H&SS Senior Honors Research Thesis Paper

Probabilistic Analysis for Economic Models – Applying Game Theory to Horse Racing

Kah Kien Ong
Carnegie Mellon University
kko@cmu.edu
Research Advisor: Dr. Patrick Sileo

I. INTRODUCTION

Economic models can appear in real-life scenarios in common but unexpected places. One commonly associates the study of economic theory with big money ventures such as the stocks and securities market. In fact, economic theory can also be applied to the field of gambling, where the manipulation of numbers and odds determine the big decisions to be made. Casino gambling games such as blackjack and roulette have been long well explored under game theory due to their inherent mathematical nature. Lottery and other games of chance too, are big havens for game theorists.

But another popular game where we can apply economic principles to that has not received relatively less attention is the sport of horse racing. Like securities markets, horse racing provides an opportunity to study the rationale behind decision-making choices under conditions of risk and uncertainty.

Game theory deals with numerical and probabilistic analysis in order to help us calculate, and perhaps devise a way of maximizing or optimizing the expected returns. For this case, based on betting odds and other information from a gambling game such as horse racing, we can perform an analysis of our expected return from the game, and come up with a strategy for betting. The breakthrough in the analysis problem would be to find a viable methodology for the generic case.

There have been several research papers touching on various aspects of the mechanics and economic theory of horse racing over the years. My goal for this project is to do a research and investigation of how game theory and probabilistic analysis can be used for practical applications in a real-life scenario like horse race betting. Based on actual information from horse races, we can do a probabilistic risk analysis of the preference of the bettors, thereby providing us with a much greater understanding of the game.

II. PARI-MUTUEL BETTING

Horse race betting is based on the ***Pari-Mutuel*** Wagering System. This is a special kind of wagering system developed by Frenchman Pierre Olle in late 19th century. The name of the system means literally “amongst ourselves” in French, and sums up the gist of how the odds are calculated.

Under the Pari-Mutuel Wagering System, when you wager, you are betting not against the track, but against everyone who is making that type of wager in a particular race. In other words, the ***odds*** or ***payoff*** of a particular horse is never fixed, and depends solely on how much money all the other bettors are placing on each horse.

Naturally, a strong horse (termed a ***favorite***) that is viewed as more likely to win a particular race would draw more bets, and thereby reduce its odds and also the payoff for all the bettors if it wins the race. Hence the betting system is self-adjusting, the racetrack can usually accept as many bets as it wishes without worrying about losing money even if a favorite or ***longshot*** (a horse with low odds of winning) comes in first.

After each race, the racetrack will collect a profit margin (typically 17-25%, depending on the type of wager) from the wagers, termed the track take or ***overround***. This covers the race expenses like maintaining the track and paying the stewards. The remaining pool is divided amongst the winners for the race. The only possible scenario where the racetrack might lose money is when a favorite that is heavily betted on wins and the shared winnings round down to nothing. In this case, the racetrack must still pay back the winners the minimum payoff of $2.10 per $2.00 bet, which results in a slight net loss.

6 Runners	Odds Offered	% Percentage Terms
Horse 1	6/4	40
Horse 2	3/1	25
Horse 3	4/1	20
Horse 4	11/2	15.38
Horse 5	9/1	10
Horse 6	12/1	7.69
		118.07 - 100 Overround 18.07% Margin

Figure 1: Odds calculations based on pool totals.

There are three common types of bets offered by racetracks that a bettor can wager on, which I will cover here in this research for the sake of uniformity and simplicity. They are ***Win*** bets for the horse to come in first place in the race, ***Place*** bets for the horse to come in either first or second, and ***Show*** bets for the horse to come in either first, second or third place. Each of these three types of bets maintains a separate ***pool***, which is the total sum of the worth of bets made on all the horses. At the end of the race, the pool totals for each type of bet (after subtracting the cut for the overround) are divided amongst all the winners. This implies that Place and Show pools generally give much smaller payoffs, since they are split amongst two and three winners respectively. Hence less people also bets on these two types of bets, further reducing the pool totals.

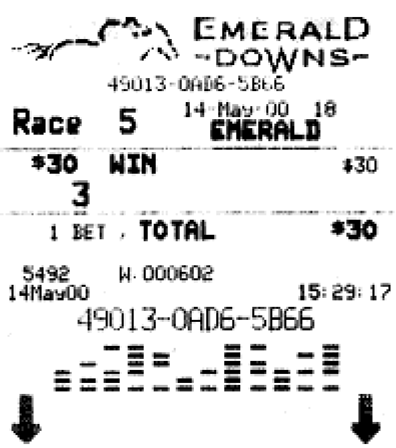

Figure 2: Typical tote ticket showing the race number, wager type and bet amount

For my research on applying game theory to the horse race betting, I would employ these pool totals for probabilistic analysis. I would investigate the possibility of finding an optimal strategy of betting such that a winning ticket can be formed based on the current track odds, which are influenced by the current pool totals.

The observation made is that these pool totals fluctuate greatly during the minutes before ***Post-Time***, which is the time when bets for the race are closed and horses line up at the posts (starting line). This is because a large number of bettors put in last-minute bets based on the current odds and pool total information, both at the racetrack and also online. These additional bets just minutes before post-time dramatically affect the payoff of the winners of the race, and hence could provide an optimal strategy for a bettor who analyses these trends.

Fair Grounds			Time:	11:59	
Race	1	83 mins. to Post			
Runner	M/L	Odds	Win	Place	Show
1	8	11	7	0	0
2	8/5	4/5	48	21	5
3	6	3	20	0	0
4	8	4	16	14	10
5	5/2	25	3	1	3
6	4	6	12	1	0
		Pool Total:	106	37	18

Go to Fair Grounds Exactas, Quinellas, Doubles, Entries, Changes or Results

Figure 3: Typical online tote board showing race number, horse numbers, odds and pool totals for each type of wager. The time to post is on the top right. (Source : Racingchannel.com)

III. BACKGROUND

Generally, there are less academic research papers and books that introduce mathematical methods to horse race betting as compared to other forms of gambling. The reason for this is probably due to the special nature of this gambling sport, which uses pari-mutuel betting as explained before. Hence, the bettors are actually wagering against each other, and odds are not fixed or known. And for the same reason, there has been quite a bit of research that touch on the economic theories underlying the sport too.

Early publications generally investigate the utility preferences of bettors and how it affects their betting decisions. This includes works by Friedman and Savage [1] and Markowitz [2], who came up with their hypothesis of the average bettor's utility of wealth curve. Later on, Weitzman [3], Rosset [4], and Ali [5] did more empirical analysis using a lot of racetrack data that they collected, and generally agreed with the earlier findings.

Following up on their work, Synder [6] investigated the preference of bettors for low probability-high return bets, as compared to high probability-low return combinations. He attributed this to the utility function and risk preferences of the bettor, and tried to show whether the bettor can take advantage of this knowledge. Many others like Figlewski [7], Asch, Malkiel and Quandt [8], Ziemba and Thaler [10], and Ziemba, Hausch and Lo [11] all covered the topic of this anomaly and tried to show market inefficiency without success. It I generally agreed that the advantage gained from the information was not sufficient for the bettor to overcome the track take and make more profit. These are all areas that I will cover briefly later on in this paper.

IV. DATA COLLECTION

In order to perform my probabilistic analysis, I first need to collect the necessary data from various horse races over a span of time. For this information, I decided to employ online horse race betting websites as they provide convenient sources of relevant material that are updated real-time for each race. It is also feasible for me to collect the needed information from these websites, as I could write a program script which downloads the online data on the website at the specific timings before post-time of each race. The data that I would specifically need to extract for each race would be the pool totals for Win, Place and Show for each race at various timings before and at post-time.

Using the programming language Perl, I wrote a web script that could download an entire webpage from the Internet when executed. This script is in turn set to execute at various fixed timings based on the race times using the Windows Scheduler.

When executed, the Perl script itself first performs a series of initialization steps and checks for the default operating system and web browser types based on the headers. If no errors occur, it will send a socket open request over the Internet through the port number provided to the remote hostname. If this is successful, it will use the GET HTTP command to download the entire webpage to the local disk based on the provided web address (url).

After each race, I parsed these html files on the local disk manually and extracted the relevant information that I needed, namely the pool totals. The figures obtained are then tabulated into an Excel spreadsheet for ease of comparison.

V. DATA SOURCES

As mentioned earlier, I chose to use online horse race betting websites as my primary source of data. Due to the great popularity of horse race betting to both serious and casual gamblers alike, many websites have sprung up providing real-time online betting options for these people. These websites have many appealing advantages over the traditional means of going to a race to lay the bets, even though the bettor is missing out on the visual action. Online betting websites provide a wealth of information to a bettor, ranging from odds, pool tables, track information, horse information and even betting strategies. This allows bettors to have the convenience of not having to go to the racetrack in person, while still having the ability to make a probabilistic analysis of the odds of winning based on the data available.

The data available varies greatly between the online betting websites. A long but non-comprehensive list of these websites that have turned up during my research appears in the Appendix A at the end of this paper. Most of them are actually very business-oriented, and provide just the minimal information needed for people to be confident enough to lay their bets. Few provide detailed information such as pool totals that allows bettors to make a thorough probabilistic analysis to come up with an optimal betting strategy. From this list of websites, I eventually picked racingchannel.com as my source of data, as it provides the required information in a clear format that updates itself frequently before a race starts.

Some of these websites actually provide other forms of ***handicapping*** information popular to experience bettors, such as the hardness of the track, shoes of the horse, appearance of the horse, and weight of the jockey etc. As I will elaborate at the end, these are also pieces of information that can affect a bettor’s strategy greatly. Unfortunately, data such as these are difficult to quantify and are open to different interpretations. It is almost impossible to assign weights objectively to these factors in order to do a probabilistic analysis involving them.

VI. EVALUATION OF DATA

Based on the data collected from thoroughbred races between December and February at Fair Grounds, a few observations were made and a preliminary evaluation performed on the proposed betting strategy.

The most obvious conclusion drawn from the data is that there is a really dramatic increase in betting activity as time to post approaches zero. From Post Time (PT) to OFF alone, the bets doubled in quantity. A sample table of data from one of the races is shown in Figure 4 in Appendix B at the end of this paper. For this particular case, I collected data up to 20 minutes before post time, and the big jumps in the pool totals can be clearly seen.

However, despite the much larger quantity of bets being placed in those last few minutes before Post Time, the fluctuations in the odds are minimal. In other words, the distribution of bets in those last few minutes is closely related to the odds or subjective probability that had been prevailing before.

At this point, I decided I will only use the data obtained for the Win pools, as this is the most common type of bet made by bettors, and also avoids the anomalies of inefficiency displayed by the Place and Show pools according to Ziemba [10].

Based on my understanding of the horse race betting system and also the data at hand, I tried to investigate the possibility of forming a winning ticket based on the additional last minute pool totals information I have. If this could be done, then it would show that the horse race betting market is not an ***efficient*** one.

There are actually three different degrees of conditions that define market efficiency. Generally, a market is efficient if no bettor can gain profits using different sources of information. Weak form tests check that no bets have a positive expected value, meaning that knowledge about the subjective probability cannot be used to get above average returns. Semi-strong form tests check whether publicly available information, such as the ***morning-line*** odds provided by handicappers, is fully discounted by the market. Strong form tests check whether all information, public or private, can be made use of by the bettor to make a profit.

However, after a mathematical proof (which I have included in Appendix C at the end of this paper), I arrived at the conclusion that it is not possible to form a winning ticket using the additional information. Hence in this sense, the horse race betting market is an efficient one. This result also agrees with findings by Snyder [6] and Ziemba [11].

VII. FAVORITE-LONGSHOT BIAS

The proportion of the money in the Win pool that is bet on any given horse (or the win bet fraction) can be seen as the ***subjective probability*** that the horse will win the race. Taking the sum over many races, one can observe that there is a strong correlation between the subjective and ***objective probability*** (or true win probability). However, studies by Ali [5] and Snyder [6], amongst others, have shown that favorites win more often than their subjective probabilities indicate, and longshots win less often. Intuitively, this means that favorites are actually much better bets than longshots. This anomaly in horse race betting is known as the ***favorite-longshot bias***.

To demonstrate this behavior, I classified my data according to the percentage the Win pools for each horse in each race constituted. In particular, I isolated the data that made up between 30-35% of the total Win pool (favorites), and those that made up between 2-5% of the total Win pool (longshots). I chose to group together a range of Win pool percentages (the subjective probability) so as to increase the number of horses/pools that could be considered given my relatively smaller data size. At the same time, I was careful not to set too large a range that would introduce too much variance. Within the two subclasses, I then noted the percentage of them that were actually winners (the objective probability). The observation is that for the first class of favorites, the objective probability is 38.3%, which is actually higher than 35%. Conversely, for the class of longshots, the objective probability is found to be 1.6%, which is lower than 2%. This agrees with the results of previous findings that indicate the presence of a favorite-longshot bias.

The presence of the favorite-longshot bias can be explained very simply. Why do most people gamble? Not so much just to win money by betting on favorites, but more rather to get the thrill of the big wins from longshots. From a social and practical point, winning on a longshot also gives the bettor much desired bragging rights. The tradeoff is the ***risk*** involved in the bet. Risk refers to the variability of the outcomes of some uncertain activity. Longshots are riskier as they have a lower chance of winning, and also a lower expected value of return, since their objective probability is lower than their subjective probability. ***Risk-seeking*** bettors clearly generate this bias because bettors will demand a higher expected return for favorites that have a lower variance of return than do longshots. It is understandable that in a gambling sport like horse racing, the representative average bettor would be considered risk-seeking, and perhaps also over a larger range of states of wealth.

So how do we explain a risk-seeking bettor wagering on a longshot with a negative expected value of return? According to Bernoulli's explanation of the St. Petersburg paradox, individuals do not care directly about the dollar prizes of a game, but rather they respond to the ***utility*** these dollars provide. Utility refers to a relative ranking of levels of preference to an individual. Hence we need to look at the utility provided by the prize money rather than the dollar amount of the prize money itself. So like the bettor, we should not look at the Expected Value, EX, of a race, but instead the ***Expected Utility***, EU(X). Therefore, bettors are not ***expected value maximizers*** with ***rational expectations***. Instead, they will maximize expected utility, based on their individual ***utility of wealth functions***.

VIII. RISK PREFERENCES

Friedman and Savage [1] were the first to theorize about the risk behavior of bettors. They came up with a utility of wealth curve with sections of different curvature corresponding to different states of present wealth. Markowitz [2] later came up with an amended version that is shown in Figure 5 below.

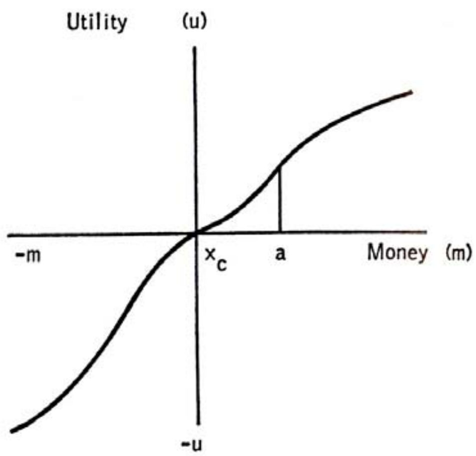

Figure 5: Markowitz's utility of wealth curve [2]

The origin x_c represents the present state of wealth. Up to the point a, bettors are risk-seeking, as can be seen from the convex curve which indicates increasing marginal utility. Beyond a, the curve becomes concave and there is decreasing marginal utility. This shows that bettors eventually become ***risk-adverse*** after winning lots of money. The more risk-seeking the bettor is, the further a will be along the m-axis. Similarly, on the negative side of the m-axis, bettors are risk-adverse when they are losing money. But beyond a certain point when they have lost most of their money, they become more risk-seeking again, hoping for a big win to recoup their losses. Studies have proven this theory by investigating the behavior differences of bettors at the end of the day and during the Depression years.

Qualitatively, we can demonstrate the risk-seeking nature of the representative average bettor using a similar method as Ali [5]. Ali uses the odds for the race for his calculation, while I will use the payoffs instead. He also considers the utility of the state of wealth of the bettor, while I choose to consider only the utility from the payoff (or loss). This is valid as I can assume that the bettor started off with a capital equivalent to one bet, which is what Ali did eventually too.

The basic assumption made is that all bettors have the same preference (homogeneous). Hence they are all identical and choice of bet will shift until they are all equal. So in theory, there is only one representative person betting in the system that we have to consider.

Consider a single race with 8 horses:
π_i is the objective probability of winning for horse i, where $8 \geq i \geq 1$.
P_i is the payoff for horse i when it wins.
Take is the track take for the race.
W_i is the wager on horse i.
W_t is the total wagers for the race.

Therefore, $W_t = \Sigma W_i$

$$P_i = \frac{Wt - Take}{Wi}$$

Now, the expected utility of a bet on horse i is equal to the objective probability of horse i winning X utility of the payoff for horse i + objective probability of horse i losing X utility of losing the $1 bet.

$$\begin{aligned} EU_i &= \pi_i U(P_i) + (1-\pi_i)U(-1) \\ &= \pi_i U(P_i) + U(-1) - \pi_i U(-1) \\ EU_i - U(-1) &= \pi_i [U(P_i) - U(-1)] \end{aligned}$$

We know that utility functions are unique only up to positive linear transformations. Hence I can assume that Without Loss of Generality (WLOG):

$$U(-1) = 0$$

Therefore, $EU_i = \text{л}_i\, U(P_i)$ ---- 1

Assuming the market system is perfect, there should be no difference between any of the horses. That is, no preference can be made between any two horses by the bettor.

$$EU_1 = EU_2 = EU_3 = \ldots = EU_8$$

Say we rank the horses in order such that horse 8 has the highest odds.

$$U(P_i) = 1 \quad \text{--- 2}$$

From 1 and 2: $U(P_i) = \text{л}_8 / \text{л}_i$

Hence the values from the data can be plotted on a utility of wealth graph and linear regression used to find the best-fit curve for the points. The resultant graph will be an upward sloping convex curve corresponding to the risk-seeking region between x_c and a from Markowitz's utility of wealth graph.

Next, I would want to estimate the risk preference of bettors, based on the data. To do this, I need to perform a Maximum Likelihood Estimation on the Log-Likelihood Function. I chose to follow a recent strategy employed by Blough [12] for comparison. The steps are replicated below.

Consider a sample set of M races:
л_i is the objective probability of the winning horse for race i, where $M \geq i \geq 1$.
P_j is the subjective probability of horse j in a race, where $8 \geq j \geq 1$.

Thus the likelihood of the sample is

$$L = \text{л}_1\, \text{л}_2\, \text{л}_3 \ldots \text{л}_M$$

The log-likelihood function is then:
$\ln L =$
$\Sigma\{(\delta Sm+\beta)\cdot\ln(\text{л}_i) - \ln[\Sigma(P_j)^{\delta Sm+\beta}]^{-1}\}$

δSm is the contribution of differences of opinion as measured using the morning line odds to explain the relation between the subjective and objective probabilities. We should expect $\delta \geq 0$, but close to 0.

Sm was calculated slightly different from Blough, since he used a combination of the morning line odds from both the Chronicle and Tribune for the Golden Gate Fields. For my calculations, I only used the morning line odds (MLODDS) provided by racingchannel.com. I also modified the derivation of the parameter accordingly as such:

$$Sm = 0.125\Sigma(\ln(MLODDS))^2$$

β is the contribution of risk preferences. A value of $\beta>1$ would indicate a risk-seeking nature. We should expect $\beta>0$, but also $\beta>1$, since bettors need to be risk-seeking to overcome the track take.

By maximizing the log-likelihood function with respect to δ and β, I can get the maximum likelihood estimates of these parameters. I set a test of the null hypothesis that $\beta=1$ and $\delta=0$ and used a statistical software Eviews 3.1 to carry out the maximum likelihood estimation of the parameters. The results of the estimation are shown in Figure 6 in the Appendix D.

The estimated coefficient of β obtained is $1.18 > 1$, indicating a risk-seeking nature. This is pretty close to the value Blough obtained as well. But the estimated coefficient of δ obtained is $-1.14 < 0$, which indicates too high a contribution by the differences of opinion. This shouldn't be the case as δ should be positive. Interestingly, Blough also obtained a negative result for δ in his test, although it was not as much negative.

IX. CONCLUSION

The horse race betting market is not an area where a common person would usually think of applying economic theories. It provides an unusual environment for investigation due to its special wagering system. My first attempt to find a way to form a winning ticket based on additional track information was unsuccessful. It was proven that even with the knowledge of pool totals close to Post-Time, bettors cannot take advantage to make a higher than average profit. But in a way, this also demonstrates the efficiency of the market.

The shift in direction to the utility function and risk preference aspect of bettors is yet another effort at analyzing the underlying theories behind the horse race betting market. Bettors are in general risk-seeking, which contributes to the favorite-longshot bias anomaly in the market. While bettors can actually take advantage of this bias to increase their expected value of return, their risk preferences indicate that they would instead continue to maintain this bias so as to maximize their expected utility from betting. In fact, the pari-mutuel wagering system ensures that the bettors self-correct whenever there is a shift in preferences in betting, thus maintaining an equilibrium.

This is an interesting phenomenon, and they are all new things that I've learnt during the course of this research. Hopefully, more people will continue to interest and investigate these more obscure applications in the field of economics, rather than always be limited in the traditional genres.

ACKNOWLEDGEMENTS

I would like to thank my research advisor, Dr. Patrick Sileo for invaluable assistance and advice without which this work would not have been possible. I would also like to thank the Carnegie Mellon University Humanities and Social Sciences faculty for giving me the opportunity to undertake this research project, which has benefited me immensely.

REFERENCES

[1] M. Friedman and L. Savage, "The Utility Analysis of Choices Involving Risk", Journal of Political Economy, Vol. 56, No. 4, pp 279-304, 1948.

[2] H. Markowitz, "The Utility of Wealth", Journal of Political Economy, Vol. 60, No. 2, pp 151-158, 1952.

[3] M. Weitzman, "Utility Analysis and Group Behavior: An Empirical Study", Journal of Political Economy, Vol. 73, No. 1, pp 18-26, 1965.

[4] R. Rosett, "Gambling and Rationality", Journal of Political Economy, Vol. 73, No. 6, pp 595-607, 1965.

[5] M. Ali, "Probability and Utility Estimates for Racetrack Bettors",

Journal of Political Economy, Vol. 85, No.4, pp 803-815, 1977.

[6] W. Snyder, “Horse Racing: Testing the Efficient Markets Model”, Journal of Finance, Vol 33, No. 4, pp 1109-1118, 1978.

[7] S. Figlewski, “Subjective Information and Market Efficiency in a Betting Market”, Journal of Political Economy, Vol. 87, No. 2, pp 75-88, 1979.

[8] P. Asch, B. Malkiel, R. Quandt, “Racetrack Betting and Informed Behaviour”, Journal of Financial Economics, Vol. 10, No. 6, pp 187-194, 1982.

[9] W. Ziemba, D. Hausch, “Beat The Racetrack”, New York, Morrow, 1987.

[10] W. Ziemba, R. Thaler, “Anomalies – Parimutuel Betting Markets: Racetracks and Lotteries”, Journal of Economic Perspectives, Vol. 2, pp 161-174, 1988.

[11] W. Ziemba, D. Hausch, V. Lo, “Efficiency of Racetrack Betting Markets (Economic Theory, Econometrics, and Mathematical Economics)”, 1994.

[12] S. Blough, “Differences of Opinion at the Racetrack”, Academic Press, Inc, 1994.

[13] R. Epstein, “The Theory of Gambling and Statistical Logic”, pp 287-294, 1995.

APPENDIX A

Horse Racing Tote Boards and Websites:

Racing Channel
http://www.racingchannel.com/

BRIS Super Tote
http://www.brisnet.com/

The Supertote Board from BRIS
http://206.24.34.156/cgi-bin/tracks.cgi

Daily Racing Forum (DRF)
http://www.drf.com/index2.html

Handicappers' Daily
http://www.itsdata.com/

Pace Advantage
http://www.paceadvantage.com/racedayinfo.htm

Trackpro.com
http://www.trackpro.com/how_to_b.htm

AboutHorseRacing.com
http://www.abouthorseracing.com/www/quick_guide.htm

HorseRacingGold.com
http://www.horseracinggold.com/Horse%20Racing%20Links.htm

Casino-Info.com
http://www.casino-info.com/tips/racebook_tips/history.html

APPENDIX B

Sample Statistics

	Race	
Fair Grounds	10	2/10/2003

	Odds (Horse number)								Pool Totals		
Time to Post	1	2	3	4	5	6	7	8	Win	Place	Show
M/L	12	9/2	8	8/5	12	7/2	6	15			
20	13	5	9/2	4/5	25	7	17	20	5592	1688	371
15	17	7/2	6	3/5	35	7	20	20	9045	2776	679
10	20	7/2	7	3/5	30	7	19	22	11335	3404	1050
5	20	5/2	8	1	30	5	15	25	15224	4937	1978
4	18	5/2	8	1	25	5	15	25	16348	5327	2109
3	20	5/2	9	4/5	30	5	15	25	18526	5819	2241
2	21	5/2	9	4/5	25	5	16	30	20321	6305	2510
1	22	5/2	10	4/5	30	5	15	30	22576	6895	2721
PT	25	3	10	4/5	30	9/2	12	35	27066	8088	2901
OFF	30	3	9	4/5	40	9/2	13	50	69933	20664	7927

Figure 4 : Sample data from a race at Fair Grounds, showing pool totals for the various types of bets at various times

APPENDIX C

Mathematical proof:

Let n be the total number of horses in the race, where $6 \leq n \leq 12$ (commonly 8).
Let $x_1, .. ,x_n$ be the amount bet on each horse on the winning ticket.
Therefore, the total bet on the winning ticket $x = \sum_{i=1}^{n} x_i$

Let $o_1, o_2, o_3, .. ,o_n$ be the closing odds for each horse. For clearer algebra, the odds in this case refer to the total payoff received for a \$1 bet, including the bet itself. Hence in actual fact, it is equal to the track odds + 1 (for the initial bet itself).
Let $p_1, p_2, p_3, .. ,p_n$ be the closing Win pools for each horse.
Further let P represent the total Win pool, where $P = \sum_{i=1}^{n} p_i$
Assuming the racetrack claims a takeoff of t%, the remaining Win pool that is returned to the winning bettors, $P' = (1 - t/100)P$

Now, we know that for $1 \leq i \leq n$, closing odds $o_i = P'/ p_i$ --------- 1

And in order for the ticket to always land a profit for any winning horse,
We want $x_i o_i > x$ for all $1 \leq i \leq n$

Substituting 1 into the equation:
$x_i P'/ p_i > x$ for all $1 \leq i \leq n$
$x_i P' > x p_i$ for all $1 \leq i \leq n$

Summing together for all i from 1 to n for both the left and right hand sides:

$$P'\sum_{i=1}^{n} x_i > x\sum_{i=1}^{n} p_i$$

=> $P'x > Px$

This is a contradiction, since we know from earlier that $P' = (1 - t/100)P$, hence $P' < P$.
Therefore, it is proven that there cannot be a way to form a winning ticket based purely on the Win bets alone.

APPENDIX D

LogL: RISK
Method: Maximum Likelihood (Marquardt)
Date: 04/07/03 Time: 10:35
Sample: 1 100
Included observations: 100
Evaluation order: By observation
Initial Values: C(2)=0.00000, C(1)=1.00000
Convergence achieved after 7 iterations

	Coefficient	Std. Error	z-Statistic	Prob.
C(2)	-1.148684	0.536958	-2.139245	0.0324
C(1)	1.166091	2.109945	0.552664	0.5805

Log likelihood	460.3354	Akaike info criterion	-9.166708
Avg. log likelihood	4.603354	Schwarz criterion	-9.114605
Number of Coefs.	2	Hannan-Quinn criter.	-9.145621

Figure 6: Maximum likelihood estimation results using Eviews

Note:
C(1) corresponds to the estimated coefficient of the parameter β·
C(2) corresponds to the estimated coefficient of the parameter δ·