392.

ASCH, P., MALKIEL, B. & QUANDT, R. (1982) Racetrack betting and informed behavior, *Journal of Financial Economics*, 10, pp. 187–194.

BACON-SHONE, J., LO, V. S. Y. & BUSCHE, K. (1992) Logistic analyses for complicated bets, *Research Report 11*, Department of Statistics, University of Hong Kong.

CRITCHLOW, D. E., FLIGNER, M. A. & VERDUCCI, J. S. (1991) Probability models on ranking, *Journal of Mathematical Psychology*, 35, pp. 294–318.

DANIELS, H. E. (1950) Rank correlation and population modes, *Biometrika*, 33, pp. 129–135.

DOWIE, J. (1976) On the efficiency and equity of betting markets, *Economica*, 43, pp. 139–150.

FABRICAND, B. F. (1965) *Horse Sense* (New York, David McKay).

FIGLEWSKI, S. (1979) Subjective information and market efficiency in a betting model, *Journal of Political Economy*, 87, pp. 75–88.

HARVILLE, D. A. (1973) Assigning probabilities to the outcomes of multi-entry competitions, *Journal of the American Statistical Association*, 68, pp. 312–316.

HAUSCH, D. B., ZIEMBA, W. T. & RUBINSTEIN, M. (1981) Efficiency of the market for racetrack betting, *Management Science*, 27, pp. 1435–1452.

HENERY, R. J. (1981) Permutation probabilities as models for horse races, *Journal of the Royal Statistical Society, Series B*, 43, pp. 86–91.

HENERY, R. J. (1983) Permutation probabilities for gamma random variables, *Journal of Applied Probability*, 20, pp. 822–834.

LO, V. S. Y (1992) *Statistical Modelling of Gambling Probabilities*, PhD dissertation, Department of Statistics, University of Hong Kong.

LO, V. S. Y. & BACON-SHONE, J. (1994) A comparison between two models for predicting ordering probabilities in multiple-entry competitions, *The Statistician*, 43, pp. 1–11.

LO, V. S. Y., BACON-SHONE, J. & BUSCHE, K. (1994) The application of ranking probability models to racetrack betting, *Research Report 18*, Department of Statistics, University of Hong Kong.

LUCE, R. D. (1959) *Individual Choice Behavior* (New York, Wiley).

MARQUARDT, D. W. (1963) An algorithm for least squares estimation of non-linear parameters, *Journal of the Society for Industrial and Applied Mathematics*, 11, pp. 431–441.

MOSTELLER, F. (1951) Remarks on the method of paired comparisons. I. The least squares solution assuming equal standard deviation and equal correlations, *Psychometrika*, 16, pp. 3–9.

PLACKETT, R. L. (1975) The analysis of permutations, *Applied Statistics*, 24, pp. 193–202.

SNYDER, W. (1978) Horse racing: testing for efficient markets model, *Journal of Finance*, 33, pp. 1109–1118.

STERN, H. (1990) Models for Distributions on permutations, *Journal of the American Statistical Association*, 85, pp. 558–564.

THURSTONE, L. L. (1927) A law of comparative judgement, *Psychological Reviews*, 34, pp. 273–286.

WEITZMAN, M. (1965) Utility analysis and group behavior, *Journal of Political Economy*, 73, pp. 18–26.