


# Probability and Optimization Models for Racing

Victor S. Y. Lo

University of British Columbia

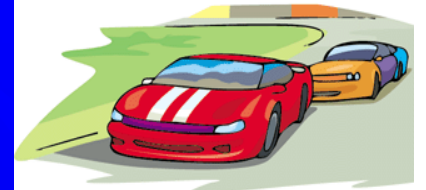
Fidelity Investments



---

Disclaimer: This presentation does not reflect the opinions of Fidelity Investments. The work here was completed at University of British Columbia and the University of Hong Kong.

# Outline



- ✦ Areas to Discuss in Racing:

- ✦ Favorite-longshot Bias (Economics, Statistics)



- ✦ Ordering Probabilities and Optimal Investment (Probability, Statistics, Finance)

# Part 1: Favorite-Longshot Bias

- ✦ Favorites are underbet and longshots are overbet – Busche & Hall (1988), Ali (1977)
- ✦ A well-known phenomenon in economic literature - Ziemba (2004); Creating opportunities – Bolton & Chapman (1986), Bentor (1994)
- ✦ Economic interpretation: Risk-loving behavior

# Favorite-Longshot Bias

Define:

$P_i$  = Bet fraction on horse  $i$ , i.e. consensus win probability,  $i = 1, \dots, n$

=  $(1 - \text{track take}) / (1 + O_i)$ , where  $O_i$  = Odds on  $i$

$\pi_i$  = *objective (true) win probability of  $i$*

$\pi_i < P_i$  if  $\pi_i$ 's are low

$\pi_i > P_i$  if  $\pi_i$ 's are high

# Statistical Model

- ◆ Many techniques mentioned in the literature – Ali (1977), Asch and Quandt (1984)
- ◆ Propose to use a simple logit model, Bacon-Shone, Lo, and Busche (1992a):

$$\pi_i = \frac{P_i^\beta}{\sum_{j=1}^n P_j^\beta}$$

# Statistical Model (continued)

✦  $\beta > 1 \rightarrow$  risk-prefer

✦  $\beta = 1 \rightarrow$  risk-neutral

✦  $\beta < 1 \rightarrow$  risk-averse

$$\pi_i = \frac{P_i^\beta}{\sum_{j=1}^n P_j^\beta}$$

# Universal Comparisons

US racetracks consistently have a risk-prefer bias with  $\beta > 1$

Racetrack	# races	Estimated $\beta$	p-value for H1: $\beta \neq 1$	Average pool size
US (Quandt's 83-84):				
Atlantic City	712	1.10	0.08	unknown
Meadowlands	705	1.12	0.02	\$52K
US (Ali's 70-74):				
Saratoga	9,072	1.16	~0	\$25K
Roosevelt	5,806	1.13	~0	\$218K
Yonkers	5,369	1.13	~0	\$228K
Japan (90)	1,607	1.07	0.01	\$168K
Hong Kong (81-89):				
Happy Valley	2,212	1.04	0.25	\$1.1M
Shatin	1,943	0.94	0.04	\$1.1M
China (23-35):				
Shanghai	730	1.03	0.38	unknown

# Utility Function Interpretation

- Expected utility maximizer is indifferent between betting on any horses in a race:

See Ali (1977), Lo (1992)

- It can be shown:

$$E(U_i) = K \quad \forall i = 1, \dots, n.$$

$$U(1 + O_i) = K \left[ 1 + \frac{\sum_{j \neq i} P_j^\beta}{(1 + t)^\beta} (1 + O_i)^\beta \right] \propto \underbrace{(1 + O_i)^\beta}_{\text{Power utility}}$$

Then, Arrow-Pratt Measure of Absolute Risk Aversion

$$= \frac{-U''(x)}{U'(x)} = -(\beta - 1) / x \quad (2)$$

$< 0$ , and increases with wealth, if  $\beta > 1$

$\Rightarrow$  Bettors take more risk as capital decline "Risk-lovers"

# Conclusion and Research Opportunities

- ◆ Favorite-longshot bias exists in many US racetracks (but not huge)...
- ◆ ...but does not exist in some Asian racetracks – would it depend on the *Pool Size*?
- ◆ Bias in other investment areas – see Ziemba (2004)
- ◆ Opportunity to understand bias or accuracy in complicated bets, e.g. Lo and Busche (1994)
- ◆ Opportunity to apply similar logit models in other applications and other sports, e.g. Lo (1994a), Willoughby (2002)

# Part 2: Ordering Probabilities

- Running time distribution ( $T_i$ 's) is key to determine ordering probabilities:

$$\begin{aligned} \pi_i &= P(T_i < \underset{r \neq i}{\text{MIN}} \{T_r\}) \\ &= \int_0^{\infty} \prod_{r \neq i} [1 - F(t_i | \theta_r)] f(t_i | \theta_i) dt_i, \end{aligned} \quad (3)$$

where  $\theta_i = E(T_i)$  or location parameter ,  
and  $f(\cdot)$  and  $F(\cdot)$  are pdf and cdf , resp .

Given  $\pi_i$ 's, solve for  $\theta_i$ 's. Then compute :

$$\begin{aligned} \pi_{ij} &= P(T_i < T_j < \underset{r \neq i, j}{\text{MIN}} \{T_r\}) \\ &= \int_0^{\infty} F(t_j | \theta_i) \prod_{r \neq i, j} [1 - F(t_j | \theta_r)] f(t_j | \theta_j) dt_j \end{aligned} \quad (4)$$

# Running Time Distribution

- ◆ The following types have been considered in literature, all assuming *independent* running times:

*Exponential – Harville(1973):*  $f(t_i | \theta_i) = \frac{1}{\theta_i} \exp(-t_i / \theta_i)$

*Normal – Henery(1981):*  $T_i \sim N(\theta_i, 1)$

*Gamma – Stern(1990):*  $T_i \sim \text{Gamma}(r, \theta_i)$ ,  
where  $r$  is the shape parameter

# Exponential Running Time

- Strictly speaking, we only need  $g(T) \sim$  Exponential, where  $g(\cdot)$  is a monotonically increasing function

$$\begin{aligned} \pi_{ij} &= P(i \text{ finishes 1st and } j \text{ finishes 2nd}) \\ &= \frac{\pi_i \pi_j}{1 - \pi_i}, \end{aligned} \quad (5)$$

where  $\pi_i$  can be estimated by bet fraction  $P_i$

$$\begin{aligned} \pi_{ijk} &= P(i \text{ finishes 1st, } j \text{ finishes 2nd, } k \text{ finishes 3rd}) \\ &= \frac{\pi_i \pi_j \pi_k}{(1 - \pi_i)(1 - \pi_i - \pi_j)} \end{aligned} \quad (6)$$

# Normal and Gamma Running Time

- ◆ The formulas are complex, as one has to solve (3), a system of integral equations, for  $\theta_i$ 's, and then compute (4)
- ◆ Henery(1981) proposed to use a first-order Taylor series approximation under normal running time
- ◆ Lo and Bacon-Shone (2007) proposed a simple approximation...

# Simple Approximation

Lo and Bacon-Shone (2007):

$$\pi_{ijk} = \pi_i \frac{\pi_j^\lambda}{\sum_{s \neq i} \pi_s^\lambda} \frac{\pi_k^\tau}{\sum_{t \neq i, j} \pi_t^\tau} \quad (7)$$

where  $\pi_i$ 's can be estimated by bet fractions  $P_i$ 's,  
 $\lambda$  and  $\tau$  are parameter values in Lo and Bacon – Shone(2007).  
Note that for Exponential time,  $\lambda = \tau = 1$ ,  
(7) reduces to (5) and (6).

# Running Time Distribution Competition

- ◆ So, which distribution should be used?
- ◆ Lo and Bacon-Shone (1994) found that Harville model has a systematic bias in estimating ordering probabilities based on Hong Kong data and Henery model is clearly superior
- ◆ Bacon-Shone, Lo, and Busche (1992b) had a similar conclusion using Meadowlands data, however...
- ◆ ... Lo (1994b) found that Stern model with  $r=4$  is better than both Henery and Harville using Japan data!

# Correlated Running Times

Constant correlation, i.e.  $\text{Corr}(T_i, T_j) = \rho \forall i \neq j$ ,

reduces to Henery; more complicated cases :

A) Non - constant correlation :  $\rho_{ij} = \psi_i \psi_j \forall i \neq j$ ,

where  $\log\left(\frac{\psi_i}{1-\psi_i}\right) = -\delta - \gamma(\theta_i - \bar{\theta})$ ,  $\bar{\theta} = \frac{1}{n} \sum_i \theta_i$ ,

i.e. correlations higher for stronger pairs.

B) Non - constant variance :  $\sigma_i = \exp[\kappa(\theta_i - \bar{\theta})]$ ,

i.e., if  $\kappa > 0$ , weaker horses will have higher variance.

If  $\gamma = \kappa = 0$ , it reduces to Henery.

# Empirical Results

◆ First order Taylor series approx employed for

Model	Estimates	p-value of Lik ratio test rel to Henery
A) Non-constant correlation ( $\gamma$ only)	$\gamma = 0.58$	0.06
A) Non-constant correlation ( $\gamma$ and $\delta$ )	$\gamma = 0.60, \delta = 0.05$	0.18
B) Non-constant variance	$\kappa = 0.08$	0.06

Non-constant correlation with slope  $\gamma$  only or non-constant variance shows some promise

# Kelly Criterion for Optimal Investment

- ✦ Instead of mean-variance criterion, we maximize expected log wealth → growth rate

Wages on all opportunities in race  $t$ ,  $(X_{t1}^*, \dots, X_{tm}^*)^T$

$$= \arg \max_{X_{t1}, \dots, X_{tm}} \{E(\log W_t) \mid \sum_i X_{ti} \leq W_{t-1}, X_{ti} \geq 0 \forall i\}$$

where  $W_t$  = total wealth at the end of race  $t$ .

- ✦ Breiman(1960), Thorp(1971), Algoet & Cover(1988) show long-run asymptotic optimality
- ✦ Adopted by Hausch, Ziemba, & Rubinstein(1981) using exponential running times, and Lo, Bacon-Shone, & Busche(1995) and Hausch, Lo, & Ziemba (1994) using other running time distributions, all showed promises

# Conclusion and Research Opportunities

- ◆ Knowing the appropriate running time distribution is key to determining ordering probabilities
- ◆ There appears to be no universal best distribution but Henery (Normal) and Stern (Gamma) are competitive
- ◆ Simple approximation is available for Henery and Stern
- ◆ Correlated running time model is more complex but may be better
- ◆ Other approximation methods may be considered especially for more complicated models
- ◆ (Fractional) Kelly is promising for optimal betting

# References for Favorite-Longshot Bias

- ◆ Ali, M.M. (1977) "Probability and Utility Estimates for Racetrack Bettors," *J. of Political Economy*, 84, p.803-815.
- ◆ Asch, P., Malkiel, B., and Quandt, R. (1984) "Market Efficiency in Racetrack Betting," *J. of Business* 57, p.165-174.
- ◆ Bacon-Shone, J., Lo, V.S.Y., and Busche, K. (1992a) "Modelling the Winning Probability," *Research Report 10*, Dept. of Statistics, the University of Hong Kong.
- ◆ Benter, W. (1994) "Computer Based Horse Race Handicapping and Wagering Systems: A Report," in Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. ed. (1994) *Efficiency of Racetrack Betting Markets*, Academic Press, p.183-198.
- ◆ Bolton, R.N. and Chapman, R.G. (1986) "Searching for Positive Returns at the Track, A Multinomial Logit Model for Handicapping Horse Races," *Management Science*, 32, p.1040-1059. Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. ed. (1994) *Efficiency of Racetrack Betting Markets*, Academic Press, p.237-247.
- ◆ Busche, K. and Hall, C.D. (1988) "An Exception to the Risk Preference Anomaly," *J. of Business*, 61, p.337-346.
- ◆ Lo, V.S.Y. (1992) "Statistical Modelling of Gambling Probabilities," *PhD Thesis*, Dept. of Statistics, The University of Hong Kong
- ◆ Lo, V.S.Y. (1994a) "Application of Logit Models in Racetrack Data," in Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. ed. (1994) *Efficiency of Racetrack Betting Markets*, Academic Press, p.307-314.
- ◆ Lo, V.S.Y. and Busche, K. (1994) "How Accurately do Bettors Bet in Doubles?," in Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. ed. (1994) *Efficiency of Racetrack Betting Markets*, Academic Press, p.465-468.
- ◆ Willoughby, K.A. (2002) "Winning Games in Canadian Football: A Logistic Regression Analysis," *The College Mathematics Journal*, 33, No.3, p.215-220.
- ◆ Ziemba, W.T. (2004) "Behavioral Finance, Racetrack Betting and Options and Futures Trading," *Mathematical Finance Seminar*, Stanford University.

# References for Ordering Probabilities

- ✦ Bacon-Shone, J.H., Lo, V.S.Y., and Busche, K. (1992b) "Logistic Analyses of Complicated Bets," *Research Report 11*, Dept. of Statistics, the University of Hong Kong.
- ✦ Harville, D.A. (1973) "Assigning Probabilities to the Outcomes of Multi-Entry Competitions," *J. of American Statistical Association*, 68, p.312-316.
- ✦ Hausch, D.B., Ziemba, W.T., and Rubinstein, M. (1981) "Efficiency of the Market for Racetrack Betting," *Management Science*, 27, p.1435-1452.
- ✦ Henery, R.J. (1981) "Permutation Probabilities as Models for Horse Races," *J. of Royal Statistical Society B*, 43, p.86-91.
- ✦ Henery, R.J. (1985) "On the Average Probability of Losing Bets on Horses with Given Starting Price Odds," *J. of Royal Statistical Society A*, 148, p.342-349.
- ✦ Lo, V.S.Y. (1994b) "Application of Running Time Distribution Models in Japan," in Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. ed. (1994) *Efficiency of Racetrack Betting Markets*, Academic Press, p.237-247.
- ✦ Lo, V.S.Y. and Bacon-Shone, J. (1994) "A Comparison between Two Models for Predicting Ordering Probabilities in Multiple-Entry Competitions," *The Statistician*, 43, No.2, p.317-327.
- ✦ Lo, V.S.Y. and Bacon-Shone, J. (2007) "Approximating the Ordering Probabilities of Multi-Entry Competitions By a Simple Method," To appear in: Hausch, D.B. and Ziemba, W.T. ed. (2007) *Handbook of Investments: Efficiency of Sports and Lottery Markets*, Elsevier.
- ✦ Stern, H. (1990) "Models for Distributions on Permutations," *J. of American Statistical Association*, 85, p.558-564.

## References for Optimal Investment Strategy

- ◆ Algoet, P.H. and Cover T.H. (1988) "Asymptotic Optimality and Asymptotic Equipartition Properties of Log-optimum Investment," *The Annals of Probability*, 16, No.2, p.876-898.
- ◆ Benter, W. (1994) "Computer Based Horse Race Handicapping and Wagering Systems: A Report," in Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. ed. (1994) *Efficiency of Racetrack Betting Markets*, Academic Press, p.183-198.
- ◆ Breiman, L. (1960) "Investment Policies for Expanding Businesses Optimal in a Long-run Sense," *Naval Research Logistics Quarterly*, 7, p.647-651.
- ◆ Haigh, J. (2000) "The Kelly Criterion and Bet Comparisons in Spread Betting," *The Statistician*, 40, Part 4, p.531-539.
- ◆ Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. (1994) "Pricing Exotic Racetrack Wagers," in Hausch, D.B., Lo, V.S.Y., and Ziemba, W.T. ed. (1994) *Efficiency of Racetrack Betting Markets*, Academic Press, p.469-483.
- ◆ Hausch, DB., Ziemba, W.T., and Rubinstein, M. (1981) "Efficiency of the Market for Racetrack Betting," *Management Science*, 27, p.1435-1452.
- ◆ Lo, V.S.Y., Bacon-Shone, J., and Busche, K. (1995) "The Application of Ranking Probability Models to Racetrack Betting," *Management Science*, 41, p.1048-1059.
- ◆ Thorp E.O. (1971) "Portfolio Choice and the Kelly Criterion," *Business and Economics Statistics Section, Proceedings of the American Statistical Association*.
- ◆ Ziemba, W.T. (2004) "Behavioral Finance, Racetrack Betting and Options and Futures Trading," *Mathematical Finance Seminar*, Stanford University.