The Journal of Gambling Business and Economics (2007) **1**, 13–29

THE OPTIMALITY AND STATISTICAL DETECTION OF PRICE RIGGING IN BETTING MARKETS

Adi Schnytzer† and Yuval Shilony‡

Department of Economics, Bar-Ilan University, Ramat Gan, Israel 52900

Final version received August 2006.

The purpose of this paper is to determine empirically whether or not there is systematic price rigging in three Australian betting markets: Horse, harness and greyhound racing. We present a simple model which shows the conditions under which it is optimal for insiders to rig prices by deliberate underperformance in some races. We then show how an empirical analysis of the relationship between win and place probabilities in conjunction with observed patterns of betting behavior, may be used to establish the presence of price rigging. It is shown that there is no significant systematic price rigging in these markets.

I. INTRODUCTION

The purpose of this paper is to determine empirically whether or not there is systematic price rigging in three Australian betting markets: horse, harness and greyhound racing. We present a simple model which shows the conditions under which it is optimal for insiders to rig prices by deliberate underperformance in some races. We then show how an empirical analysis of the relationship between win and place probabilities in conjunction with observed patterns of betting behavior, may be used to establish the presence of price rigging. It is shown that there is no significant systematic price rigging in these markets.

There has been recent research into corruption in other spheres,[1] but, to the best of our knowledge, this is the first to test for *systematic* corruption in the racing industry. That there is a perception of cheating is well-known. As we write, the British riders Robert Winston, Luke Fletcher, Robbie Fitzpatrick and Fran Ferris have been informed that they will be charged with corruption.[2] In an unrelated affair, Kieren Fallon has been charged with race fixing.[3] And such cases have been observed in all countries where animals race and betting is permitted on the outcome. In this paper, we attempt to determine whether cheating is the exception or the rule in the Australian case.[4]

Our method rests on the isolation of a group of horses and dogs which may be deemed as candidates for deliberate underperformance. The purpose of such underperformance today is to get better odds on the animal in its next race. In terms of the betting market, animals may be divided into the following categories:

---

†Email: schnyta@mail.biu.ac.il
‡Email: shilony@mail.biu.ac.il

1. Those which are plunged to win in the current race. In other words, those whose odds have shortened in the win betting market.
2. Those which are plunged to place[5] in the current race. In other words, those whose odds have shortened in the place betting market.
3. Those animals which have never been plunged.
4. Those animals who are plunged for either the win or place in other races, but not in today's race.

To the extent that an animal's connections have an incentive to rig prices in their favor by deliberate underperformance, it is unlikely that they would engage in such behavior *systematically* while, at the same time, betting conspicuously on the animal to perform well. We therefore exclude animals in categories 1 and 2 from consideration. Similarly, animals which are *never* plunged do not provide connections with large profits via betting and thus are, likewise, not evident candidates for deliberate underperformance. This leaves us with category 4. The fact that these animals are sometimes plunged indicates that their connections seek to gain from betting. This, according to the model presented in the next section, may provide them with an incentive for occasional deliberate underperformance. To the extent that this is the case, we would expect it to happen when the animals are not plunged. It is thus the animals in this group which are our candidates for being, in Australian racing parlance, "not on the job" (NOTJ).

In order to test for deliberate underperformance, we run a series of regressions designed to compare the performance of NOTJ animals with members of the other three groups. In section III we show that, not only don't these animals systematically under perform relative to the animals in other categories, they always *outperform* significantly the animals that are never plunged.

II. THE MODEL

Consider a cohort of n new racing horses. Their racing life spans over a number of seasons, where they compete against each other, and their owners wish to maximize profits over their lives. The quality of the horses is represented by their true winning probabilities, $p_1, \ldots, p_n$, each of which is known only to its respective owner, $\sum_{i=1}^{n} p_i = 1, \quad p_i \geq 0, \quad i = 1, \ldots, n$. In the theory and the empirical part below we make use of the probabilities of each horse to gain 2nd and 3rd place. In order to calculate such probabilities, in particular from the winning probabilities p_i, one needs a probabilistic model of the race. For every such model there is an exact relationship between the win and place probabilities. See appendix 1 for an example of such a model and the implied relationship. In general the probability of a horse getting second or third as a function of the probability of winning has the shape of an inverted U. A hopeless horse has a tiny chance of getting second or third and so also has a super star that nearly always wins. Alas, we don't know the actual model of the races because it is part of the question we pose. *Inter alia*, a rigged race has a different model from a fair one.

The profits come from prize money and from betting. The price of betting on horse i is determined in the betting market by its winning record which can be manipulated by its owner by not always trying to win. Suppose in each racing season there are k races where the same n horses always compete. We look now at a season as the constituent game that is repeated over the life of the horses. A strategy for owner i in this game is a k-tuple $\left((f_i^1, s_i^1, W_i^1, T_i^1), \ \ldots \ \ldots, (f_i^k, s_i^k, W_i^k, T_i^k)\right)$

$$\text{where } f_i^j = \begin{bmatrix} 1 & i \text{ tries to win race } j \\ 0 & \text{otherwise} \end{bmatrix} \quad \text{and}$$

$$s_i^j = \begin{bmatrix} 1 & i \text{ tries to come second or third, but not first, in race j} \\ 0 & i \text{ tries to come fourth or worse in race j} \end{bmatrix}.$$

Note that $0 \le f_i^j + s_i^j \le 1$ for every i and j. If $f_i^j + s_i^j = 0$ the horse is made to run out of a place, if possible. W_i^j is the amount the owner of horse i bets on its horse for the win only in race j and T_i^j is the amount he bets "each way" in that race. Thus, in total he spends $W_i^j + 2T_i^j$ in race j, where $W_i^j + T_i^j$ has been bet in total for the win and T_i^j has been bet for the place.

Suppose that in race j the effort vector is $\left((f_1^j, s_1^j), (f_2^j, s_2^j), \ \ldots \ \ldots, (f_n^j, s_n^j)\right)$. This has, of course, strong implications for the winning probability of each horse i. This probability falls to zero for a horse that does not try and thereby raises that of the other horses. (See Appendix 1 for an example). Denote by $\tilde{p}_i^j$, $\tilde{\tilde{p}}_i^j$, and $\tilde{\tilde{\tilde{p}}}_i^j$ the probabilities of horse i to win race j, to run second and to run third, respectively, given the effort vector in that race.

The probabilities that horse i arrive second and third in race j are developed in the appendix. The place probability of horse i in race j, donated $\dot{p}_i^j$, which is the probability to come either first, second or third is the sum of the three i.e. $\dot{p}_i^j = \tilde{p}_i^j + \tilde{\tilde{p}}_i^j + \tilde{\tilde{\tilde{p}}}_i^j$

While p_i is private information, we assume that the actual expected rates of success of horse i over the season given the owners' strategies, i.e. $\frac{1}{k}\sum_{j=1}^{k} \tilde{p}_i^j$ for the win and, $\frac{1}{k}\sum_{j=1}^{k} \dot{p}_i^j$ to run a place, are known to the betting public. Therefore the former would be the price of the win bet in every race of the season i.e. the price of the contingent claim offering \$1 if horse i wins and zero otherwise, while the price of the place bet would be a quarter of that. This is because, in the Australian market, for all races where number of starters is at least eight[6], the place odds offered by bookmakers are 1/4 the win odds.

Deliberate underperformance also involves a risk for the owner. The races are videoed, closely monitored and investigated if a suspicion arises. Assume S is the expected penalty for the "small" offence of any i in race j of $f_i^j = 0$, $s_i^j = 1$. That is, S is the product of the penalty times the probability of being caught not trying to win race j but trying to run a place. Similarly let L

be the expected penalty for the "large" offence of any i in race j of $f_i^j = 0$, $s_i^j = 0$. In practice, the penalties for the two offences are identical, the guilty party being barred from the racecourse for some period and so losing any income associated with training, riding or driving and on-course betting for the relevant period, but the probabilities are different.

On the basis of these assumptions we can write the expected profit of owner i, which he/she attempts to maximize if risk-neutral, as

$$\Pi_i = \sum_{j=1}^{k}\left[\tilde{p}_i^j\left(\Pr ize + \frac{W_i^j + T_i^j}{(1/k)\sum_{l=1}^{k}\tilde{p}_i^l}\right) + \dot{\tilde{p}}_i^j\frac{W_i^j + 2T_i^j}{(4/k)\sum_{l=1}^{k}\tilde{p}_i^l} - \left(W_i^j + 2T_i^j\right)\right]$$

$$- S\sum_{l=1}^{k}\left(1 - f_i^l\right)s_i^l - L\sum_{l=1}^{k}\left(1 - f_i^l - s_i^l\right)$$

The n owners of the horses play between them a game the structure of which was laid out heretofore. The strategy for each owner i is how strongly to perform i.e. whether to attempt a win ($f_i^j = 1$, $s_i^j = 0$), run 2nd or 3rd ($f_i^j = 0$, $s_i^j = 1$) or "nowhere" ($f_i^j = 0$, $s_i^j = 0$) and how much to bet for the win and each way in each race. The full details of the equilibrium of the game are not of interest to us, but the following aspects are of importance in determining whether or not prices are actually rigged.

Claim 1: All owners trying to win every race is not a Nash equilibrium of the game if at least one owner i can bet in each race an amount more than the expected prize plus the expected penalty S i.e. $W_i + 2T_i > p_i\text{Prize} + S$.

Proof: Consider the point of everybody always trying to win. All races are identical so we can suppress the race index. Owner i's profit is then:

$$\Pi i = k\left[p_i\left(\Pr\text{ize} + (W_i + T_i)/p_i\right) + \dot{p}_i\frac{W_i + 2T_i}{4p_i} - (W_i + 2T)\right]$$

$$= k\left[p_i \Pr\text{ize} + \frac{\dot{p}_i}{4p_i}\mathrm{W_i} + \left(\frac{\dot{p}_i}{2p_i} - 1\right)T_i\right]$$

where $\dot{p}_i$ is calculated according to a fairly run race. Owner i can gain by deviating to the strategy of, say, trying to win only a fraction α of the races, randomly selected, but still trying to run a place in the others. By doing so he will win on average only a fraction α of the races and the betting price of his horse to win will cut to α times its value. It will not change, though, his place probability because probability is shifted from first to second place but the sum is intact. Consequently he will participate in αk of the races in the betting

TABLE 1

ESTIMATING THE WINNING PROBABILITIES (CONDITIONAL LOGIT)

	Greys		Horses		Trots	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
wprb_c	10.08177*	0.835169	8.901113*	1.68866	10.38679*	0.804567
wprb_15	−0.05748	0.625725	2.525649	1.473831	−0.13031	0.634898
pprb_c	13.57284*	2.39944	23.10703*	2.960943	23.71622*	1.92919
pprb_15	0.330737	0.890258	−4.20214*	2.108371	−2.7268*	1.121537
$wprb\_15^2$	0.760397	1.100999	−3.21202	3.166071	0.733238	1.028067
$pprb\_15^2$	−2.2371	2.301526	7.214373	6.185939	4.91481	2.92945
$wprb\_c^2$	−9.63494*	1.514528	−7.85037*	3.375522	−9.39499*	1.322413
$pprb\_c^2$	−38.3075*	7.304914	−70.4299*	9.787542	−69.2923*	6.028706
NOTJ	0.306879*	0.04637	0.193204*	0.048056	0.269164*	0.039215
Barrier	−0.00231	0.006988	0.014134*	0.004596	0.003255	0.004808
Weight			−0.01704	0.012106		
Log Likelihood	−6358.02		−7366.766		−8820.6765	
No. Obs.	28792		41787		49561	

(Coefficients significant at better than 5% are marked with a *.)

where wpbr_c is the probability equivalent of the tote win odds at the close of betting,[9]

ppbr_c is the probability equivalent of the tote place odds at the close of betting,

wpbr_15 is the probability equivalent of the tote win odds 15 minutes from the close of betting,

ppbr_15 is the probability equivalent of the tote place odds 15 minutes from the close of betting,

NOTJ is a dummy variable equal to 1 when the animal has been plunged in any race in the sample other than the present race and 0 otherwise,

Barrier is the barrier drawn by the animal in this race

and Weight is the weight carried by the horse in the race.

and risk S for the $(1 - \alpha)k$ races of not trying. The new profit is

$$\Pi_i(\alpha) = \alpha k\left[p_i\left(\text{Pr}ize + \frac{W_i + T_i}{\alpha p_i}\right) + \dot{p}_i\frac{W_i + 2T_i}{4\alpha p_i} - (W_i + 2T_i)\right] - (1-\alpha)kS$$

$$= k\left[\alpha p_i \text{Pr}ize + \left[1 + \frac{\dot{p}_i}{4p_i} - \alpha\right]W_i + \left[1 + \frac{\dot{p}_i}{2p_i} - 2\alpha\right]T_i\right] - k(1-\alpha)S$$

and $\Pi i(\alpha) > \Pi_i = \Pi_i(1)$ as well as $\frac{\partial \Pi_i(\alpha)}{\partial \alpha} < 0$ *if* $W_i + 2T_i > p_i\text{Prize} + S$.

Note that:

1. The condition in the claim, $W_i + 2T_i > p_i$Prize $+ S$ may be hard to meet, not only in lucrative races, but also if the insider has few sources of income away from the track.
2. Profit increases in W_i and in T_i so betting will be increased to its limit.
3. Deliberate underperformance can be prevented by either limiting the sums insiders may bet $W_i + 2T_i$ or by increasing the prize or the expected penalty.

Corollary: The game has no equilibrium in pure strategies.

Proof: The deviation in the proof of the claim must be randomized or else the betting public and the other owners will take advantage of it by betting on the horse when it tries to win. The same type of deviation is advantageous to every owner from every point of pure strat egies by increasing the return on a win and saving on the cost of betting.

III. RESULTS

In order to test the hypothesis that there is systematic deliberate underperformance at Australian race tracks, we need to show that NOTJ animals systematically underperform other categories of animals. We first derive the best unbiased estimate of winning probabilities that we can get. In order to do this, we run conditional logit regressions[7] to explain winning as a simple polynomial function of ALL prices and other information at our disposal. The results are shown in Table 1. We then use these "objective" winning probabilities to explain the performance of the four categories of animals noted in the Introduction, viz; those never plunged, those plunged for the win (and, perhaps also the place) in the current race, those plunged for the place (but not for the win) in the current race and, those that are plunged some other time, but not today and may thus not be on the job. As explained in the appendix, we use tote odds throughout in spite of the fact that insiders certainly prefer betting with bookies. In doing so, we rely on Schnytzer and Shilony

TABLE 2

WINNING PERFORMANCE OF ANIMAL CATEGORIES (CONDITIONAL LOGIT)

	Greys		Horses		Trots	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
win	17.13305	1.156762	19.22411	0.792799	18.29505	0.870272
win^2	−23.1431	2.548625	−31.3436	2.223816	−22.3958	1.822271
WPLUN	1.019669	0.201843	0.897221	0.34825	1.402432	0.17686
iw1	−9.12113	1.75783	−9.11869	2.786341	−10.7018	1.432788
iw2	15.92236	3.363757	20.13368	5.299094	15.32515	2.533903
PPLUN	0.476501	0.158465	0.82346	0.339763	0.644873	0.156909
ip1	−3.75901	1.532798	−7.2682	3.055989	−3.78303	1.391417
ip2	6.08787	3.08951	13.10791	6.159852	4.930984	2.618116
NOTJ	0.628823	0.12436	0.540424	0.125876	0.633941	0.085772
in1	−6.69595	1.383136	−6.64205	1.499319	−5.86209	1.001735
in2	12.84466	2.872715	15.53242	3.702625	9.472396	2.009417
Log Likelihood	−6393.3292		−7438.7992		−8936.2264	
No. Obs.	28792		41787		49561	

(All coefficients are significant at better than 5%)

where win is the winning probability of the animal as predicted from the appropriate regression in Table 1,

WPLUN is a dummy variable equal to 1 when the animal has been plunged for the win in the current race, 0 otherwise,[10]

iw1 = win*WPLUN,

iw2 = win^2*WPLUN,

PPLUN is a dummy variable equal to 1 when the animal has been plunged for the place in the current race, 0 otherwise,

ip1 = win*PPLUN,

ip2 = win^2*PPLUN,

in1 = win*NOTJ

and in2 = win^2*NOTJ.

TABLE 3

2nd PLACE PERFORMANCE OF ANIMAL CATEGORIES (CONDITIONAL LOGIT)

	Greys		Horses		Trots	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
win	13.94686*	1.18303	17.48696*	0.813349	18.45268*	0.944242
win^2	−24.8731*	2.976755	−34.307*	2.495261	−33.7868*	2.487736
WPLUN	0.73543*	0.200297	1.012543*	0.381238	1.378501*	0.177437
iw1	−7.86341*	1.874667	−9.08593*	3.215258	−13.2886*	1.575383
iw2	15.44818*	3.907268	17.1935*	6.432177	25.52112*	3.235856
PPLUN	0.474644*	0.142744	0.350142	0.373246	0.54093*	0.14957
ip1	−3.64382*	1.557473	−5.49558	3.494675	−5.55007*	1.473103
ip2	5.796798	3.400394	13.46737	7.325567	9.261069*	2.961949
NOTJ	0.545298*	0.109276	0.739281*	0.113903	0.767073*	0.071777
in1	−4.9979*	1.4252	−5.92267*	1.569746	−8.37203*	1.070354
in2	10.1616*	3.370294	10.55971*	4.382479	18.50454*	2.677386
Log Likelihood	−7200.1464		−8067.3917		−10532.555	
No. Obs.	28792		41787		49561	

(Coefficients significant at better than 5% are marked with a *.)

TABLE 4

3rd PLACE PERFORMANCE OF ANIMAL CATEGORIES (CONDITIONAL LOGIT)

	Greys		Horses		Trots	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
win	9.3168*	1.242517	13.84843*	0.827829	12.83561*	0.911213
win^2	−21.4447*	3.456913	−31.1122*	2.718934	−23.9877*	2.422955
WPLUN	0.775207*	0.202966	1.05035*	0.39462	0.756992*	0.195444
iw1	−8.37022*	1.9884	−9.89053*	3.389464	−8.1027*	1.72962
iw2	18.69824*	4.418566	20.41902*	6.897514	14.25374*	3.466919
PPLUN	0.250421	0.139468	0.601341	0.359563	0.875405*	0.142819
ip1	−2.07115	1.603288	−7.54269*	3.446001	−8.70241*	1.527699
ip2	4.910649	3.606017	19.70249*	7.296461	15.32773*	3.212778
NOTJ	0.48992*	0.102909	0.558811*	0.11173	0.682109*	0.064414
in1	−5.26079*	1.492952	−5.04405*	1.629187	−6.17261*	1.065253
in2	13.14489*	3.858621	11.215*	4.753141	10.57815*	2.716102
Log Likelihood	−7406.4783		−8390.7012		−11062.147	
No. Obs.	28792		41787		49561	

(Coefficients significant at better than 5% are marked with a *.)

TABLE 5

OUT OF THE MONEY PERFORMANCE OF ANIMAL CATEGORIES (LOGIT)

	Greys		Horses		Trots	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
win	−17.1082	0.812409	−21.272	0.53264	−20.5823	0.609351
win^2	24.75806	1.991192	35.32756	1.661574	27.57	1.439126
WPLUN	−1.12168	0.148197	−1.21162	0.272484	−1.51829	0.134621
Iw1	11.63551	1.415378	11.60243	2.313097	13.92465	1.178018
Iw2	−23.3494	2.909375	−24.249	4.602323	−22.6947	2.220336
PPLUN	−0.45661	0.100042	−0.79473	0.25	−0.83067	0.103104
Ip1	3.627291	1.160794	9.048583	2.444103	7.599791	1.092169
Ip2	−6.66192	2.638423	−20.2056	5.320585	−12.4113	2.308207
NOTJ	−0.61088	0.072671	−0.65218	0.07349	−0.71063	0.044998
In1	6.429518	1.001281	6.321904	1.042235	7.266962	0.715898
In2	−13.8049	2.306787	−13.4198	2.857508	−13.1294	1.617152
constant	2.117483	0.0536178	2.589708	0.0314151	2.518214	0.0347988
Log Likelihood	−17374.316		−20718.08		−25915.908	
No. Obs.	28792		41787		49561	

(All coefficients are significant at better than 5%)

(1995), where we showed that plunges with bookmakers are transferred to tote prices via outsiders who observe the inside trading.

As to our a priori expectations, the theoretical model presented in the previous section provides a condition under which deliberate underperformance is likely to occur. Whether or not this condition is met is an empirical question about which we had no strong priors. However, it does seem reasonable to argue that, to the extent that there is deliberate underperformance and consequent price rigging, it would manifest itself differently in the three different types of racing. Thus, in our model we distinguish between, on the one hand, underperformance of a strong kind, whereby the aim is to run out of the money and, on the other, a weak form, where the idea is not to win but still run a place. It is difficult to imagine how such a distinction might be managed *systematically* in the case of greyhound racing, where there is no human intervention once the dogs are on the track, there are no legal constraints once the race is underway, and underperformance is largely a matter of feeding (and possibly drug) routines. By contrast, in harness racing, not only do drivers often own or train their charges, they are also permitted to bet and horses drawn on the second row at the start can easily be run into pockets from which they can never again be expected to see the light of day. Thus, fine tuning is evidently possible, provided that monitoring and punishment regimes are sub-optimal. The case of thoroughbred racing is somewhere in between: there is human intervention on the track but jockeys generally do not own or train their horses and are never permitted to bet. They therefore, at least in principle, provide at least something of a principal-agent problem to the connections who wish to indulge in price rigging.

Consider first, as an aside, one of the interesting features of Table 1. The market prices for these races – neither for the win nor the place, nor early in the betting nor late - take no account of NOTJ animals, in spite of their significant winning proclivities. This type of weak form inefficiency has never, to our knowledge, been pointed out and warrants further study.

Our strategy for comparing the performance of the animals in different categories is very straight forward. We run conditional logit regressions to explain, respectively, winning, running second and running third, and we run simple logit regressions[8] to explain running out of the money. In all cases, these are run as functions of our Table 1 estimates of winning probability and its square (in accordance with the theoretical results presented in the appendix) as well as dummies for plunged and NOTJ animals and interaction terms. The results are presented in Tables 2 through 5.

Perhaps the most remarkable thing about the results is how little they differ between the different types of racing. All coefficient signs are identical and magnitudes (albeit not easy to interpret in these kinds of regressions) are very similar. The only difference at all rests in the fact that the dummies for place plunges and their interaction terms are not uniformly significant or insignificant in the regressions for a place. Further, these differences make intuitive sense: Just as deliberate underperformance would be more difficult to fine tune in the case of greyhounds than in harness racing, it is more difficult to predict that that a dog will run a place than that a pacer or trotter would finish in the money. As to evidence of deliberate underperformance, there simply is none. Even after allowing for the quality of the animal, as the regressions do, NOTJ animals significantly outperform animals who are never plunged *in every respect and every regression*! Furthermore, their performance is more or less on a par with those animals plunged for a place, the coefficients relating to NOTJ and PPLUN being very similar. They are outperformed only by animals plunged for a win and that is not surprising since the latter are the direct beneficiaries of inside information.

IV. CONCLUSIONS

The is a wealth of anecdotal evidence to the effect that cheating occurs in animal racing industries world-wide. Whether or not such corruption is truly as common as media reports sometimes imply has not hitherto been considered from an econometric viewpoint. Our purpose has been to demonstrate whether or not there is systematic price rigging in three Australian betting markets: horse, harness and greyhound racing. We have presented a simple model which derives the conditions under which it is optimal for insiders to manipulate prices via deliberate under-performance in some races. We have then shown how an empirical analysis of the relationship between win and place probabilities in conjunction with observed patterns of betting behavior, may be used to establish the presence of price rigging. We find no evidence that there is significant systematic price rigging in these markets.

V. APPENDIX

An Example Model of a Race.

Suppose the animals run independently of each other as if on different tracks or in separate lanes. Each animal i has a distribution $F_i(t)$ where t is the difference between the time it takes animal i to complete the given distance and the best time it takes the best animal of the trade, not necessarily running in any given race. We assume this distribution is exponential, i.e., $F_i(t) = 1 - e^{-\lambda_i t}$ whose density is $f_i(t) = \lambda_i e^{-\lambda_i t}$. It is not an attractive assumption but it facilitates much the exposition of the example.[11] The mean time of animal i is $1/\lambda_i$. The probability that animal i wins the race is:

$$p_i = \int_0^\infty \prod_{j \neq i} (1 - F_j(t)) f_i(t) dt = \int_0^\infty \prod_{j \neq i} e^{-\lambda_j t} \lambda_i e^{-\lambda_i t} dt$$

$$= \lambda_i \int_0^\infty e^{-t \sum_{j=1}^n \lambda_j} dt = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \tag{A1}$$

Let $p = (p_1, p_2, \ \ldots \ \ldots, p_n)$ be the true winning probabilities of the n animals in a race. Now suppose deliberate underperformance and that the animals' effort vector in race j is $((f_1^j, s_1^j), (f_2^j, s_2^j), \ \ldots \ \ldots, (f_n^j, s_n^j))$. A animal that does not try has a zero chance to win and for the trying ones it is the conditional probability given that the non-trying animals are out of the race provided at least one animal is trying. If none is trying the probabilities are undefined and one may define them as *1/n* for all animals. We assume such a case never arises. So

$$\text{(A2)} \qquad \tilde{p}_i^j = \frac{f_i^j p_i}{\sum_{m=1}^n f_m^j p_m} = \frac{f_i^j \lambda_i}{\sum_{m=1}^n f_m^j \lambda_m}$$

where the second equality follows for our special case of exponential distributions.

What is the "place" probability of animal *i,* i.e. that it runs first, second or third in the race? We shall derive it from p in steps.

First, what is the probability that animal i runs second? For that to happen another animal k has to win the race and i has to win among the remaining $n - 1$ animals. That event is exactly identical to the event that i wins given that k does not participate in the race. The conditional probability of that, given the absence (or victory) of k is, applying Bays' rule, $\frac{p_i}{1-p_k} = \frac{\lambda_i}{\sum_{j \neq k} \lambda_j}$. The numerator is the probability of the intersection of the two events "i wins" and "j does not win". The denominator is the probability of the event "j does not win". Considering the probable identities the winning animal k may assume we get the probability for animal i to run second

$$\ddot{p}_i = \int_0^\infty \sum_{k \neq i} \left[F_k(t) \prod_{j \neq i,k} (1 - F_j(t)) \right] f_i(t) dt$$

$$= \int_0^\infty \sum_{k \neq i} \left[\prod_{j \neq i,k} (1 - e^{-\lambda_k t}) e^{-\lambda_j t} \lambda_i e^{-\lambda_i t} dt \right]$$

$$= \lambda_i \int_0^\infty \sum_{k \neq i} \left[e^{-\sum_{j \neq k} \lambda_j t} - e^{\sum_{j=1}^n \lambda_j t} \right] = \lambda_i \sum_{k \neq i} \left[\frac{1}{\sum_{j \neq k} \lambda_j} - \frac{1}{\sum_{j=1}^n \lambda_j} \right]$$

$$\text{(A3)} \qquad = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j} \sum_{k \neq i} \frac{\lambda_k}{\sum_{j \neq k} \lambda_j} = p_i \sum_{k \neq i} \frac{p_k}{1 - p_k}$$

What is the probability that animal i runs third? That event happens when some animal j wins the race, some animal k wins among the remaining animals and runs second and, finally, animal i wins among the rest and runs third. Applying the reasoning as in (A3) we obtain now

$$\dddot{p}_i = \sum_{j\neq i} p_j \sum_{k\neq i,j} \frac{p_k}{1-p_j}\frac{p_i}{1-p_j-p_k}$$

(A4) $$= p_i \sum_{j\neq i} \frac{p_j}{1-p_j} \sum_{k\neq i,j} \frac{p_k}{1-p_j-p_k}$$

Now we are equipped to calculate the place probability of animal *i,* i.e. that it runs first, second or third. Since these three events are mutually exclusive the place probability is the sum of its win probability plus the probability of running second plus the probability of running third, i.e.

(A5) $$place_i = p_i + \ddot{p}_i + \dddot{p}_i$$

which sum up to three over the race.

Now enter deliberate underperformance, (A3) and (A4) are modified respectively to

$$\tilde{\ddot{p}}_i^j = \sum_{l\neq i}\left(\frac{f_l^j p_l}{\sum_{m=1}^n f_m^j p_m}\frac{(f_i^j + s_i^j)p_i}{\sum_{m\neq l}(f_m^j + s_m^j)p_m}\right) = \frac{(f_i^j + s_i^j)p_i}{\sum_{m=1}^n f_m^j p_m}\sum_{l\neq i}\frac{f_l^j p_l}{\sum_{m\neq l}(f_m^j + s_m^j)p_m}$$

(A3′)

$$\tilde{\dddot{p}}_i^j = \sum_{l\neq i}\left(\frac{f_l^j p_l}{\sum_{m=1}^n f_m^j p_m}\sum_{k\neq l,i}\frac{(f_k^j + s_k^j)p_k}{\sum_{m\neq l}(f_m^j + s_m^j)p_m}\frac{(f_i^j + s_i^j)p_i}{\sum_{m\neq l,k}(f_m^j + s_m^j)p_m}\right)$$

(A4′) $$= \frac{(f_i^j + s_i^j)p_i}{\sum_{m=1}^n f_m^j p_m}\sum_{l\neq i}\left(\frac{f_l^j p_l}{\sum_{m\neq l}(f_m^j + s_m^j)p_m}\sum_{k\neq l,i}\frac{(f_k^j + s_k^j)p_k}{\sum_{m\neq l,k}(f_m^j + s_m^j)p_m}\right)$$

2.The Inverted U-shaped Place-only Curve

As animal *i* in a given race gets better, i.e. its probability p_i of winning the race increases at the expense of all other animals in the race, its probability of running second or third changes accordingly. For small values of p_i, when it increases, the chances of second or third place grow as well but eventually for large enough p_i they decline since the chances of winning the race must be in conflict with those of running second or third.

One can see this pattern even in a simple example of four runners with probabilities 0.1, 0.2, 0.4 and 0.5. Assuming exponential distributions, their respective probabilities of running second or third are 0.3488, 0.5587, 0.5702 and 0.5221. The best animal has a lower probability of running a place than the inferior animals.

To show formally this property of the probability of running second or third one must first specify how the probability of each of the other animals changes when p_i changes. We shall now assume that they change proportionally, i.e. that the ratio of the probabilities of any pair of two other animals remains constant. One simple way to do it is to change only λ_i.

Let $p_1, \ldots \ldots, p_n$ be the initial probabilities of the n animals and $x_1, \ldots \ldots, x_n$ the new ones. Define the new probabilities as the following n − 1 functions of x_i:

$$x_j(x_i) = p_j \frac{1 - x_i}{1 - p_i} \quad \text{for all } j \neq i$$

This functions pass through $(p_1, \ldots, p_n)$, they sum up to $1 - x_i$ as required and are the only ones that satisfy our assumption of proportionality. Denoting by P_i the probability of i running second or third, we get by applying the functions:

$$\begin{aligned} P_i(x_i) &= x_i \sum_{j \neq i} \left[\frac{p_j(1 - x_i)}{1 - p_i - p_j(1 - x_i)} \left(1 + \sum_{k \neq i,j} \frac{p_k \frac{1-x_i}{1-p_i}}{1 - p_j \frac{1-x_i}{1-p_i} - p_k \frac{1-x_i}{1-p_i}} \right) \right] \\ &= x_i(1 - x_i) \sum_{j \neq i} \left[\frac{p_j}{1 - p_i - p_j(1 - x_i)} \left(1 + \sum_{k \neq i,j} \frac{p_k}{\frac{1-p_i}{1-x_i} - p_j - p_k} \right) \right] \end{aligned}$$

(A6)

To ascertain the inverted U shape differentiate

$$\begin{aligned} \frac{\partial}{\partial x_i} P_i(x_i) &= (1 - 2x_i) \sum_{j \neq i} \left[\frac{p_j}{1-p_i-p_j(1-x_i)} \left(1 + \sum_{k \neq i,j} \frac{p_k(1-x_i)}{1-p_i-(1-x_i)(p_j+p_k)} \right) \right] \\ &\quad - x_i(1 - x_i) \sum_{j \neq i} \left[\begin{array}{l} \frac{p_j^2}{[1-p_i-p_j(1-x_i)]^2} \left(1 + \sum_{k \neq i,j} \frac{p_k(1-x_i)}{1-p_i-(1-x_i)(p_j+p_k)} \right) + \\ \frac{p_j}{1-p_i-p_j(1-x_i)} \sum_{k \neq i,j} \frac{p_k(1-x_i)}{[1-p_i-(1-x_i)(p_j+p_k)]^2} \end{array} \right] \end{aligned}$$

(A7)

Since the two squared brackets are positive, for small enough x_i the P_i curve is rising while for $x_i > 0.5$ it is declining.•

3. The Data

The data set was compiled from pre- and post- greyhound, horse and harness race postings onto the Victoria region TABCORP web site (www.tabcorp.com.au) and comprises 3599 greyhound races with 28792 dogs (May 1998 through April 1999), 3569 thoroughbred horse races with 41787 horses (May 1998 through May 1999) and 4983 standardbred horse races with 49561 horses (January 1998 through April 1999). Race data were obtained from the remote site using a command driven http browser (LYNX) and PERL operating on the university's UNIX network. Starting between 4 and 7 hours before the start of the day's races a list of available races were downloaded and start times for harness races extracted. Starting from 70 minutes prior to the posted start time each race's information was then saved from the remote site in Victoria onto the local host at Bar Ilan University. Each file was updated periodically so that any new information between 2 hours to the final post-race results could usually be obtained. Due to the dynamic nature of the data acquisition, disruptions in internet access caused by overload of either the local (Bar Ilan) or remote site (Victoria) resulted in loss of information to the data set. This loss was without any discernible pattern and therefore should have no systemic influence on the analysis.

During the tabulation of the data from individual races downloaded into the final data set, updates were expressed according to their occurrence relative to the actual rather than the posted start time for each race for posting times less than 30 minutes before the listed start time. This adjustment was necessary since in 20.4% of the races the actual start time of the race was up to 10 minutes later than the listed start time displayed on the Victoria TABCORP web page. We assume that bettors on-course adjust their betting behavior to delays in the start of a race.

NOTES

1. See, for example, Duggan and Levitt (2002), Jacob and Levitt (2003), Mauro (1995) and Porter and Zona (1993).
2. See The Independent, 17 August, 2006. http://sport.independent.co.uk/general/article1218382.ece.
3. See http://www.buzzle.com/editorials/7-27-2006-103739.asp.
4. The most famous incidence of corruption in the Australian racing industry in recent years concerns the bookmaker, Robbie Waterhouse. See, for example, http://www.smh.com.au/articles/2003/04/21/1050777213845.html.
5. Here and throughout this paper, we use the Australian meaning of the word "place"; namely, that the animal runs first, second or third. To run a place is thus equivalent to what is known in US racing as "showing".
6. In the empirical part of the paper, we consider only races with 8 or more starters.
7. The form of regression was developed by Mc Fadden (1973) and first used in the context of explaining winning probabilities by Figlewski (1979). It has subsequently become the most popular regression for this purpose.
8. See, for example, Ramanathan (2002).
9. The data were obtained as prospective win and place tote payouts and thus include breakage of 10 cents. Since rounding causes a larger percentage error for small odds than for large odds, we follow Griffith (1949) and assume continuous payouts rather than payouts falling into 10 cent intervals. The easiest way

to accomplish this is to assume that for a sufficiently large number of observations, the mean payout before rounding will fall half way between the actual payout and the next payout up. In practice, this amounts to adding 5 cents to the projected payouts before calculating probability equivalents.

10. An animal is said to have been plunged (either for the win or the place) if he probability equivalent of the prospective tote odds rises by no less than 5 percent between 15 minutes before the race and the close of betting.
11. For similar, but alternative, approaches to the modeling of outcome probabilities in horse racing, see Harville (1973), Henery (1981) and Lo and Bacon-Schone (1994).

REFERENCES

M Duggan and S D Levitt 'Winning isn't everything: corruption in sumo wrestling' *American Economic Review* (2002) 95(5) 1594–1605.

S Figlewski 'Subjective information and market efficiency in a betting market' *Journal of political Economy* (1979) 87(1) 75–88.

D A Harville 'Assigning probabilities to the outcomes of multi-entry competitions' *Journal of the American Statistical Association* (1973) 68 312–316.

R J Henery 'Permutation probabilities as models for horse races' *Journal of the Royal Statistical Society B* (1981) 43 86–91.

B A Jacob and S D Levitt 'Rotten apples: and investigation of the prevalence and predictors of teacher cheating' *Quarterly Journal of Economics* (2003) 118(3) 843–887.

V S Y Lo and J A Bacon-Schone 'Comparison between two models for predicting ordering probabilities in multiple-entry competitions' *The Statistician* (1994) 43(2) 317–327.

D Mc Fadden 'Conditional logit analysis of qualitative choice behavior' in Paul Zaremba (ed) *Frontiers in Econometrics* (New York, Academic Press, 1973).

P Mauro 'Corruption and growth' *Quarterly Journal of Economics* (1995) 110(2) 681–711.

R H Porter and J D Zona 'Detection of bid rigging in procurement auctions' *Journal of political Economy* (1993) 101(3) 518–538.

R Ramanathan *Introductory Econometrics with Applications* 5th Edition (South-Western, 2002), ch. 12.

A Schnytzer and Y Shilony 'Inside-information in a betting market' *Economic Journal* (1995) 105 963–971.