RESEARCH REPORT

Serial No. 11

May 1992

LOGISTIC ANALYSES FOR COMPLICATED BETS

by

J. Bacon-Shone, V.S.Y. Lo and K. Busche

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

LOGISTIC ANALYSES FOR COMPLICATED BETS

John Bacon-Shone, Dept. of Statistics
Victor S. Y. Lo, Dept. of Statistics
Kelly Busche, School of Economics
University of Hong Kong

Abstract

The problem discussed is estimating the probabilities of finishing order in a horse race based on simple winning probabilities only. Some models have been proposed based on different assumptions of running time distributions of horses for this problem. However, no detailed data analyses for comparing these models can be found. In this paper, we apply logit models and utilize several data sets and bet types to study the goodness of these models in detail. These complicated bet types include exacta, trifecta and quinella bets. Formal tests for non-nested models are applied whenever possible. Our empirical results suggest that the model based on independent normal running times is better than the others.

To predict the winning probabilities of horses, many previous studies suggested that the win bet fractions are reasonable estimates. We utilize this information of winning probabilities to predict the ordering probabilities. Harville (1973) predict the ordering probabilities. Harville (1973) proposed a simple and convenient model that bettors can easily use in practice. In fact, the betting system proposed by Hausch, Ziemba & Rubinstein (1981) used the Harville model in determining the optimal bet amounts to place and show. The Harville model is equivalent to assuming that the running times are exponentially distributed. Henery (1981) and Stern (1990) assumed normal and gamma distributions respectively for the running times. Based on a likelihood approach, this paper considers the comparison among these models and the particular bet fractions. One conclusion is that in exacta and trifecta bets, no method based on the win bet fractions can outperform the exacta and trifecta bet fractions in predicting the relevant ordering probabilities.

KEY WORDS : Logit models; Cox's test; Horse races; Ordering probability; Running time distribution.

I. Introduction

It is always the dream of any bettor to be able to predict the chances of his bets in pari-mutuel betting of horse-racing. Previous empirical studies showed that win bet fraction is quite consistent with the true winning probability although a favourite-longshot bias usually exists (e.g. Griffth (1949); McGlothlin (1956); Hoerl & Fallin (1974); Ali (1977); Synder(1978); Fabricand (1979); Hausch, Ziemba & Rubinstein (1981); Asch,Malkiel & Quandt (1982)). Absence of this bias was also reported for racetrack in Hong Kong (see Busche & Hall (1988)). A simple model was suggested by Bacon-Shone, Lo and Busche (1991) to fit the true winning probability on the bet fraction and measure the favourite-longshot bias. Nevertheless, we may assume that win bet fraction is a good estimate of winning probability.

Another small group of researchers studied the estimation of probability of some finishing order, e.g. P(horse i finishes 1st and j finishes 2nd) based on the knowledge of simple winning probability, i.e. P(horse i wins) for all i. These models were suggested by Harville (1973), Henery (1981) and Stern (1988,1990) respectively. One reasonable estimate of winning probability is the win bet fraction. McCulloch & Zijl (1985) gave a test for the Harville model using Australian races. However, they used the show bet fraction which was assumed to be a good estimate of show probability (i.e. P(horse i finishes 1st, 2nd or 3rd)) instead of using the observed finishing order in their analysis (the computation rule of returns for show bet in Australia is different from the other countries). Apart from McCulloch & Zijl, no detailed empirical studies can be found for the purposes of analysis and comparisons of these models. In fact, Hausch, Ziemba & Rubinstein (1981) and Hausch & Ziemba (1985) used win bet fraction together with the simple model proposed by Harville (1973) to estimate more ordering probabilities and then used those probabilities for optimization of bet amounts.

In this paper, we will report the analysis of some more complicated horse-racing bets - exacta, trifecta and quinella. We use logit models to compare the models proposed by Harville (1973), Henery (1981) and Stern (1988,1990). Moreover, data based on different bet types can also be used to estimate these ordering probabilities. Thus, comparisons are made among both models and bet types. Part II will describe the previously proposed models for estimations of complicated probabilities. The logit model for comparison purposes will be given in part III followed by empirical analyses in part IV, V and VI. Our available data sets include Hong Kong and Meadowlands (U.S.). The main findings will be summarised in part VII and a conclusion is given in part VIII.

II. Description of some proposed models

Harville model

The simplest model to estimate the ordering probabilities is the one proposed by Harville (1973). For instance, to predict P(horse i wins and horse j finishes 2nd), we may simply use :

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

if π_i and π_j are known (π_i's can be estimated by bet fractions). A similar idea was also mentioned in Plackett (1975). Moreover, it is the ranking model proposed by Luce & Suppes (1965) in the study of choice behaviour. In fact, the conditional probability that horse j finishes second given i wins, $\pi_{j|i}$ may not equal $\pi_j/(1-\pi_i)$ in general. One common argument is mentioned in Hausch, Ziemba and Rubinstein (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 shown that if the running times follow exponential distribution independently with different mean running times, (1) will be obtained. That is,

running time of horse i, $T_i \overset{\text{ind}}{\sim} \exp(1/\theta_i)$ independently,

or $f(t_i|\theta_i) = \frac{1}{\theta_i}\exp(-t_i/\theta_i)$, $t_i>0$.

<u>Henery model</u>

Henery (1981) argued that a perhaps unrealistic feature of the Harville model was that (1) did not depend on the number of horses in the race. He suggested to assume 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 (i.e. $Var(T_i)$ = constant $\forall$ i). Under the Henery model,

$$P\,[T_1 < T_2 < \ldots < T_n] = \int_{-\infty}^{\infty}\phi(t_1-\theta_1)\ \ldots \int_{t_{n-1}}^{\infty}\phi(t_n-\theta_n)\ dt_n \ldots dt_1 \qquad (2)$$

where $\phi(.)$ is the pdf of standard normal distribution,

However, computing (2) is difficult and even computing π_{ij} is not easy because, unlike the Harville model, no closed form solution can be found. Henery suggested to use a Taylor series expansion about $\underset{\sim}{\theta} = \underset{\sim}{0}$:

$$P\,[T_1 < T_2 < \ldots < T_n] \simeq P\,[T_1 < T_2 < \ldots < T_n]\Big|_{\underset{\sim}{\theta} = \underset{\sim}{0}} +$$

$$\sum_{i=1}^{n}\theta_i \left\{ \frac{\partial\, P\,[T_1 < T_2 < \ldots < T_n]}{\partial\, \theta_i} \Bigg|\ \underset{\sim}{\theta} = \underset{\sim}{0} \right\}$$

$$= \frac{1}{n!} + \frac{\Sigma\, \theta_i\, \mu_{i;n}}{n!} \qquad (3)$$

where $\mu_{i;n}$ is the expected value of the ith standard normal order statistic in a sample of size n.

Alternatively,

$$P\,[T_1 < T_2 < \ldots < T_n] = \Phi\,\{\Phi^{-1}\,P\,[T_1 < T_2 < \ldots < T_n]\,\}$$

$$\simeq \Phi\,[\,\xi + \frac{1}{n!\phi(\xi)} \sum_i \theta_i\,\mu_{i:n}] \qquad (4)$$

where $\xi = \Phi^{-1}(1/n!)$,
by Taylor's expansion about $\underset{\sim}{\theta}=\underset{\sim}{0}$ for the term inside the large bracket. Using similar methods,

$$P\,[T_1 \text{ is smallest}] \simeq 1/n + \theta_1\mu_{1;n}/(n-1) \qquad (5)$$

$$\text{or } \Phi\,[\,z_0 + \frac{\theta_1\,\mu_{1:n}}{(n-1)\phi(z_0)}\,] \qquad (6)$$

where $z_0 = \Phi^{-1}(1/n)$.
Hence, by using (5) or (6), we can have estimates of $\underset{\sim}{\theta}$ if $\underset{\sim}{\pi}$ is known or the win bet fractions are good estimates of $\underset{\sim}{\pi}$. Then, we may substitute the estimated values of $\underset{\sim}{\theta}$ to appropriate equations to obtain estimates of ordering probabilities. From our experience, using methods similar to (3) and (5) produce a large number of negative probabilities. Therefore, we concentrate on the idea of (4) and (6) in our analyses. For example,

$$\pi_{ij} = P\,(T_i < T_j < \text{others})$$

$$\simeq \Phi\,\{\,a + \gamma\,[\,\theta_i\mu_{1;n} + \theta_j\mu_{2;n} + \frac{(\theta_i + \theta_j)(\mu_{1;n} + \mu_{2;n})}{n-2}\,]\} \qquad (7)$$

where $a = \Phi^{-1}[\frac{1}{n(n-1)}]$ and $\gamma = \frac{1}{n(n-1)\phi(a)}$ here.

In practice, to satisfy the unit-sum constraint, simple adjustment is usually necessary. Though the title of Henery (1981) involves the word horse races, he did not analyse any horse-racing data by his method.

Stern model

To extend the Harville model, a natural choice is the Gamma running times model proposed by Stern (1990). In his paper, the probability of a permutation was set equal to the probability that k independent gamma random variables with common shape parameter and different scale parameters are ranked according to the permutation. This distribution was motivated by considering a competition in which k players, scoring points according to independent Possion processes, were ranked according to the time until r points were scored. Gamma models can be used to estimate the ordering probabilities in horse-racing when only the probability of winning was given for each horse, i.e. $T_i \sim \text{Gamma}(r,\lambda_i)$ independently or,

$$f(t_i|\lambda_i) = \frac{1}{\Gamma(r)} \lambda_i^r t_i^{r-1} \exp(-\lambda_i t_i) \qquad t > 0.$$

where r is predetermined and λ_i can be estimated from π_i. Obviously, when r = 1, the Stern model becomes the Harville model, and when $r \longrightarrow \infty$, it becomes the Henery model. However, even when r = 2, the formula for computing π_{ij}, etc. is very complicated and no closed form solution can be found.

As a very small empirical study, Stern analysed 47 races and he found that the use of r=1 (which is the Harville model) for estimating ordering probabilities is less accurate than that of r=2.

III. Logit model for more complicated bets

Bacon-Shone, Lo & Busche (1991) suggested to fit the constant-β model in order to analyse the win bet data, i.e.

$$\pi_i = \frac{P_i^{\beta}}{\sum_r P_r^{\beta}}$$

where π_i = P(horse i wins),

P_i = Win bet fraction of horse i,

i.e. the proportion of win bet on horse i, and

β is a parameter estimated by maximum likelihood assuming that the win event follows multinomial distribution.

The above model can be rewritten as follows :

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

which means the multivariate logit of the winning probability depends on the logit of the bet fractions in a very simple way. Using a similar model structure for conditional probabilities, we have :

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

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

$$= \frac{\text{P (horses i \& j finish first \& second resp.)}}{\text{P (horse i wins)}}$$

$$= \pi_{ij} / \pi_i$$

Hence, we have :

$$\pi_{ij} = \pi_i \ \pi_{j|i}$$

$$= \frac{P_i^{\beta}}{\sum_r P_r^{\beta}} \ \frac{P_{j|i}^{\mu}}{\sum_{s \neq i} P_{s|i}^{\mu}} \tag{8}$$

where P_i = win bet fraction

$P_{j|i} = P_j / (1-P_i)$

β and μ are parameters to be estimated.

The parameter β is used to test the favourite-longshot bias while μ is to test whether the previous models (discussed in the previous part) have systematic bias or not.

$$\text{Let} \quad Y_{ij} = \begin{cases} 1 & \text{if horse i wins and horse j finishes 2nd} \\ 0 & \text{otherwise} \end{cases}$$

Assuming $\underset{\sim}{Y}^{(2)} = (Y_{12}, \ldots, Y_{n,n-1})^T \sim$ Multinomial ($\underset{\sim}{\pi}^{(2)}$)

where $\underset{\sim}{\pi}^{(2)} = (\pi_{12}, \ldots, \pi_{n,n-1})^T$, we have the following log likelihood:

$$l = \sum_{l=1}^{m} \sum_{i} \sum_{j} y_{ij} \ln \pi_{ij,l}$$

$$= \sum_{l=1}^{m} \ln \pi_{[12]l}$$

where $\pi_{[12]l}$ denotes the probability that the winning horse wins and the horse finishing second finishes second in race l.

Hence, Maximum Likelihood Estimators for the parameters can be easily obtained. Note that $\hat{\beta}$ and $\hat{\mu}$ will be asymptotically uncorrelated because $\dfrac{\partial^2 l}{\partial\hat{\beta}\ \partial\hat{\mu}} = 0$.

The above model based on win bet fraction is hereby called Bacon-Shone, Lo & Busche (BLB) model. In the following, we mainly use the idea of this logit model to analyse our data with different bet types and different models.

IV. Empirical analysis for exacta bet

In this part of study, we choose the exacta bet data in Meadowlands. Exacta bet means the type of bet that the bettors are required to guess which horses will finish first and second, respectively, in exact order. To estimate the exacta probability (i.e. π_{ij}), the exacta bet fraction itself is thought to be a good choice. Another choice is to use the win bet data but this involves estimation of π_{ij} from estimates of π_i. The Harville, Henery and Stern models are three alternatives for doing this task. One more alternative is our BLB model which should improve over the Harville model.

To facilitate the direct comparison between the exacta bet fractions and the estimated probabilities estimated by using different models, the marginal exacta bet fraction is also used. Define ${}^{E}P_{ij}$ = exacta bet fraction for horse i and j, the marginal exacta bet fraction ${}^{E}P_{i}$ is defined to be $\sum_{j\neq i} {}^{E}P_{ij}$.

Note that unlike the simple win bet, even on-track bettors cannot see the instant changes of odds of the exacta bet.

To check how good the estimation of π_{ij} using the Henery model is, the following model is fitted :

$$\pi_{ij} = \frac{\hat{\pi}_i^{\beta}}{\sum_r \hat{\pi}_r^{\beta}} \quad \frac{\hat{\pi}_{j|i}^{\mu}}{\sum_{s\neq i} \hat{\pi}_{s|i}^{\mu}} \tag{9}$$

where $\hat{\pi}_i = \sum_{j\neq i} \hat{\pi}_{ij}$, $\hat{\pi}_{ij}$ is estimated by the Henery model and $\hat{\pi}_{j|i} = \hat{\pi}_{ij} / \hat{\pi}_i$.

To estimate π_{ij} by the Henery model, we first estimate $\underline{\theta}$ by (6). Then two strategies are employed :

(i) Use (7), i.e. the approximation formula proposed by Henery to estimate π_{ij} ;

(ii) Numerical integration, in particular Gauss-Hermitian Quadrature, is used.

For the Stern model, numerical integration has been employed. The empirical results are shown in Table 1.

From Table 1, exacta bet fraction is the best for estimating the ordering probailities. If we concentrate on the model based on win bet fractions, our BLB model appears to be the best but it is very close to the Henery model. And the Henery model is slightly better than the Stern (r = 2) model but close to the Stern (r = 5) model. Moreover, the Harville model appears to be the worst one among all.

Cox's tests (Cox (1961,1962)) have been applied to test the above arguments more formally. The Cox's test involves two steps. E.g. To test the Henery vesus the Harville model, the Henery model is associated with the null hypothesis and the Harville model is associated with the alternative hypothesis. After computing the test statistic, the procedure is reversed, i.e. the Harville model is treated as the model under the null hypothesis and another test statistic is computed again. In general, to test two models under H_f and H_g, the test statistics are :

Table 1

Comparisons among models for estimating exacta probabilities (510 races)

Bet type / method	$l(1,1)$	$\hat{\beta}$	$\hat{\mu}$	$l(\hat{\beta},\hat{\mu})$
Exacta bet	-1844.46	1.1457	0.9090	-1840.27
Win bet :				
Harville	-1875.77	1.1255	0.6910	-1855.96
Henery :				
approx	-1872.81	1.1591	0.9224	-1868.41
almost exact	-1859.63	1.0828	0.9071	-1857.50
Stern :				
r = 2	-1870.86	1.0819	0.7368	-1858.29
r = 5	-1864.86	1.0804	0.7917	-1857.50
Marginal exacta :				
Harville	-1863.10	1.1457	0.7362	-1847.80
Henery :				
almost exact	-1852.09	1.0961	0.9620	-1850.39
Stern :				
r = 2	-1860.45	1.0963	0.7805	-1851.28
r = 5	-1855.36	1.0949	0.8402	-1850.33

(N.B. : l stands for log likelihood; the Harville model with the estimated parameters is our BLB model.)

Under H_f,

$$T_f = \frac{(\text{loglik}_f - \text{loglik}_g) - E(\text{loglik}_f - \text{loglik}_g \mid H_f)}{\sqrt{V(\text{loglik}_f - \text{loglik}_g \mid H_f)}}$$

$\sim N(0,1)$

and under H_g,

$$T_g = \frac{(\text{loglik}_g - \text{loglik}_f) - E(\text{loglik}_g - \text{loglik}_f \mid H_g)}{\sqrt{V(\text{loglik}_g - \text{loglik}_f \mid H_g)}}$$

$$\sim N(0,1)$$

where $E(.\mid H_k)$ and $V(.\mid H_k)$ are expectations and variances under H_k, loglik_k=log likelihood value for the model under H_k

(k=f,g),

$$= \sum_l \sum_i \sum_j Y_{ij,l} \ln {}^{k}\hat{\pi}_{ij,l}$$

${}^{k}\hat{\pi}_{ij,l}$ = estimated probability of horses i & j finishing first & second, respectively, under H_k

$$E(\text{loglik}_f - \text{loglik}_g \mid H_f) = \sum_l \sum_i \sum_j {}^{f}\hat{\pi}_{ij,l} \left(\ln \frac{{}^{f}\hat{\pi}_{ij,l}}{{}^{g}\hat{\pi}_{ij,l}} \right)$$

$$V(\text{loglik}_f - \text{loglik}_g \mid H_f)$$

$$= \sum_l \left\{ \sum_i \sum_j \left(\ln \frac{{}^{f}\hat{\pi}_{ij,l}}{{}^{g}\hat{\pi}_{ij,l}} \right)^2 \text{Var}(Y_{ij,l} \mid H_f) + \right.$$

$$\left. \sum_{\substack{i\ j\ r\ s \\ \text{not}(i=r\ \&\ j=s)}} \sum\sum\sum \left(\ln \frac{{}^{f}\hat{\pi}_{ij,l}}{{}^{g}\hat{\pi}_{ij,l}} \right) \left(\ln \frac{{}^{f}\hat{\pi}_{rs,l}}{{}^{g}\hat{\pi}_{rs,l}} \right) \text{Cov}(Y_{ij,l}, Y_{rs,l} \mid H_f) \right\}$$

$$V(Y_{ij,l} \mid H_f) = {}^{f}\hat{\pi}_{ij,l}(1 - {}^{f}\hat{\pi}_{ij,l})$$

$$\text{Cov}(Y_{ij,l}, Y_{rs,l} \mid H_f) = -{}^{f}\hat{\pi}_{ij,l}\, {}^{f}\hat{\pi}_{rs,l}$$

Similarly for $E(\text{loglik}_g - \text{loglik}_f \mid H_g)$ and $V(\text{loglik}_g - \text{loglik}_f \mid H_g)$.

The results of Cox's tests are shown in Table 2. In this table, T_i is the test statistic value under H_i. The following preference order can be concluded at 5% significance level :

Exacta bet fraction > Henery model > Stern model (r=5)

> Stern model (r=2) > Harville model

V. Empirical analysis for Trifecta bet

Another type of bet in Meadowlands - Trifecta bet means that the bettors have to guess the three winning horses in correct order and thus, it is more complicated than the exacta bet.

Table 2

Cox's tests for exacta comparison

H_f	H_g	T_f	T_g
Exacta bet fractions	Henery model	-0.4048	-5.5133
Exacta bet fractions	Harville model	0.4210	-8.2143
Exacta bet fractions	Stern model (r=2)	0.2331	-7.5051
Exacta bet fractions	Stern model (r=5)	-0.0613	-6.5238
Henery model	Harville model	1.0143	-6.0441
Henery model	Stern model (r=2)	1.6316	-5.1844
Henery model	Stern model (r=5)	1.3219	-3.5934
Stern model (r=2)	Harville model	1.8387	-3.7273
Stern model (r=5)	Harville model	2.7069	-5.6157
Stern model (r=5)	Stern model (r=2)	4.2778	-5.5838

Our model for trifecta bet will be similar to the exacta, i.e. model (8). Define :

π_{ijk} = P(horse i wins, j finishes 2nd & k finishes 3rd), and

$\pi_{k|ij}$ = P(horse k finishes 3rd | horses i & j finish 1st & 2nd)

We have the following logit model :

$$\pi_{ijk} = \pi_i \; \pi_{j|i} \; \pi_{k|ij}$$

$$= \frac{P_i^{\beta}}{\sum_r P_r^{\beta}} \quad \frac{P_{j|i}^{\mu}}{\sum_{s\neq i} P_{s|i}^{\mu}} \quad \frac{P_{t|ij}^{\omega}}{\sum_{t\neq ij} P_{t|ij}^{\omega}} \qquad (9)$$

where P_i = win bet fraction

$P_{j|i} = P_j / (1-P_i)$, $\quad P_{k|ij} = P_k / (1-P_i-P_j)$

β, μ and ω are parameters to be estimated.

However, we only have 120 races for Trifecta bet in Meadowlands and hence, the result here is not as reliable as that for the exacta bet in part 4. Some results are shown in Table 3.

In addition, we have obtained trifecta data for the winning combinations in Hong Kong.Although we only have the trifecta bet fractions for the winning combinations, we can still compute the log likelihood for the trifecta bet fractions without estimating any parameter. Some results for 1986-1989 Hong Kong data (1809 races) are shown in Table 4.

Table 3

Estimations of Trifecta probabilities in Meadowlands (120 races)

Bet type / method	$l(1,1,1)$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\omega}$	$l(\hat{\beta},\hat{\mu},\hat{\omega})$
Trifecta bet fraction	-695.17	1.0849	1.1681	0.6875	-692.14
Win bet :					
Harville	-711.50	1.0268	0.8517	0.4353	-697.78
Henery:approx	-703.12	1.0644	1.1563	0.7022	-701.07
almost exact	-699.83	——	——	——	——

(For almost exact computation of the Henery model, only the estimated probabilities for the winning combinations are computed due to the large amount of time required in estimating all π_{ijk} and thus, no parameter has been estimated; the Harville model with estimated parameters is our BLB model.)

Table 4

Estimations of Trifecta probabilities in Hong Kong (1809 races)

Bet type / method	$l(1,1,1)$	$\hat{\beta}$	$\hat{\mu}$	$\hat{\omega}$	$l(\hat{\beta},\hat{\mu},\hat{\omega})$
Trifecta bet fraction	-10656.75	—	—	—	—
Win bet :					
Harville	-10747.98	0.9651	0.7634	0.6466	-10666.93
Henery:approx	-10689.61	—	—	—	—
almost exact	-10667.25	—	—	—	—

(For the Henery model, no parameter has been estimated due to the large amount of memory required to store all the estimated probabilities for 1809 races; the Harville model with estimated parameters is our BLB model.)

Table 5

Cox's tests for trifecta et comparisons

H_f	H_g	T_f	T_g
Trifecta bet fractions	Henery model	- .7303	-4.0754
Trifecta bet fractions	Harville model	- .2562	-6.0896
Henery model	Harville model	- .3253	-4.3355

Cox's tests (Cox (1961,1962)) have also been applied to test the goodness of different models and data in Table 3 more formally. The results are shown in Table 5. The following preference order can be concluded at 5% significance level :

Trifecta bet fraction > Henery model > Harville model

Formal test for Table 4 is not possible due to lack of complete data for each race. However, it would also be of interest to directly compare the log likelihood values (Cox (1962)).

VI. Empirical analysis for the Quinella bet

For the Quinella bet, the bettors have to guess which two horses finish first and second, regardless of their order and thus, it is not the same as the exacta bet. This type of bet is not available in Meadowlands but it exists in Hong Kong. Similar to the win, exacta and trifecta bets, the quinella bet fraction should be a choice for estimating the quinella probability. We define some new notation for it :

$$Z_{ij} = \begin{cases} 1 & \text{if horse } i \text{ wins and } j \text{ finishes 2nd, or} \\ & \text{if horse } j \text{ wins and } i \text{ finishes 2nd} \\ 0 & \text{otherwise} \end{cases}$$

for $i>j$.

and Q_{ij} = Quinella bet fraction for horses i and j.

Also,

$$^{q}\pi_{ij} = P (Z_{ij} = 1) = \pi_{ij} + \pi_{ji} \qquad (i>j)$$

Similar to the constant-β model for the win bet, the logit model for the quinella bet is :

$$^{q}\pi_{ij} = \frac{Q_{ij}^{\eta}}{\sum_r \sum_s Q_{rs}^{\eta}} \qquad \text{for } i>j \qquad (10)$$

Assuming $\underset{\sim}{Z} \sim$ Multinomial $(^{q}\underset{\sim}{\pi})$. The maximum likelihood estimator of η can be obtained by maximizing the following log likelihood with respect to η :

$$l = \sum_{i>j} z_{ij} \ln {}^{q}\pi_{ij}$$

To compare the quinella bet fraction with the win bet fraction using the Harville and Henery models, we can simply replace the Q_{ij}

in (10) by the quinella probabilities estimated by the Harville and Henery models. 369 races in H.K. provided by the Hong Kong Jockey Club are used for model fitting and the results are presented in Table 6. The empirical result again suggests that the Henery model is better than the Harville model.

Table 6

Estimation of Quinella probabilities
(369 races)

Bet type / Model	$l(1)$	$\hat{\eta}$	$l(\hat{\eta})$
Quinella bet fraction	-1222.43	0.9702	-1222.33
Win bet fraction :			
Harville	-1225.11	0.8363	-1221.05
Henery:			
approx	-1222.55	0.9732	-1222.48
almost exact	-1221.39	0.9559	-1220.52

Table 7

Cox's tests for quinella bet comparisons

H_f	H_g	T_f	T_g
Quinella bet fractions	Henery model	-2.4595	-3.4878
Quinella bet fractions	Harville model	-2.8045	-2.4231
Henery model	Harville model	-2.8475	0.5008

Again, Cox's tests have been applied to test the above arguments. The results are shown in Table 7. However, this case is not so lucky. By looking at the two test statistic values for each case of the first two tests, we do not have any conclusive results since the two steps of the Cox's tests reject the null hypotheses at 5% level and thus contradictory results follow. For the final test (third row in the table), we still conclude that the Henery model is better than the Harville model at 5% level.

VII. Main findings

From all the above analyses, we have the following findings on comparing log likelihoods based on our data sets :

(i) For the exacta bet in Meadowlands, no systematic method can be found that can beat the exact bet fraction. Similarly for the trifecta bet in Meadowlands.

(ii) Again for the exacta bet, the marginal exacta bet fractions appear to be better than the win bet fractions for estimating exacta probabilities using any model.

(iii) The Henery model is better than the Harville model.

(iv) Logistic models show that the Harville model overestimates P(horse j finishes 2nd | i wins) if j is a favourite horse and P(horse k finishes 3rd | i wins and j finishes 2nd) if k is a favourite horse. This is known as the Silky Sullivan phenomenon.

(v) The Harville model plus estimated parameters and the Henery model have similar precision.

VIII. Conclusion

The common Harville model is empirically shown to be worse than the Henery and Stern model. It may be because the assumption of exponential distribution of running times is not valid. A logit model may be used to improve the Harville model without considering more complicated model. Another result is that the bet fractions of the particular bet type appear to be good estimates for the probabilities of getting return for the same bet type. We have empirically shown that no method based on win bet fractions can strongly beat these.

REFERENCES

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

Asch,Peter;Malkiel,B. and Quandt,R. (1982) "Racetrack betting and informed behavior." *Journal of Financial Economics* 10 (June) : 187-94.

Asch,Peter and Quandt,R. (1987) "Efficiency and profitability in exotic bets." *Economica* 54 : 289-98.

Bacon-Shone,J.H.; Lo,V.S.Y. and Busche,K. (1992) "Modelling winning probability." Technical Report, 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. (1961) "Tests of separate families of hypotheses." *Proc. 4th Berkeley Symp.*, 1 : 105-123.

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

Fabricand,B.P. (1979) The science of winning : A random walk on the road to riches. Van Nostrand Reinhold.

Griffth,R.M. (1949) "Odds adjustments by American horse-racing bettors." *American Journal of Psychology* 62 : 290-294.

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.

Hausch,D.B. and Ziemba,W.T. (1985) "Transactions costs, extent of inefficiencies, entries and multiple wagers in a racetrack betting model." *Management Science* 31 (April) : 381-394.

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

Hoerl,A.E. and Fallin,H.K. (1974) "Reliability of subjective evaluations in a high incentive situation." *Journal of Royal Statistical Society, Series A* 137 : 227-230.

Li, W.K. (1991) "Discriminating nonlinear time series models using a bootstrapped Cox test." Technical report, Department of Statistics, University of Hong Kong.

Luce, R. Duncan and Suppes, P. (1965) "Preference, utility and subjective probability" in Luce,R.D., Bush,R.R. and Galanter,E., ed., Handbook of Mathematical Psychology Vol. III, Chapter 19 : 249-410.

McCulloch, B. and Zijl, T.V. (1985) "Direct test of Harville's mult-entry competitions model on race-track betting data." *Journal of Applied Statistics* 13, No.2 : 213-220.

McGlothlin,W.H. (1956) "Stability of choices among uncertain alternatives." *American Journal of Psychology* 69 : 604-619.

Plackett,R.L. (1975) "The analysis of permutations." *Applied Statistics*, 24 : 193-202.

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

Stern, Hal (1988) "Gamma Processes, paired comparisons and ranking." Computer Science and Statistics : Proceedings of the 20th symposium on the Interface : 635-639.

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