A Comparison Between Two Models for Predicting Ordering Probabilities in Multiple-Entry Competitions

Victor S. Y. Lo; John Bacon-Shone

The Statistician, Vol. 43, No. 2. (1994), pp. 317-327.

Stable URL:
http://links.jstor.org/sici?sici=0039-0526%281994%2943%3A2%3C317%3AACBTMF%3E2.0.CO%3B2-P

The Statistician is currently published by Royal Statistical Society.

The Statistician (1994)
43, *No.* 2, *pp.* 317–327

A comparison between two models for predicting ordering probabilities in multiple-entry competitions

By VICTOR S. Y. LO†

University of British Columbia, Vancouver, Canada

and JOHN BACON-SHONE

University of Hong Kong

[Received January 1993. Final revision June 1993]

SUMMARY

To predict ordering probabilities of a multiple-entry competition (e.g. a horse-race), two models have been proposed. Harville proposed a simple and convenient model that can easily be used in practice. Henery proposed a more sophisticated model but it has no closed form solution. In this paper, we empirically compare the two models by using a series of logit models applied to horse-racing data. In horse-racing, many previous studies claimed that the win bet fraction is a reasonable estimate of the winning probability. To consider complicated bet types which involve more than one position, ordering probabilities (e.g. P(horse i wins and horse j finishes 2nd)) are required. The Harville and Henery models assume different running time distributions and produce different sets of ordering probabilities. This paper illustrates that the Harville model is not always as good as the Henery model in predicting ordering probabilities. The theoretical result concludes that, if the running time of every horse is normally distributed, the probabilities produced by the Harville model have a systematic bias for the strongest and weakest horses. We concentrate on the horse-racing case but the methodology can be applied to other multiple-entry competitions.

Keywords: Horse-races; Logit model; Ordering probability; Running time distributions

1. Introduction

In the pari-mutuel betting system of horse-racing, it is useful to predict P(horse i wins and horse j finishes 2nd) from the simple knowledge of the winning probabilities, i.e. P(horse i wins). For more complicated bets such as the exacta and trifecta where predictions of more than one finishing order are required, even on-track betters cannot observe the changes of odds. Thus another source of information is required to predict the finishing order probabilities. Important information comes from the win bet market which is the simplest bet market where the gamblers simply need to guess which horse will win. A reasonable estimate of the win probability is the win bet fraction (i.e. the proportion of money bet on a particular horse in the win bet market). Previous empirical studies showed that the win bet fraction is quite consistent with the true winning probability although a favourite–long shot bias sometimes exists (e.g. Griffith (1949), McGlothlin (1956), Hoerl and Fallin (1974), Ali (1977), Snyder (1978), Fabricand (1979), Hausch *et al.* (1981), Asch *et al.* (1982), Henery (1985) and Busche and Hall (1988)). For more complicated bets which involve more than one finishing position, ordering probabilities such as P(i wins and j finishes 2nd) are required. Two alternative models based on different distributional assumptions of running times have been proposed by Harville (1973) and Henery (1981). A thorough collection of academic papers in racing research is given in Hausch *et al.* (1994).

† *Address for correspondence*: Division of Management Science, Faculty of Commerce and Business Administration, University of British Columbia, 2053 Main Mall, Vancouver, British Columbia, V6T 1Z2, Canada.
E-mail: vlo@unixg.ubc.ca

 0039–0526/94/43317

In this paper, we compare the two models proposed by Harville (1973) and Henery (1981). The former is simple and easy to use but the latter is much more complicated. We shall consider the estimation of ordering probabilities by using the two models, including theoretical discussion of the difference between the Harville and Henery models. Section 2 will briefly review the Harville and Henery models. Empirical analysis and theoretical discussion will be given in Sections 3 and 4 respectively, with conclusions in Section 5.

2. Description of some proposed models

2.1. *Harville model*

The simplest and most commonly used model to estimate ordering probabilities is the model proposed by Harville (1973). The basic idea is simple. For instance, to predict P(horse i wins and horse j finishes 2nd), we may use

$$P_{ij} = \frac{P_i P_j}{1 - P_i}$$

if P_i and P_j are known. A similar idea was also mentioned in Plackett (1975). Moreover, it is the ranking model proposed by Luce and Suppes (1965) in the study of choice behaviour. As interpreted by Harville (1973), this model assumes that the event that horse j ranks ahead of all the other horses, save possibly horse i, is independent of the event that horse i wins. These values of P_i can be estimated by the win bet fractions F_i. At a first glance, the above formula seems reasonable and thus it has been used by Hausch *et al.* (1981), Tuckwell (1981), Asch and Quandt (1987), Ziemba and Hausch (1985, 1987) and other researchers. It is also known that some gamblers use this method. However, $P_{j|i}$ may not be equal to $P_j/(1 - P_i)$ in general. One argument is mentioned in Hausch *et al.* (1981):

> 'no account is made of the possibility of the Silky Sullivan problem; that is, some horses generally either win or finish out-of-the-money; for these horses the formulas greatly over-estimate the true probability of finishing second or third'.

One reasonable way to find these ordering probabilities is to assume an underlying probability distribution for the running times of horses. It can be easily shown that, if the running times follow exponential distributions independently with different mean running times, the above formula will be obtained (see Dansie (1983)).

McCulloch and Van Zijl (1986) gave a direct test for the Harville model and indicated that the model had a bias. However, their paper depended on the assumption that the show bet fraction for their New Zealand data was the same as the corresponding true ordering probabilities.

2.2. *Henery model*

Henery (1981) assumed that the running times are independent normal with unit variance, i.e. $T_i \sim N(\theta_i, 1)$ independently. The resulting probabilities are obviously the same as that of a general constant variance model. Under the Henery model,

$$P[T_1 < T_2 < \ldots < T_n] = \int_{-\infty}^{\infty} \phi(t_1 - \theta_1) \ldots \int_{t_{n-1}}^{\infty} \phi(t_n - \theta_n)\, \mathrm{d}t_n \ldots \mathrm{d}t_1$$

where $\phi(\ )$ is the density function of the standard normal distribution.

However, computing this probability is difficult and even computing P_{ij} is not easy because, unlike the Harville model, no closed form solution exists. An approximation suggested by Henery is

$$P[T_1 < T_2 < \ldots < T_n] = \Phi(\Phi^{-1} P[T_1 < T_2 < \ldots < T_n])$$
$$\simeq \Phi\left\{\xi + \frac{1}{n!\,\phi(\xi)} \sum_i \theta_i \mu_{i;n}\right\} \tag{1}$$

where $\xi = \Phi^{-1}(1/n!)$, $\mu_{i;n}$ is the expected value of the ith standard normal order statistic in a sample of size n and $\Phi(\ )$ is the cumulative density function of the standard normal distribution.

Approximation (1) is obtained by using Taylor's expansion about $\theta = 0$ for the term inside the braces.

Using similar methods,

$$P[T_i \text{ is smallest}] \simeq \Phi\left\{z_0 + \frac{\theta_i \mu_{1;n}}{(n-1)\,\phi(z_0)}\right\} \tag{2}$$

where $z_0 = \Phi^{-1}(1/n)$.

Henery (1981) also suggested another approximation method but that method produces many negative probabilities in our experience; thus we only consider the kind of approximations mentioned above in this paper.

Hence, by using approximation (2), we can have estimates of $\boldsymbol{\theta}$ if the true win probabilities P_i are known or the win bet fractions are good estimates of the P_i. Then, we may substitute the estimated values of $\boldsymbol{\theta}$ in the appropriate equations to obtain estimates of the ordering probabilities. For example,

$$\pi_{ij} = P[T_i < T_j < \text{others}]$$
$$\simeq \Phi\left[a + \gamma\left\{\theta_i \mu_{1;n} + \theta_j \mu_{2;n} + \frac{(\theta_i + \theta_j)(\mu_{1;n} + \mu_{2;n})}{n-2}\right\}\right] \tag{3}$$

where

$$a = \Phi^{-1}\left\{\frac{1}{n(n-1)}\right\}$$

and

$$\gamma = \frac{1}{n(n-1)\,\phi(a)}$$

here.

In practice, to satisfy the unit sum constraint, simple scaling is usually necessary.

Another model which extends Harville's exponential running time model was proposed by Stern (1990) who suggested gamma running times with a fixed shape parameter.

3. Logistic analysis for Harville and Henery models

On the basis of a complicated model fitting process, Bacon-Shone *et al.* (1992a) suggested the use of the simple constant β model to analyse win bet data, i.e.

$$\pi_i = F_i^\beta / \sum_r F_r^\beta \tag{4}$$

where $\pi_i = P(\text{horse } i \text{ wins})$, F_i is the win bet fraction of horse i (i.e. the proportion of win bet on horse i) and β is a parameter to be estimated by maximum likelihood assuming that the win event follows a multinomial distribution.

This model can be rewritten as

$$\ln(\pi_i/\pi_k) = \beta \ln(F_i/F_k) \qquad \text{for any } i, k\ (i \neq k)$$

which means that the multivariate logit of the win probability depends on the logit of the bet fractions in a very simple way. This indicates a particular form of the favourite–long shot bias raised in the literature. In particular, if $\beta = 1$, there is no favourite–long shot bias. If $\beta > (<) 1$ the win bet fractions underestimate (overestimate) the winning probabilities associated with the favourite horses and overestimate (underestimate) the winning probabilities associated with the long shots, i.e. the public underbet (overbet) the favourite horses and overbet (underbet) the long shots.

Using a similar structure for conditional probabilities, we have

$$\ln(\pi_{j|i}/\pi_{k|i}) = \mu \ln(P_{j|i}/P_{k|i}) \qquad \text{for any } i, j, k\ (i \neq j \neq k)$$

where

$$\begin{aligned}\pi_{j|i} &= P(\text{horse } j \text{ finishes 2nd} \mid \text{horse } i \text{ wins}) \qquad \text{based on this logit model} \\ &= \frac{P(\text{horses } i \,\&\, j \text{ finish 1st \& 2nd respectively})}{P(\text{horse } i \text{ wins})} \\ &= \pi_{ij}/\pi_i\end{aligned}$$

and

$$P_{j|i} = P(\text{horse } j \text{ finishes 2nd} \mid \text{horse } i \text{ wins})$$

based on the Harville or Henery model using the win bet fractions F_i. For the Harville model,

$$P_{j|i} = F_j/(1 - F_i) \tag{5}$$

and, for the Henery model,

$$P_{j|i} = P_{ij}/F_i,$$

where

$$\begin{aligned}P_{ij} &= P\left[T_i < T_j < \min_{r \neq i,j} \{T_r\}\right] \qquad T_i \sim N(\theta_i, 1) \\ &= \int_{-\infty}^{\infty} \Phi(u + \theta_j - \theta_i) \prod_{r \neq i,j} \{1 - \Phi(u + \theta_j - \theta_r)\}\, \phi(u)\, du,\end{aligned}$$

which can be approximated by equation (3), the θ_i can be estimated by using approximation (2) with the win probabilities and the P_i estimated by the win bet fractions F_i. If the favourite–long shot bias of the win bet fractions is considered to be strong, we may simply estimate the P_i by the model-based win probabilities π_i using the constant β model (4) instead of using F_i.

Similarly, we define

$$\pi_{ijk} = P(\text{horse } i \text{ finishes 1st}, j \text{ finishes 2nd and } k \text{ finishes 3rd})$$

based on a logit model. Other notation such as $\pi_{k|ij}$, $\pi_{m|ijk}$, $P_{k|ij}$ and $P_{m|ijk}$ etc. is self-explanatory.

To study how good the Harville and Henery models are, we fitted the following series of models for conditional probabilities:

$$\begin{aligned} \text{logit}(\pi_{j|i}) &= \mu \,\text{logit}(P_{j|i}), \\ \text{logit}(\pi_{k|ij}) &= \omega \,\text{logit}(P_{k|ij}), \\ \text{logit}(\pi_{1|ijk}) &= \zeta \,\text{logit}(P_{1|ijk}), \\ &\vdots \qquad\qquad \vdots \end{aligned} \tag{6}$$

where all the logits are multivariate. Again, the conditional Ps on the right-hand side are based on either the Harville or the Henery model. Unlike the constant β model (4), which is used to test the favourite–long shot bias of the win bet fractions, models (6) are mainly used to test the distributional assumption associated with the conditional probabilities P. If all the parameters are equal to 1, the underlying probability model does not cause any systematic bias. However, if the parameters are less (greater) than 1, this indicates that the probability model overestimates (underestimates) the conditional probability associated with a favourite horse and underestimates (overestimates) the conditional probability associated with a long shot. A parameter value of 0 corresponds to uniform conditional probabilities.

To simplify our analysis, the approximation method proposed by Henery (1981) is employed in this section since exact computations involve many high dimensional integrations. For each race,

$$\begin{aligned} P_{q|q-1} &= P[T_{i1} < \ldots < T_{iq} | T_{i1} < \ldots < T_{i,q-1}] \\ &\simeq \frac{\Phi\left[C_q + v_q \left\{ \sum_{r=1}^{q} \theta_{ir}\mu_{r;n} + \sum_{r=1}^{q} \theta_{ir} \sum_{s=1}^{q} \mu_{s;n}/(n-q) \right\} \right]}{\Phi\left[C_{q-1} + v_{q-1} \left\{ \sum_{r=1}^{q-1} \theta_{ir}\mu_{r;n} + \sum_{r=1}^{q-1} \theta_{i} \sum_{s=1}^{q-1} \mu_{s;n}/(n-q+1) \right\} \right]} \\ &\qquad \text{for } q = 2, 3, \ldots, n-1 \quad \text{and} \quad i1, \ldots, iq \in \{1, \ldots, n\}, \end{aligned} \tag{7}$$

where $C_q = \Phi^{-1}(1/{}_nP_q)$, $v_q = 1/\phi(C_q){}_nP_q$, ${}_nP_q = n(n-1)\ldots n-q+1$, $\mu_{i;n}$ is the ith expected standard normal order statistic and n is the total number of horses in the race.

Scaling is required to adjust formula (7) so that all conditional probabilities sum to 1. For each q, the following log-likelihood function conditional on the previous $q-1$ finishing order is to be maximized with respect to the parameter of interest:

$$l_q = \sum_{k=1}^{m} \log\left(\prod_{i=q}^{n_k} \pi_{i|[q-1],k}^{y_{ik}} \right) = \sum_{k=1}^{m} \log \pi_{[q]|[q-1],k}, \qquad q = 2, 3, \ldots, n-1,$$

where k denotes the race number, $y_{ik} = 1$ if horse i finishes qth in race k, $y_{ik} = 0$ otherwise, n_k is the number of horses in race k, m is the total number of races in the data set and the subscript $[i]$ denotes the horse finishing in the ith position. From models (6), if the parameter of interest is equal to 1, $\pi_{i|[q-1]}$ (the model-based probability) will be the same as $P_{i|[q-1]}$ which is estimated by either the Henery or the Harville model.

In addition, we define the log-likelihood of the results of the first q to finish as follows:

$$\begin{aligned} L_q &= \sum_{i=1}^{q} l_i \qquad \text{where } l_1 \text{ is now defined to be } \sum_{k=1}^{m} \log \pi_{[1],k} \\ &= \sum_{i=2}^{q} \sum_{k=1}^{m} \log \pi_{[i]|[i-1],k} + \sum_{k=1}^{m} \log \pi_{[1],k} \\ &= \sum_{k=1}^{m} \log \pi_{[1,2,\ldots,q],k} \end{aligned}$$

where $[1, 2, \ldots, q]$ denotes the first q horses to finish, $q = 2, 3, \ldots, n - 1$. It is easy to see that maximizing l_q and L_q is equivalent because each parameter is only contained in one of the l_q.

We have chosen 600 races involving eight horses from a Hong Kong (1981–89) data set for this analysis. The data were collected from the Royal Hong Kong Jockey Club's racing records for the years 1981–90. In our case, $n = 8$ and $q = 1, 2, 3, \ldots, 7$, where $q = 1$ corresponds to the simple constant β model (4). The results are shown in Tables 1 and 2. The estimated value of β (the parameter associated with $q = 1$), 0.965, is very close to 1, and it is not significantly different from 1 at any reasonable level by a likelihood ratio test. Thus, from this analysis, the favourite–long shot bias does not exist in Hong Kong; this result is consistent with Busche and Hall's (1988) findings. Hence, we simply use the win bet fractions F_i for computing the Harville or Henery conditional probabilities P for $q = 2, \ldots, 7$. If β was significantly different from 1, we might use the model-based win probability estimates π using model (4) instead of F_i. Note that $l_q(1)$ and l_q(para. est.) in these tables are log-likelihood values when the appropriate parameter is 1 (thus the model-based probabilities will reduce to the probabilities associated with the Harville or Henery model) and the maximum likelihood estimate respectively, and similarly for $L_q(1)$ and L_q(para. est.).

It may be easier to observe the pattern of systematic differences between the two models in Figs 1–3. Fig. 1 indicates that the difference in log-likelihood values l_q(Henery) $- l_q$(Harville) between the two models increases with the level q. This shows that the Henery model is generally better especially when q is large, i.e. if the conditional probability that a horse finishes in a lower position is to be predicted. This can also be judged by looking at $l_q(1)$ or $L_q(1)$ in Tables 1 and 2. Fig. 2 shows that the estimated parameters for the Harville model decrease with q. This means that the bias caused by the Harville model (i.e. the deviation of the estimated parameter from 1) is more serious for larger values of q. In contrast, as shown in Fig. 3, the estimated parameters for the Henery model are quite close to 1 for various

TABLE 1
Likelihood analysis for the Harville model

	Values for the following values of q:						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Parameter estimates	0.9653	0.8551	0.6675	0.5271	0.4480	0.3616	0.2369
$l(1)$	−1102.7	−1055.2	−1015.5	−949.3	−827.8	−689.5	−463.9
l(para. est.)	−1102.6	−1052.6	−1000.8	−917.9	−791.9	−639.8	−409.3
$L(1)$	−1102.7	−2157.9	−3173.4	−4122.7	−4950.5	−5640.0	−6103.9
L(para. est.)	−1102.6	−2155.2	−3156.0	−4073.9	−4865.8	−5505.6	−5914.9

TABLE 2
Likelihood analysis for the Henery model

	Values for the following values of q:						
	1	*2*	*3*	*4*	*5*	*6*	*7*
Parameter estimates	0.9653	1.1358	1.1002	1.0372	1.1747	0.9681	0.7134
$l(1)$	−1102.7	−1061.5	−1007.2	−921.8	−793.0	−641.8	−411.3
l(para. est.)	−1102.6	−1060.1	−1006.7	−921.8	−792.2	−641.8	−410.5
$L(1)$	−1102.7	−2164.2	−3171.4	−4093.2	−4886.2	−5528.0	−5939.3
L(para. est.)	−1102.6	−2162.7	−3169.4	−4091.2	−4883.4	−5525.2	−5935.7

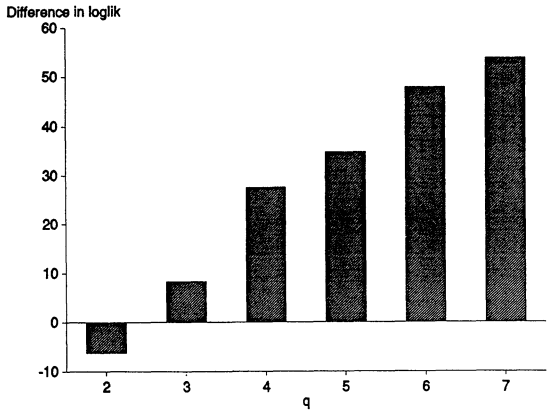

Fig. 1. Log-likelihood difference (Henery value minus Harville value)

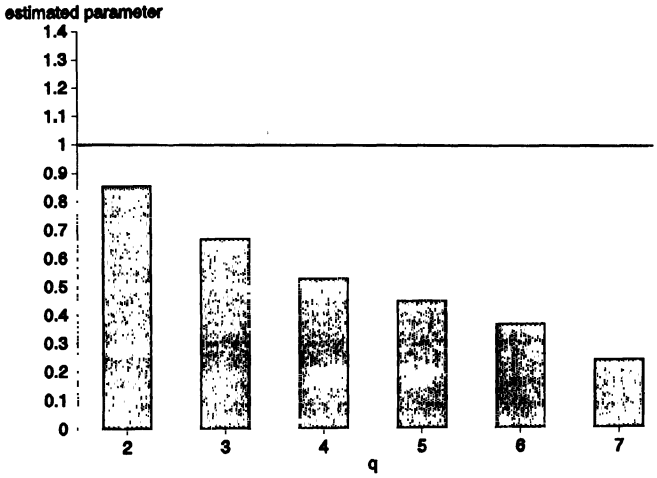

Fig. 2. Estimated parameters under the Harville model

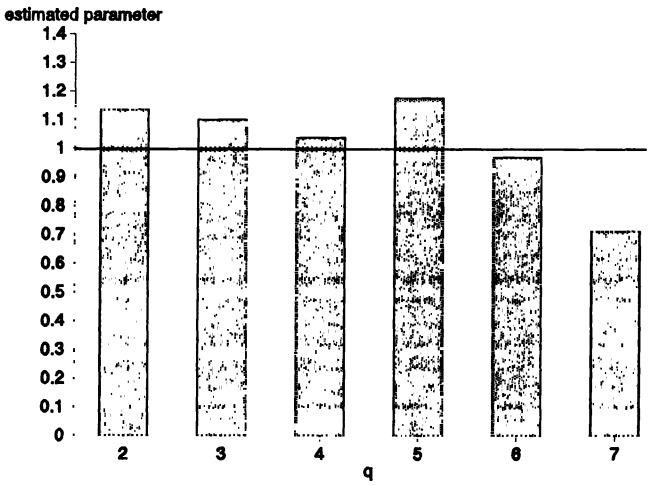

Fig. 3. Estimated parameters under the Henery model

values of q. (The log-likelihood for discrete events is simply the logarithm of the probability that the observed event occurs under a particular probability model. Thus a higher log-likelihood implies a higher chance associated with what we observed and thus the model is closer to reality. Some people may consider that the method of comparing log-likelihood values directly is not sufficiently rigorous given the non-nested models; detailed Cox (1962) tests reported in Bacon-Shone *et al.* (1992b) for simpler comparisons support these conclusions.)

4. Theoretical investigation of Harville and Henery models

In this section, the difference in estimating the conditional probability that horse j finishes second given that horse i finishes first (denoted by $P_{j|i}$) by the Harville and Henery models will be investigated theoretically under the assumption that the Henery model is correct, i.e. we shall study the following difference:

$$P_{j|i} - \frac{P_j}{1 - P_i} \tag{8}$$

where $P_{j|i}$ is now estimated by the Henery model and $P_j/(1 - P_i)$ is the corresponding expression under the Harville model. Let θ_i, the expected running times of horse i, equal $E(T_i)$. (Note that $n \geqslant 3$. Otherwise, there is no need to discuss $P_{j|i}$.)

Theorem 1.

$$P_{a|i} - \frac{P_a}{1 - P_i} \leqslant 0 \qquad \text{and} \qquad P_{b|i} - \frac{P_b}{1 - P_i} \geqslant 0$$

where

$$\theta_a = \min_{r \neq i} (\theta_r) \qquad \text{and} \qquad \theta_b = \max_{r \neq i} (\theta_r).$$

The proof appears in Appendix A.

When $n > 3$, we have only shown that the above result is valid for extreme values of j (the strongest and weakest horses measured in expected mean running times). But for j such that $\theta_a < \theta_j < \theta_b$, the difference may be greater than or smaller than 0 depending on the particular set of $(\theta_1, \theta_2, \ldots, \theta_n)$.

Theorem 1 means that, if the running times satisfy the assumptions of the Henery model, the Harville model will overestimate the conditional probability that the favourite horse will finish second and underestimate the conditional probability that the long shot will finish second. In the previous section, we have shown that the Henery model generally fits our data better; therefore, according to theorem 1, a systematic bias is expected to appear when we use the Harville model instead. This is also consistent with the parameter estimates of the logit models (6) (see Tables 1 and 2). Recall that the Harville model is more commonly used because of its simplicity. A parallel theoretical result for the comparison between the Stern (1990) model and the Harville model appears in Lo (1994).

5. Conclusion

The results obtained in this paper support the conclusion that the Harville model has a systematic bias in estimating ordering probabilities based on our data set. In contrast, according to our data analysis, the Henery model does not cause any systematic bias and thus it appears to be more reliable. Our theoretical result in section 4 analytically supports the systematic bias caused by the Harville model when the Henery model holds.

Acknowledgements

We acknowledge Kelly Busche for the Hong Kong data. In addition, we are grateful for the comments of a referee.

Appendix 1: proof of theorem 1

We need the following lemmas.

Lemma 1. Let u, v and w be non-negative functions. Moreover, u is non-decreasing and v/w is non-increasing; then,

$$\int uv \Big/ \int uw \leqslant \int v \Big/ \int w.$$

For the proof, see Gutmann and Maymin (1987).

Lemma 2. Define the following function:

$$J(v; \theta_j) = \frac{\phi(v + \theta_j - \theta_i) \prod_{r \neq ij} \Phi(v - \theta_i + \theta_r)}{\sum_{s \neq ij} \phi(v + \theta_s - \theta_i) \prod_{t \neq is} \Phi(v - \theta_i + \theta_t)}$$

where $\phi(\ )$ and $\Phi(\ )$ are the probability density function and the cumulative density function of the standard normal distribution respectively. Then,

$$J(v; \theta_a) \text{ is non-decreasing in } v$$

and

$$J(v; \theta_b) \text{ is non-increasing in } v,$$

where $\theta_a = \min_{r \neq i}(\theta_r)$ and $\theta_b = \max_{r \neq 1}(\theta_r)$.

Lemma 2 can be proved by differentiating $J(\ )$ with respect to v. As the proof is very tedious, it is not reported here. It can be found in chapter 5 of Lo (1992).

Proof of theorem 1. Consider the difference (8),

$$P_{j|i} - \frac{P_j}{1 - P_i} = \frac{1}{P_i}\left(P_{ij} - \frac{P_i P_j}{1 - P_i}\right)$$

$$\begin{aligned} P_{ij} &= \int_{-\infty}^{\infty} \{1 - \Phi(u - \theta_j + \theta_i)\} \prod_{s \neq ij} \Phi(u - \theta_j + \theta_s)\, \phi(u)\, du \\ &= \int_{-\infty}^{\infty} \prod_{s \neq ij} \Phi(u - \theta_j + \theta_s)\, \phi(u)\, du - \int_{-\infty}^{\infty} \prod_{r \neq j} \Phi(u - \theta_j + \theta_r)\, \phi(u)\, du \\ &= P_{j(i)} - P_j \end{aligned}$$

where

$$P_{j(i)} = P\left[T_j < \min_{s \neq i}\{T_s\}\right],$$

i.e. the probability that horse j wins if horse i is removed from the race.

Therefore,

$$P_{j|i} - \frac{P_j}{1 - P_i} = \frac{1}{P_i}\left(P_{j(i)} - \frac{P_j}{1 - P_i}\right).$$

Define

$$g_{j|i} = P_{j(i)} - \frac{P_j}{1 - P_i}.$$

Thus, it suffices to show that $g_{a|i} \leqslant 0$ and $g_{b|i} \geqslant 0$. Rewrite

$$g_{j|i} = \frac{P_{j(i)}\left(\sum_{s \neq ij} P_s + P_j\right) - P_j}{1 - P_i}$$

$$= \frac{\sum_{s \neq ij} P_s \sum_{s \neq ij} P_{s(i)}}{1 - P_i}\left(\frac{P_{j(i)}}{\sum_{s \neq ij} P_{s(i)}} - \frac{P_j}{\sum_{s \neq ij} P_s}\right).$$

Now, we consider

$$\frac{P_{j(i)}}{\sum_{s \neq ij} P_{s(i)}} - \frac{P_j}{\sum_{s \neq ij} P_s} = \frac{\int_{-\infty}^{\infty} \prod_{r \neq ij} \Phi(u - \theta_j + \theta_r)\, \phi(u)\, du}{\sum_{s \neq ij} \int_{-\infty}^{\infty} \prod_{t \neq si} \Phi(u - \theta_s + \theta_t)\, \phi(u)\, du} - \frac{\int_{-\infty}^{\infty} \prod_{r \neq j} \Phi(u - \theta_j + \theta_r)\, \phi(u)\, du}{\sum_{s \neq ij} \int_{-\infty}^{\infty} \prod_{t \neq s} \Phi(u - \theta_s + \theta_t)\, \phi(u)\, du}$$

$$= \frac{\int_{-\infty}^{\infty} \prod_{r \neq ij} \Phi(v - \theta_i + \theta_r)\, \phi(v - \theta_i + \theta_j)\, dv}{\int_{-\infty}^{\infty} \sum_{s \neq ij} \prod_{t \neq si} \Phi(v - \theta_i + \theta_t)\, \phi(v - \theta_i + \theta_s)\, dv}$$

$$- \frac{\int_{-\infty}^{\infty} \Phi(v) \prod_{r \neq ij} \Phi(v - \theta_i + \theta_r)\, \phi(v - \theta_i + \theta_j)\, dv}{\int_{-\infty}^{\infty} \Phi(v) \sum_{s \neq ij} \prod_{t \neq si} \Phi(v - \theta_i + \theta_t)\, \phi(v - \theta_i + \theta_s)\, dv}$$

by a change of variables using $v = u - \theta_j + \theta_i$ in the numerator and $v = u - \theta_s + \theta_i$ in the denominator. This expression is less than or equal to 0 when $j = a$ and greater than or equal to 0 when $j = b$ by using lemma 2 together with lemma 1. Hence, $g_{a|i} \leqslant 0$ and $g_{b|i} \geqslant 0$ and the required result follows.

References

Ali, M. M. (1977) Probability and utility estimates for racetrack bettors. *J. Polit. Econ.*, **84**, 803–815.

Asch, P., Malkiel, B. and Quandt, R. (1982) Racetrack betting and informed behavior. *J. Fin. Econ.*, **10**, 187–194.

Asch, P. and Quandt, R. (1987) Efficiency and profitability in exotic bets. *Economica*, **54**, 289–298.

Bacon-Shone, J., Lo, V. S. Y. and Busche, K. (1992a) Modelling the winning probability. *Research Report 10.* Department of Statistics, University of Hong Kong, Hong Kong.

———(1992b) Logistic analyses for complicated bets. *Research Report 11.* Department of Statistics, University of Hong Kong, Hong Kong.

Busche, K. and Hall, C. D. (1988) An exception to the risk preference anomaly. *J. Bus.*, **61**, 337–346.

Cox, D. R. (1962) Further results on tests of separate families of hypotheses. *J. R. Statist. Soc.* B, **24**, 406–424.

Dansie, B. R. (1983) A note on permutation probabilities. *J. R. Statist. Soc.* B, **45**, 22–24.

Fabricand, B. P. (1979) *The Science of Winning: a Random Walk on the Road to Riches.* New York: Van Nostrand Reinhold.

Griffith, R. M. (1949) Odds adjustments by American horse-racing bettors. *Am. J. Psychol.*, **62**, 290–294.

Gutmann, S. and Maymin, Z. (1987) Is the selected population the best? *Ann. Statist.*, **15**, 456–461.

Harville, D. A. (1973) Assigning probabilities to the outcomes of multi-entry competitions. *J. Am. Statist. Ass.*, **68**, 312–316.

Hausch, D. B., Lo, V. S. Y. and Ziemba, W. T. (eds) (1994) *Efficiency of Racetrack Betting Markets.* New York: Academic Press. To be published.

Hausch, D. B., Ziemba, W. T. and Rubinstein, M. (1981) Efficiency of the market for racetrack betting. *Mangmnt Sci.*, **27**, 1435–1452.
Henery, R. J. (1981) Permutation probabilities as models for horse races. *J. R. Statist. Soc.* B, **43**, 86–91.
———(1985) On the average probability of losing bets on horses with given starting price odds. *J. R. Statist. Soc.* A, **148**, 342–349.
Hoerl, A. E. and Fallin, H. K. (1974) Reliability of subjective evaluations in a high incentive situation. *J. R. Statist. Soc.* A, **137**, 227–230.
Lo, V. S. Y. (1992) Statistical modelling of gambling probability. *PhD Thesis.* Department of Statistics, University of Hong Kong, Hong Kong.
———(1994) Application of running time distribution model in Japan. In *Efficiency of Racetrack Betting Markets* (eds D. B. Hausch, V. S. Y. Lo and W. T. Ziemba). New York: Academic Press. To be published.
Luce, R. D. and Suppes, P. (1965) Preference, utility and subjective probability. In *Handbook of Mathematical Psychology* (eds R. D. Luce, R. R. Bush and E. Galanter), vol. III, ch. 19, pp. 249–410. New York: Wiley.
McCulloch, B. and Van Zijl, T. (1986) Direct test of Harville's multi-entry competitions model on race-track betting data. *J. Appl. Statist.*, **13**, 213–220.
McGlothlin, W. H. (1956) Stability of choices among uncertain alternatives. *Am. J. Psychol.*, **69**, 604–619.
Plackett, R. L. (1975) The analysis of permutations. *Appl. Statist.*, **24**, 193–202.
Snyder, W. W. (1978) Horse racing: testing the efficient markets model. *J. Fin.*, **33**, 1109–1118.
Stern, H. (1990) Models for distributions on permutations. *J. Am. Statist. Ass.*, **85**, 558–564.
Tuckwell, R. H. (1981) Anomalies in the gambling market. *Aust. J. Statist.*, **23**, 287–295.
Ziemba, W. T. and Hausch, D. B. (1985) *Betting at the Racetrack.* New York: Strauss.
———(1987) *Dr. Z's Beat the Racetrack.* New York: Morrow.