The Favourite-Longshot Bias Across Different Classes of Horse Races: Evidence From the 2003 UK Flat Turf Season.

Hui-Fai Shing[1]

Royal Holloway University

Draft: 27th March 2005

ABSTRACT

The purpose of this paper is to re-examine the Favourite-Longshot for bookmaker first show and starting odds data taken from the 2003 UK flat season by comparing normalized implied win-probabilities against the realised percentage of winners in the same manner as Cain, Law and Peel (2003). Despite the new information mines available to bettors and the birth of betting exchanges, the nature of the Favourite-Longshot Bias for flat racing remains the same. In accordance with the Efficient Markets Hypothesis and models which show the effect of information revelation on prices, implied starting win-probabilities seem to predict the outcome of races better than the win-probabilities implied by the first show of odds. The secondary objective of this paper is to see how this bias varies across different classes of races (essentially based on the prize money they offer), if the top class races are subject to fewer insiders, the Shin (1991, 1992 and 1993) model would predict that these races will exhibit a less apparent bias, the results support this claim for opening odds data. In addition to this, starting odds are comparatively accurate across the spectrum of classes, this coupled with the finding that finding that opening odds demonstrate a more apparent Favourite-Longshot Bias for lower classes of races would suggest a greater presence of insiders in lower class races who tend to place their bets earlier.

Keywords: Favourite-Longshot Bias, Betting, Shin Model, Market Efficiency.
JEL Classification Numbers: D81, G14, L83

---

[1] The author wishes to thank Urs Haegler for his helpful comments and getting this project started, and Jonathan Wadsworth and Alexander Koch for their help and perseverance. Any errors are of course my own.
Correspondence address: Hui-Fai Shing, Economics Dept. Royal Holloway University, Egham, Surrey, TW20 0EX h.shing@rhul.ac.uk

I. INTRODUCTION

The Favourite-Longshot Bias in betting is the anomaly whereby the relative returns on bets on favourites are superior to longshots. This paper has two inter-related objectives; firstly I wish to examine the degree of the Favourite-Longshot bias in the UK horse racing betting market for the 2003 flat season. My second objective is to see how this Favourite-Longshot bias varies between different classes of horse races; one might intuitively expect that high class races will offer more evidence of 'efficiency' than their lower grade counterparts. Based on the Shin (1991, 1992) model and the assumption that more insiders operate in lower class races, one would expect the prices in lower class races to be less efficient; or more precisely, exhibit a larger Favourite-Longshot Bias.

Market efficiency is based on the notion that "all prices reflect all relevant information" Fama (1970). Fama defines three types of market efficiency,

1) Weak form efficiency. No investor can earn excess returns by developing trading rules based on historical price or return information. In other words, the information in past prices or returns is not useful or relevant in achieving excess returns.
2) Semistrong-form efficiency. No investor can earn excess returns from trading rules based on any publicly available information, e.g. annual reports of companies etc.
3) Strong-form efficiency. No investor can earn excess returns using any information, whether publicly available or not.

What does betting market efficiency imply? One can say that:

Weak form efficiency implies that one cannot make excess returns by looking at prices (odds) and price movements.
Semi-Strong form efficiency implies that in addition to the above, bettors cannot make excess returns by studying the form of horses/jockeys/trends etc.
Strong form efficiency implies that in addition to the above, bettors with insider information cannot make excess returns.

If bettors are risk neutral[2] and there was competition amongst bookmakers, an implication of strong-form efficiency is that the expected returns on all bets, be it on favourites or longshots should be the same (Thaler and Ziemba, (1988))[3], intuitively, if any bets, for example bets on longshots, yielded more negative returns, a risk neutral bookmaker would wish to offer more of these bets in order to earn greater profits; he would be undeterred by the potentially large payouts. However, economists have found a well-known anomaly where the relative return (although usually negative) from backing favourites is greater (less negative) compared with backing longshots[4]; this is known as the Favourite-Longshot bias.

In what follows, I shall be looking at new data from the 2003 UK flat season. Betting markets in the UK have changed significantly since the last major studies[5] have been conducted. Since the 6th of October 2001, off track betting taxes in the UK have been abolished, and currently, bookmakers must pay a tax based on their gross profits. Previously, bettors who were not betting with bookmakers at the racetrack had to either pay a 9% tax on their stake or a 9% tax on their winnings,[6] (though some off track bookmakers moved some of their operations off-shore so some bettors were able to bet tax-free, and this was one of the primary reasons for the British government to take this initiative). One could argue that one of the implications of having betting duty was that it discouraged betting on favourites, a 9% addition to the stake would be relatively large compared with the winnings of a horse with short odds (and it would be pointless betting in this manner on a horse whose odds are less than 1/10 as the tax paid will be less than the return), and if the bettor chose to pay the tax on his winnings as opposed to his stake, then it would arguably make his relatively small returns (larger in expected terms compared with backing a longshot) even smaller. However the effect of this is debateable since starting prices (the odds at which bettors at the betting shop are paid out on if they do

[2] A person is said to be risk neutral if he is indifferent among all alternatives with the same expected value, if a horse has true win-probability p, a fair bet, where the expected value of the net winnings is zero, would have (fair) odds of $[(1/p)-1]$, if these odds were set, then on average, the bookmakers' profits will be zero and the bettors' profits will also be zero. If we denote state 0 as the state where the bettor does not bet, state 1 where the bettor bets and loses, and state 2 where the bettor bets and wins, then for a fair bet, $E(W) = pW_1 + (1-p)W_2 = W_0$, where the subscripts denote the states, W denotes wealth, W_1 denotes wealth in state 1, i.e. $W_0 - Stake$, and $W_2 = W0 - Stake + Stake/p$. In other words, for the risk neutral bettor or bookmaker, $U(W_0) = pU(W_1) + (1-p)U(W_2)$.

[3] This was also suggested by Snyder (1978), however Vannebo (1980) argues that in the face of "positive skewness in the distribution of the outcome of a bet", biased expected returns are a "result inherent in rational behaviour towards risk, and will be incurred even when the wagering market is efficient".

[4] This phenomenon was first observed by psychologists Griffith (1949) and McGlothlin (1956). We could also interpret this as the probability implied by the odds to the too low for favourites, and too high for longshots in relation to the frequency of winning outcomes (Cain, Law and Peel (2003)).

[5] For example the one by Cain, Law and Peel (2003), which compared data from the 1978 season to the 1987 season.

[6] Please refer to http://www.hm-treasury.gov.uk/budget/budget_2001/press_notices/bud_bud01_pressbet.cfm for more details.

not take the odds offered by the bookmaker) and the odds available to all bettors once the betting market has formed are determined by the on track bookmakers[7]; though off track bookmakers have a large influence via a presence (having a stand) at the race track or having representatives who place bets with other on track bookmakers in order to contract the odds of certain horses which the off-track bookmakers may have large liabilities on.

A more important change is the birth of betting exchanges and also the punter has witnessed the advent of information mines such as the internet and (at least for the 2003 season) the satellite/cable specialist horse racing television channel *attheraces*. One might intuitively expect that because of the removal of the off/on track tax distortion and more information being available to the bettor will result in markets being more efficient and hence, the degree of the Favourite-Longshot bias will have been narrowed.

In the next section, I shall look at the UK betting market, and focus in particular on why the 2003 flat season betting market was so different to the seasons preceding this that other economists have looked at when investigating market efficiency of betting markets. I shall also talk about the dataset. Section III, will look at betting market efficiency in general, going into the Favourite-Longshot bias with a little more detail. Finally, the results of this investigation are presented; the win-probabilities implied by the odds and the actual relative win-frequencies are compared, this will be conducted on the probabilities implied by the odds when the market opened and also the odds available at the start of the race, and shall provide clues as to whether starting odds are more accurate than opening odds. 95% confidence intervals for the theoretical number of winners based on the (normal approximation to the) binomial distribution are also calculated and compared with the observed number of winners. This is followed by a comparison of the results with the results from the 1978 and 1987 flat seasons. The secondary objective of this paper will then be put under the spotlight by examining the Favourite-Longshot bias across different classes of races.

II. THE UK BETTING MARKET

[7] Who bettors always bet with tax free.

This section will provide a whistle-stop tour of the UK flat racing scene and the betting market in the UK. The domestic flat racing season runs from late March to early November[8], there are usually two to three (though there can be more than half a dozen) meetings a day, with more during the middle of the season when there are evening meetings). Each meeting has a minimum of six races (there are nine races on some occasions), and it is usually always the same jockeys, trainers, and owners that are involved. In the 2003 flat season there were seven classes of races in the UK (A-G), and these are fundamentally classified by the prize money they offer, hence the quality of horses they attract (clearly there are restrictions that apply, and horses running in class G races are normally run under the condition that they are available to be sold afterwards), a table elaborating on this is included in the appendix. There are dozens of racecourses around the UK, each one of them physically unique, some racecourses 'specialize' in staging higher class races and there are racecourses that specialize in staging lower class races. There are also a variety of races, most races are handicap races, where the weight each horse carries is based on their official rating, which is based on their past performances, thus, theoretically each horse has an equal chance of winning if they performed to their rating[9] and in other types of races, the weights carried depend on the horse's record, their age and their gender.

It is useful to begin by explaining some betting terminology. Firstly, odds correspond to the return of the bet if the horse wins; a winning £10 bet on a horse at odds of 5-1 will pay £60, a £50 net profit. Prices correspond to the cost of purchasing a state contingent claim that pays £1 in the event that horse *i* wins the race. For example, if a horse's odds are 5-1, then the price of the state contingent claim paying in the event that the horse wins is £1/6[10]. Prices are often misinterpreted as probabilities, however, the prices do not sum up to one unless the bookmaker is offering a book of fair bets which the bookmaker will not do as he needs to cover his costs and try to make a profit. The bettor has two main ways to place a bet, and recently (in the last two to three years) a third way has been born. The returns from betting are not directly subject to taxation anymore:

[8] There were 217 days of racing during the flat season in 2003.

[9] Though of course, the odds for horses participating in these races vary substantially because horses which are in good form or are 'ahead of the handicapper' will be favoured because they are carrying less weight relative to their rivals, and of course the conditions of the racetrack and jockey bookings play an important role.

[10] $\Pr ice_i = 1/(Odds_i + 1)$

1) Bookmaker odds – bettors choose to take the odds offered by the bookmaker, or have bets settled on the starting price (SP) of the horse (it would be more accurate if starting prices were actually called *starting odds*!). SPs are determined by the on track bookmakers, there are around a dozen bookmaker stalls at the small meetings, and around 50 at the large meetings. Off track bookmakers (who usually are also represented on track), who would include the High Street and telephone betting firms can influence the odds by sending money to the track. Clearly the SP is the same for every bookmaker.

How are these odds determined? Many factors are involved. every evening before the race, 'tissues'; the odds forecast from some 'gurus' are published[11], these odds are formulated from the win probabilities[12] estimated by the gurus, and then a margin/mark-up is added so that the bookmaker is not offering a book of fair bets, the nature of this margin is what this paper shall look into. Empirically and theoretically, this mark-up should be larger for longshots because the potential losses for laying bets to insiders for outsiders is much larger; hence in bookmaker betting markets, there is a Favourite-Longshot Bias. Clearly the prices in a book should sum to more than one and is called the *over-round*, bookmaker over-rounds in the UK typically vary from 110% (for small fields) to 130% for larger fields[13]. In the absence of bettors with inside information, the over-round indicates the theoretical profit the booker should make, for example, if the prices sum up to (i.e. the over-round is) 120%, the bookmaker will on average expect to pay out £100 for every £120 he takes in, his profits are (20/120) 16.7%. To give a simple example, imagine a race with 3 horses that have an equal chance of winning, a bookmaker offering fair bets would set the odds of each horse at 2/1, a bookmaker does not offer fair bets because he needs to cover his costs and attempt to earn a profit and would give less generous odds, of say 7/4 on each horse. Odds of 7/4 correspond to a price of 4/11, in turn leading to an over-round of 12/11 or 109.09% implying a theoretical profit margin of 8.3%. To see why this is so, if there were 3 bettors, each with £10 and each of them bet on a

[11] See McCririck (1991) pp. 105 for a more detailed explanation; this book provides a very useful introduction to the world of betting.

[12] The (fair) odds would just be: $Odds_i = (1/\Pr obability_i) - 1$

[13] From a random sample of 30 races from that data I obtained, with small-regular size fields of 6-10 runners, the mean over-round was 114.5% and for another random draw of 30 races with 15 or more runners, the mean was 130%.

different horse, the bookmaker receives £30 of bets but he would need to pay out £27.50, £10 plus the lucky bettor's winnings of £17.50. He makes a profit of £2.50, which is 8.3% of £30. With insiders and competition amongst bookmakers, the over-round reflects the odds consistent with them making zero profits; there is a chance that an insider places a bet with them.

Bookmakers can look at these tissue odds and the behaviour of betting exchanges (which shall be explained later) before they post/offer their odds. On the subject of when 'on track' bookmakers offer their odds, this usually occurs about 15-20 minutes before the start of the race, these odds are referred to as the 'first show', or opening odds. For some races, usually about 4-6 a day, odds are available in the morning (these are called *morning line* odds, but these are not too competitive because fewer firms offer these odds and they leave themselves as the first port of call for insiders, who may deem these bets as good value). Obviously, for big races, one can obtain odds well in advance[14]. After the market opens, trading takes place and bookmakers will adjust the odds of the horses such that they do not have too much liability on too few horses, bookmakers need to lengthen the odds of horses which are not being backed in order to 'balance their books' so and/or to make a larger expected profit.

2) Tote (Pari-mutuel) betting – this is where all bets are placed into a pool, which is shared (after deductions) proportionately by the holders of winning tickets. This is the system used by most of the rest of the world, where there is a Tote monopoly. The pools typically open in the morning. In the UK, these pools are very small compared with the rest of the world, and it is usually fairly easy to affect prices. Curiously, tote dividends are usually better for longshots, and worse for hot favourites, and there are many peculiar circumstances where it would be wiser to use the Tote/take bookmaker odds, for example one should not back grey horses at Tote odds[15]. Tote odds are solely determined by the volume of bets, the volume of correct bets, and the percentage of money taken out of the pool as taxes and rounding the odds down to the nearest 10p.

[14] Once again, see McCririck (1991) pp. 77-9 to learn about the pros and cons of *ante-post* betting.
[15] McCririck (1991) pp. 48-9 talks about some more of these anomalies.

Since Tote dividends are only known after the race, it is argued that[16] insiders are not likely to bet on the Tote, but rather take the odds that bookmakers have on their boards to lock in any supernormal rate of return that is on offer. Gabriel and Marsden (1990) found that Tote odds are generally higher than bookmaker odds (for 4 odds categories) supporting this proposition. Tote 'odds' are expressed as gross returns, i.e. bookmaker odds of 5-1 are equivalent to a tote dividend of £6.00.

3) Betting exchanges – these are websites where the public (and bookmakers) can back or lay horses, i.e ordinary punters can be the bookmaker; the exchanges have been in existence for a few years and are rapidly growing. Winning punters have to pay around a 5% deduction on these bets. Because of the very competitive nature of betting exchanges, the odds offered to backers are typically better (more so for longshots), the profit margins (for backers and layers) are close to zero. The odds for each race are determined by a double auction where all participants (the account holders) know how much money is available for backing and laying at all odds; thus odds are determined purely by demand (backers) and supply (layers) all the liabilities are held by the account holders (the bettors), and the bettor must have the money in their account to make/accept a bet.

Betting exchanges open a book for a race about 24 hours before the race, and their odds are a good guide to the final odds posted by bookmakers, and of course the movements of odds in betting exchanges can prove to be very informative; these provide the insider with the first opportunity to back (lay) the horse for which they posses positive (negative) information. The amount of money available on them is typically smaller, and it is not too difficult for agents to affect prices. Betting exchanges characterize a low-friction market where many tests of market efficiency and human behaviour can be investigated.

III. BETTING MARKET EFFICIENCY

When investigating weak form efficiency economists have looked at the odds of runners and looked at the return if you back horses at those odds (see Hausch, Lo and Ziemba (1994) for a

[16] Gabriel and Marsden (1990).

selection of these papers and in particular Snyder (1978) where the results from early empirical work are shown). Generally there are no positive returns (after taxes/deductions), but as was drawn upon earlier, there is a Favourite-Longshot Bias; where returns of favourites (lower variance bets) are relatively better than those of longshots (higher variance bets), in other words, the incidence of the over-round falls on the longshots. This is antipodal to capital markets, where the investor would generally expect to have a higher return in order to compensate for holding a more risky asset.

The explanations suggested for this bias include[17] risk attitudes of bettors, Weitzman (1965) and Ali (1977) estimated the utility function of the representative bettor and found that they were risk loving, the bettor prefers an uncertain prospect with a particular expected value to the prospect of obtaining the same expected value with certainty, their utility functions are convex. This means that in the absence of the bias (i.e. the expected returns on favourites and longshots was the same), bettors will prefer to bet on longshots, thus contracting their odds and leading to the Favourite-Longshot Bias.

An explanation for fixed-odds bookmaker betting markets is the existence of bettors with inside information, in particular Shin (1991, 1992) constructed a model of optimal odds determination with a fixed-odds risk neutral bookmaker facing a fraction (z, which can be estimated, see Shin (1993)) of 'insiders' who always bet on the right horse. Shin found for profit maximizing odds, the mark-up between the available odds and the fair odds increased as the fair odds increased, i.e. a Favourite-Longshot bias existed and this was subject to there being more insider trading when a longshot was subject to positive insider information, which goes with intuition. The intuition behind this can be given by the fact that a £1 bet on a favourite exposes the bookmaker to a smaller potential loss than a £1 bet on a horse at 33/1, "the bookmaker will require a greater risk premium to insure himself against the possibility of inside information on a longshot"[18] in other words, there will be a greater markup (the difference between the fair price/probability and the actual price) on the prices of longshots; this is achieved by reducing their odds thus leading to a Favourite-Longshot Bias; the incidence of the over-round falls predominately on longshots. Cain, Law and Peel (2001) find that this measure of insider trading is closely related to another measure suggested by Crafts

[17] This paragraph and the following paragraph runs parallel to Hausch, Lo and Ziemba (1994), pp. 252-3
[18] Fingleton and Waldron (1996)

(1985) which is based on significant contractions in odds of a particular horse on the day of the race (a plunger).

Another consequence of the Shin model is that the over-round should be positively correlated with the number of competitors in the event; "*ceteris paribus*, a larger field of competitors leads to higher odds against any individual winning the event and thus higher winnings for insiders. In these, circumstances bookmakers need enhanced margins to protect themselves". Cain, Law and Peel (2003)) examined the Favourite-Longshot bias, the Shin measure of insider trading and over-rounds for a variety of sports and found compelling evidence in favour of the Shin model, sports with more insider trading generally seemed to exhibit a larger Favourite-Longshot Bias, this is linked closely to the secondary objective of this paper because one may expect different levels of insider activity prevailing in different classes of horse races in the UK; if lower class races are subject to more insider trading, the odds present in these markets will exhibit a larger Favourite-Longshot Bias.

Other authors have gone on to develop the Shin model, Fingleton & Waldron (1996) relaxed Shin's assumption that the insiders know which horse is going to win to them just having positive information, as well as allowing the bookmakers to make a profit. They solve the model for a profit maximising bookmaker, and an infinitely risk averse bookmaker (one who equalizes liabilities across all horses) and still find a Favourite-Longshot Bias. Also, they found that bookmaker behaviour in Ireland adheres to the infinitely risk averse form.

In other tests of weak form efficiency, authors have also looked at odds movements of horses, and formed filter rules, i.e. back horses if their odds change significantly. Crafts (1985) looked at odds movements of UK races and found that bets settled at SP were not profitable, but bets struck at forecasted odds (i.e. the tissue odds and I must stress are not guaranteed to be available to bettors, though one may expect to be able to obtain comparable odds on the betting exchanges) did have a positive return; this finding is not surprising if SPs are weak form efficient. In pursuit of profitable wagering rules, Crafts (1994) proposed some wagering rules and found that it was possible to earn a rate of profit of 0.558 at SP for a rule based on horses which have not started for that year; this rate of profit was 2.619 for bettors who placed their bets at forecast prices.

For pari-mutuel betting, Lo (1994) ran a simple logit model where pari-mutuel win bet fractions (implied win-probabilities) are fitted against the binary dependent variable of whether the horse wins or not, this was done for morning line (tissue odds) and odds recorded at intervals before the off. Data was obtained from 712 races in Atlantic City in 1978. As well as obtaining log likelihood values to determine the goodness of fit of the various sets of odds, a parameter β is also calculated which illustrates the strength of the Favourite-Longshot Bias, if β was greater than 1 a Favourite-Longshot Bias was present in the odds data. Two important findings are reported, firstly morning line odds (β = 1.5655, standard error = 0.086) demonstrate a larger Favourite-Longshot Bias than the starting odds (β = 1.1023, standard error = 0.059) and it also seem to fall for the last 6 minutes before the start of the race. Also the log-likelihood of the odds increased as the start of the race drew closer. In this article, Cox's (1961, 1962) test for non-nested hypothesis was employed to show that the probabilities implied by the starting odds were superior to the implied probabilities at any other time. It is difficult to know how much to draw from these findings into a bookmaker odds framework because in a pari-mutuel market one would expect informed bets to be placed very late in order to see whether the bet will potentially offer enough value because the dividends are only known after the race has started, whereas in bookmaker based markets, informed betters are likely to take the odds offered by bookmakers at the first opportunity if they believe that the bookmaker is offering a good bet[19]. This is backed up by a finding of Law and Peel (2002) for 971 races in the UK where they found that z had mean value of 4.4% for opening odds and 2.7% for starting odds and fell for 902 of theses 971 races, this is consistent with the hypothesis that insiders tend to bet early.

To test for Semi-strong form efficiency, the most obvious method would be to see whether one can earn excess returns by studying the form of the horses; an intuitive way is to see whether horse racing pundits earn superior returns to ordinary bettors; this was done by Snyder (1978), where it was found that the returns from pundits (track handicappers) demonstrated a different bias to that of ordinary bettors[20]. By running a multinomial logit model, some authors[21] have found that excess returns can be made.

[19] This is different to conventional wisdom where the informed should place their bets during periods with higher trade volumes so as to hide beneath the uninformed traders. The argument for bookmaker betting is that insiders like to lock in the high returns on offer early.

[20] Snyder originally claimed that this was a test of strong form efficiency because the track handicapper could possess inside information, however, the track handicapper's predictions were published (and readily available) before the race.

[21] See Bolton and Chapman (1986)

To test for Strong form efficiency, one would need to know when a horse being backed is subject to real inside information, and the returns of insiders would have to be examined. Bettors can call premium rate tipping lines, but when does this inside information (if it is indeed inside information) become public knowledge? Quite interestingly, Law & Peel (2002) find that "significant positive betting returns are achieved when shortening odds are accompanied by a rise in the Shin measure (of insider trading, z)" and when shortening odds are accompanied by a fall in z, returns are negative, suggesting herd behaviour. This is evidence against Strong-form betting market efficiency.

III. THE DATASET AND THE 2003 SEASON

The analysis in this paper shall be based on a new dataset covering the 2003 flat season in the UK, the data was downloaded from the Racing Post website and includes the opening (first show) and starting odds of the horses in every race, the class of the race (and whether or not it was a handicap), the date and location of the race and the going. The dataset covers all the races, however, omissions include all the 'novelty' races (such as Amateur/Ladies/Apprentice races, the Shergar Cup, etc) which apart from the Shergar Cup are very low class races. Also omitted are races in which there were withdrawn horses in between the formation of the market and the start of the race; if a horse is withdrawn after the market was set up and before the start of the race, the probabilities of the other horses will be affected, and there is not usually enough time for a new market to be set up. If this occurs, a 'Rule 4'; where a fixed proportion (based on the withdrawn horse's odds) on all bets would be deducted, will be enforced[22]. If a new market is formed, then only bets struck at fixed odds would be affected by the Rule 4. Races with new markets formed were excluded from the dataset as usually these were of short duration and had few movements. It was also not clear for many races whether certain horses were withdrawn after the market formed, in these situations, I usually included the observations in the dataset, unless it was absolutely clear that the horse was withdrawn after the market was set up. Other observations removed from the dataset include races which were subject to a boycott by trainers in response to low prize money. The dataset covers 3,573 races in which there were a total of 38,954 entries.

[22] Whether or not a Rule 4 will need to be enforced, and the size of the Rule 4 depends on the odds of the horse which is withdrawn. To give the reader a rough idea, if a horse with odds of over 14/1 is withdrawn, then no Rule 4 will be enforced, and if a horse of odds 2/1 is withdrawn, then a deduction of 30p per pound is enforced.

In the next section, I shall be comparing the probabilities implied by the odds with the observed probabilities (the relative win frequency of the horses of that odds category). The probabilities in the dataset correspond to the respective price divided by the over-round of that particular race so that the probabilities sum up to one; in effect, in addition to saying that it is wrong to compare prices with probabilities (since probabilities sum up to one and prices sum up to over one), I am saying that it is unfair to compare the expected return of a bet at odds of X/1 to another bet of X/1 when the over-rounds are different; one only needs to consider a bet on a horse with a 50% chance of winning with a low margin added to its odds and another bet on another horse also with a 50% chance of winning but with a large margin added to its odds (perhaps because it is in a large field, or the perceived presence of insiders betting on that race). The probabilities of the two horses winning are the same, they will both win the race half the time, but the odds of the second horse are likely to be much lower. The probabilities would imply that bookmakers are adding the same margin to each horse, but in reality this is not true, there is a bigger margin incorporated in the prices of longshots, for evidence of this, one just has to look at the odds prevailing in betting exchanges. A potential problem of using these probabilities is that the probabilities may be distorted, we are taking more percentage points off the probability of favourites compared with longshots. One could of course carry out the same investigation using probabilities conditional on the Shin model; i.e. the margins taken out for longshots will be proportionately larger than for favourites. However these probabilities will remove any Favourite-Longshot Bias and one of the aims of this paper is to study this in some detail.

As touched upon earlier, there are diverse motivations for investigating the 2003 flat season. I believe that the main reason would be the growth of betting exchanges. Betting exchanges should affect market efficiency through two dimensions. Firstly there is a competitive element, if there are a subset of bettors (the bettors with access to the internet or telephone and an account with a betting exchange) with access to more competitive odds, hence bookmakers must offer more competitive odds. One would intuitively expect that the odds of longshots would be affected more since the divergence of the odds between bookmakers and betting exchanges of longshots is larger; though clearly an investigation is needed to say this more concretely.

Secondly and more importantly would be the informational aspect. Betting exchanges provide bookmakers with a wealth of information; bookmakers can observe overnight and early morning movements in the odds offered/taken by betters on the exchanges, instead of bookmakers being the first port of call for bettors with inside information, the exchanges are now the first port of call for any better with inside information who wishes to take advantage of any uninformed bettor's actions based on the uninformed bettors' difference of opinion. One would expect that the birth of betting exchanges would lead to less on-track market moves (plunges) since the exchanges are the first port of call, any market moves should have already taken place before the on-track market is set up. However, this argument could be somewhat overstated as the amounts available on the exchanges are relatively small compared to what a better would stake compared with an on-track bookmaker, and there is also the fact that any layer of a horse on the exchanges would of course offer bets taking into account that the person on the other side is an insider. But overall, one would expect that exchanges should make the on-track market opening odds closer to their starting odds, i.e. there is a smaller chance that opening odds are too far-off their starting odds.

Within the informational dimension, there is also the case of the unscrupulous bettor with negative inside information, who lays a horse on the exchanges. Layers who are sure that a horse will not put in a respectable performance can be pretty sure of winning, however, bettors with positive inside information only know that their bet is very good value, their horse still has to beat the other contenders. Any horse whose odds are drifting on the exchanges will also provide the bookmaker with useful information; this should further steer opening odds closer to starting odds because there should be fewer drifters on the on-track market. However, one could argue that many drifts occur because a horse turned out badly in the parade-ring or moved badly to the post.

IV. RESULTS – WHOLE DATASET

The tables and graphs below are presented in order for us to explore the Favourite-Longshot Bias, these allow us to compare the implied win-probabilities (IP_i), given by,

$$IP_i = \frac{1}{(Odds_i + 1)\beta}$$

where $Odds_i$ correspond to the odds of horse i, and β is the overround for that particular race; i.e. the prices are all scaled down by the same factor in each race so that the sum of the

probabilities in each race is equal to unity, these probabilities are calculated for the opening odds (the first show of odds available at the racetrack) and the starting odds, this allows us to investigate how the accuracy of the implied probabilities changed from the formation of the market to the cessation of the market, where the probabilities should be the most accurate because bettors and bookmakers would be placing bets based on more information such as how the horse looked in the paddock, how it moved to the post as well as any 'enlightening' betting patterns. The observed probabilities (P_i) are based on the percentage of winners with the respective implied probabilities, it is the relative win frequency.

The column titled *error* indicates how much the observed probability (%) exceeds the implied probability. In order to analyse the results more critically, 95% confidence intervals of the theoretical number of winners per implied probability group (g) were calculated, these were based on the number of runners in that group, the probability of the midpoint of that group and then by using the normal approximation to the binomial distribution; more formally, the 95% confidence interval for the theoretical number of winners for a probability group with midpoint p_g is given by:

$$n_g p_g \pm 1.96\sqrt{n_g p_g (1 - p_g)}$$

only for g where $n_g p_g > 5$ and $n_g(1-p_g) > 5$

The tables only show whether the observed number of winners fell outside the 95% confidence interval, a * next to the error value denotes that the number of winners was too low and a ^ denotes that the number of winners was too high, if we have the traditional Favourite-Longshot Bias, then we would expect * for lower probabilities and ^ for higher probabilities. The figures show these results graphically, the implied probabilities (for opening odds and starting odds) are plotted against the observed probabilities, and a 45º line is included to judge the nature of the bias. Once again, if the probabilities based on the way they were calculated were 100% accurate, we would expect the curve to coincide with the 45º line, and if there was the traditional Favourite-Longshot bias, we would expect the curve to lie below the 45º line for low probabilities and then we would expect it to lie over the 45º degree line for favourites.

In addition to the figures and the 95% confidence intervals, a logit model, with a binary variable of whether a horse wins or not is run against the implied opening and starting win-

probabilities, and the pseudo R^2s are compared to corroborate the finding of whether starting implied win-probabilities provide a better fit than opening implied win-probabilities.

Before I begin analyzing the results, it would be useful to distinguish between a favourite and a longshot; what makes a favourite? and what makes a longshot? I would say that it depends on the number of runners in the race. Clearly an *odds-on* horse in a several runner race with all the other horses' odds above 3/1 would be a classified as a favourite, yet I would also define a 6/1 chance in a 20 runner race as a favourite, though the chances of this horse winning are relatively low. In general I would arbitrarily define a favourite as a horse with odds of less than 7/4 (that's a horse with about a 35% chance of winning) and I would say that a longshot would be a horse with odds of 10/1 or greater (less than 10% chance of winning). Cain, Law and Peel (2003) arbitrarily define a favourite as horses with "prices greater than 0.8"[23], this reflects odds of 1/4, I would refer to these horses as '*dead certs*'; in fact there are less than 20 horses with probabilities (normalized prices) of greater than 80% in the dataset.

TABLE 1

Implied and observed probabilities for the whole dataset.

		Opening				Starting			
PrbA Range	Class Midpoint	Runners	WIN%	Error (ovr)		Runners	WIN%	Error (ovr)	
"0-2"	1	4864	0.68	-0.32	*	5732	0.61	-0.39	*
"2-4"	3	8326	2.11	-0.89	*	7885	2.44	-0.56	*
"4-6"	5	6303	4.55	-0.45		6107	4.21	-0.79	*
"6-8"	7	4442	6.64	-0.36		4304	6.48	-0.52	
"8-10"	9	3158	8.49	-0.51		3125	9.18	0.18	
"10-12"	11	2643	11.01	0.01		2538	10.68	-0.32	
"12-14"	13	1752	12.73	-0.27		1727	12.10	-0.90	
"14-16"	15	1413	14.65	-0.35		1452	16.39	1.39	
"16-18"	17	1121	19.62	1.96	^	1087	17.66	0.66	
"18-20"	19	921	20.96	1.96		927	21.14	2.14	
"20-22"	21	677	21.86	0.86		694	23.05	2.05	
"22-24"	23	543	23.76	0.76		551	22.87	-0.13	
"24-26"	25	493	27.38	2.38		427	28.34	3.34	
"26-28"	27	351	31.05	4.05		374	27.81	0.81	
"28-30"	29	296	30.74	1.74		311	35.37	6.37	^
"30-32"	31	262	33.97	2.97		260	33.08	2.08	
"32-34"	33	199	37.69	4.69		206	35.44	2.44	
"34-36"	35	175	38.86	3.86		177	41.24	6.24	
"36-40"	38	283	41.70	3.70		260	41.15	3.15	
"40-45"	42.5	207	46.86	4.36		211	48.34	5.84	
"45-50"	47.5	195	57.44	9.94	^	197	54.82	7.32	^

[23] pp. 265

"50-60"	55	198	58.08	3.08	240	55.83	0.83
"60-80"	70	118	68.64	-1.36	143	66.43	-3.57
"80-100"	90	14	92.86	2.86	19	94.74	4.74

**Indicates that the observed number of winners was lower than the 95% confidence interval of predicted winners based on the number of runners, the implied probability and the normal approximation to the binomial distribution.*
^ Indicates that the observed number of winners was greater than the 95% confidence interval of predicted winners based on the number of runners, the implied probability and the normal approximation to the binomial distribution.

Table and Figure 1 show the results for the whole dataset, i.e. all classes of races, they seem to suggest that the opening and starting win-probabilities implied by the odds are an accurate reflection of the actual outcome. The graph shows a relatively large deviation, a hump, from the 45º line for horses with implied win-probabilities of around 45-50%; this reflects fair odds or around 6/5 to *evens* (1/1). In fact the 95% confidence intervals for the forecasted number of winners are only violated at the very low win-probability end of the spectrum (where the horses win less often than they should); horses with less than 4% of winning probability for opening odds, and horses with less than 6% at the starting odds, and the only other place where the 95% confidence intervals are violated is at the 45-50% win-probability region and for starting win-probabilities of 28-30% (fair odds of in between 9/4-5/2), these horses win more often than they should compared with their implied win-probabilities. For opening win-probabilities, the 95% confidence interval is exceeded at opening win-probabilities of 16-18% (or fair odds of just under 5/1), but this disappears for the implied starting win-probabilities. Here bookmakers seem to have corrected this apparent mis-pricing.

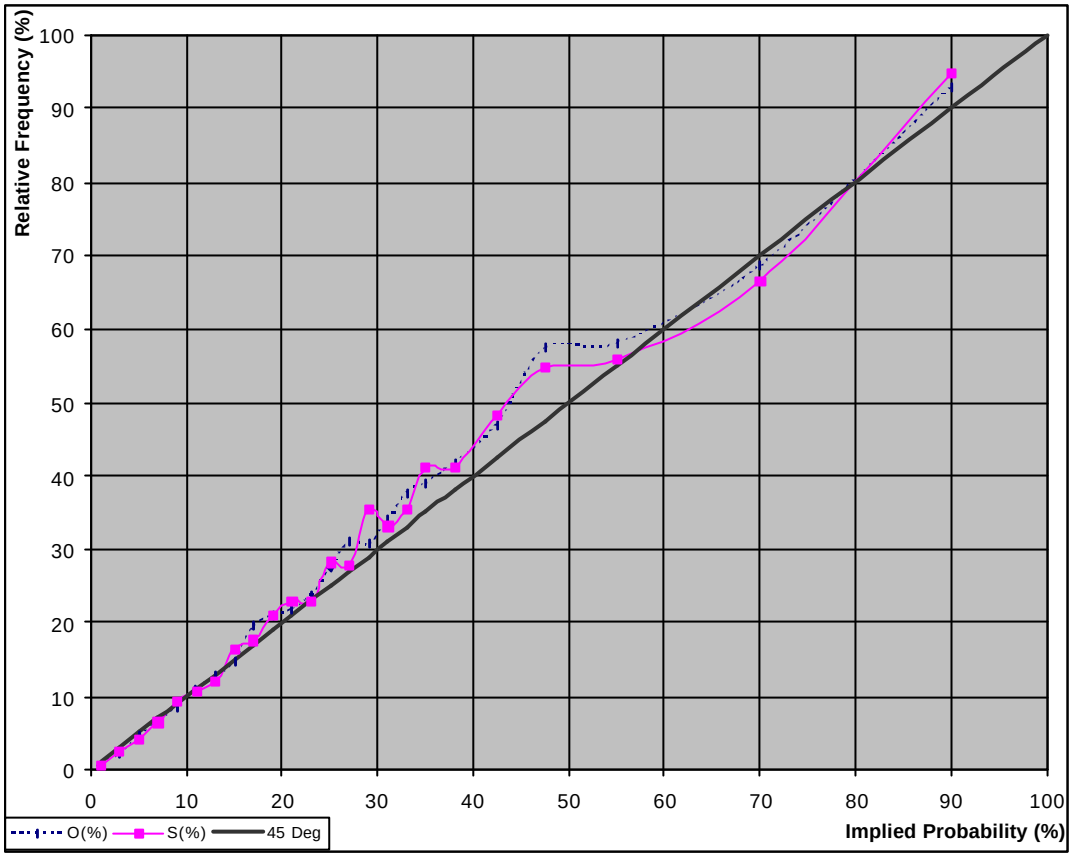

FIGURE 1 - *Plot of implied win probabilities (at opening and starting odds) against the percentage of winners. (38,954 Observations.)*

Even though the probabilities implied by the odds seem to be fairly accurate, the Favourite-Longshot Bias does still seem to persist, though perhaps in a more dampened form; in the next section I shall compare Figure 1 with the same graphs for the 1987 and 1978 flat seasons, this will ensure that our insight is less blinkered and more concrete. Going back to our analysis, on average, the error is always negative (though not necessarily significant) for horses with less than 14% win-probability (fair odds of greater than 6-1) and always positive for horses with win-probabilities greater than 24%[24] (except for horses with implied win probabilities of 60-80%, fair odds of 4/6 to 1/4, and once again this positive error is not significant). One should also note that for higher implied probabilities, there were far less observations compared with the low probability observations, so we would expect these points to be less accurate.

The analysis continues by looking at how the Favourite-Longshot Bias has changed from the formation of the market to the start of the race. One would expect starting odds to be more accurate than opening odds because between the opening and the cessation of the market a

[24] This represents fair odds of about 3/1

vast volume of information flows with respect to each horse's chances, examples of such information would include how the horse looks in the paddock, how the horse gallops to the starting stalls, interviews with relevant players (such as jockeys and trainers), and perhaps more importantly, betting patterns at the track (and on betting exchanges) during the existence of the market; on-track bookmakers pay particular attention to bets struck by well known professional bettors and bettors in general like to back horses which are well supported in the hope that the catalyst of the odds contractions was money from agents with inside information[25]. If the opening win-probabilities were a more accurate reflection of actual outcomes, then the new information has been used in a 'negative way' or bettors are subject to some sort of betting bias, for example, they prefer to bet on longshots no matter what.

At first glance, the nature of the Favourite-Longshot Bias is the same for opening and starting win-probabilities. The win-probabilities of longshots are underestimated, the underestimation is significant from 0-4% implied opening probabilities and from 0-6% for starting probabilities, taken literally, this means that the 'longshot' part of the bias is even more significant for starting probabilities; horses that start at fair odds of just below 20/1 win less often than horses whose fair odds opened at just below 20/1. This can occur for a variety of reasons, if opening odds were a reflection of the true win-probabilities of the horses, then bettors could be too keen on horses with probabilities opening at less than 4% (over 25/1), and backing them so that their odds contract to make their implied probabilities around the 5% mark. Alternatively, there may be a proportion of bettors who know which horses with opening probabilities of 4-6% have a greater chance than that, and bet on them so that the horses left in the 4-6% category are the bad bets. One shouldn't pay too much attention to this as the greater degree for the 'longshot' part of the bias is only greater by two percentage points. Apart from this difference, the patterns are very similar, there is the 'hump' from 25-50% (though as touched upon earlier, it is only significant at 45-50%). The only difference is at implied win-probabilities of 28-30%, where the number of winners exceeds the 95% confidence interval; horses with starting odds of 13/8 (1.625/1) do better than horses with opening odds of 13/8. This could arise from bookmakers lengthening the odds of horses who have odds of around 11/8-6/4 (or horses with a probability of around 30-35%) because bettors, for whatever reason, are not keen to bet on them. These two points aside, the pattern exhibited

[25] McCririck (1992, pp. 51-2) writes "perversely, many punters turn up at a racecourse or betting shop intending to back a horse, see that its price is much shorter than they had expected, and then lump on even more in the belief that because of support it must have a far better chance then they'd though".

is very similar; suggesting that the information flows between the opening and the cessation of the market are not too revealing, or that significant information flows do not occur too often enough to make a significant impact on our results.

To corroborate these findings, the pseudo-R^2 for the logit model of the binary win variable on implied opening win-probabilities was 0.1504 (Log likelihood = -10143.89) compared with 0.1534 (Log likelihood = -10107.75) for implied starting win probabilities. This bodes well with the visual findings and the 95% confidence interval calculations, starting probabilities seem to be more accurate than winning probabilities, though only slightly better. This analysis will be repeated for the results from each class of horserace, but before this, it will be beneficial to look at the pattern of the Favourite-Longshot Bias for other flat seasons.

Figures 2 and 3 correspond to the same diagram as Figure 1 but are for the 1978 and 87 flat seasons respectively and are solely for starting odds. Even though the number of observations is small compared with our dataset, the results are very similar, for very low implied win-probability horses, their observed win-probabilities are consistently lower than their theoretical value, and there is also a hump present, it seems to prevail from implied win-probabilities from 25-55% for the 1978 season and from 20-35% in the 1987 season, this is very similar to the hump from 25-50% which is present for the 2003 season. In addition, just as for the 2003 season, after the hump, the curve converges back to the 45º line and shoots off above it for very high probability horses; however since the number of very hot favourites here must have been less than the number from the 2003 dataset, one should not pay too much attention to these very hot favourites, though it is a relief that the same phenomenon is being observed for the three seasons. Overall, the pattern seems to be very similar, but that 'favourite' part of this bias seems to be much more apparent. It will be interesting to see how the results change when the sample is split according to the class of the race.

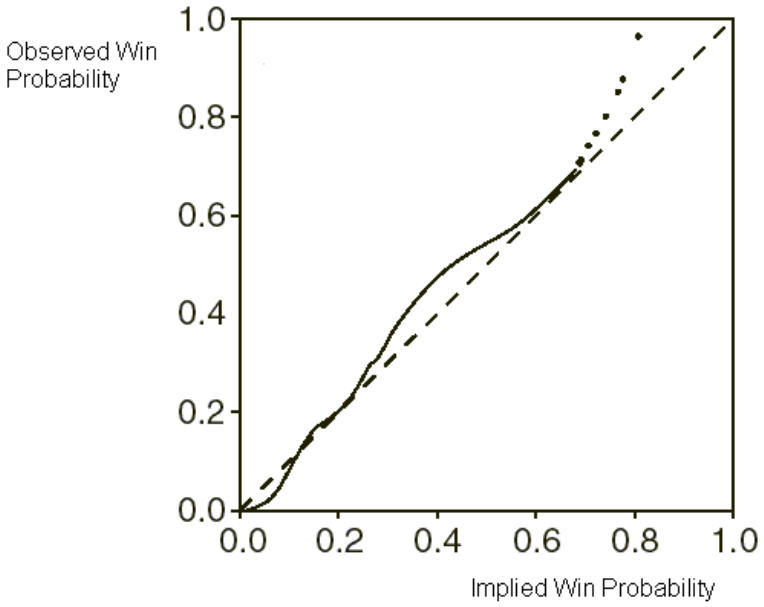

FIGURE 2 – *Non-parametric regression of win probabilities (at starting odds) against the percentage of winners for the **1978** flat season. (1,430 Obs.). Source: Cain, Law & Peel (2003).*

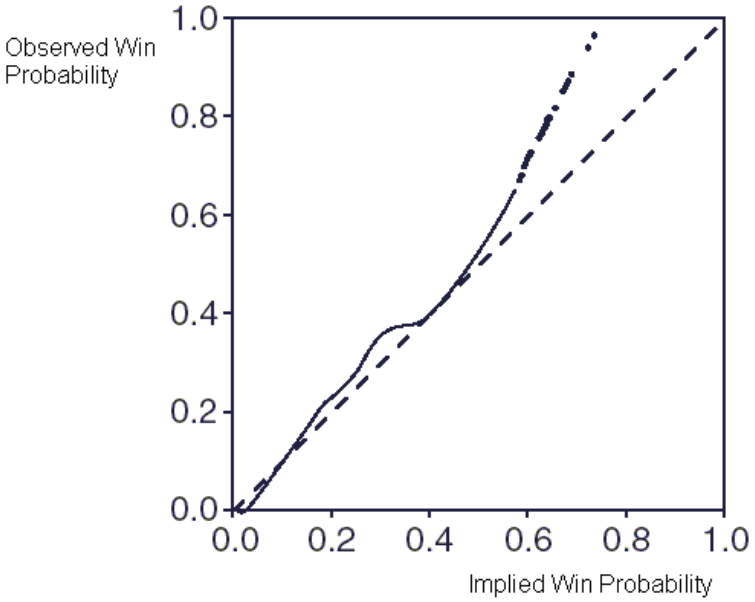

FIGURE 3 – *Non-parametric regression of win probabilities (at starting odds) against the percentage of winners for the **1987** flat season. (1,047 Obs.). Source: Cain, Law & Peel (2003).*

V. RESULTS – DIFFERENT CLASSES

The rationale behind investigating whether the Favourite-Longshot Bias varies across different classes of races is clear. Intuitively, one would expect betting markets of high class races to be more 'efficient'; one would expect deviations from the norm or biases to be more

prevalent for lower class races. Bookmakers and bettors are better informed and more familiar with high class horses, and if there were to be any corrupt practices in horse racing, one would expect it to occur in the lower class races; it would be easier to influence the result, or more precisely prepare a horse for a '*coup*' in the 4.20 at Folkstone than a race at Royal Ascot. In terms of the Shin model, one may expect more insiders to operate in the lower classes, thus bookmakers will price their book with this in mind hence the odds for lower class races would exhibit a larger Favourite-Longshot Bias. This would go hand in hand with Cain, Law and Peel's (2003) observation that U.S. baseball odds do no exhibit a Favourite-Longshot bias, the curve follows the 45º line, they argue that this is because baseball bettors are the most knowledgeable and sophisticated of all sports bettors[26]. The absence of a varying Favourite-Longshot Bias across different classes of races would indicate that bookmakers perceive the proportion of insiders for races of each class to be similar.

TABLE 2

Implied and observed probabilities for Class A and B races.

		Opening				Starting			
PrbA Range	Class Midpoint	Runners	WIN%	Error (ovr)		Runners	WIN%	Error (ovr)	
"0-2"	1	538	0.37	-0.63		610	0.33	-0.67	
"2-4"	3	968	2.27	-0.73		913	2.08	-0.92	
"4-6"	5	777	3.35	-1.65	*	776	3.48	-1.52	
"6-8"	7	544	7.90	0.90		531	7.91	0.91	
"8-10"	9	363	9.37	0.37		378	10.58	1.58	
"10-12"	11	318	12.89	1.89		298	11.07	0.07	
"12-15"	13.5	314	14.33	0.83		315	15.87	2.37	
"15-20"	17.5	374	16.31	-1.19		362	16.57	-0.93	
"20-25"	22.5	187	19.25	-3.25		192	18.75	-3.75	
"25-30"	27.5	134	32.84	5.34		135	27.41	-0.09	
"30-40"	35	121	35.54	0.54		124	35.48	0.48	
"40-60"	50	64	61.90	10.94		68	63.24	13.24	^
"60-100"	80	9	55.56	-24.44		9	55.56	-24.44	

Table and Figure 2 illustrate the results of only Class A and B races these are the top races which offer prize money in excess of £12,000, however for most owners the real reward is the potential stud value of their horse[27] should they be fortunate enough to win such a race; these results are based on 4711 runners. Since these are high class races, there are relatively few of them, and one would expect that these odds will be more efficient than the odds for lower classes, bookmakers and bettors are better informed about these horses, thus the presence of

[26] This is the hypothesis of Woodland and Woodland (1994) cited in Cain, Law and Peel (2003).
[27] Particularly if it's a female, if it is a male, then they must win the very top races to be a potential top sire.

inside information is less likely, and the larger amounts of prize money on offer should deter any wrong doing. The results are very encouraging, the 95% confidence intervals for the expected number of winners are only violated once for implied opening and starting win-probabilities, this is much better than for the pooled data even though these results are based on a much smaller sample size[28]. The implied probabilities are very good estimates of the actual outcome of the races compared with the whole dataset. It is also not surprising to see that where these 95% confidence intervals are violated, the direction of the violation is the direction that one would expect. For opening probabilities of 4-6% the observed probability is too low, and for starting probabilities of 40-60% the observed probability is too high.

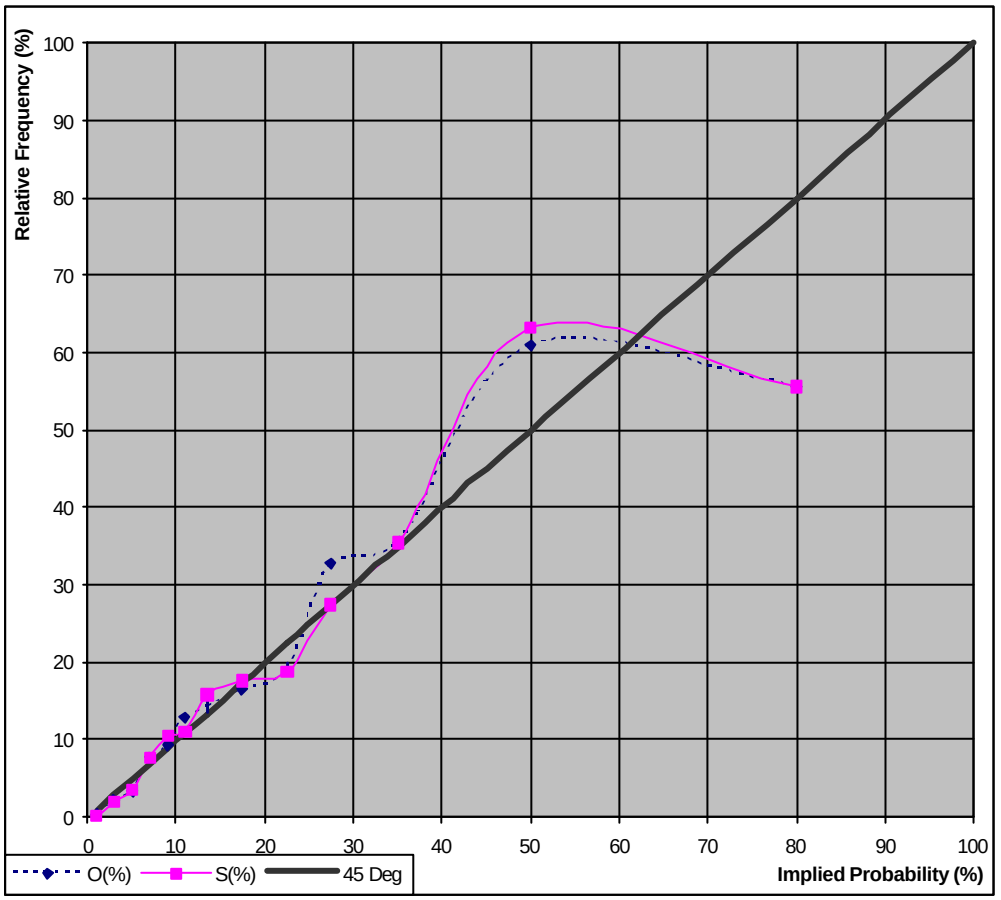

FIGURE 4 - *Plot of implied win probabilities (at opening and starting odds) against the percentage of winners. Class A and B races only (4,711 Observations).*

The curves in Figure 4 appear to be a lot more wild than compared with Figure 1, but this is purely based on the fact that there are far less observations. Also, one should not pay too much attention to the high probability observations as they were not based on many observations. In answer to the question of whether opening or starting probabilities are more

[28] However, one must bear in mind that there are less odds classes here.

efficient, Figure 4 would seem to suggest that it does not seem too clear cut. The only noticeable difference would be the disappearance of a hump at implied win-probabilities of 25-35% for starting prices, on the other hand, since the presence of a Favourite-Longshot Bias was very limited for opening probabilities in the first place, it is no surprise that the nature of the curve has not changed by too much. The pseudo-R^2 for the opening probabilities logit model is 0.1436 (Log likelihood = -1255.87), and for starting probabilities is 0.1454 (Log likelihood = -1253.22), once again the accuracy of implied starting win-probabilities is slightly better and the results are in agreement with the earlier findings.

TABLE 3

Implied and observed probabilities for Class C and D races.

		Opening				Starting			
PrbA Range	Class Midpoint	Runners	WIN%	Error (ovr)		Runners	WIN%	Error (ovr)	
"0-2"	1	2261	0.80	-0.20		2625	0.72	-0.28	
"2-4"	3	3338	2.34	-0.66	*	3144	2.77	-0.23	
"4-6"	5	2619	4.62	-0.38		2554	3.99	-1.01	*
"6-8"	7	1902	6.31	-0.69		1845	6.88	-0.12	
"8-10"	9	1366	8.93	-0.07		1332	9.01	0.01	
"10-12"	11	1100	10.82	-0.18		1107	10.84	-0.16	
"12-14"	13	823	13.85	0.85		774	12.14	-0.86	
"14-16"	15	636	15.09	0.09		690	15.65	0.65	
"16-18"	17	498	16.06	-0.94		502	15.34	-1.66	
"18-20"	19	445	17.75	-1.25		411	20.44	1.44	
"20-25"	22.5	733	22.78	0.28		729	23.73	1.23	
"25-30"	27.5	471	30.15	2.65		451	28.82	1.32	
"30-35"	32.5	299	34.78	2.28		312	35.90	3.40	
"35-40"	37.5	214	42.06	4.56		188	42.55	5.05	
"40-50"	45	238	50.00	5.00		234	48.29	3.29	
"50-70"	60	187	60.96	0.96		222	57.21	-2.79	
"70-100"	85	34	88.24	3.24		44	90.91	5.91	

The results for Class C and D races (races with prize money of more than £4,800) in Table 3 are similar to the results for Class A and B races. These results are based on 17,164 runners over 1713 races; much more than for the Class A and B races. There is only one violation of the forecasted number of winners for opening and starting implied win-probabilities, and this occurs at very low implied win-probabilities. Looking at Figure 5 confirms this story, the deviations from the 45°line are of a minor nature. For opening implied win-probabilities of larger than 20% (fair odds of less than 4/1), the curve is always above the 45° line. This is also reflected in the implied starting win-probabilities; however for starting probabilities between 50-70% (fair odds of evens to just less than 4/9) the curve is under the 45° line,

previously it was very close to it, one would not expect this if information is being used efficiently but since this deviation is not significant, this matter should not draw too much attention. The pseudo-R^2 for the opening probabilities logit model is 0.1618 (Log likelihood = -4670.81), and for starting probabilities is 0.1636 (Log likelihood = -4660.84), once again the accuracy of implied starting win-probabilities is slightly better. Overall the pattern for the top 4 classes of races seems to be very similar.

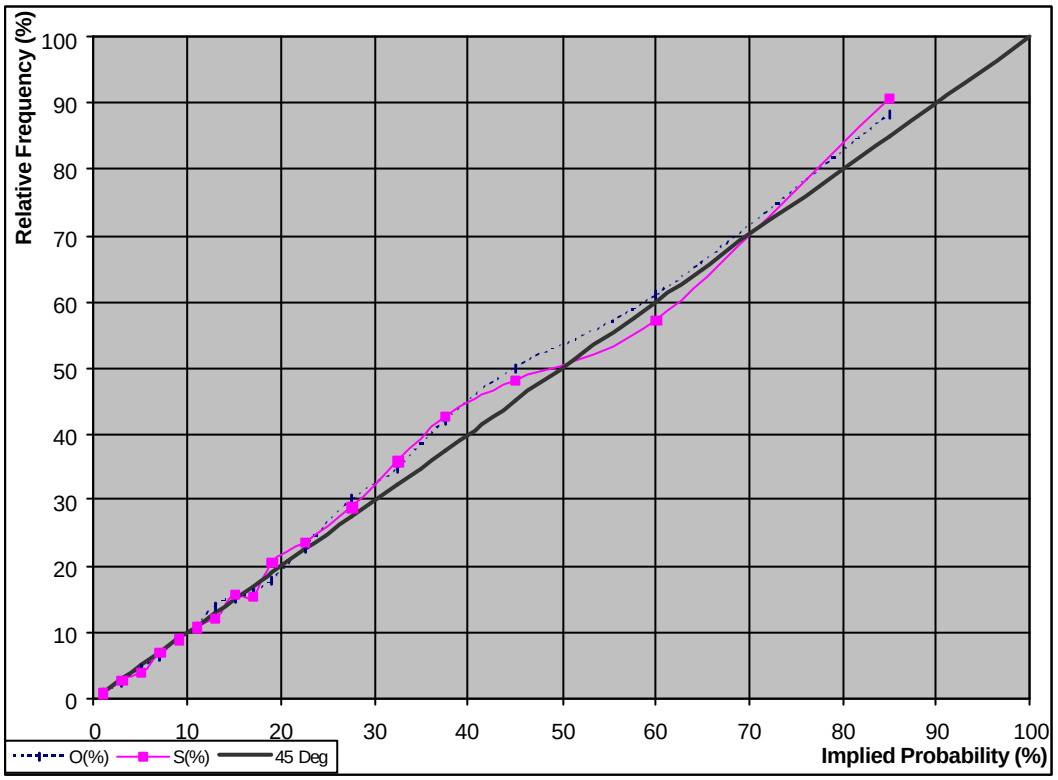

FIGURE 5 - *Plot of implied win probabilities (at opening and starting odds) against the percentage of winners. Class C and D races only (17,164 Observations).*

Table 4 shows the results for Class E races, this are the most common class of race, for this sub-sample alone, there are 862 races covering 10,074 entries, these races offer between £3,800-4,850 in prize money. The evidence suggests that the opening implied win-probabilities are less accurate than the starting probabilities; this bodes well with the belief that information is used efficiently. For opening implied win-probabilities, there are three violations of the 95% confidence intervals of the theoretical number of winners, the first occurs at implied win-probability of 2-4% (odds of greater than 25/1) where the number of winners is over-estimated, and implied win-probabilities of 16-20% (horses around the 4/1-5/1 mark) where they win more often than their implied probabilities. This forms the basis of a hump from win-probabilities of 15-30%, which still exists for starting win-probabilities.

TABLE 4

Implied and observed probabilities for Class E races.

		Opening				Starting			
PrbA Range	Class Midpoint	Runners	WIN%	Error (ovr)		Runners	WIN%	Error (ovr)	
"0-2"	1	1145	0.70	-0.30		1415	0.71	-0.29	
"2-4"	3	2381	1.97	-1.03	*	2216	2.35	-0.65	
"4-6"	5	1661	4.76	-0.24		1621	4.38	-0.62	
"6-8"	7	1162	6.54	-0.46		1122	5.70	-1.30	
"8-10"	9	852	8.33	-0.67		850	8.82	-0.18	
"10-12"	11	718	10.86	-0.14		688	10.17	-0.83	
"12-14"	13	447	10.51	-2.49		429	13.05	0.05	
"14-16"	15	377	14.85	-0.15		352	16.19	1.19	
"16-18"	17	275	23.64	6.64	^	278	21.22	4.22	
"18-20"	19	210	27.14	8.14	^	235	20.00	1.00	
"20-25"	22.5	345	22.32	-0.18		355	27.32	4.82	^
"25-30"	27.5	187	33.16	5.66		192	33.33	5.83	
"30-40"	35	182	36.26	1.26		168	33.93	-1.07	
"40-60"	50	112	50.89	0.89		127	49.61	-0.39	
"60-100"	80	20	80.00	0.00		26	76.92	-3.08	

Looking at pseudo-R^2 for the logit model with opening win-probabilities, the value is 0.1345 (Log likelihood = -2544.84), and for starting probabilities is 0.1414 (Log likelihood = -2527.17), once again the accuracy of implied starting win-probabilities is slightly better. With the pattern that is developing for comparisons of the logit model for opening and starting implied win-probabilities, it seems quite clear that starting probabilities are more accurate than opening probabilities, as one would expect if information is being used efficiently during the formation of the market. The improvement is small, but it seems safe to conclude that starting probabilities are definitely not more accurate than opening probabilities.

Figure 6 shows this more clearly, it also demonstrates how accurate the probabilities are when implied win-probabilities are above 30%. For starting probabilities, horses with 20-25% (horses with odds of around 10/3) implied win-probabilities' chances are underestimated. Whilst there is some evidence that for Class E races, the opening implied win-probabilities may be less efficient than for the higher class races, there is no evidence that the starting probabilities are less accurate, there is still only just one violation.

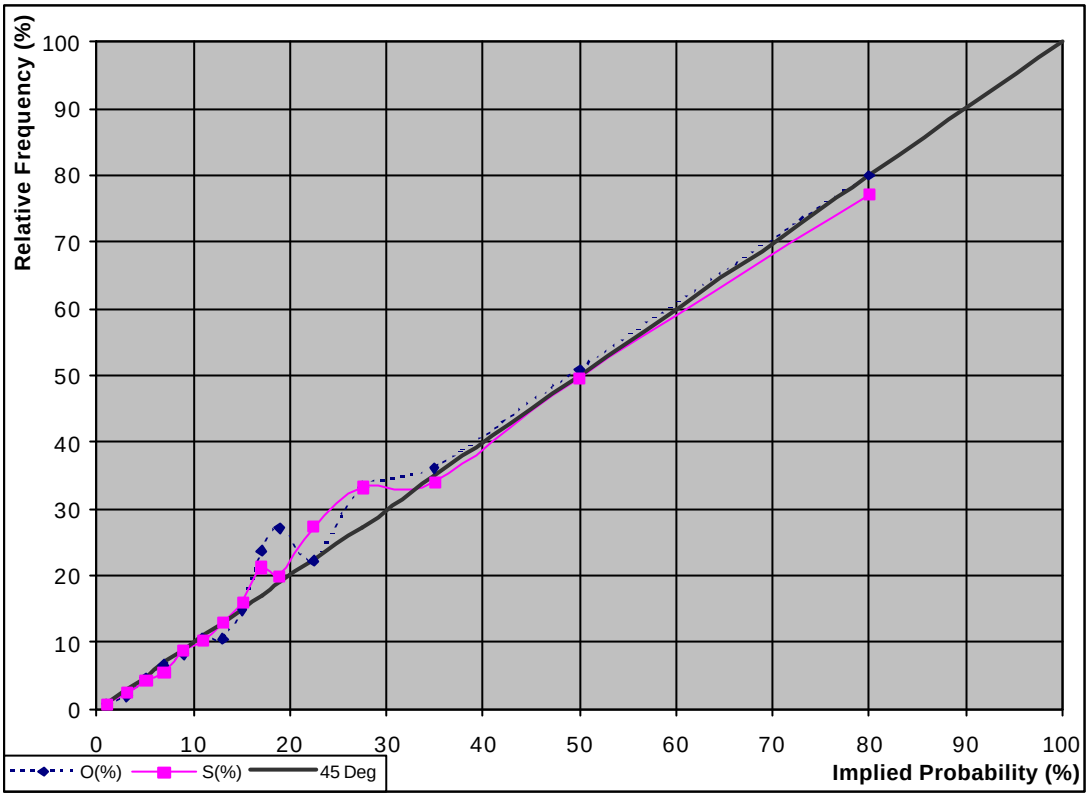

FIGURE 6 - *Plot of implied win probabilities (at opening and starting odds) against the percentage of winners. Class E races only (10,074 Observations).*

Having three violations of the 95% confidence interval for the forecasted number of winners provides some evidence of opening prices being less efficient for Class E races compared with the classes above, however one could argue that because the intervals of the implied win-probability ranges are different for each set of results that this comparison is unfair but I believe that the intervals are similar enough. This bodes well with our earlier hypothesis that for lower class races, prices will probably be less efficient and will probably exhibit a larger Favourite-Longshot Bias because of the higher proportion of insiders, in addition to this, the results for Class E also demonstrate that information flows efficiently in the market, overall, the implied starting win-probabilities are more accurate than the implied win-probabilities formed at the opening of the market.

The finding that implied starting win-probabilities are more accurate as well as exhibiting a narrower Favourite-Longshot Bias than implied opening win-probabilities is consistent with the finding by Law and Peel (2002) that the estimated degree of insider trading declines between opening and starting odds which was mentioned earlier. It would seem that bookmakers begin by setting odds which offer them greater protection against insiders at first, and the implications of the Shin (1991, 1992) model will mean that these odds will exhibit a

larger Favourite-Longshot Bias. Once initial bets have been struck, bookmakers revise their odds with the belief that the insiders have struck their bets, i.e. they formulate their odds, maximizing expected profits taking account of the bets which have been struck and subject to a lower value of z, and this in turn will reduce the Favourite-Longshot Bias and generate the results the data has produced.

TABLE 5

Implied and observed probabilities for Class F and G races.

		Opening				Starting			
PrbA Range	Class Midpoint	Runners	WIN%	Error (ovr)		Runners	WIN%	Error (ovr)	
"0-2"	1	931	0.54	-0.46		1097	0.36	-0.64	*
"2-4"	3	1648	1.76	-1.24	*	1614	2.17	-0.83	
"4-6"	5	1240	5.00	0.00		1146	4.97	-0.03	
"6-8"	7	821	6.70	-0.30		804	5.72	-1.28	
"8-10"	9	584	7.02	-1.98		562	9.07	0.07	
"10-12"	11	496	10.69	-0.31		446	10.99	-0.01	
"12-14"	13	265	12.08	-0.92		309	11.33	-1.67	
"14-16"	15	219	13.70	-1.30		226	13.72	-1.28	
"16-20"	18	328	24.70	6.70	^	308	23.38	5.38	^
"20-25"	22.5	189	29.10	6.60	^	200	23.50	1.00	
"25-30"	27.5	112	25.00	-2.50		104	35.58	8.08	
"30-40"	35	104	45.19	10.19	^	109	42.20	7.20	
"40-60"	50	55	52.73	2.73		64	56.25	6.25	
"60-100	80	13	69.23	-10.77		16	62.50	-17.50	

Table 5 shows the results for the lowest class races, Classes F and G, these races offer prize money of below £3,100. There are 556 races of these races in the dataset covering 7,005 runners. The results seem to suggest that the trend of a more evident Favourite-Longshot Bias is continuing as we move towards lower class races. For opening implied win-probabilities, the pattern is roughly the same as Class E races but in addition, there is an extra violation (underestimation) at win-probabilities of 30-40% (horses with odds of around 9/4-11/8). The violations at implied win-probabilities of 20-25% and 30-40% are corrected for starting implied win-probabilities, as is the underestimation at 2-4%. There is also an additional underestimation at 0-2%, implied win-probabilities (these are the horses with fair odds of greater than 50/1).

Figure 7 illustrates the results from Table 5, compared with all the curves that have been explored, this seems to be the most inaccurate; a comparison with Class A and B races would be a little unfair as those results were based on half as many observations. The hump prevails

over a longer range for these low class races, here it runs from implied win probabilities of 20-50% (win an exception at 25-30% for opening probabilities), and is similar to the one found for the whole dataset. The Figure itself demonstrates little evidence that implied starting probabilities are more efficient than opening probabilities, there is evidence for this from the number of violations of the 95% confidence intervals of the theoretical number of winners and the improvement of the pseudo-R^2 from moving from opening probabilities (0.1434, Log likelihood = -1663.47) to starting probabilities (0.1496, Log likelihood = -1656.74). It is a shame that there are relatively few observations for the highest implied win-probability class as it would be interesting to see whether the curve heads upwards or downwards for high implied win probabilities.

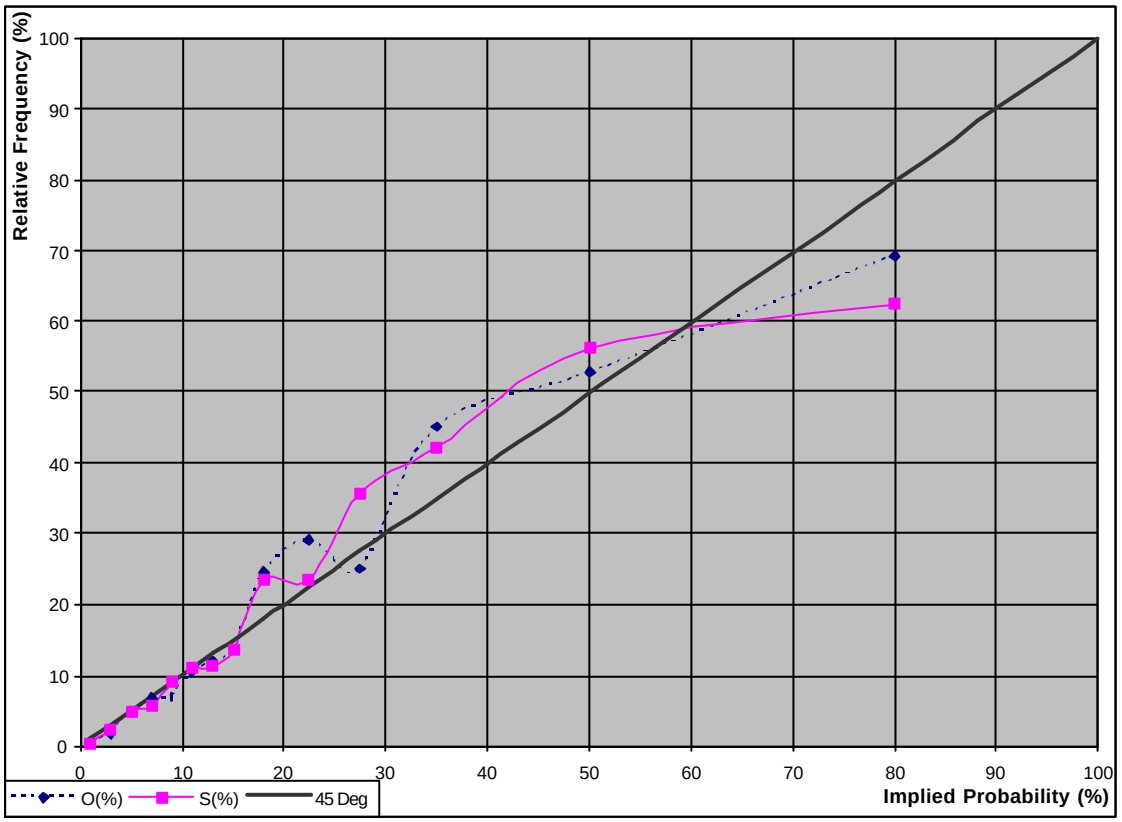

FIGURE 7 - *Plot of implied win probabilities (at opening and starting odds) against the percentage of winners. Class F and G races only (7,005 Observations).*

The evidence suggests that the win-probabilities implied by the prices are accurate for all classes of races and there is arguably a minor Favourite-Longshot Bias. For opening implied win-probabilities, there is evidence that for the lowest class races the Favourite-Longshot Bias is more apparent. Looking at whether the observed number of winners is consistent with the forecasted number of winners for the odds intervals, there were more violations for Class E, F

and G races; with Class F and G races showing the largest number of violations. For every class (but not for the whole dataset pooled together) the number of violations for implied opening win-probabilities is less than or equal to the number of violations for starting win-probabilities, this is evidence in favour of the hypothesis that starting win-probabilities are more accurate, i.e. information is being used efficiently and it is being incorporated into prices, the result is also consistent with the belief that insiders tend to bet earlier to lock in the 'high' returns on offer, thus forcing bookmakers to set odds with a stronger Favourite-Longshot bias.

VI. CONCLUSIONS

This paper has analysed the nature of the Favourite-Longshot Bias in British Horse Racing in the modern era of Betting Exchanges, tax free off course betting and full television coverage. One of the central findings is that the nature of the Favourite-Longshot Bias is the same as it was for the 1978 and 1987 Flat seasons. For longshots the implied win-probabilities, based on the price of the horse normalized by the over-round of the race, are consistently underestimated; this is consistent with the Shin model, and win-probabilities were underestimated (there was a hump above the 45 line for horses with implied win-probabilities of between 25-50% though the width of this interval varied from class to class. Overall the implied win-probabilities are very accurate, when grouping horses by their implied win-probabilities and comparing the actual number of winners with the 95% confidence interval for the number of winners, there were only one or two violations out of about 15 classes and these occurred at either the very low probability end or where the "hump" was. There is no evident substantiation that betting markets are any more efficient in the modern era than they were in 1978 and 1987 based on looking at the nature of the Favourite-Longshot Bias, the analysis will need to be repeated for the older data in order to conduct a fair comparison.

The results provide evidence of the intuitive hypothesis of weak-form market efficiency that implied win-probabilities at the off provide a better estimate of a horse's real chance of winning than the win-probabilities by the opening sets of odds. There are less violations of the 95% confidence interval of the predicted number of winners based on the binomial model for implied starting win-probabilities and the results of the logit model backs up this observation. The results are in agreement with Lo's (1994) findings for pari-mutuel markets and the finding is true for every class; there are more correct odds movements than incorrect ones.

As for variation amongst different classes, subtle differences do seem to exist. Opening and starting implied win-probabilities are very accurate for races down to Class D races, where visually, the Favourite-Longshot bias seems to exist with a very weak manifestation. For the lowest class races, and ignoring the very high probability horses, the plots of implied win-probability against relative win frequency does indeed show a relatively more severe Favourite-Longshot Bias. The most noticeable result from this investigation has been that for opening prices, there is evidence from the number of violations of the 95% confidence intervals of the forecasted number of winners that there is a more evident Favourite-Longshot Bias for lower class (E, F and G) races compared with higher class races, this is in agreement with the hypothesis that there are more insiders operating in these classes of races. Quite interestingly for these lower classes, even though the patterns for opening and starting implied win-probabilities is visually similar, the number of violations does reduce for the starting probabilities, this is what one would expect if information is being used efficiently. The evidence suggests that implied starting probabilities are of similar accuracy for all classes, yet there seems to be a difference for opening prices, it would seem that the market mechanism brings the nature of the Favourite-Longshot Bias of the probabilities to a unique structure; this is consistent with the empirical observation that insiders strike their bets early to lock in the generous odds offered by the less knowledgeable bookmakers and once these bets have been struck the remaining market will consist of less insiders which bookmakers will take into account and price their books accordingly. Having fewer insiders left in the market after the early bets have been struck would explain why the Favourite-Longshot bias has narrowed between the formation and the cessation of the market and why the results seem to suggest that the degree of accuracy of implied starting win-probabilities are of similar accuracy for every class.

VII. APPENDIX

TYPES OF FLAT RACES

THE CLASSES (A TO G)

A Pattern (Group1/2/3) & Listed.
B Ratings band 0-100+; min value £14,900 (2yo £12,550; nurseries £20,000).
C 0-85/100; £9,750 (2yo £9,000).
D 0-75/85; £4,850.
E 0-65/75; £3,800.
F 0-60/65; £3,100.
G selling, apprentice & amateur races below £3,100.

THE CATEGORIES

PATTERN: the top races, Groups 1/2/3
LISTED: races just below Pattern standard, including the top Rated Stakes.
RATED STAKES: valuable handicaps with 10-14lb weight range. Max rating 110, but usually 100/95
CONDITIONS STAKES: non-handicaps, intended for horses just below class A.
CLASSIFIED STAKES: non-handicaps with max rating, for horses that have run at least 3 times, or run twice and won at least once.
STANDARD HANDICAP: (including nurseries)
MAIDEN HANDICAP: for 3yo + maidens, max rating 70, that have run at least four times.
OPEN MAIDEN
RATING-RELATED MAIDEN: non-handicaps with max rating, for horses that have run at least 3 times.
NOVICE FLAT RACE: for 2/3 yo only which have not won more than twice (some class restrictions)
AUCTION MAIDEN: for 2yo originally sold by auction at specified sales.
MEDIAN AUCTION MAIDEN: for 2yo by stallions that established a median price for their yearlings at specified sales.
SELLING, CLAIMING, APPRENTICE, AMATEUR

GENERAL DEFINITIONS

HANDICAP: weights are determined by official ratings. Flat horses are rated on scale 0-140 and most handicaps have specified ability range.
NURSERY: handicap for 2yo
CLAIMING RACE: connections nominate the minimum price for which the horse can be claimed (bought on demand); this determines the weight to be carried.
SELLING PLATE: for low-class horses (often called platers); winner must be offered for auction and any other runner may be claimed for a price defined in race conditions.
MAIDEN: horse that has not won a race.

Source: Betfair Sports Diary 2003.

REFERENCES

Ali, M. (1977). 'Probability and Utility Estimates for Racetrack Bettors', *Journal of Political Economy*, 85, 803-15.

Bolton, R. and R. Chapman (1986). 'Searching for Positive Returns at the Track: A Multinomial Logit Model for Handicapping Horse Races', *Management Science*, 32, 1040-60.

Cain, M., D. Law, and D. Peel (2001). 'The Relationship Between Two Indicators of Insider Trading in British Racetrack Betting', *Economica*, 68, pp. 97-104.

Cain, M., D. Law, D. Peel (2003). 'The Favourite-Longshot Bias, Bookmaker Margins and Insider Trading in a Variety of Betting Markets', *Bulletin of Economic Research*, 55:3, pp. 263-73.

Cox, D.R. (1961). 'Tests of Separate Families of Hypotheses', *Proc. 4th Berkley Symposium*, 1, pp. 105-23.

Cox, D.R. (1962). 'Further Results on Tests of Separate Families of Hypotheses", *Journal of Royal Statistical Society B* 24, pp. 406-24.

Crafts, N.F.R. (1985). 'Some Evidence of Insider Knowledge in Horse Race Betting in Britain', *Economica*, 52, pp. 295-304.

Crafts, N.F.R. (1994). 'Winning Systems? Some Further Evidence of Insiders and Outsiders in British Horse Race Betting', in Hausch, D., Lo, V., and Ziemba, W., eds. (1994). *Efficiency of Racetrack Betting Markets*, pp. 545-9.

Fama, E. (1970). 'Efficient Capital Markets: A Review of Theory and Empirical Work', *Journal of Finance*, 25, pp. 383-417.

Fingleton, J., and P. Waldron (1996). 'Optimal Determination of Bookmakers; Betting Odds: Theory and Tests', *Trinity Economic Papers*, Technical Paper no. 9, December.

Gabriel, P.E. and J.R. Marsden (1990). 'An Examination of Market Efficiency in British Racetrack Betting', *Journal of Political Economy*, 98, pp. 874-85.

Hausch, D., V. Lo, and W. Ziemba, eds., (1994). *Efficiency of Racetrack Betting Markets*, Academic Press, San Diego.

Griffith, R. M. (1949), "Odds Adjustments by American Horse-Racing Bettors", *American Journal of Psychology*, 62, pp. 290-294

Law, D., and D. Peel (2002). 'Insider Trading, Herding Behaviour and Market Plungers in the British Horse-race Betting Market', *Economica*, 69, pp. 327-38.

Lo, V.S.Y., (1994). 'Applications of Logit Models in Racetrack Data', in Hausch, D., Lo, V., and Ziemba, W., eds. (1994). *Efficiency of Racetrack Betting Markets*, pp. 307-14.

McCririck, J. (1991). *World of Betting*. London: Stanley Paul.

McGlothlin, W. H. (1956), "Stability of Choices Among Uncertain Alternatives", *American Journal of Psychology*, 69, pp. 604-619.

Shin, H.S. (1991). 'Optimal Betting Odds Against Insider Traders', *Economic Journal*, 101, pp. 1179-85.

Shin, H.S. (1992). 'Prices of State-Contingent Claims with Insider Traders, and the Favourite-Longshot Bias', *Economic Journal*, 102, pp. 426-35.

Shin, H.S. (1993). 'Measuring the Incidence of Insider Trading in a Market for State Contingent Claims', *Economic Journal*, 103, pp. 1141-53.

Snyder, W. (1978). 'Horse Racing: Testing the Efficient Markets Model", *Journal of Finance*, 33, pp. 1109-118.

Thaler, R., and W. Ziemba (1988). 'Parimutuel Betting Markets: Racetracks and Lotteries', *Journal of Economic Perspectives*, 2, pp. 161-74.

Vannebo, O. (1980). 'Horse Racing: Testing the Efficient Markets Model: Comment', *Journal Of Finance*, 35, pp. 201-2.

Weitzman, M. (1965). 'Utility Analysis and Group Behaviour: An Empirical Study', *Journal of Political Economy*, 73, pp. 18-26.