# RESEARCH REPORT

Serial No. 16

June 1992

## AN APPROXIMATION TO ORDERING PROBABILITIES OF MULTI-ENTRY COMPETITIONS

by

Victor S.Y. Lo and John Bacon-Shone

THE UNIVERSITY OF HONG KONG

DEPARTMENT OF STATISTICS

Pokfulam Road, Hong Kong. Tel: (852) 859 2466; Fax: (852) 858 9041
E-Mail: STATIST@HKUCC.HKU.HK or STATIST@HKUCC.BITNET

AN APPROXIMATION TO ORDERING PROBABILITIES OF MULTI-ENTRY COMPETITIONS

Victor S.Y. Lo John Bacon-Shone

*Department of Statistics,*

*University of Hong Kong*

## Abstract

To predict the ordering probabilities of multi-entry competitions (e.g. horse-races), Harville (1973) proposed a simple way of computing the ordering probabilities based on the simple winning probabilities. This simple model essentially assumes the underlying model (e.g. running time in horse-racing) is independent exponential. Henery (1981) and Stern (1990) respectively proposed to use normal and gamma distributions for the running time. However, both the Henery and Stern model are too complicated to use in practice. Bacon-Shone, Lo & Busche (1992,b) have shown that the Henery model fits better in horse-racing using particular data sets. In this paper, we propose to use a simple way of computing ordering probabilities which approximate both the Henery and Stern model quite well. Using Hong Kong, U.S. and Japanese data, a large scale empirical investigation is undertaken.

*Keywords* : Ordering probabilities; Horse-races; Running time distributions

## I. Introduction

In multi-entry competitions, Harville (1973) proposed to use the following formula to compute the ordering probabilities :

$$\pi_{ij} = \frac{\pi_i \pi_j}{1 - \pi_i} \qquad (1)$$

where $\pi_{ij}$ = P(i wins and j finishes second), and

$\pi_i$ = P(i wins).

In horse-racing, i and j are two horses, and the value of $\pi_i$ can be estimated by the win bet fraction (see Ali (1977), Snyder (1978), Busche & Hall (1988) and Bacon-Shone, Lo & Busche (1992,a) for details of using the win bet fractions). Similarly for more complicated ordering probabilities.

This simple formula (1) is implied by the assumption of independent exponential distributions for running times with different parameters for each horse in each race (Dansie (1983)). Henery (1981) proposed assuming independent normal distribution for the running times (hereafter called the Henery model). However, numerical integration or an approximation method has to be used. Similarly, Stern (1990) proposed to use gamma distribution with fixed integral shape parameter r. Similar to the Henery model, we have to find the parameters using the win bet fractions by solving a complicated set of nonlinear equations. For descriptions of the three models, see Bacon-Shone, Lo & Busche (1992,b). Bacon-Shone, Lo & Busche (1992,b) reported many empirical analyses of different complicated bets. Our conclusion is that using the information from win bet fractions alone, for the analyses of exacta bet (in Meadowlands), trifecta bet (in Meadowlands and Hong Kong) and quinella bet (in Hong Kong), the Henery model was found to be better than the others in predicting the relevant ordering probabilities for those bets according to a likelihood approach. The results are confirmed by Cox's test (Cox(1962)). For details, see Bacon-Shone, Lo & Busche (1992,b).

In this paper, we consider a simple approximation to the Henery model in section II and extend the approximation to the Stern model in section III. Conclusion will be given in section IV.

## II. A simple approximation of the Henery model

For the Henery model, computation of $\pi_{ij}$ is not simple because it involves integrations and approximations. We now propose a simple way based on the fact that a function of the ordering probabilities is close to a constant.

Define :

$$\lambda^{Hen}_{ijl} = \frac{\ln(\hat{\pi}_{ij}/\hat{\pi}_{il})}{\ln(\hat{\pi}_j/\hat{\pi}_l)} \qquad (2)$$

We can compute this term for any combination of i,j,l in all the races. This $\lambda_{ijl}^{Hen}$ is very close to a constant (say, $\lambda^{Hen}$) for different i,j,l and in different races. Thus, based on (2), $\hat{\pi}_{ij}$ can be estimated directly instead of using numerical integrations by

$$\hat{\pi}_{ij} = \hat{\pi}_i \frac{\hat{\pi}_j^{\lambda^{Hen}}}{\sum_r \hat{\pi}_r^{\lambda^{Hen}}}$$

We may take $\hat{\pi}_i$ as the win bet fraction.

Consider the following two models :

$$\text{logit } \pi_{j|i} = \mu_{Harv} \text{logit } P_{j|i} \qquad \text{(win bet)} \qquad j \neq i \qquad (3)$$

$$\text{logit } \pi_{j|i} = \mu_{Hen} \text{logit } \hat{\pi}_{j|i}^{Hen} \qquad (4)$$

where $\hat{\pi}_{j|i}^{Hen} = \hat{\pi}_{ij}^{Hen} / \hat{\pi}_i^{Hen}$ $\qquad j \neq i$

and $\hat{\pi}_{ij}^{Hen}$ are obtained by the Henery model using win bets,

$\hat{\pi}_i^{Hen} = \sum_{r \neq i} \hat{\pi}_{ir}^{Hen}$,

$\pi_{j|i}$ = P(horse j finishes second | horse i wins),

$P_{j|i} = P_j / (1-P_i)$ based on the Harville model,

$P_i$ = Win bet fraction of horse i.

Empirical results for the models in (3) and (4) were reported in Bacon-Shone, Lo & Busche (1992,b). We may estimate the constant $\lambda^{Hen}$ by $\hat{\mu}_{Harv} / \hat{\mu}_{Hen}$ using maximum likelihood because :

If (3) and (4) are true,

$$\mu_{Harv} / \mu_{Hen} = \text{logit } \hat{\pi}_{j|i} / \text{logit } P_{j|i}$$

$$= \frac{\ln(\hat{\pi}_{j|i} / \hat{\pi}_{l|i})}{\ln(P_{j|i} / P_{l|i})} = \frac{\ln(\hat{\pi}_{ij} / \hat{\pi}_{il})}{\ln(P_j / P_l)} \qquad \text{for any } l$$

$$= \frac{\ln(\hat{\pi}_{ij} / \hat{\pi}_{il})}{\ln(\hat{\pi}_j / \hat{\pi}_l)} = \lambda_{ijl}^{Hen}$$

It can be found in the Hong Kong and Meadowlands data sets that the $\lambda^{Hen}_{ijl}$ is close to $\lambda^{Hen}$ = 0.76 which is itself close to the value of $\hat{\mu}_{Harv} / \hat{\mu}_{Hen}$ obtained by maximum likelihood method.

Some summary values of $\lambda^{Hen}_{ijl}$ in Hong Kong, Meadowlands and Japan are shown in Table 1.

Table 1

Summary values of $\lambda^{Hen}_{ijl}$

| Racetrack | mean | standard deviation |
|---|---|---|
| Hong Kong (89) | 0.76694 | 0.023841 |
| Meadowlands | 0.75609 | 0.021646 |
| Japan | 0.77929 | 0.033924 |

(Note : In the above table, i,j are horses finishing first and second, respectively. Horse l (≠i,j) varies.)

Further, similar observation can be found for $\tau$, where

$$\tau^{Hen}_{ijkl} = \frac{\ln(\hat{\pi}_{ijk}/\hat{\pi}_{ijl})}{\ln(\hat{\pi}_k/\hat{\pi}_l)} \qquad (5)$$

Again, from our data, we observe that this $\tau^{Hen}$ is estimated to be 0.62 by using a ratio of maximum likelihood estimators of two parameters similar to the above. The summary values of $\tau^{Hen}_{ijkl}$ are shown in Table 2.

Table 2

Summary values of $\tau^{Hen}_{ijl}$

| Racetrack | mean | standard deviation |
|---|---|---|
| Hong Kong (89) | 0.65182 | 0.034036 |
| Meadowlands | 0.63932 | 0.042666 |
| Japan | 0.66051 | 0.038304 |

(In the above table, i,j and k are horses finishing first, second and third, respectively. Horse l (≠i,j,k) varies.)

Although the mean of $\tau^{Hen}_{ijl}$ is close to 0.64 or 0.65 in the above table, using 0.62 does not make much difference. This can be seen by comparing the use of 0.62 and 0.65 as shown in Table 3. In this table, $\lambda^{Hen}$ is fixed at 0.76 and the following model is used (hereafter called the "discount model") :

$$\hat{\pi}_{ijk} = \hat{\pi}_i \frac{\hat{\pi}_j^{\lambda^{Hen}}}{\sum_{s \neq i} \hat{\pi}_s^{\lambda^{Hen}}} \frac{\hat{\pi}_k^{\tau^{Hen}}}{\sum_{t \neq i j} \hat{\pi}_t^{\tau^{Hen}}} \tag{6}$$

We can see that there is not a big difference in the log likelihood values when $\tau^{Hen}$ is set to 0.62 and 0.65.

Table 3

Empirical comparisons for $\tau^{Hen}$ = 0.65 & 0.62

| Racetracks | $\tau^{Hen}$ | $l(1)$ |
|---|---|---|
| Hong Kong | 0.65 | -700.20 |
| | 0.62 | -699.68 |
| Meadowlands | 0.65 | -10667.47 |
| | 0.62 | -10667.80 |

To understand the effect of race size on $\lambda^{Hen}$ and $\tau^{Hen}$, we can compute the summary values for different race sizes. This is shown in the following tables.

Table 4 (a)

Summary values of $\lambda^{Hen}$ and $\tau^{Hen}$ for different race sizes in Hong Kong (89)

| | | $\lambda^{Hen}$ | | $\tau^{Hen}$ | |
|---|---|---|---|---|---|
| race size | no.of races | mean | s.d. | mean | s.d. |
| 4 | 4 | .66849 | .040495 | .51855 | .012520 |
| 5 | 6 | .69144 | .036893 | .55803 | .020846 |
| 6 | 20 | .72090 | .021310 | .58774 | .016298 |
| 7 | 30 | .74073 | .020627 | .61269 | .018581 |
| 8 | 78 | .75068 | .019147 | .62511 | .021593 |
| 9 | 54 | .76199 | .016006 | .63991 | .022372 |
| 10 | 88 | .76540 | .015475 | .65187 | .024506 |
| 11 | 28 | .77256 | .017335 | .65686 | .023166 |
| 12 | 42 | .77764 | .014179 | .66479 | .026024 |
| 13 | 28 | .78205 | .017591 | .67245 | .026179 |
| 14 | 43 | .78585 | .017244 | .67855 | .033779 |

Table 4 (b)

Summary values of $\lambda^{Hen}$ and $\tau^{Hen}$ for different race sizes in Meadowlands

| race size | no.of races | $\lambda^{Hen}$ mean | $\lambda^{Hen}$ s.d. | $\tau^{Hen}$ mean | $\tau^{Hen}$ s.d. |
|---|---|---|---|---|---|
| 6 | 10 | .70887 | .029418 | .58171 | .021856 |
| 7 | 16 | .72873 | .025289 | .60771 | .023279 |
| 8 | 59 | .74547 | .022320 | .62140 | .025569 |
| 9 | 119 | .75221 | .020152 | .63577 | .028272 |
| 10 | 275 | .76051 | .018940 | .64514 | .030223 |
| 11 | 20 | .75881 | .018647 | .62615 | .134344 |
| 12 | 11 | .75728 | .031161 | .66438 | .032847 |

Table 4 (c)

Summary values of $\lambda^{Hen}$ and $\tau^{Hen}$ for different race sizes in Japan

| race size | no.of races | $\lambda^{Hen}$ mean | $\lambda^{Hen}$ s.d. | $\tau^{Hen}$ mean | $\tau^{Hen}$ s.d. |
|---|---|---|---|---|---|
| 5 | 16 | .71797 | .018695 | .56012 | .016663 |
| 6 | 48 | .73225 | .025915 | .58574 | .020141 |
| 7 | 78 | .74290 | .027168 | .60243 | .025457 |
| 8 | 148 | .75414 | .028855 | .62272 | .024623 |
| 9 | 186 | .76282 | .030605 | .63234 | .027260 |
| 10 | 213 | .76952 | .030276 | .64505 | .029224 |
| 11 | 181 | .77911 | .029083 | .65454 | .028384 |
| 12 | 221 | .77880 | .031315 | .66154 | .029983 |
| 13 | 102 | .78498 | .030542 | .66749 | .030749 |
| 14 | 109 | .78867 | .026733 | .67742 | .029808 |
| 15 | 66 | .79424 | .028137 | .68049 | .035557 |
| 16 | 188 | .79718 | .031614 | .68721 | .032792 |
| 17 | 5 | .82112 | .030213 | .68467 | .032733 |
| 18 | 22 | .79706 | .035295 | .69389 | .033436 |

From the above tables, we see that $\lambda^{Hen}$ and $\tau^{Hen}$ have an increasing trend as the race size n increases but the values do not vary a lot.

Simulation results

We will present some simulation results to support our idea of $\lambda^{Hen} \approx 0.76$.

Assume that, for each race, $\theta_i \overset{iid}{\sim} N(\mu_0,\sigma_0)$ for i=1,2,...,n, where $\mu_0$ is an arbitrary constant (since ordering probabilities depend on the difference between $\theta_i$'s only) and $\sigma_0^2$ is a prespecified value. The value $\sigma_0$ can be interpreted as a measure of dispersion of the mean running times of the horses in the same race. In other words, it measures the variation of abilities (or winning probabilities) of the horses in the race. Based on this assumption, we can use Monte Carlo simulation of $\underline{\theta} = (\theta_1,\ldots,\theta_n)^T$ and then compute $\lambda^{Hen}$ using (3).

We have set n = 10 for our simulation purpose as the average number of horses in both of our data sets are about 10. We try different $\sigma_0$ to observe its influence on the $\lambda^{Hen}_{ijl}$. Fifty races are simulated for each $\sigma_0$. For each race, we have fixed two horses for i & j but l is varying over the other horses and thus, we have eight $\lambda^{Hen}_{ijl}$'s for each race. Therefore, there are 50x8 = 400 $\lambda^{Hen}_{ijl}$'s for each simulation. The simulation results are shown in Table 5.

Table 5

Simulations of $\lambda^{Hen}_{ijk}$ for n=10 and 50 races

| $\sigma_0$ | mean of $\lambda^{Hen}$ | s.d. of $\lambda^{Hen}$ |
|---|---|---|
| 0.2 | 0.7697 | 0.00913 |
| 0.4 | 0.7666 | 0.02477 |
| 0.6 | 0.7571 | 0.03806 |
| 1.0 | 0.7513 | 0.05679 |
| 1.5 | 0.7319 | 0.09233 |

From Table 5, we observe that the mean value of $\lambda^{Hen}$ is close to 0.76 for all $\sigma_0$ values though the standard deviation depends on $\sigma_0$. This means the the accuracy of our discount model depends on $\sigma_0$.

Large $\sigma_0$ may affect our approximation of $\pi_{ij}$. Simulation results for other extreme values of n are shown in Table 6 and 7. Though the mean values deviate from 0.76 a little, the differences are quite small and thus using 0.76 should not have any serious adverse effect on the estimation of $\pi_{ij}$.

Table 6

Simulations of $\lambda_{ijk}^{Hen}$ for n=7 and 50 races

| $\sigma_0$ | mean of $\lambda^{Hen}$ | s.d. of $\lambda^{Hen}$ |
|---|---|---|
| 0.2 | 0.7475 | 0.00969 |
| 0.4 | 0.7400 | 0.02292 |
| 0.6 | 0.7437 | 0.03401 |
| 1.0 | 0.7486 | 0.04834 |
| 1.5 | 0.7467 | 0.09244 |

Table 7

Simulations of $\lambda_{ijk}^{Hen}$ for n=14 and 50 races

| $\sigma_0$ | mean of $\lambda^{Hen}$ | s.d. of $\lambda^{Hen}$ |
|---|---|---|
| 0.2 | 0.7865 | 0.01972 |
| 0.4 | 0.7912 | 0.02625 |
| 0.6 | 0.7881 | 0.03918 |
| 1.0 | 0.7738 | 0.05914 |
| 1.5 | 0.7774 | 0.09160 |

## Empirical analysis using the discount model

In this section, we will use fixed $\lambda$ and $\tau$ to compare with different models. Some empirical results are shown in Table 8.

Table 8

Comparison among different models in different bet types

| Models | loglik | Models | loglik |
|---|---|---|---|
| Exacta : 510 races (Meadowlands) | | Quinella : 4153 races (Hong Kong) | |
| Harville | -1875.77 | Harville | -13619.28 |
| Henery | -1859.63 | Henery | -13589.55 |
| discount | -1859.25 | discount | -13586.95 |
| Trifecta : 120 races (Meadowlands) | | Trifecta : 1809 races (Hong Kong) | |
| Harville | -711.50 | Harville | -10747.98 |
| Henery | -699.83 | Henery | -10667.25 |
| discount | -699.68 | discount | -10667.80 |

From Table 8, it is clear that the accuracy (measured by the log likelihoods) of the discount model is close to that of the Henery model.

Comparison of probability estimations using a closeness measure

In this subsection, we aim at comparing the closeness of probability estimations produced by different models by assuming the Henery model is correct. That is, we want to show that our discount model is a relatively close to the Henery model. We apply the following well-known closeness measure for our comparison purpose.

$$I(\underline{\hat{\pi}} \; ; \; \underline{\pi}^{*}) = \sum \sum \hat{\pi}_{ij} \ln(\frac{\hat{\pi}_{ij}}{\pi^{*}_{ij}})$$

where $\hat{\pi}_{ij}$ is the exacta probability (i.e. P(i wins and j finishes second) ) produced by the Henery model, and $\pi^{*}_{ij}$ is the associated probability produced by other model. This is called the Kullback-Leibler quantity of information (hereafter called the KL information quantity) which has the following properties :

(i) $I(\hat{\underline{\pi}};\underline{\pi}^*) \geq 0$,

(ii) $I(\hat{\underline{\pi}};\underline{\pi}^*) = 0$ iff $\hat{\pi}_{ij} = \pi^*_{ij}$ (i,j = 1,...,n)

(see Sakamoto, Ishiguro & Kitagawa (1986) for details)

We adopt the above quantity to compare the closeness of two distributions. Namely, the smaller the value of $I(\hat{\underline{\pi}};\underline{\pi}^*)$, the closer we consider the model for $\pi^*_{ij}$ to the Henery model. Other approximations to the Henery model are also considered for comparison purposes. Application of the Henery model involves two stages :

(i) Compute $\underline{\theta}$ based on the win bet fractions,

(ii) Compute more complicated probabilities based on the $\underline{\theta}$ obtained in (i).

Here, we include both first and second order Taylor series approximations in both stages for comparisons. The first order Taylor series approximation is due to Henery (1981). The second order Taylor series approximation formula is developed by the first author and is available upon request. The result of comparisons for exacta probability in Meadowlands (510 races) is shown in the following table.

Table 9

Comparisons using KL information quantity

| Model | average KL | s.d. of KL |
|---|---|---|
| (i) 1st order Taylor series for $\underline{\theta}$ (Henery) | | |
| a) 1st order for $\pi_{ij}$ | 0.013105 | 0.011799 |
| b) 2nd order for $\pi_{ij}$ | 0.014162 | 0.015276 |
| c) Numerical integration for $\pi_{ij}$ | 0.002217 | 0.003137 |
| (ii) 2nd order Taylor series for $\underline{\theta}$ (Henery) | | |
| a) 2nd order for $\pi_{ij}$ | 0.000415 | 0.000601 |
| b) Numerical integration for $\pi_{ij}$ | 0.000723 | 0.000679 |
| (iii) Fixed $\lambda$ | 0.000339 | 0.000262 |
| (iv) Harville | 0.019989 | 0.008370 |

In the above table, inclusion of the Harville model is to show the relative large difference between the Harville and the Henery models. Each result is a comparison between the stated model with the exact Henery model. The exact Henery model is based on a numerical method for computing $\underline{\theta}$ and numerical integration for $\pi_{ij}$. We can see that the discount model is closest to the Henery model. Besides, (ii) a) is quite good but that still involves a lot of computation time when compared to the discount model. Hence, the discount model is very close to the exact Henery model and very convenient to use in practice.

## III. Approximation to a more general model

In this section, we discuss the approximation to a more general model - the Stern model (Stern(1990)) which assumes that the running times follow the Gamma distribution with a fixed shape parameter, r. It is more general in the sense that when the shape parameter r=1, it reduces to the Harville model; when r=∞, it becomes the Henery model. Maximum likelihood estimation of r in Japan will also be reported. By using a likelihood-based argument, we will show that Stern's Gamma model with maximum likelihood estimate of r is better than both the Harville (r=1) and Henery (r=∞) models in Japan. Since predicting complicated probabilities under the Stern model is computationally intensive, we will propose a simple approximation and give numerical evidence.

### Fitting the Stern model

Stern's Gamma model (Stern (1990)) is motivated by considering a competition in which n players, scoring points according to independent Poisson processes, are ranked according to the time until r points are scored. Thus r should be an integer under this assumption. Whether this assumption is reasonable or not when applied to horse-racing problem is an open question. But we can consider it as an alternative model to the Harville and Henery models. Let the running time of horse i, $T_i \sim \text{Gamma}(r,\theta_i)$ independently or;

$$g_r(t_i|\theta_i) = \frac{1}{\Gamma(r)}\, \theta_i^r t_i^{r-1} \exp(-\theta_i t_i) \qquad t_i > 0.$$

where r is predetermined and $\theta_i$ can be estimated from $\pi_i$ (or the bet fraction, $P_i$).

We may try to estimate this r by comparing the log likelihood : $-\sum_l \ln \pi_{[123]l}$, where [123]l denotes the 3 top horses in race l,

with different values of r. The result for Japanese data is shown in Table 10. The computations are done by using Gauss-Laguerre integration for the Stern model and Gauss-Hermitian integration for the Henery model (i.e. r=∞). For the Stern model, we need to find $\underline{\theta}$ first by solving the following equation :

$$P_i = \int_0^\infty \prod_{s\neq i}[1-G_r(t_i|\theta_s)]\ g_r(t_i|\theta_i)\ dt_i$$

where $G_r(t_s|\theta_s)$ is the cumulative distribution function associated with $g_r(t_s|\theta_s)$ .

Table 10

Log likelihood values under the Stern model for Japanese data

| r | log likelihood |
|---|---|
| 1 (i.e. Harville) | -8977.57 |
| 2 | -8954.57 |
| 3 | -8950.60 |
| 4 | -8950.35 |
| 5 | -8950.94 |
| 6 | -8951.82 |
| 7 | -8952.65 |
| 8 | -8953.44 |
| ∞ (i.e. Henery) | -8986.88 |

From the above table, the log likelihood is maximized when r=4 Thus the Gamma distribution with r=4 is a better distributional assumption of running time in Japan. We may also fit the Stern model in Hong Kong and Meadowlands and the results are shown in Table 11 and 12 respectively.

Table 11

Log likelihood values under the Stern model for Hong Kong (89) data

| r | log likelihood |
|---|---|
| 1 | -2523.37 |
| 10 | -2504.55 |
| 20 | -2503.58 |
| 30 | -2503.44 |
| 40 | -2503.72 |
| ∞ (Henery) | -2502.74 |

Table 12

Log likelihood values under the Stern model for Meadowlands data

| r | log likelihood |
|---|---|
| 1 | -2845.93 |
| 10 | -2800.87 |
| 20 | -2798.02 |
| 30 | -2796.90 |
| 40 | -2795.78 |
| ∞ (Henery) | -2792.94 |

We can see that r=∞ (i.e. the Henery model) appears to be the best in both data sets.

The discount method proposed in section II is simple enough to apply the Henery model in practice. In this section, we will extend the idea for the Stern model. We define :

$$\lambda^{r}_{ijl} = \frac{\ln(\pi^{(r)}_{ij} / \pi^{(r)}_{il})}{\ln(\pi_j / \pi_l)}$$

$$\tau^{r}_{ijkl} = \frac{\ln(\pi^{(r)}_{ijk} / \pi^{(r)}_{ijl})}{\ln(\pi_k / \pi_l)} \qquad (7)$$

where $\pi^{(r)}_{ij}$ is P(horse i wins and horse j finishes 2nd) under the

Stern model with shape parameter r. Similarly for $\pi^{(r)}_{ijk}$. If we can assume that the above two values are close to two constants, denoted by $\lambda^r$ and $\tau^r$ respectively, the model for approximation of the Stern model is :

$$\pi^{(r)}_{ijk} = \pi_i \frac{\pi_j^{\lambda^r}}{\sum_{s \neq i} \pi_s^{\lambda^r}} \frac{\pi_k^{\tau^r}}{\sum_{t \neq i,j} \pi_t^{\tau^r}} \qquad (8)$$

where $\pi_i$'s are estimated by the win bet fraction $P_i$'s ; $\lambda^r$ and $\tau^r$ can be estimated by the mean of $\lambda^r_{ijl}$ and $\tau^r_{ijkl}$ based on a large number of races or by the ratio of maximum likelihood estimators. Here, we choose the first method because the second one requires to compute $\pi^{(r)}_{ijk}$ for all combinations and for each race. We have large number of races in Japan and r is varying and thus, the second method will be too tedious.

The summary values of $\lambda^r$ and $\tau^r$ based on 1583 races in Japan are shown in Table 13.

Table 13

Summary values of $\lambda^r$ and $\tau^r$

| r | $\lambda^r$ | | $\tau^r$ | |
|---|---|---|---|---|
| | mean | s.d. | mean | s.d. |
| 2 | 0.9336 | 0.01346 | 0.8920 | 0.02685 |
| 3 | 0.9021 | 0.01683 | 0.8423 | 0.02495 |
| 4 | 0.8836 | 0.01859 | 0.8140 | 0.02663 |
| 5 | 0.8712 | 0.01976 | 0.7953 | 0.02778 |
| 6 | 0.8623 | 0.02064 | 0.7819 | 0.02859 |
| 7 | 0.8555 | 0.02135 | 0.7717 | 0.02919 |
| 8 | 0.8500 | 0.02193 | 0.7636 | 0.02967 |

( Note : In this table, we have set i,j,k equal to the horses finishing in the top three positions in each race and l is varying for $\lambda^r_{ijl}$ and $\tau^r_{ijkl}$. )

From the above table, the standard deviations are quite small, in general. Hence, we expect the mean values are good approximations to the $\lambda^r_{ijl}$ and $\tau^r_{ijkl}$ for different combinations of horses in different races. The relation between $(\lambda^r, \tau^r)$ may be approximated as :

$$\hat{\lambda}^r \simeq \lambda^\infty + (\lambda^1 - \lambda^\infty)/\sqrt{r} \quad \text{and} \quad \hat{\tau}^r \simeq \tau^\infty + (\tau^1 - \tau^\infty)/\sqrt{r}$$

where $\lambda^1 = \tau^1 = 1$, $\lambda^\infty = 0.76$ and $\tau^\infty = 0.62$.

We also compute the summary statistics of KL information quantities for different r to compare the true $\pi_{ij}$ (obtained by numerical integration) with the above discount model (using the values in Table 13) in the following table. Also, the KL information quantities for comparing the true $\pi_{ij}$ with those predicted by the Harville model is treated as a control for comparison.

Table 14

Comparison between the discount model and the Harville model using KL information quantity (500 races)

| | Discount | | Harville | |
|---|---|---|---|---|
| r | Ave. KL | sd. KL | Ave. KL | sd. KL |
| 2 | .000286 | .000520 | .002098 | .001078 |
| 3 | .000085 | .000097 | .003961 | .001534 |
| 4 | .000124 | .000090 | .005670 | .002129 |
| 5 | .000142 | .000105 | .006983 | .002594 |
| 6 | .000155 | .000117 | .008046 | .002985 |
| 7 | .000166 | .000128 | .008912 | .003286 |
| 8 | .000176 | .000136 | .009651 | .003557 |

Clearly, by comparing the KL information quantities above, our discount model is much more accurate than the naive Harville model for predicting the complicated probabilities based on the Stern model.

Moreover, we compare the log likelihood values of Stern models using numerical integrations (i.e. from Table 10) with our discount model for predicting $\pi_{ijk}$ in the following table. Note that we have chosen $\lambda^\infty = \lambda^{Hen} = 0.76$ and $\tau^\infty = \tau^{Hen} = 0.62$. We observe that the log likelihood values based on two methods do not have big differences.

Table 15

Comparisons of log likelihood values for Japanese data

| r | numerical integrations | Discount |
|---|---|---|
| 1 (Harville) | -8977.57 | 8977.57 |
| 2 | -8954.57 | 8956.22 |
| 3 | -8950.60 | 8952.38 |
| 4 | -8950.35 | 8952.00 |
| 5 | -8950.94 | 8952.50 |
| 6 | -8951.82 | 8953.23 |
| 7 | -8952.65 | 8954.00 |
| 8 | -8953.44 | 8954.75 |
| ∞ (Henery) | -8986.88 | 8986.45 |

## IV. Conclusion

We have proposed to use the discount model in (8) with different values of $(\lambda^r,\tau^r)$ for different r. This model has been shown to provide good approximations to both the Henery and Stern model. It also includes the Harville model (r=1). To apply the model in practice (e.g. betting), we suggest to collect relevant data and find out what value of r is most appropriate and then apply (8) using appropriate parameter values. The effect of this improved probability estimation on betting strategy (e.g. the Dr.Z system proposed by Hausch, Ziemba & Rubinstein (1981)) is usually to improve the strategy. This is investigated in Lo & Bacon-Shone (1992).

## REFERENCES


Ali, M.M. (1977) "Probability and utility estimates for racetrack bettors." *Journal of Political Economy* 84 (August) :803-15.

Bacon-Shone,J.H.; Lo,V.S.Y. and Busche,K. (1992,a) "Modelling winning probability." Research report #10, Department of Statistics, University of Hong Kong.

Bacon-Shone,J.H.; Lo,V.S.Y. and Busche,K. (1992,b) "Logistic analyses of complicated bets." Research report #11, Department of Statistics, University of Hong Kong.

Busche,K. and Hall,C.D. (1988) "An exception to the risk preference anomaly." *Journal of Business* 61 : 337-46.

Cox. D.R. (1962) "Further results on tests of separate families of hypotheses." *Journal of Royal Statistical Society B* 24, No.2 : 406-424.

Dansie, B.R. (1983) "A note on permutation probabilities." *Journal of Royal Statistical Society B*, 45, No.1 : 22-24.

Harville,D.A. (1973) "Assigning probabilities to the outcomes of multi-entry competitions." *Journal of the American Statistical Association* 68 (June) : 312-16.

Hausch,D.B.;Ziemba,W.T. and Rubinstein,M. (1981) "Efficiency of the market for racetrack betting." *Management Science* 27 (December) : 1435-52.

Henery,R.J. (1981) "Permutation probabilities as models for horse races." *Journal of Royal Statistical Society B* 43, No.1 : 86-91.

Lo,V.S.Y., Bacon-Shone,J. & Busche, K. (1992) "A modified racetrack betting system." Research report, Department of Statistics, University of Hong Kong.

Sakamoto Y., Ishiguro, M. & Kitagawa,G. (1986) "Akaike information criterion statistics." D.Reidel publishing co..

Snyder, W. W. (1978) "Horse racing : Testing the efficient markets model." *Journal of Finance* 33 (September) : 1109-1118.

Stern, H. (1990) "Models for distributions on permutations." *Journal of the American Statistical Association* 85, No.410 (June) : 558-564.