Parabola Volume 45, Issue 3 (2009)

Picking Winners

Bruce Henry[1]

Mathematics is largely concerned with finding and describing patterns in a logically consistent way, and where better to look for patterns than in Melbourne Cup races [2]. Can we use a little mathematics to help us win? Most people bet on a particular horse, or a select few, in the hope that one of their selections will win and that the amount recovered in winnings will exceed the amount outlaid. Most of the risk here is concerned with picking the winner. Usually the amount to be won is known exactly (in on-course betting) or approximately in TAB betting.

There is a another strategy to consider. What if we backed all horses? This strategy eliminates the risk of picking the winner but if you followed this strategy you would surely lose money. Or would you? If you had followed this strategy in 2008 then you would have won. There were twenty-two horses that raced and the payout for a win on the TAB was $41.00 so that equates to a $19.00 profit for a 'sure thing'. You would have won on the trifecta too. To win on the trifecta you need to pick the horses that will finish first, second and third in the correct order. If there are N horses in the race then there are N ways to pick the first, $N-1$ ways to pick the second and then $N-2$ ways to pick the third; so there are $N(N-1)(N-2)$ possibilities to cover. In 2008 there were 22 starters with 9,240 combinations (which we cover at $1 each) for the trifecta which paid out $22,324 on SuperTAB. A cool $13,084 for a 'sure thing'. Or how about the first four (picking the first four in correct finishing order) which would have cost $175,560 to cover all possibilities but paid out a staggering $589,778. That equates to a profit of $414,218 for a 'sure thing'.

But of course the 'sure thing' is not really a sure thing. It was only a sure thing in picking the winner. The pay out was not a sure thing. Many other possibilities may have occurred resulting in a much smaller payout, less than the $22 needed to cover all possibilities for the winner, and less than $9,240 needed to cover all possible combinations for the trifecta. In 2005 there were 23 starters and the winner paid only $3.60 so that the 'cover all possibilities' strategy would have resulted in a $19.40 loss. The possible loss is even greater on the trifecta. Consider 1992 with 21 starters and thus 7,980 trifecta combinations. The winning combination only paid out $144 so that the 'cover all possibilities' approach would have resulted in a $7,836 loss. In 2006 there

[1]Bruce Henry is an Associate Professor in the Department of Applied Mathematics, School of Mathematics and Statistics, University of New South Wales.

[2]The idea for this article arose out of a group discussion on possible applications of mathematics between A/Professor Bruce Henry and Professor Tony Dooley from the School of Mathematics and Statistics, UNSW, and Ivana Tranchini and Jenna Frost, from Horizon Communication Group.

Year	N	W	C	T
2008	22	46.50	9,240	22,324
2007	21	22.60	7,980	1,377
2006	23	17.50	10,626	1,100
2005	24	3.60	12,144	4,951
2004	24	4.20	12,144	2,504
2003	23	9.10	10,626	13,203
2002	23	6.00	10,626	2,665
2001	22	10.00	9,240	5,210
2000	22	16.90	9,240	3,570
1999	24	6.20	12,144	32,131*
1998	24	12.00	12,144	2,030
1997	22	5.70	9,240	1,246
1996	22	6.80	9,240	24,507
1995	21	9.90	7,980	1,519
1994	24	20.40	12,144	3,220
1993	24	16.80	12,144	55,893
1992	21	5.30	7,980	144
1991	23	4.80	10,626	7,180
1990	24	11.50	12,144	1,142
1989	23	27.20	10,626	13,520
	456	263.00	208,278	199,443

Table 1: The table lists the numbers of horses N that raced in each year of the Melbourne Cup over the past twenty years together with the TAB payouts for a win W and the trifecta T. The cost of covering all possible combinations C in the trifecta is also shown.

* In 1999 it was a dead heat for third with two paying combinations in the trifecta: one that paid \$19,242 and the other that paid \$12,889.

were 23 starters which would have cost \$212,520 to cover all possibilities for the first four but the return was only \$9,616.

The table below shows more complete data over the past twenty years for the 'cover all possibilities' strategy. The raw data was obtained from the *Millers Guide* website `http://www.millersguide.com.au/melbourne-cup-results`

The table lists the number of starting horses N, the payout for the winner W, the number of combinations for the trifecta C, and the payout for the trifecta T over the past twenty years. The totals for each column are also shown. The payouts are based on a \$1 bet in the case of picking a winner and a \$1 bet for a trifecta combination.

The totals from the table reveal that to cover all possibilities for a win over the past twenty years would have cost \$456, for a return of \$263, equating to a loss of \$193. In the case of the trifecta an outlay of \$208,278 was needed to cover all possibilities over the twenty-year period and this would have returned \$199,443 for a net loss of

$8834. Clearly the stategy of covering all possibilities is not a winner. But there is some hope in this data. The average player return percentage through the TAB has been approximately 84%. The strategy of covering all possibilities for a win has returned 58%, much less than the average, but the strategy of covering all possibilities in the trifecta, whilst not profitable, has returned almost 96%. This is close enough to consider a little tweaking.

Suppose we forgo the certainty of picking the winning combination by eliminating all horses whose starting odds are longer than a given threshold amount. Here we want to forget about the rank outsiders. Again we use raw data from the *Millers Guide* web site.

Summary analysis is shown in the table folllowing for two cases: (i) eliminating all horses whose odds are longer than 100/1 and (ii) eliminating all horses whose odds are longer than 50/1.

The table lists the number of starting horses with odds of 100/1 or shorter $N(100)$, the number of combinations for the trifecta in this case $C(100)$, and the pay out for the trifecta in this case $T(100)$. The number of starting horses, numbers of combinations and trifecta payout for horses with odds of 50/1 or shorter are shown in columns headed $N(50)$, $C(50)$ and $T(50)$. The totals for each column are also shown.

The results in the table are interesting. The strategy of eliminating all horses with odds longer than 100/1 resulted in a bigger loss than the strategy of covering all possibilities, returning 86% of the outlay. There were only two Melbourne Cups in the past twenty years in which horses with longer odds than 100/1 finished in the final three. But these years were precisely the years in which the payout was largest. We saved a bit by not covering so many possibilities but we lost more by not picking up the big winnings with these long odds horses.

On the other hand the strategy of eliminating all horses with odds longer than 50/1 and covering all remaining combinations on the trifecta would have proven very profitable over the past twenty years. We will refer to this strategy as the $MCT50$ scheme. There were four Melbourne Cup races in the past twenty years in which a horse with odds longer than 50/1 finished in the top three (including 1999 when there was a dead heat for third and one of those horses was at odds longer than 50/1). So the total payout is less, but the outlay with this strategy is considerably less. This strategy would have returned $40,885 profit over the past twenty years. The 50/1 cut-off is optimal for this historical data but other cut-offs would have also proved profitable. For example, the strategy of covering all possibilities except those including horses with odds longer than 60/1 would have returned a profit of $33,133 over the past twenty years, and the strategy of covering all possibilities except those including horses with odds longer than 160/1 would have returned a profit of $25,059. But the cut-off at 45/1 would have only just broken even and the cut-offs of 40/1 and 100/1 would have both returned losses of about $20,000.

One of the interesting features of the strategy of eliminating all horses with odds longer than 50/1 and covering all remaining combinations over the past twenty years is that after the initial outlay of $4896 in 1989 no further funds were needed. The growth of this initial investment compares favourably with the average growth in the

Year	$N(100)$	$C(100)$	$T(100)$	$N(50)$	$C(50)$	$T(50)$
2008	19	5,814	22,324	15	2730	22,324
2007	17	4,080	1,377	13	1716	1,377
2006	18	4,986	1,100	14	2184	1100
2005	20	6840	4951	14	2184	0
2004	24	12144	2504	17	4080	0
2003	19	5814	13203	18	4896	13203
2002	23	10626	2665	17	4080	2665
2001	22	9240	5210	18	4896	5210
2000	19	5814	3570	16	3360	3570
1999	20	6840	12889	15	2730	12889
1998	21	7980	2030	19	5814	2030
1997	20	6840	1246	16	3360	1246
1996	20	6840	24507	17	4080	24507
1995	21	7980	1519	17	4080	1519
1994	21	7980	3220	17	4080	3220
1993	19	5814	0	16	3360	0
1992	19	5814	144	15	2730	144
1991	21	7980	7180	18	4896	7180
1990	21	7980	1142	19	5814	1142
1989	20	6840	13520	18	4896	13520
		144,156	124,307		75966	116851

Table 2: The table lists the numbers of horses with starting odds of 100/1 or less $N(100)$ and the number of horses with starting odds of 50/1 or less $N(50)$ together with the costs of covering all these horses $C(100)$ and $C(50)$ respectively in the trifecta. The trifecta payouts based on these combinations are also shown.

share market over the same period and it has performed better than average bank interest rates. The plot shows the growth of the $MCT50$ scheme (circles) based on the initial outlay of $4896 with growth of the same initial asset with compound interest accruing at an annual rate of 11.23% (solid line). The vertical axis is in units of $10,000.

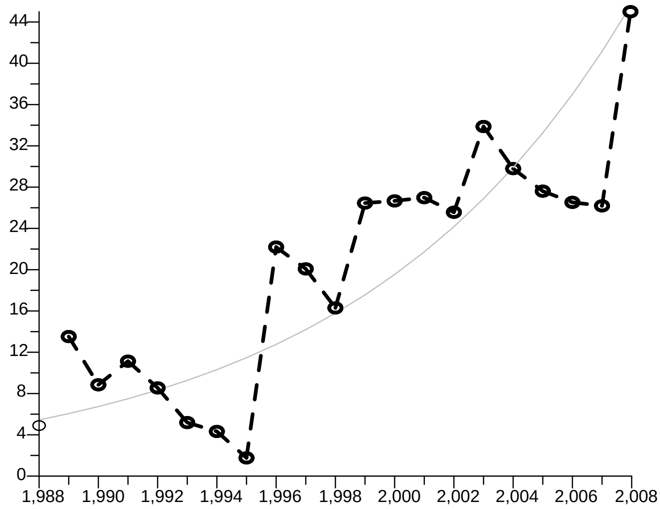

Figure 1: The plot shows the growth of the initial outlay of $4896 (circles) if it had been invested in the Back All Combinations with Odds of 50/1 or Less in the Melbourne Cup Trifecta ($MCT50$) over the past twenty years, compared with growth of the same initial asset with compound interest at an annual rate of 11.23%. The vertical axis is in units of $10,000.

One of the things that financial analysts like to know for a given investment strategy is the risk. There are many ways to try to define and measure risk. There are different questions that can be addressed here. How much could you lose? What is the probability of losing a given amount? How robust is the scheme? For the $MCT50$ scheme it is easy to determine the maximum possible loss in any given year but the other issues are not as clear cut. What is the risk in covering all possibilities except those including horses with odds longer than fifty to one in the Melbourne Cup Trifecta with $1 bet on each combination? In any given year the maximum outlay is \$ $N(N-1)(N-2)$ where N is the number of horses with starting odds fifty to one or shorter. The upper bound on N is 24 so that you could outlay as much as $12,144 but if you did outlay this amount you would be covering all possibilities and you would win something back. So $12,144 is an upper bound on the possible loss but not the least upper bound. It is a simple matter to compute the profit or loss in each year of the past twenty years based on the $MCT50$ scheme, and then do some simple stats to compute the mean, the median and the standard deviation. The values are $2044 for the mean, -$971 for the median, and $7280 for the standard deviation. The positive value for the mean indicates an overall win but the negative value for the median indicates that in most years there will be a loss. This sort of information is easy to visualize in a frequency histogram of the data.

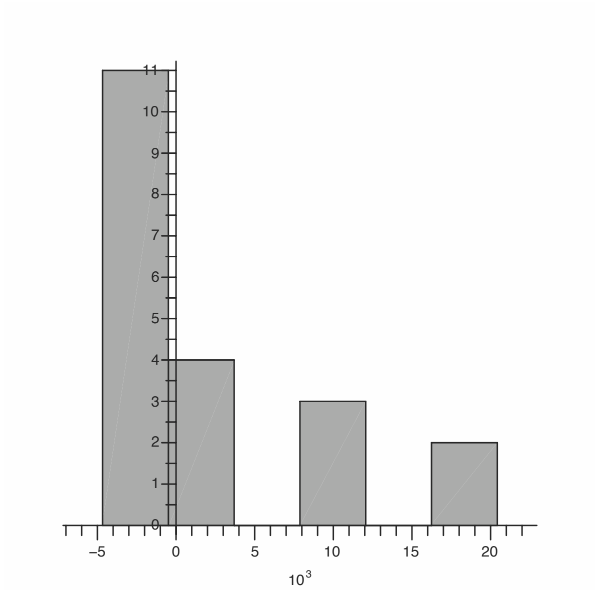

Figure 2: The figure shows the histogram for the number of times that the net return has been in a given range based on Back All Combinations with Odds of 50/1 or Less in the Melbourne Cup Trifecta $(MCT50)$ over the past twenty years.

The central problem in using this data to project forward risk is that we need to make assumptions about how likely it is that future trends will be similar to past trends. The horse betting scheme is not special in this sense. Can we be sure that past growth in the share market will continue or that house prices will continue to rise? If we do assume that the future will be the same as the past, at least in a statistical sense, then we can make a risk measurement called Value at Risk. This lets you know the probability of making a profit less than a given amount in any given year. This is readily obtained from the Empirical Cumulative Distribution Function shown in Figure 3.

The vertical axis represents probability and the horizontal axis represents the annual net profit (if positive) or loss (if negative). From this plot we can deduce that the $MCT50$ scheme has a sixty-five percent probability of returning a loss in any given year. There is a ten percent probability that the loss will be more than $4,000 (capped to a maximum loss of approximately $5,000) and there is a ten percent probability that there will be a profit of more than $10,000 (capped to a maximum of approximately $20,000).

A different measure of risk is one that is related to the robustness of the scheme. One way to estimate this sort of risk is through sensitivity analysis. How much would our profit or loss change if we changed the threshold cut-off from 50/1 to some other cut-off? We showed results of this sort of analysis earlier. The profit/loss based on past data was very sensitive to changes in this threshold, although some comfort might be taken in the fact that a net profit would have ensued with a cut-off of 60/1.

In recent years the TAB has introduced flexi betting. You no longer need to outlay

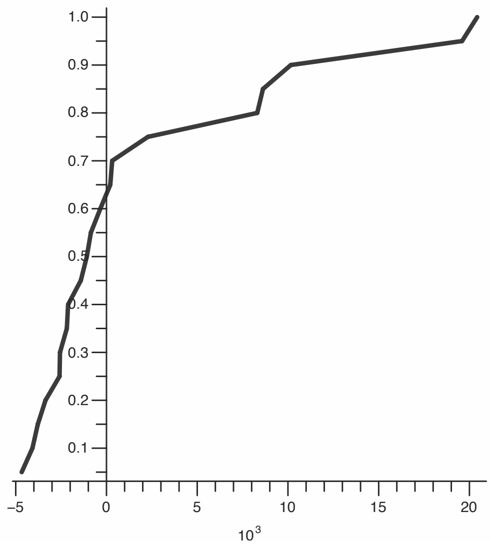

Figure 3: The figure shows the Empirical Cumulative Distribution Function for the net returns based on Back All Combinations with Odds of 50/1 or Less in the Melbourne Cup Trifecta over the past twenty years.

$1 for each combination in a trifecta. You can outlay a fraction of this and if you win then you win the same fraction of the winning payout. This means that you don't need to outlay a few thousand dollars to test the $MCT50$ strategy you can outlay any fraction of this, which raises another issue. If you had a total of $X dollars that you were willing to bankroll on the $MTC50$ scheme how much of this should you outlay? You don't want to put it all out there in the first year because you remember that there is roughly a two in three chance that you will lose in any given year. Even worse, if you started the $MCT50$ in 1992 you would have suffered four consecutive years of losses so you need to keep some of your funds in reserve. The optimal fraction of your current bankroll that you should wager is given by the mathematical formula called the Kelly Bet. The fraction is

$$f = \frac{bp - q}{b}$$

where p is the probability of winning, $q = 1 - p$ is the probability of losing and b is the odds that you receive on the bet if you win. Kelly's formula is famous and widely used in betting (including investing). A popular account is given in the book *Fortune's Formula: The untold story of the scientific betting system that beat casinos and Wall Street* by William Poundstone (Hill and Wang, 2005). To apply this to the $MTC50$ scheme we estimate $p = 0.35$ with $q = 0.65$. To compute b we refer to the tabulated data and compute the ratio of amount won to amount outlaid in each year that the scheme returned a win (seven of the past twenty years). This yields the result $b = 3.84$ and then you find $f = 0.18$.

The above rudimentary anaylsis does not promote this or any other betting strategy, whether it is based on horses or shares. It is always possible to analyze past history and find a strategy that would have worked better for that set of outcomes; but the past performance of a betting strategy does not necessarily correlate with future performance. Here is a simpler scheme that would have returned positive over the past twenty years. Cover all horses for a win except those horses with odds longer than 17/1. For a total outlay of $161 you would have recovered $189. The strategy of covering almost all possibilities shifts the uncertainty in picking a winner to the uncertainity in picking the cut-off. With historical data there is no uncertainity in picking the cut-off to maximize the profit. This is a very simple optimization problem.

A footnote to this analysis: The 2009 Melbourne Cup had 14 horses with closing odds at 50/1 or shorter. The first three place-getters and their closing odds were Shocking (10/1), Crime Scene (40/1) and Mourilyan (28/1). The trifecta paid $8694. Another nice win if you do the maths.

A Simple Derivation of the Kelly Bet

The formula for the Kelly Bet can be derived using simple calculus.

Let $C(n)$ denote the current amount available after n bets, let b denote the payout odds for a win, let p denote the probability for a win, let q denote the probability for a loss, and let f denote the fixed fraction of the current amount to outlay in each bet.

First note that if the last bet was a win then

$$C(n) = C(n-1) + bfC(n-1) = (1+bf)C(n-1)$$

and if the last bet was a loss then

$$C(n) = C(n-1) - fC(n-1) = (1-f)C(n-1).$$

It follows that in a sequence of n bets with n_w wins and n_l losses

$$C(n) = (1+bf)^{n_w}\,(1-f)^{n_l}\,C(0).$$

We wish to find the fraction f to maximize $C(n)$. Thus compute

$$\begin{aligned}\frac{dC(n)}{df} &= n_w b(1+bf)^{nw-1}(1-f)^{n_l}C(0) - n_l(1+bf)^{n_w}(1-f)^{n_l-1}C(0)\\ &= (1+bf)^{n_w-1}(1-f)^{n_l-1}C(0)\,(n_w b(1-f) - n_l(1+bf))\end{aligned}$$

and then $\frac{dC(n)}{df} = 0$ when

$$f = \frac{n_w b - n_l}{b(n_w + n_l)}.$$

This can be simplified by noting $n_w + n_l = n$, so that

$$f = \frac{b\frac{n_w}{n} - \frac{n_l}{n}}{b}.$$

Finally we expect that if n is sufficently large then $p = \frac{n_w}{n}$ and $q = \frac{n_l}{n}$ and we obtain the result

$$f^\star = \frac{bp - q}{b}.$$

It is a simple matter to check that the second derivative of $C(n)$ with respect to f is negative at $f = f^\star$.