RESEARCH REPORT

Serial No. 18

June 1992

A MODIFIED RACETRACK BETTING SYSTEM

by

Victor S.Y. Lo, John Bacon-Shone and Kelly 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

A MODIFIED RACETRACK BETTING SYSTEM

Victor S.Y. Lo, John Bacon-Shone and Kelly Busche

Department of Statistics, University of Hong Kong,
Pokfulam Road, Hong Kong

Department of Statistics and Social Sciences Research
Centre, University of Hong Kong
Pokfulam Road, Hong Kong

School of Economics, University of Hong Kong,
Pokfulam Road, Hong Kong

Abstract

Hausch, Ziemba & Rubinstein (1981) developed a betting system (the Dr.Z system) that empirically demonstrated positive profits in two racetracks. The Dr.Z system assumes running times are distributed exponentially but the use of other distributions for running times (Henery (1981) and Stern (1990)) produces a better fit in some racetracks although the better fit is at the cost of severely increased complexity in computing ordering probabilities. Lo & Bacon-Shone (1992) proposed a simple model of computing ordering probabilities which is a good approximation to those based on the Henery model and the Stern model. In this paper, we add this model to the Dr.Z system and use a simple way of computing ordering probabilities which is a good approximation to those based on the Henery model. With data sets in the U.S. and Hong Kong, we show improved profit over the Dr.Z system at lower levels of risk. With the special return formula in Japan, our model does not have a big difference in profits from the Dr.Z system.

Keywords : RACETRACK BETTING; RUNNING TIME DISTRIBUTIONS; BETTING SYSTEM

Acknowledgements

We thank Professor Richard Quandt for his U.S. data and Professor Junji Shiba for his help in acquiring the Japanese data.

I. Introduction

A mathematical model was developed by Hausch, Ziemba and Rubinstein (1981) (the Dr.Z system) to search for possible profits in racetrack place and show betting pools (see also Hausch & Ziemba (1985)). Assuming win bet fractions are accurate estimates of winning probabilities, the Dr.Z system uses the Harville model (Harville(1977)) to predict the ordering probabilities for both selection of appropriate horses and optimisation of bet amounts in multi-horse betting pools. For example, to compute π_{ij} = P(horse i wins and horse j finishes second),

$$\pi_{ij} = \frac{\pi_i \pi_j}{1 - \pi_i} \tag{1}$$

where π_i = P(horse i wins),
and the value of π_i is estimated by the win bet fraction (see Ali (1977), Snyder (1978), Busche & Hall (1988) and Bacon-Shone, Lo & Busche (1992a) for details of using the win bet fractions). More complicated ordering probabilities are calculated in an analogous way.

Relationship (1) is implied by assuming running times are distributed as independent exponential with unique parameters for each horse in each race (Dansie (1983)). An alternative to the Harville model was proposed by Henery (1981), who assumes independent normal distributions for the running times. Then $\pi_{ij} = P(T_i < T_j < \min_{r \neq i,j}(T_r))$, where T_i = running time of horse i which is distributed as $N(\theta_i,1)$. Here, θ_i's are obtained by solving the equations :

$$\pi_i = P(T_i < \min_{r \neq i}(T_r)) \quad \text{for } i = 1,\ldots,n \text{ (no. of horses)}$$

where π_i's are assumed to be known or estimated by the bet fractions. However, this requires numerical integration or an approximation method.

In addition, Stern (1990) proposed to assume independent gamma distributions with a fixed shape parameter r for the running times. However, his model also requires using numerical methods to evaluate the ordering probabilities. His model 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. We can consider it as an alternative model to the Harville and Henery models. Let the running time of horse i, $T_i \sim$ Gamma(r,θ_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 θ_i can be estimated from π_i (or the bet fraction, P_i). Note that when r=1, it reduces to Harville's exponential model; when $r=\infty$, it is the same as Henery's normal model.

In another paper (1992b), the authors compared the three models in several betting pools. Using the information from win bet fractions alone, the Henery model was found to be better (by likelihood measure and Cox's (1962) test) than the Harville model and Stern model with finite positive integer shape parameters in predicting the relevant ordering probabilities for Hong Kong and Meadowlands data. In section II, we consider the simple model for predicting the ordering probabilities (proposed by Lo & Bacon-Shone (1992)). We apply it to the Dr.Z system in section III. Conclusion is given in section IV.

II. A simple approximation of the Henery model

Lo & Bacon-Shone (1992) proposed a class of simple models (hereafter called the discount model) which can approximate the Henery and Stern model quite well. Let π_i = P(horse i wins), π_{ijk}= P(horse i wins , horse j finishes second and horse k finishes third), the discount model gives a relation between the ordering probability π_{ijk} and the simple winning probability π_i as follows :

$$\hat{\pi}_{ij} = \pi_i \frac{\pi_j^{\lambda^r}}{\sum_s \pi_s^{\lambda^r}} \frac{\pi_k^{\tau^r}}{\sum_t \pi_t^{\tau^r}} \qquad (2)$$

where $\wedge$ indicates an estimate, π_i is estimated by the win bet fraction of horse i, and λ^r & τ^r are values such that the resulting estimated ordering probabilities are close to the ordering probabilities associated with the Stern model with shape parameter r over a range of race sizes (see Lo & Bacon-Shone (1992) for details). Hence, our model can be considered as an approximation to the Stern class model with shape parameter r. Note that when r=1, the Stern

Table 1

Parameter values for the discount model

r	λ^r	τ^r
1	1.0000	1.0000
2	0.9336	0.8920
3	0.9021	0.8423
4	0.8836	0.8140
5	0.8712	0.7953
6	0.8623	0.7819
7	0.8555	0.7717
8	0.8500	0.7636
∞	0.7600	0.6200

model reduces to the Harville model and thus $\lambda^1 = \tau^1 = 1$ are chosen to match the Harville formula. When $r = \infty$, the Stern model reduces to the Henery model. The parameter values for some r are given in Table 1.

Among the Harville, Henery and Stern models, Lo & Bacon-Shone (1992) concluded that the Henery model is the best in both Meadowlands and Hong Kong data. However, in Japan, the Stern model with shape parameter $r = 4$ fits the best.

Without the discount model, we need to rely on numerical integrations, and then the procedure becomes too complicated to use in practice as it requires solving a large system of nonlinear equations involving integrals. The discount model is much simpler to use. Now, we apply the discount model in our racetrack data.

III. Applying the discount model to the Dr.Z system

Hausch, Ziemba and Rubinstein (1981) (HZR) developed the Dr.Z system to exploit differences in predictions implied by different betting pools. Assuming the win pool produces accurate estimates, the system searches for underbet horses in place and show pools. Using data from Santa Anita (Los Angeles) and Exhibition Park (Vancouver), pure profit was reported. Additional evidence of profits can be found in Ziemba & Hausch (1987). The details of the Dr.Z system are described below.

1. Selection of horses :

Using the simple Harville model (equation (1) above), the Dr.Z system replaces the winning probabilities by the bet fractions and then estimates the more complicated probabilities. Where offered odds on the complicated bets exceed those implied by the win odds by an arbitrary "fudge factor", the system indicates a bet and, the bet amount (described below) is calculated. The "fudge factor" used by the Dr.Z system restricts bets to those where the estimated expected return per dollar is above 1.16 or above 1.20.

Assuming that the win pool is nearly efficient, this selection procedure produces profit from any market inefficiencies that exist in place and show pools.

2. Optimization of bet amounts:

After selecting those horses which ensure at least a minimum expected return, the Dr.Z system maximizes the "Kelly" criterion by choosing the bet amounts for the selected horses. The Kelly criterion maximizes the rate of growth of return. For references to the Kelly criterion and its comparison with Markowitz's mean-variance approach, see Kelly (1956); Breiman (1960); Thorp (1971); Roll (1973); Rosner (1975); Maier, Peterson & Weide (1977) and Grauer (1981).

Using the Kelly criterion, the Dr.Z system maximizes the expected logarithm of final wealth for each race subject to a constraint on bet size to avoid ruination (see their original paper for details). This maximization typically takes too much time to solve at a racetrack, so HZR suggest a simpler procedure using a set of linear approximations (See HZR (1981) for details) which we have not used in this paper.

Expected wealth under the Dr.Z system is increasing over time, and is always better than random betting (i.e. we bet on every horse with proportion of money equal to the win bet fraction determined by the public) provided the underlying probability model is sufficiently accurate.

We apply the Dr.Z system to three racetracks : Meadowlands (U.S.), Hong Kong and Japan. The first step is to select the horses to be bet on. The expected return from $1 bet to show[1] on horse l is :

$$E\,(X^s_l) = \sum_{l\neq j\neq k}\sum\sum (\hat{\pi}_{ljk} + \hat{\pi}_{jlk} + \hat{\pi}_{jkl})\ Ret_{ljk}$$

where

$\hat{\pi}_{ijk}$ = estimated probability of horse i wins, horse j finishes second and horse k finishes third,

Ret_{ljk} = Show return on horse l per dollar if horses l, j and k finish in the first three positions.

$$= 1 + INT\left[\frac{(1-t)(Sh+1)-(1+Sh_l+Sh_j+Sh_k)}{3\,(1+Sh_l)} * 10\right]/10$$

in the U.S.

$$= 1 + round\left[\frac{(1-t)(Sh+1)-(1+Sh_l+Sh_j+Sh_k)}{3\,(1+Sh_l)} * 100\right]/100$$

in Hong Kong

$$= INT\left[0.838 + \frac{0.246\,(Sh - Sh_l - Sh_j - Sh_k)}{Sh_l}\right]/10$$

in Japan

Sh_j = total dollar amount bet to show on horse j,

$Sh = \sum_j Sh_j$ and t = track take.

The INT is a truncation function; round is a rounding function.

[1]We use U.S. terminology : a place bet wins if the horse comes in the first or second; a show bet wins if the horse comes in the first, second or third.

The differences are due to the differential use of Breakage; in Hong Kong, returns are "rounded" and in the U.S., returns are "broken" to the lower 10¢. The track take is 17% and 18% in Hong Kong and Meadowlands (U.S.) respectively. The Dr.Z system The return formula for Japan is very different from the other two and its implied track take is not a constant. Formulas for place bets are similar.

To investigate the Japanese return formula, we note that if horse i finishes in the first three positions, the return for show bet on horse i per yen is (ignoring the truncation) :

$$0.838 + \frac{0.246\ (Sh - Sh_{[1]} - Sh_{[2]} - Sh_{[3]})}{Sh_i}$$

$$= 1 + \frac{(0.738\ Sh + 0.262 \sum_{r=1}^{3} Sh_{[r]} - 0.486\ Sh_i) - \sum_{r=1}^{3} Sh_i}{3\ Sh_i}$$

$$= 1 + \frac{(1-t_i)\ Sh - \sum_{r=1}^{3} Sh_{[r]}}{3\ Sh_i}$$

where t_i = the implied track take, $Sh_{[r]}$ = show bet on the horse which finishes in the rth order, and Sh = sum of show bets for all horses. Note that the last expression is the traditional return formula if t_i is a constant for all i. Here, t_i is not a constant, and

$$1 - t_i = 0.738 + \frac{0.262 \sum_{\substack{r=1 \\ r\neq i}}^{3} Sh_{[r]} - 0.224\ Sh_i}{Sh}$$

i.e. the implied track take (and thus the return) of horse i depends on the show bet of that horse relative to the other winning horses. In fact, when $Sh_i / \sum_{r=1}^{3} Sh_{[r]} > 0.539$, the return is less than that under the traditional formula. Hence, the special Japanese formula gives a lower return to the most favourites and higher return to the longshots. Thus, according to the selection rule of the Dr.Z system, we pick up more longshots and less favourites and the bet becomes more risky unless we apply the probability restriction to screen out the longshots. The case for the return formula of place bet is similar.

To illustrate the effect of using Japanese return formula more clearly, we consider an extreme numerical example. Suppose the proportion of show bets on the first three horses, $\sum_{r=1}^{3} Sh_{[r]}/Sh = 0.6$ and also $Sh_{[1]}/Sh = 0.5$, $Sh_{[2]}/Sh = 0.01$. Then the implied track take for the more favourite horse [1] is 0.348 with return per dollar bet = 1.03 and the track take for the longshot [2] is 0.110 with return per dollar bet = 10.68. If track take is a constant (1−0.738 = 0.262), then the returns per dollar bet for [1] and [2] are 1.09 and 5.60 respectively. This indicates that the Japanese return formula increases the return for longshots considerably when compared to the traditional formula.

The Dr.Z system indicates a show (place) bet on horse i if $E(X_i^s)$ ($E(X_i^p)$) $\geq \alpha$. The "fudge factor" α is chosen "subjectively". In HZR (1981, p.1444), "Intuition suggests that Santa Anita with its larger betting pools would have more accurate estimates of the q_i (winning probabilities) than would be obtained at Exhibition Park. Hence positive profits would result from lower values of α.". Hence, they use α=1.20 for Exhibition Park and α=1.16 for Santa Anita. Pool sizes per race in Hong Kong are greater than those in Meadowlands so although we follow HZR and use 1.16 for Hong Kong, we also try α = 1.12 and 1.20. We vary the Dr.Z system in two ways :

(i) For the selection rule of the Dr.Z system, in addition to the constraint that expected return $\geq \alpha$, we can use an additional probability restriction that P(having a profit) $\geq \beta$ (also arbitrarily chosen) in order to screen low probability horses. The two restrictions we use are :

a. P(i finishes 1st or 2nd) $\geq$ 0.10 for place bet, and
   P(i finishes 1st, 2nd or 3rd) $\geq$ 0.20 for show bet

b. P(i finishes 1st or 2nd) $\geq$ 0.20 for place bet, and
   P(i finishes 1st, 2nd or 3rd) $\geq$ 0.30 for show bet

These sets of restrictions screen out the longshot horses so that where more variant longshots might produce inaccurate probability estimates, they will not be selected for our optimization. Note that b is more restrictive than a.

(ii) For the estimation of ordering probabilities π_{ijk} (horses i,j and k in order) using win bet fractions, the Dr.Z system uses the

Harville model. In the following, we use the discount model with appropriate r for the selection of horses and optimization of bet amounts.

The results of final wealth for Hong Kong (1985-90, 2291 races), Meadowlands (1984, 705 races) and Japan (1990-91, 1583 races) are shown in the following three tables. From the analysis in the previous section, we choose $(\lambda^{\infty},\tau^{\infty})$ = (0.76,0.62) in Hong Kong and Meadowlands and (λ^{4},τ^{4}) = (0.88,0.81) in Japan. Apart from the final wealth, we also show the total bet amounts and minimum capital during the wealth history in the tables. Intuitively, minimum capital is an indicator of risk associated with the betting method.

Table 2

The Dr.Z system and its variants in Hong Kong 1985-89

	Harville			Discount ($r = \infty$)		
	original	prob. restr. a	prob. restr. b	original	prob. restr. a	prob. restr. b
	US $	US $	US $	US $	US $	US $
$\alpha = 1.12$						
F.W.	5	5	5	55	179	806
T.B.	31172	31184	30947	41216	37110	36859
M.C.	2	14	14	38	941	2277
$\alpha = 1.16$						
F.W.	2	35	26	674	1222	3332
T.B.	34926	35063	33297	55010	38757	34446
M.C.	2	4	3	440	651	1672
$\alpha = 1.20$						
F.W.	55	53	62	944	1601	2569
T.B.	22270	22139	21210	28627	23281	17449
M.C.	14	13	16	69	879	1574

Table 3

The Dr.Z system and its variants in Meadowlands 1984

	Harville			Discount ($r = \infty$)		
	original	prob. restr. a	prob. restr. b	original	prob. restr. a	prob. restr. b
	US $	US $	US $	US $	US $	US $
$\alpha = 1.12$						
F.W.	2007	2076	1358	17388	18470	16888
T.B.	215251	215887	210843	62762	61862	59479
M.C.	1268	904	1321	7182	7170	7537
$\alpha = 1.16$						
F.W.	4296	4448	3378	15631	16449	14539
T.B.	161525	161733	158735	36466	35555	33763
M.C.	3895	4051	3008	8371	8464	8388
$\alpha = 1.20$						
F.W.	6397	6569	9921	14273	15029	13871
T.B.	128097	101888	101348	21153	21153	19980
M.C.	3471	3519	4829	8218	8218	8330

Table 4

The Dr.Z system and its variants in Japan 1990-91

	Harville			Discount ($r = 4$)		
	original	prob. restr. a	prob. restr. b	original	prob. restr. a	prob. restr. b
	US $	US $	US $	US $	US $	US $
$\alpha = 1.12$						
F.W.	7370	9611	9745	6199	10234	10306
T.B.	18648	16832	15687	10471	6164	5487
M.C.	6392	8389	8724	6018	9033	9033
$\alpha = 1.16$						
F.W.	7927	9985	10356	6078	9569	9936
T.B.	11960	10300	9455	6394	2812	2476
M.C.	7190	8847	8847	6078	9041	9230
$\alpha = 1.20$						
F.W.	7772	9438	9804	5755	8723	9084
T.B.	5127	3615	3283	4777	1847	1491
M.C.	7199	8827	9004	5587	8381	8739

Note :

(1) In Table 2, HK$ have been transferred at US$ 1= HK$ 7.8.

Similarly, in Table— 4, Japanese yens have been transferred at US$ 1= ¥ 137.8 .

(ii) The initial wealth for all three data sets : US $ 10,000.

(iii) F.W. = Final Wealth; T.B. = Total Bet amount; M.C. = Minimum Capital during the betting history.

(iv) In Hong Kong, only US style place bets are available for 4-6 horse races, and for races with more horses, US style show bets are available. In Japan, only place bets are available for 4-7 horse races, and for races with more horses, show bets are available.

From the previous tables, there are no profits based on the Dr.Z system with the Harville model in Hong Kong or Meadowlands. In contrast, the discount model with $r = \infty$ produces positive profits in Meadowlands. In Hong Kong, profits are greater than the Harville model but still negative.

In Japan, the difference between the Harville model and the discount model with $r = 4$ is relatively small when compared to the difference in Hong Kong and Meadowlands. And the discount model does not appear to perform better than the Harville model. This may be due to the fact that the return formula is quite different from the others. The Dr.Z system assumes the return formula is the same as those in Hong Kong and Meadowlands and hence, we modified the nonlinear programming model slightly for this difference in Japan.

We indicated above that using the Japanese return formula, the Dr.Z system picks up more longshots and less favourites and this increases the riskiness of our bet. However, with the Harville model in Japan, the ordering probabilities for favourites are overestimated and those for longshots are underestimated (see Lo & Bacon-Shone (1992) for details) and thus the number of longshots picked up are reduced even if probability restriction is not applied. The above is a possible reason why the discount model with $r = 4$ (which is expected to be more accurate by log likelihood measure) does not appear to be better. Further, we note that improving probability estimation does not always imply higher profits although we expect our optimal bets should be placed more accurately. In addition, it is possible that the fact that long shots are much better rewarded in

Japan may be a partial explanation of why the underlying probability distribution of running times is different in Japan.

IV. Conclusion

Ziemba and Hausch (1987) report that the Dr.Z system produced positive profits at racetracks in the U.S. and Canada. The discount model presented in this paper offers a simple modification to the Dr.Z system in order to fit the running times distribution of a particular data set. It produces significantly better returns with both our Hong Kong and Meadowlands data sets. Further investigation is required to explain why the Dr.Z system cannot produce profits in Hong Kong. We speculate that the larger pool sizes in Hong Kong may attract more professional gamblers with a resulting more efficient market. While in Japan, the discount model does not improve the profits, we suspect it is due to the special return formula of place and show bets. In conclusion, we recommend the use of the discount model with appropriate r in the Dr.Z system which improves estimation of ordering probabilities. This way allows the system to bet more accurately, at least in expectation. We suggest to use $r = \infty$ in Hong Kong and Meadowlands and $r = 4$ in Japan.

REFERENCES

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

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.

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

Breiman, L. (1960) "Investment policies for expanding businesses optimal in a long-run sense." *Naval Research Logistics Quarterly* 7 :647-651.

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.

Grauer R.R. (1981) "A comparison of growth optimal and mean variance investment policies." *Journal of Financial and Quantitative Analysis* 16, No.1 (March) : 1–21.

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.

Kelly, J. L. (1956) "A new interpretation of information rate." *Bell System Technical Journal,* 35 : 917–926.

Lo, S.Y. and Bacon-Shone, J. (1992) "An approximation to ordering probabilities of multi-entry competitions." Research report, Department of Statistics, University of Hong Kong.

Maier, S.F.; Peterson, D.W. & Weide, J.H.V. (1977) "A Monte Carlo investigation of characteristics of optimal geometric mean portfolios." *Journal of Financial and Quantitative Analysis* 12 (June) : 215–233.

Rosner, Bernard (1975) "Optimal allocation of resources in a pari-mutuel setting." *Management Science* 21 (May) : 997–1006.

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.

Thorp, Edward (1971) "Portfolio choice and the Kelly criterion." in Stochastic optimization models in Finance, edited by W.T. Ziemba and R.G Vickson (1975).

Ziemba, W.T. and Hausch, D.B. (1987) Dr. Z's beat the racetrack. Revised edition. William Morrow and Co., inc. New York.