3.

Market Movers in the British Bookmaker Betting Market: Evidence from the 2003 UK Flat Turf Season

3.1 – INTRODUCTION

The adjustment of odds in a betting market reflects unexpected support (or lack of) for the respective competitors in an event. The purpose of this chapter is to investigate the accuracy of odds changes in the British bookmaker betting market for the flat turf season of 2003. The finance literature abounds with evidence of investor under- and over-reaction to news potentially affecting the value of a security as well as evidence of investors engaging in herding behaviour (see Shleifer (2000)). The purpose of this study is to investigate whether this manner of behaviour is prevalent in this bookmaker betting market. Empirical evidence (see Chapter 2 for a summary) indicates that starting odds are a more accurate indicator of true winning probabilities than the odds quoted in the opening stages of the market. This is consistent with weak-form market efficiency and models where trades themselves generate information.

Previous investigations (see Section 1.3) into market movers focus mainly on the returns from betting on market movers. Conclusions drawn using this methodology are flawed if returns exhibit a favourite-longshot bias. Also these previous investigations did not analyse in detail how accurate these market moves are. The new approach taken in this chapter involves investigating how the changes in the win-probabilities implied by the odds between the formation of the market (opening probability, *OPR*) and the start of the race (starting probability, *SPR*) relate to the horses' chances of winning using a linear probability model (LPM). The analysis is first conducted for the whole dataset and then repeated for every class of horse race that the dataset comprises of. The motivation behind this is to investigate whether over-reaction is more likely to occur in the lower class races, (classes are essentially distinguished by the prize money on offer). In these events, there are typically fewer bookmakers present at the racecourse than in higher class races and hence less competition. Moreover, the market is less liquid so it is easier for large bets to affect prices, and this may make over-reaction more likely.

The results indicate that the degree of reaction in the market is consistent with the market reacting correctly. A move assigning a one percentage point increase in the implied win-probability translates into an increase of the horse's chances of winning by one percentage point. This marginal effect also seems to be constant across all magnitudes of changes, across all classes of races and independent of attendance levels and the number of bookmakers present.

Section 3.2 discusses the relevance of market movers in betting markets, followed in Section 3.3 by an examination of the evidence amassed so far, including discussion of the findings in Section 1.3. Section 3.4 discusses the methodology used in this chapter. Section 3.5 presents the results based on linear specifications and robustness checks are provided in Section 3.6. Non-linear specifications are tested in Section 3.7 and the conclusions are presented in Section 3.8.

3.2 – THE RELEVANCE OF MARKET MOVERS IN BETTING MARKETS

This investigation focuses on the informational efficiency of changes in quoted odds. In traditional financial markets, many behavioural anomalies persist: investors have been observed to under-react to individual news announcements, over-react to a string of positive/negative news announcements, and are known to engage in herding behaviour which cause price bubbles (see Shleifer (2000) for a summary). Herding or crowd behaviour is where uninformed investors follow price trends, without any form of coordination, their trades help to sustain the price trend encouraging more traders join in, pushing prices further away from their fundamental values. It would not be surprising if these phenomena were to exist in betting markets since insider trading is acknowledged to take place, (see Crafts (1985) page 303, and later on in this chapter), and for reasons of greed, deceit, etc.; *uninformed betters are thus more likely to follow trends*.

As discussed in Chapter 1, the market lasts for 15-20 minutes before the start of the race. The market odds are available to all bettors (on-course and off-course) and are defined as the best odds at which a 'significant' amount of money can be traded at on the course. At the start of the market, bookmakers post odds based on information in the public domain, the odds posted by their rivals, and any inside information they might have (or believe to have). Odds movements occur if there are excessive numbers of (or a lack of) bets on particular horses. Bookmakers react by shortening (lengthening) the odds in order to discourage (attract) bets so as to reduce the variance of their expected returns. For example, when laying a book of bets, if a bookmaker receives excessive bets on horse *i*, he will contract the odds on horse *i* and increase the odds of the other horses, especially ones for which he has not received many bets on. This practice is known as *balancing the books*.

In general, it is difficult to establish what news trickles through during the market phase that market participants can react to. In between the posting of opening odds and the start of the race, news events should play a small role because the most important news that a bettor receives is revealed prior to the formation of the market, namely the form, draw biases, and the state of the going[1]. This news will have already been incorporated prior to formation of the on-course market. Market movements in on-course horse race betting markets therefore will be mostly influenced by the distribution of bets at the racecourse. Moreover, movements occurring on betting exchanges before the formation of the on-course market mean that there is a lesser role of price discovery for the on-course market.

[1] A horse drawn to start from a better stalls position has to cover less distance and/or has the advantage of better ground. For UK racecourses, it is not uncommon to see strips of ground where horses can run faster; probably because they are watered less rigorously than the other parts. The going corresponds to the state of the ground and horses have different preferences towards the going.

However, there can still be incidents of news events that potentially affect the odds of the competitors during the existence of the market phase. These include how the horse turns out in the parade ring, jockey and trainer comments, and any bets placed by agents associated with a certain horse/stable. However, the average bettor is seldom informed about these bets. The other significant forms of information concern how a horse cantors to post, if it is deemed to be unruly then it could have expended too much energy prior to the race or it will be too keen once the race starts. There are still many opportunities for bettors to under- or over-react to information. This investigation does not distinguish between moves caused by news or ‘irregular’ betting patterns.

Bettors’ attentions are drawn to market moves, when one watches the racing on television, betting pundits take their fair share of the airtime (such as John McCririck of Channel 4 Racing and Angus Loughran of the BBC). Betting pundits discuss odds movements in the on-course market and the betting exchanges. In essence, the parallel from financial markets would be the chart analysts on Bloomberg Television who make recommendations on whether to buy or sell based on past price movements of securities. This parallel has limits though; financial analysts’ aims are to predict the future movement of prices as opposed to betting pundits who primarily provide an idea of which horses have been or have not been subject to support.

Changes in odds are indicative of unexpected support for a horse. When bettors see the odds of a horse contract they know that barring bookmakers artificially contracting the odds[2], there is somebody who believes that the horse’s original odds were good value. Market participants are also aware of betting *coups* which occur in racing since some stables also rely on betting as a source of income so that insufficiently paid trainers and their staff can make ends meet (see Crafts (1985), page 303). Typically such a stable intent on executing a *coup* or ‘cheating’ would set up a horse for a ‘gamble’ by initially running a horse in conditions where they believe to be least favoured (distance, going, and ensuring that it is not hard ridden etc.). This ‘deceives’ the handicapper and observers, who thus perceive the horses’

[2] There is nothing to prevent bookmakers doing this except for competition amongst the bookmakers.

chances to be worse than they are in reality for a subsequent race, where higher odds will be quoted. The stable and associated parties take advantage of this inside information and place heavy bets onto the horse, contracting its odds, and bettors see the contraction as a signal of insider trading[3].

McCririck (1991, pp. 51-2) writes "perversely, many punters turn up at a racecourse or betting shop on [*sic*] tending to back a certain 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 than they had thought – even when it no longer represents a value bet". He goes on to suggest that bettors should not jump onto the bandwagon. Chapter 7 of his book talks about many famous *coups*. The following section of this chapter discusses examples and empirical evidence concerning horses who were market movers or subject to bettor sentiment.

In the 2003 Epsom Derby, the most valuable race in the UK, Kris Kin was quoted at odds of 14/1 in the morning of the race. This reflects an implied winning probability of 6.6% less a margin of one or two per-cent. The horse opened up on-course at 10/1 (9% implied win-probability) and started the race at 6/1 (14%). The impression of this horse had been that from in its previous appearances, it was a good horse, nothing special and a quirky character. In its previous race at Chester it was an outsider of 4 horses with starting odds of 20/1 (<5%) and it won. The horse was then supplemented for the Derby (at a cost of £90,000) and it was not until the morning of the race that pundits were commenting that Sir Michael Stoute, one of the top trainers in the world would not supplement the horse for the Derby unless it had a good chance and that the horse was still open to further improvement. It seems to me that the move from 14/1 to 10/1 reflected this, (a move from 14/1 to 10/1 in such an important race is itself a very significant move), and the rest of the move was due to the feedback traders lumping on. Kris Kin went on to justify the support and win the Derby, dealing a big blow to bookmakers, but since the 3 favourites of the race (who the bookmakers also have large liabilities on in addition to the plungers) did not fare too well, the blow to the bookmakers was lessened.

[3] See Cain, Law and Peel (2001) who find that a measure of the contraction of a horse's odds is closely related to Shin's measure of the incidence of insider trading (explained earlier).

Perhaps the most memorable plunge in recent times was on Fujiyama Crest, in the seventh and final race at the British Festival of Racing at Ascot in 1996. This horse was the mount of Frankie Dettori, one of the best known jockeys in the world. Prior to this race, Dettori had carried out the amazing feat of winning all of the six previous races at that meeting. Fujiyama Crest was quoted at 12/1 in the morning newspapers reflecting chances of winning of around 7.5%. In fact, in his autobiography[4], Dettori remembers telling a weighing room colleague after the sixth race that he had no chance of winning, he was on the twelve or fourteen to one shot who was carrying far too much weight and was hopelessly out of form. However, because of bettor sentiment and the fact that many of the large bookmakers had huge potential liabilities on the horse due to multiple/accumulator bets placed on Dettori, Fujiyama Crest was sent off at odds of 2/1 (around a 30% chance of winning). One can understand the bookmakers' willingness to lay this horse. Assume that the horse's real chances of winning were 7.5%, then the expected value of laying a £1 bet for the bookmaker would be over 77 pence[5] with a variance of 38 pence. Unfortunately, for the bookmakers, Dettori gave Fujiyama Crest a beautiful ride and managed to cling on to win the race and complete his 'Magnificent Seven', a feat which any pundit would have believed to be impossible, especially at such a big meeting and will probably never be repeated again. On that day, bookmakers lost £40m, Ladbrokes, one of the world's largest bookmaking firms lost £10m alone[6], whilst William Hill was reported to have lost £8m[7]. A great deal of these losses would have been on Fujiyama Crest.

Are the starting probabilities of market movers way off the mark? Let us put plungers and drifters into perspective for the 2003 season. The horses with the biggest plunges saw their odds-implied win-probability[8] increase by an average of 15 percentage points (range is 13 to 18 percentage points) between the opening and

---

[4] Chapter 18 "*The Bookies Were Crying For Mercy*", 'Frankie: The Autobiography of Frankie Dettori, (2004) Collins Willow.
[5] The bookmaker loses £2 with 7.5% probability and wins £1 with 92.5% probability.
[6] Source: Gambling Magazine http://www.gamblingmagazine.com/articles/31/31-57.htm (Accessed June 2005)
[7] Source: BBC Website http://news.bbc.co.uk/sport1/hi/other_sports/1565232.stm (Accessed June 2005)
[8] How this is calculated will be discussed later. I am referring to the 10 horses with the largest *DPR*, the change in the implied win-probability between the formation and the cessation of the on-course market.

cessation of the market. They had an average starting probability of 38% and won five out of the ten races. As for market *drifters*, the ten biggest drifters saw their odds-implied win-probability fall by an average of 12 percentage points (range -11 to -18 percentage points). They had an average starting probability of 25%, and won three of the ten races. This crude example shows that the *starting* probabilities of market movers are quite accurate!

In any race, assuming that market moves are based only on changes to the market participants' information set, three alternative situations can arise: under-reaction, over-reaction or the market reacting correctly.

> **Hypothesis 1**: Under the hypothesis of **under-reaction**, the changes in the implied win-probabilities of the horses would be lower than the increase in the actual likelihood of the horse winning given news events and information generated by trades.
>
> **Hypothesis 2**: Under the hypothesis of **over-reaction**, the changes in the implied win-probabilities of the horses would be greater than the increase in the actual likelihood of the horse winning. This would be consistent with herding behaviour, especially if the market over-reacts to plunges.
>
> **Hypothesis 3**: If the **market reacts correctly**, the change in the implied win-probability as perceived by the market is equal to the actual change in the horse's chances of winning.

These hypotheses will be reformulated with respect to the model that will be used for the analysis in Section 3.4.

3.3 – EVIDENCE FROM BOOKMAKER BETTING MARKETS

An interesting observation made by Crafts (1985: discussed in Section 1.3) with regards to the degree of reaction to market movers (see Table 1.1) is that the expected profits of backing a big plunger at starting odds (-0.01) are greater compared to backing a very big plunger at starting odds (-0.09 - for all races). This evidence is suggestive of the market over-reacting (in relative terms) to these larger moves. Bettors may be herding and placing their bets even though this cannot be rationally justified. The data suggests that from the bettors' point of view, it is better to back a big plunger rather than a very big plunger, the odds of which have had their value taken away from them; bookmakers are not keen on having too many liabilities on one horse so they over-contract the odds. For drifters, the expected losses are smaller for the very big drifters than for big drifters, this could suggest a relative under-reaction by the market towards the big drifters (their odds should increase by more). However there is no indication of the initial odds of these horses in Craft's paper and this information is important, as will become clear later on.

Law and Peel's (2002) results (see Section 1.3) are the same for pre-market (*meo)* movers. However the lack of a 'very big' plungers category prevents a direct comparison to Crafts' results; the only conclusion is that there is under-reaction to big plungers (compared to small plungers) because they provide expected returns which are less negative. For on-course movements (Table 1.2, 1987 data, *mop*, *@St.Odds* column), it is better to back a big drifter, suggesting that the market over-reacts compared with the small drifters. For plungers and big plungers, the returns are the same, suggesting a similar degree of market reaction.

One point to note is that *the above approaches do not take into account the opening/starting prices of the horses*. An issue to bear in mind when discussing these results is the *favourite-longshot bias*. The evidence presented so far suggests that the rate of return seems to be related to the magnitude of the move, but it is also related to the probability of the horse winning. For example, in Table 1.2, *meo* or *mop* classes with the lowest mean prices (longshots) are the classes which offer the worst returns. In Table 1.2, the rate of return at starting odds using *mop* for the 1987 data is the best for big drifters, and the worst returns are from backing non-movers. Conclusions that it is better to back big drifters are premature and do not tell the full story. Big drifters have an average price of 0.18 and non-movers have an average price of 0.07. The FL-Bias would automatically dictate that the returns of the big drifters are superior to the non-movers because non-movers are longshots compared with the big drifters. The methodology adopted for this chapter's analysis attempts to control for this 'sample selection' problem, as does the analysis discussed in the following paragraph.

The secondary focus in Law and Peel (2002) is an attempt to distinguish between herd activity and insider behaviour. They use the Shin (1993) measure of insider activity (ζ) to conduct some of their analysis. The parameter ζ can be estimated using odds data and an iterative procedure, see Shin (1993) or Cain, Law and Peel (1997). Law and Peel run regressions to analyse what factors affect the rate of return at starting odds concentrating on *mop*, *meo* and other variables allowing for the FL-Bias and the estimated value of ζ. By including the horse's price as a regressor, the specifications of their regressions take into account the chances of the horses winning, so this methodology does not suffer from the problems of the previous analysis with regards to the existence of a FL-Bias. One of their most interesting findings comes from a regression run only for horses with *meo* greater than 0.05; the aim of taking this sample was to concentrate on any activity by herders who *respond to the contraction of odds occurring in the morning*[9]. The result of this regression for 1632 runners is shown below[10]:

$$\underset{\text{Return}}{\textit{Rate Of}} = \underset{(0.110)^{*}}{-0.316} + \underset{(0.302)^{**}}{0.631 Prob} - \underset{(0.525)^{*}}{0.916 mop} + \underset{(0.878)^{*}}{1.807 d^{*} mop} \tag{3.1}$$

where *d*mop* is *mop* multiplied by a dummy variable (*d*) which takes the value 1 if the Shin measure of insider trading (ζ) has increased during the market phase. The presence of insiders could increase during the market phase because bookmakers underestimated the incidence of insider trading and pick this up through large amounts of big bets on a certain horse[11]. The coefficient on *Prob* indicates the presence of a FL-Bias; horses with higher prices earn higher returns. The point to note here is that *subject to a morning move*, the bigger the plunge from opening to starting odds, the more negative the returns are. More importantly *if the Shin measure of insider trading rises during the market phase*, expected returns rise (the specification in (3.1) assumes that the effect is the same for *any* rise in ζ). The coefficient estimate suggests that (provided that *mop* is high enough,) it is enough to compensate for both the average loss (the constant, 31.6%) and the losses caused by

[9] However, opening odds significantly shorter than forecasted odds do not necessarily indicate a morning move.
[10] Robust standard errors are in parentheses, ** indicates significance at the 5% level and * indicates significance at the 10% level respectively. Without conditioning on a morning move, the estimated coefficient on *dmop* is not significant.
[11] These large bets cause market movers, so market movers are correlated to changes in ζ.

the negative *mop* coefficient. In other words *for this sub-sample of horses who were subject to support in the morning, plunges occurring on-course when ζ does not rise are consistent with herd activity hence their lower returns, but when the on-course plunge is caused by informed bettors, positive returns can be earned; this is evidence against weak-form*[12] *and even-strong form efficiency because a set of plunges are more successful in the presence of insiders*. Data from the 2003 season for the morning moves is not available for the proceeding analysis, so running this regression would not be possible.

The next section explains and justifies the measure of the market move employed in this study.

3.4 – THE MARKET MOVE MEASURE AND METHODOLOGY

The aim of this chapter is to investigate how changes in the perception of a horse's chances of winning between the formation and the cessation of the market are related to its observed chances of winning. The LPM from the previous chapter is adapted to measure this. The market move measure adopted is the change in the implied win-probability between the formation of the market and the off, *DPR*. For horse *i*:

$$DPR_i = SPR_i - OPR_i. \tag{3.2}$$

This gives the percentage point change in the implied win-probability as attributed by the market (expressed as a decimal). Using the absolute change rather than the relative change avoids running into an issue similar to the one for Crafts' (1985) measure, discussed in Section 1.3. For example, if the relative change is used, a horse with 1% *OPR* (corresponding to odds of roughly 100/1) whose *SPR* is 2% (~50/1) will have a 100% increase in the implied probability, the same as a horse whose implied win-probability increases from 20% (4/1) to 40% (6/4); however, the latter change reflects a much larger increase in support and would be comparable to a horse whose implied win-probability increased from 1% to 21% (~7/2).

[12] Observers of odds (prices) can theoretically estimate *z* and choose to place bets on plungers (whose opening odds were significantly shorter than the forecasted odds) when *z* rises during the market phase.

The following LPM is adopted:

$$P(Win) = \alpha + \beta_1 OPR + \beta_2 DPR \,. \tag{3.3}$$

This specification states that the probability of a horse winning is a linear function of the opening implied win-probability attributed by the market and the change in the implied win-probability measure between the formation and the start of the race. At first view, this specification could be subject to multicolineraity problems if there is correlation between *OPR* and *DPR*. The correlation coefficient between *OPR* and *DPR* is 0.0756, so colinearity is not an issue for this specification. The slightly positive coefficient is due to big outsiders not experiencing significant negative moves. The interactions with *n* (used in Chapter 2) have been removed for simplicity, inclusion causes colinearity problems, see Section 3.6: Table 3.4.

To derive predictions about the coefficients with respect to the hypotheses stated at the end of Section 3.2, consider first the expected win-probability of horse *i* given the information set *I* at the opening of the market ($t = 0$):

$$E\left[P(WIN_i \mid I_0)\right] = \varphi_0 + \delta_0 OPR_i \,. \tag{3.4}$$

If opening probabilities are unbiased estimates of the true win-probabilities, i.e. without a FL-Bias, then $\varphi_0 = 0$ and $\delta_0 = 1$. Thus, a horse with a one percentage point higher implied probability of winning indeed has a one percentage point higher chance of winning. At the start of the race ($t = 1$), the information set is richer, so for starting odds:

$$E\left[P(WIN_i \mid I_1)\right] = \varphi_1 + \delta_1 SPR_i \tag{3.5}$$

The change in the expected win-probability between the opening and the closing of the market is then:

$$\Delta E\left[P(WIN)\right] = \varphi_1 - \varphi_0 + \delta_1 SPR - \delta_0 OPR \,. \tag{3.6}$$

In a *perfect information world*, i.e. without any biases, $\delta_0 = \delta_1 = 1$ (hence the φs will be equal to zero), the change in the expected win-probability between the formation and the cessation of the market is *DPR*; equation (3.6) simplifies to equation (3.2). If the market reacts correctly, as in Hypothesis 3, then a regression of equation (3.3) yields $\alpha = 0$, $\beta_1 = \beta_2 = 1$. A $\beta_2 \neq 1$ is indicative of the market mechanism failing to impound new information correctly. The last relation is of particular interest since it tells us whether the market under- or over-reacts to new information. Under the hypothesis of *market under-reaction* (Hypothesis 1), $\beta_2 > 1$; the actual change in the probability of a horse winning is not fully incorporated by *DPR*. Similarly for market over-reaction, $\beta_2 < 1$.

The coefficient of β_1 is also of interest, since if $\beta_1 \neq 1$, then $\beta_2 \neq 1$ is also consistent with the market correcting for an initial mispricing at the start of the market. In the presence of *a FL-Bias whose strength weakens towards the start of the race* as demonstrated in Chapter 2, $\delta_0 > \delta_1 > 1$. Given equation (3.3), in order to *be consistent with the market reacting correctly,* β_2 *can be greater than or less than unity* depending on whether the horse is a favourite or a longshot.

Under a situation with a correction of an initial mispricing, the observed *DPR* can be viewed as the sum of two constituent parts:

$$DPR = DPR_I + DPR_C, \tag{3.7}$$

where DPR_I corresponds to the part of the observed *DPR* caused by information (the part that this investigation is interested in), and DPR_C corresponds to the part which is responsible for the (partial) correction of any initial bias. (Under a situation with a full correction of the initial bias, $\delta_0 > \delta_1 = 1$).

For the sake of the following analysis, a longshot is defined as a horse whose implied win-probability is over-estimated according to a regression of the objective win-probability on the implied win-probability[13]. For *longshots*, DPR_C is negative since their implied win-probabilities are initially overestimated using I_0, and *DPR* is an under-estimate of DPR_I. If DPR_I had a one-to-one relationship with the change in the observed win-probability (i.e. the market reacting correctly), $\beta_2 > 1$ in order to

[13] Note when regressing objective and subjective probabilities, the crossover distinguishing between favourites and longshots will occur at $OPR = 1/\bar{n}$, see footnote 15 from Chapter 1.

compensate for the negative correction. Similarly for *favourites*, $\beta_2 < 1$ in order for the market to react correctly. There is a *switchover* of β_2 between favourites and longshots. In other words a move which implies a +5% point *DPR* on a favourite should predict a smaller change in the actual expected winning probabilities than the same *DPR* change on a longshot because the favourite's move partially corresponds to a correction. In what follows, this will be referred to as the *favourite-longshot effect*. Regressions of specification (3.3) will thus be run separately for favourites and longshots to test for a switchover (see Table 3.2), its existence will support the hypothesis that the degree of market reaction is correct. Further support for this is included in Section 3.A.2 where the original probabilities are converted into estimated (unbiased) probabilities to generate an adjusted version of *DPR, ADPR*.

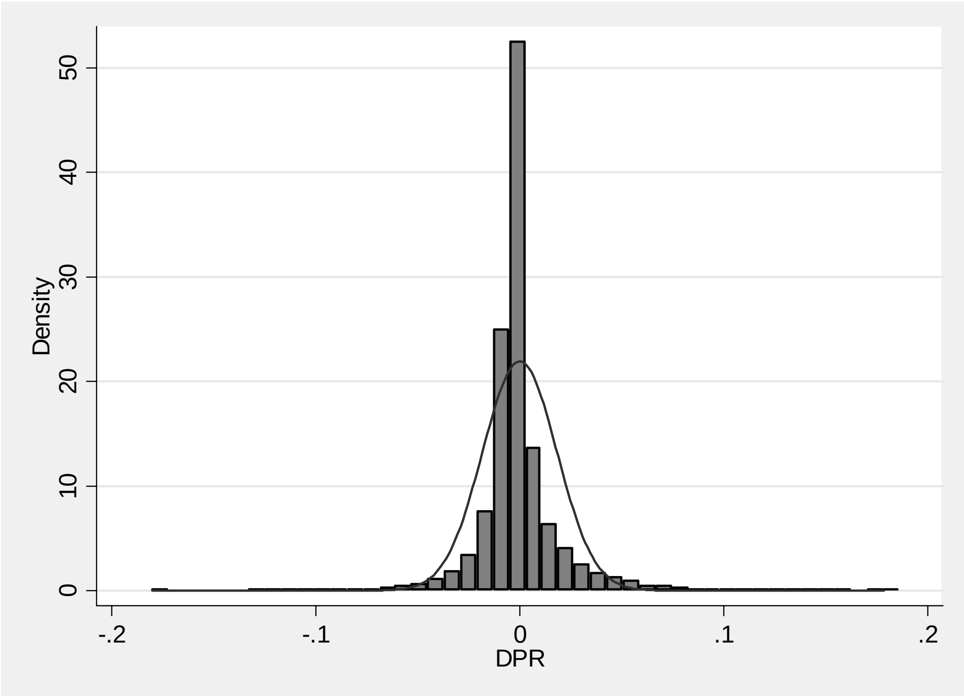
The market move measure adopted here has two additional benefits relative to Law and Peel's measure. First, the interpretation of the coefficients is more intuitive. For example, the estimated coefficients in equation (3.1) do not carry an immediately intuitive meaning; it is not clear what an increase *mop* by 0.01 causing returns to fall by 0.009 means. *DPR* is similar to Law and Peel's measure[14] yet it has an intuitive interpretation as the change in implied winning probability.

Second, the measure avoids a problem that can be best seen as an example. Consider a situation where horse *i*'s odds contracted from 4/1 to 2/1, and all the other horses in the race had odds of 5/1 and their odds remained unchanged after *i*'s move. The implied win-probabilities of these outsiders have implicitly fallen. Law and Peel's measure would capture the move on horse *i*, but record the other horses to be non-movers even though there is a negative move in the implied win-probabilities. *DPR* does not suffer from this problem because it takes into account the overround. The distribution of *DPR* is presented in Figure 3.1. There is strong kurtosis around the value of zero[15]. This indicates that the majority of horses are non-movers.

[14] Consider four moves, 100/1 to 50/1, 10/1 to 13/2, 8/1 to 5/1 and 2/1 to 6/4, the value of mop is 0.01, 0.05, 0.06 and 0.11 respectively. The value of *DPR*, ignoring the overround is 0.01, 0.04, 0.06 and 0.07 respectively. A linear regression of *dpr* on *mop* yields the relationship: $dpr = 0.002 + 0.481mop$ ($R^2 = 0.787$).

[15] *DPR* has a mean value of zero (0.0000064) with a maximum value of 0.178 and a minimum value of -0.180, the variable has a variance of 0.00033, skewness of 0.754 (the right tail is more pronounced than the left tail) and a kurtosis of 10.91; the 5th percentile has a value of -0.0247 and the 95th percentile has a value of 0.0318.

FIGURE 3.1
THE DISTRIBUTION OF DPR



The overlaying curve represents a normal density function with the same mean and variance as the sample.

The linear probability model in equation (3.3) is the starting point of the analysis. Different specifications of this model are considered in order to investigate whether the effect of *DPR* varies across the *DPR* spectrum. The data is also split to investigate whether over-reaction is more likely to occur in the lower class events. As in the previous chapter, the regressions use clustered (by the race) standard errors because of dependence of the error term; if one horse wins, the other horses in that race will be losers. The *computational issue* discussed in Chapter 2 also arises here and is dealt with using bootstrap robustness tests (see Section 3.6). In the appendix, a logit model is utilised, and a specification which eliminates the favourite-longshot effect are also presented.

3.5 – EMPIRICAL RESULTS: LINEAR PROBABILITY MODEL

Table 3.1 shows the results from the linear probability model in (3.3). The data display a FL-Bias as in the previous chapter. The coefficient of *OPR* is always significantly above unity. *Ignoring the favourite-longshot effect*, the estimates of the coefficient of *DPR* support the view that the market reaction is correct because the hypothesis that they are equal to unity cannot be rejected. However, the results are not conclusive because of the relatively large standard errors. On the one hand, the hypothesis that the coefficients are equal to unity cannot be rejected, on the other hand hypotheses of the market under- or over-reaction (for example null hypotheses of β_{DPR} = 1.30 and β_{DPR} = 0.70 respectively) cannot be rejected either. There are no significant differences across classes. The point estimates for β_{DPR} suggest that under-reaction is more prevalent in the lower classes, market participants are not understanding that moves in these events are potentially more informative.

TABLE 3.1

REACTION TO MARKET MOVERS: LINEAR PROBABILITY MODEL (1)

Dependent Variable: Win (Binary)

Independent Variable	**i) ALL**	ii) A & B	iii) C & D	iv) E	v) F & G
			Class		
OPR	1.105	1.130	1.085	1.109	1.173
	(0.018)***	(0.052)***	(0.025)***	(0.040)***	(0.054)***
DPR	0.961	0.880	0.782	1.320	1.006
	(0.126)***	(0.438)**	(0.186)***	(0.245)***	(0.284)***
constant	-0.010	-0.122	-0.008	-0.009	-0.014
	(0.002)***	(0.005)**	(0.002)***	(0.003)***	(0.004)***
Races	3590	451	1719	863	557
Runners	39137	4798	17245	10073	7021
R^2	0.140	0.130	0.154	0.125	0.127

Notes: Regressions with clustered (by race) standard errors, in parentheses. *OPR* is the opening implied win probability and *DPR* is the measure of the market move as defined in the text. *,**,*** indicate significance at the 10%, 5% and 1% level respectively. The coefficient on *OPR* is always significantly different to unity and the coefficient on *DPR* is never significantly different to unity at the 5% level for these regressions.

As mentioned in the previous section, because of the *favourite-longshot effect*, estimates of β_{DPR} are biased estimates of β_{DPRI}. As well as measuring the move caused by information, the observed *DPR* also consists of the part of the move which represents a correction of the initial bias. Under a situation where *DPR* reflects a (partial or full) correction of the initial FL-Bias, if the degree of market reaction is correct, β_{DPR}'s magnitude relative to unity will switch over for favourites (β_{DPR} less than unity) and for longshots (β_{DPR} greater than unity). In this instance, favourites and longshots are defined as horses with an *OPR* of greater than and less than the crossover point, $1/\bar{n}$, where the subjective and objective win-probabilities are equal according to the linear model.

TABLE 3.2
REACTION TO MARKET MOVERS: LINEAR PROBABILITY MODEL (2)

			Class		
	All	A & B	C & D	E	F & G
Predicted Switchover	0.092	0.094	0.100	0.086	0.080
Actual Switchover	0.092	0.094	0.100	0.086	0.079
β DPR (Below Switchover)	1.310	1.431	1.020	1.356	1.288
Longshots	(0.176)***	(0.670)**	(0.249)***	(0.382)***	(0.374)***
β DPR (Full Sample)	0.961	0.880	0.783	1.320	1.006
	(0.126)***	(0.438)**	(0.186)***	(0.245)***	(0.284)***
β DPR (Above Switchover)	0.888	0.785	0.730	1.313	0.944
Favourites	(0.145)***	(0.494)	(0.215)***	(0.283)***	(0.330)*

Rubric: Same as Table 3.1 but only the coefficient of *DPR* is displayed

The results from these regressions are shown in Table 3.2. For all classes, there is a switchover: $\beta_{DPRLongshot} > \beta_{DPRPooled} > \beta_{DPRFavourite}$. For the pooled regression, the estimate of $\beta_{DPRLongshot}$ is significantly greater than the point estimate of $\beta_{DPRFavourite}$. This supports the hypothesis that there is a *favourite-longshot effect*. All the estimates (split by class) except for the estimates for Class E races *support the hypothesis that the degree of market reaction is correct* ($\beta_{DPRLongshot} > 1$ and $\beta_{DPRFavourite} < 1$). For the point estimates from Class E races, $\beta_{DPRLongshot} > \beta_{DPRPooled} > \beta_{DPRFavourite} > 1$. Such an estimate of β_{DPR} for favourites greater than unity is inconsistent with the hypotheses that the market reacts correctly or market over-reaction. The point estimates suggest that under-reaction occurs in Class E races. However, the hypothesis that estimates from all the regressions are equal to unity cannot be rejected.

3.6 – ROBUSTNESS CHECKS OF THE LPM

In this section several robustness checks are applied. The bootstrapping method employed in Chapter 2 is used to investigate whether the results are affected by the computational issue. Additionally, the effect of different sized fields, variation among drifters and plungers, the impact of attendance levels and the number of bookmakers present, and a specification omitting non-movers are all investigated.

A *computational issue*, where it was noted that when more than one horse from the same race have similar prices, only one of the horses can win, thus the win totals (or relative win frequencies) may be unrepresentative of a typical horse with those prices, was discussed in Chapter 2. This problem recurs again under the specifications such as equation (3.3). In this instance, not only is the estimated coefficient of *OPR* subject to the problem, but so is the estimated coefficient on *DPR*. When more than one horse from the same race have similar values of *DPR*, only one of the horses can win, thus the win totals (or relative win frequencies) may be unrepresentative of a typical horse with those values of *DPR*[16]. To cater for this potential issue, as with the previous chapter, a *bootstrap* is employed where one horse is drawn from each race to form a sub-sample and parameter estimates are made using the sub-sample, the process is then repeated many times (sampling with replacement) and the mean of the estimated coefficients are recorded.

The results from using the bootstrapping method are presented in Table 3.3. The point estimates are very close to the full sample results, suggesting that the impact of the computational issue is negligible. The point estimates for β_{OPR} and β_{DPR} using the bootstrap are all lower than the respective estimate using the original full sample specification, but once again they are not significantly different compared to the original estimates and it is possible to conclude that the qualitative results once again hold.

[16] Technically this issue also potentially affects the estimates or returns using rules based on the measure of the market move (such as those discussed in Section 1.3), but since the classifications were only applied using five bins, any filtering such as the filtering used in Chapter 2 will result in most of the observations being dropped.

TABLE 3.3

REACTION TO MARKET MOVERS: BOOTSTRAP RESULTS

Dependent Variable: Win (Binary)

Independent Variable	**i) ALL**	ii) A & B	iii) C & D	iv) E	v) F & G
			Class		
OPR (OLS)	1.105	1.130	1.085	1.109	1.173
OPR (Bootstrap)	1.087	1.105	1.081	1.063	1.145
	(0.057)	(0.164)	(0.074)	(0.129)	-0.173
DPR (OLS)	0.961	0.880	0.782	1.320	1.006
DPR (Bootstrap)	0.814	0.853	0.634	1.181	0.965
	(0.285)	(0.934)	(0.407)	(0.616)	(0.743)
constant (OLS)	-0.010	-0.122	-0.008	-0.009	-0.014
constant (Bootstrap)	-0.010	-0.013	-0.010	-0.007	-0.012
	(0.005)	(0.016)	(0.008)	(0.011)	(0.014)

Notes: Interpretation of the parameter estimates are the same as Table 3.1 but only parameter estimates are shown. Original OLS point estimates, and the means and standard deviations (in parentheses) of 1000 estimates from the bootstrapping shown.

One may expect to observe larger moves in smaller field races because the value of inside information should be greater (in a large field race, if an agent knows that his horse is better than the market predicts, there are more competitors who are able to spring a surprise). Thus herding behaviour is more likely to occur in races with smaller fields. If over-reaction is more pertinent for races with large fields then bettors are engaging in herding behaviour in races which the value of inside information is lower. Whether or not under- or over-reaction is more likely to occur in small or large fields is not clear and an uninvestigated issue.

To investigate the impact of the *number of runner*s, an interaction of the number of runners and *DPR* is added to the specification in equation (3.3):

$$P(Win) = \alpha + \beta_1 OPR + \beta_2 DPR + \beta_3 n * OPR + \beta_4 n * DPR \tag{3.8}$$

This specification allows for the investigation of the marginal impact on the degree of market reaction for every extra runner. If β_4 is positive, the fewer the runners in a race, the lower the marginal effect, thus over-reaction is more pertinent. A negative β_4 suggests that over-reaction is more pertinent in the races with larger fields. The results are presented in Table 3.4, column *i*. The coefficient on *DPR* is no longer significant, only the *n*DPR* interaction term is significant and positive suggesting more over-reaction in races with smaller fields. In a 5 runner race, the estimated coefficient on *DPR* is around 0.36, whereas in a 20 runner race, its estimated value is 1.32. This result, if significant, would support the hypothesis of over-adjustment for small fields. The problem with this specification however, is multicolinearity between the two independent variables. The correlation coefficient between *DPR* and *n*DPR* is 0.927.

TABLE 3.4

THE EXTENT OF MARKET REACTION FOR DIFFERENT SIZE FIELDS

Dependent Variable: Win (Binary)

Independent Variable	i)	ii)
OPR	1.005 (0.023)***	1.002 (0.023)***
N*OPR	0.013 (0.003)***	0.014 (0.003)***
DPR	0.341 (0.345)	0.863 (0.151)***
N*DPR	0.061 (0.030)**	
HIGHN*DPR		0.350 (0.270)
constant	-0.014 (0.002)***	-0.014 (0.002)***
Races	3590	3590
Runners	39137	39137
R^2	0.141	0.141

Rubric: Same as Table 3.1. HIGHN is a dummy variable equal to unity in races with more than 12 runners.

Table 3.4, column *ii* reports the result from a regression that attempts to get around the multicolinearity issue by interacting DPR with a dummy variable indicating races with 13 runners or more. The results suggest that the degree of market reaction is the same for races with 13 or more runners and races with 12 or fewer runners. However, the point estimate of the coefficient on the interaction is also positive, thus backing up the results from specification *i*. Multicolinearity is also a problem in this instance because the correlation coefficient between the change in probability measure and the interaction term is just over 0.5.

To test whether the extent of market reaction varies for horses which experience positive and negative market moves, the regressions in Table 3.1 have been run with the sample split for horses with positive and negative *DPR*. The results are presented in Table 3.5 and indicate a pattern akin to the results in Table 3.1 suggesting that the relationship for positive and negative market movers is similar. The size of the standard errors hinders this approach and also indicates that for the smallest sub-samples (Class A&B and F&G races) the coefficients on *DPR* are not significantly different to zero even at the 10% level.

A result worthy of note is that for the Class F&G races. The point estimate on the coefficient of *DPR* for the negative movers sub-sample is 0.959 (standard error: 0.643). However, the coefficient for the positive mover sub-sample is only 0.212 (standard error: 0.464), which is much lower than for any other class suggesting over-reaction for plungers in the low class races. An increase in the win-probability implied by the odds of 1% point translates into the horse's expected win-probability improving by just 0.212% points. However the hypothesis that this coefficient equal to unity cannot be rejected either.

To investigate any differences in the degree of market reaction caused by differences in the *attendance levels* and the *number of bookmakers* present, an interaction between *DPR* and the recorded attendance level or the number of bookmakers is added to equation (3.3). A positive (negative) coefficient on *Attendance*DPR* would indicate that there is more under- (over-)reaction at meetings with higher attendance levels. A negative coefficient would be consistent with a significant number of uninformed bettors at large meetings who herd onto plungers. In terms of the effects of the number of bookmakers present, one would expect that in markets with fewer bookmakers, prices would be more responsive to large bets because of the lack of depth in these markets. Under such a situation, the coefficient on *Bookmakers*DPR* would be positive.

TABLE 3.5

REACTION TO MARKET MOVERS: LINEAR PROBABILITY MODEL (3)

Dependent Variable: Win (Binary)

DPR ≤ 0 ONLY

Independent Variable	**i) ALL**	ii) A & B	iii) C & D	iv) E	v) F & G
			Class		
OPR	1.153^	1.140	1.181^	1.089	1.120
	(0.039)***	(0.112)***	(0.053)***	(0.086)***	(0.118)***
DPR	1.142	0.869	1.132	0.910	0.959
	(0.259)***	(0.931)	(0.373)***	(0.497)*	(0.643)
constant	-0.013	-0.013	-0.011	-0.012	-0.014
	(0.002)***	(0.007)*	(0.004)***	(0.005)**	(0.006)**
Races	3590	451	1718	863	557
Runners	20851	2572	9033	5521	3725
R^2	0.117	0.118	0.135	0.088	0.098

DPR > 0 ONLY

Independent Variable	**i) ALL**	ii) A & B	iii) C & D	iv) E	v) F & G
			Class		
OPR	1.073^	1.099	1.035	1.063	1.295^
	(0.034)***	(0.108)***	(0.046)***	(0.074)***	(0.095)***
DPR	1.040	1.130	0.927	1.648	0.212
	(0.209)***	(0.745)	(0.304)***	(0.422)***	(0.464)
constant	-0.007	-0.013	-0.006	-0.009	-0.010
	(0.002)***	(0.007)*	(0.004)	(0.005)*	(0.006)*
Races	3588	451	1717	863	557
Runners	18286	2226	8212	4552	3296
R^2	0.157	0.140	0.171	0.148	0.147

Notes: As Table 3.1, but the data is split into positive and negative movers. Regressions with standard errors clustered by race in parentheses.
*,**,*** indicate significance at the 10%, 5% and 1% level respectively. ^ indicates that the coefficient is significantly different to unity at the 5% level (applicable to *OPR* and *DPR* only).

The results from these regressions are presented in Table 3.6. The correlation coefficient between the interactions and *DPR* are 0.65 for attendance and 0.82 for the number of bookmakers, so the parameter estimates are affected adversely by multicolinearity. The insignificant interaction terms for both factors suggest that the degree of market reaction is independent of the attendance levels and the number of bookmakers. Even at smaller meetings, where the race goers are more likely to be knowledgeable, bookmakers adjust odds in the same manner as at the larger meetings where the majority of bettors are likely to be uninformed. Also, at meetings with few bookmakers, there is no evidence that bookmakers over-react to significant bets placed by bettors.

TABLE 3.6
THE IMPACT OF ATTENDANCE LEVELS AND THE NUMBER OF ON-COURSE BOOKMAKERS ON THE EXTENT OF REACTION TO MARKET MOVERS

Dependent Variable: Win (Binary)

Independent Variable	Attendance	Bookmakers
OPR	1.110 (0.019)***	1.110 (0.019)***
DPR	0.964 (0.168)***	1.016 (0.219)***
ATTENDANCE*DPR	0.002 (0.020)	
BOOKMAKERS*DPR		-0.001 (0.004)
constant	-0.010 (0.002)***	-0.014 (0.002)***
Races	3498	3498
Runners	38163	38163
R^2	0.141	0.141

Rubric: Same as Table 3.1. Attendance is the recorded attendance for the meeting (measured in thousands) and bookmakers is the number of on-course bookmakers present at the meeting.

The success of the prediction that the market reacts correctly could be attributable to the majority of horses which have a small *DPR*. If these non-movers' chances of winning do not change, they will put pressure on making β_{DPR} equal to unity. In order to test this claim, regressions are run omitting runners with $-0.01 < DPR < 0.01$; (26,549 out of the 39,137 horses are omitted), two further specifications are

TABLE 3.7
THE EXTENT OF MARKET REACTION OMITTING NON-MOVERS

Pooled	DPR < -0.01 & DPR > 0.01	DPR < -0.01	DPR > 0.01
βDPR			
1.105	0.946	1.171	1.032
(0.018)***	(0.130)***	(0.367)***	(0.301)***
Runners	12588	6418	6170

Rubric: Same as Table 3.1, but only the coefficient of *DPR* is displayed. The hypothesis that the coefficient of *DPR* is equal to unity cannot be rejected at the 5% level of significance.

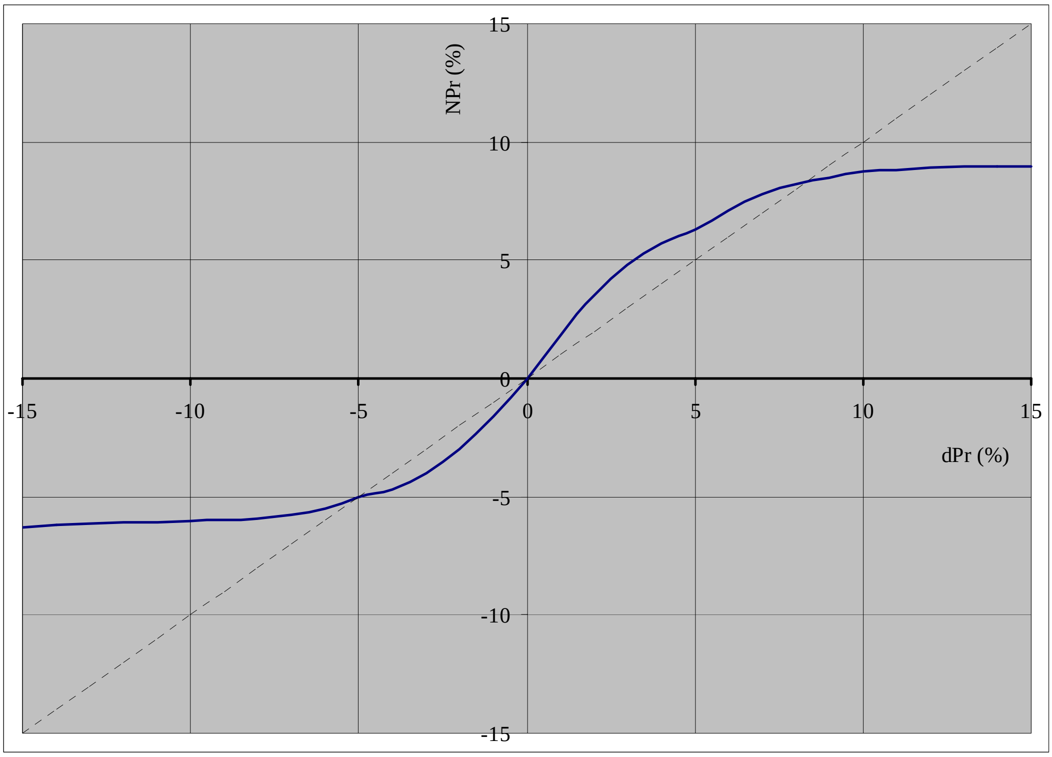
run keeping only the positive and negative movers. The results are presented in Table 3.7, and once again, the estimates of the coefficient on *DPR* are not significantly different from unity. The hypothesis that the estimates of *DPR* are close to unity because of non-movers can be rejected because the regressions omitting the non-movers yield the same results.

3.7 – NON-LINEAR SPECIFICATIONS

So far the analysis assumes that the effect of a market move is linear, which rules out the possibility that the marginal effect differs across the *DPR* spectrum. The marginal effect for large moves could be lower than the marginal effects of small moves. Given a plunger that is subject to herding behaviour, any further increases in *DPR* are unlikely to be justified, thus the slope coefficient on *DPR* in this region is likely to be lower. Any such effect can be captured using non-linear specifications of equation (3.3).

With *non-linear* models, there are two ways in which to interpret the regression output, considering the *marginal effect*, or the *total effect* of the move. In the linear probability model (3.3), the *marginal effect* (i.e. 'is the extra change an under- or over-reaction?') is simply β_2, for non-linear models it will be the derivative of the probability function with respect to *DPR*. Figure 3.2 shows a hypothetical relationship between *DPR* and the change in expected probability of the horse winning compared to a horse whose *DPR* is equal to zero, (*NPR*). In other words the point of concern here is the slope of the curve in Figure 3.2 and whether it is greater than, less than or equal to unity[17]. If the slope is greater than unity, there is under-reaction at the margin (implied probability changes fail to fully impound the increase in the chances of winning) and a slope less than unity implies market over-reaction at the margin.

FIGURE 3.2
A HYPOTHETICAL RELATIONSHIP BETWEEN *DPR* AND *P(WIN)*



Rubric: *NPR* is the change in the expected probability of the horse winning compared to a horse whose *DPR* is equal to zero. The slope of the curve measures the marginal effect of an implied probability change. In the absence of a *favourite-longshot effect*, being under the 45° line indicates (total effect) market over-reaction and being under it indicates under-reaction.

To analyse the *total effect* of the move (i.e. ‘is *DPR* itself an under- or over-reaction?’), the function from the regression output could be normalised such that a horse with *DPR* equal to zero is assigned a zero percent change in its chances of winning and plotted in normalised win-probability (*NPR*) and *DPR* space. To illustrate how to interpret such a graph, consider again Figure 3.2. The 45 degree line maps the locus of all points where *DPR* and the change in the estimated win-probability are of equal magnitude. In the absence of a FL-Bias (causing a *favourite-longshot effect*), if the curve resulting from the regression output lies above the 45 degree line, under-reaction is present. For example in Figure 3.2 at a *DPR* of 2%,

[17] In some cases, we would not be comparing against unity because *there may be a favourite-longshot effect*, but this is not as important as we are concerned with marginal changes.

NPR is 2.5%, this indicates that a horse with *DPR* of 2 percentage points has a 2.5 percentage points higher chance of winning compared to a horse with *DPR* of 0, (the market has under-reacted). The remainder of Figure 3.2 depicts a situation where there is under-reaction to small moves and over-reaction to large moves in both directions, (i.e. bookmaker over-reaction or herding behaviour *à la* Law and Peel (2002), where bets on small positive moves earn superior returns to bets on the largest positive moves).

To remove the linearity constraint, DPR^2 and DPR^3 are added to the LPM in (3.3). The results in Table 3.8 show that the squared and cubic terms are insignificant and thus the standard LPM is favoured[18]. Regressions with only *DPR* and DPR^2, and *DPR* and DPR^3 as the independent variables (not shown) also yield insignificant coefficients on the squared and cubic terms. Also regressions (not shown) with only DPR^2 or DPR^3 replacing *DPR* perform worse than (3.3) based on the R^2 values.

TABLE 3.8

REACTION TO MARKET MOVERS: CUBIC LINEAR PROBABILITY MODEL

Dependent Variable: Win (Binary)

Independent Variable	**i) ALL**	ii) A & B	iii) C & D	iv) E	v) F & G
OPR	1.109^	1.112	1.089^	1.110^	1.192^
	(0.021)***	(0.064)***	(0.029)***	(0.048)***	(0.066)***
DPR	1.032	0.723	0.880	1.195	1.279
	(0.148)***	(0.524)	(0.212)***	(0.302)***	(0.339)***
DPR2	-0.838	5.436	-0.911	-1.248	-3.750
	(2.206)	(9.736)	(2.932)	(4.516)	(5.799)
DPR3	-11.666	57.147	-10.803	52.441	-78.328
	(28.479)	(169.462)	(34.968)	(59.479)	(81.356)
constant	-0.010	-0.012	-0.009	-0.009	-0.014
	(0.002)***	(0.005)**	(0.002)***	(0.004)***	(0.004)***
Races	3590	451	1719	863	557
Runners	39173	4798	17245	10073	7021
R^2	0.140	0.130	0.155	0.125	0.127

Notes: As Table 3.1. *DPR2* and *DPR3* are the squared and cubed exponents of *DPR*. Regressions run with clustered (by race) standard errors, in parentheses.
*,**,*** indicate significance at the 10%, 5% and 1% level respectively. ^ indicates that the coefficient is significantly different to unity at the 5% level (applicable to *OPR* and *DPR* only).

[18] An unreported regression using the bootstrap method outlined earlier gives very similar point estimates of the coefficients and the quadratic and/or the cubic terms are also not statistically significant.

Another non-linear approach is to consider a logit model. The problem with any analysis utilising this approach is that a functional form (i.e. the logistic function), is imposed onto the results. The results from this approach should therefore be treated with caution. This matter is discussed further in the Appendix (3.A.1).

3.8 – CONCLUSIONS

This chapter analyses the informational efficiency of changes in quoted odds. Three hypotheses are tested for: market under-reaction, market over-reaction and the market reacting correctly.

Many previous investigations use filter rules to compare the relative returns of backing plungers and drifters. Conclusions that it is better or worse to back plungers or drifters based on this methodology are premature. In some cases, drifters are more likely to be favourites and plungers are more likely to be longshots or vice versa. In the presence of a FL-Bias, the returns from backing drifters will automatically be superior to backing longshots. To overcome this problem, a linear probability model is employed but a problem using probability measures affected by a favourite-longshot bias to investigate the extent of reaction to market moves is identified. In an environment where implied win-probabilities exhibit a favourite-longshot bias which narrows towards the start of the race, the measure of the market move is a biased estimate of the market move caused by new information because the move also consists of a correction of the initial bias in the odds; what is referred to as a *favourite-longshot* effect. Under the hypothesis that the degree of market reaction is on average correct, the presence of the bias predicts a *switchover* on the coefficient of interest between favourites and longshots. This is found to be the case from the empirical analysis, supporting the hypothesis that the market is correctly impounding new information. Moreover, a solution utilising estimated probabilities free from the bias (presented in the Appendix 3.A.2) yields similar results to the biased probability measure.

Despite the bookmaker market for bets being a place where bettor sentiment plays a role, the adjustment of odds by bookmakers does seem to be appropriate. If the market attributes an X% point increase in the probability of horse *i* winning between the formation and cessation of the market, then horse *i*'s observed probability of winning increases by X% points. This result holds for all classes of races, and meetings with different levels of attendance and on-course bookmakers. This finding remains robust even when allowing for specifications such as splitting the sample for positive and negative movers and adding squared and cubic terms into the LPM. However, because of the relatively large standard errors on the estimates, hypotheses of market under- or over-reaction cannot be rejected either.

The *computational issue* outlined in the previous chapter in which there can only be one possible winner from a race also recurs in this chapter. When two or more horses from the same race have similar values of *DPR*, only one of the horses can win, thus the win totals (or relative win frequencies) may be unrepresentative of a typical horse with those values of *DPR*. Once again, a bootstrap method involving only drawing one horse from each race is employed and it is found that the effects caused by this issue are negligible.

3.A.1 – APPENDIX: ANALYSIS WITH A LOGIT MODEL

The linear probability model in (3.3) suffers from the fact that the dependent variable is binary. This section runs the analysis with a simple logit model. The problem with using the logit model is that the functional form of the probability density function will be restricted to be the logistic function. For a simple logit model with the dependent variables in the form of (3.3), the marginal effect is:

$$MFX = B_2\left[\frac{e^{-Z}}{(1+e^{-Z})^2}\right], \text{ where } Z = A + B_1 OPR + B_2 DPR \,. \tag{3.9}$$

In this case, this marginal effect is equivalent to β_2 in the linear probability model, but it will not be constant. Running this logit regression, the following relationship is obtained from the full sample of 39,137 observations from 3,590 races (clustered, by race, standard errors in parentheses, *** indicates significance at the 1% level):

$$P(Win) = \frac{e^z}{1+e^z}$$

$$z = \underset{(0.024)^{***}}{-3.359} + \underset{(0.008)^{***}}{6.299OPR} + \underset{(0.163)^{***}}{8.507DPR} \tag{3.10}$$

TABLE 3.A.1
MARGINAL EFFECTS FOR THE LOGIT MODEL

	OPR (%)						
DPR (%)	5	10	20	30	40	50	60
-10	-	-	0.527^	0.978	1.447^	1.548^	1.171
	-	-	(0.037)***	(0.089)***	(0.176)***	(0.251)***	(0.238)***
-5	-	0.336^	0.675^	1.167	1.546^	1.450	0.981
	-	(0.035)***	(0.076)***	(0.089)***	(0.220)***	(0.239)***	(0.184)***
0	0.032^	0.439^	0.847	1.344	1.574^	1.297	0.797
	(0.045)***	(0.065)***	(0.125)***	(0.199)***	(0.235)***	(0.120)***	(0.129)***
5	0.398^	0.569^	1.034	1.483^	1.525^	1.115	0.631^
	(0.076)***	(0.106)***	(0.181)***	(0.237)***	(0.216)***	(0.145)***	(0.083)***
10	0.520^	0.725	1.223	1.563^	1.408^	0.925	0.490^
	(0.117)***	(0.157)***	(0.035)***	(0.246)***	(0.170)***	(0.092)***	(0.048)***

Standard errors, clustered by race, in parentheses. A coefficient of 1 indicates that the marginal change in the probability relates to an identical marginal change in the chances of the horse winning, a coefficient greater (less) than 1 indicates (marginal) under-reaction (over-reaction).
^ indicates that the coefficient on *DPR* is significantly different to unity at the 5% level. The censoring is due to longshots not being able to have negative probabilities.

The *MFX* for various levels of *OPR* is displayed in Table 3.A.1, the logit model suggests that there is variation of the *MFX* across different *OPR* and *DPR*. The variation is due to the logit model imposing its functional form as will be evident later on. The results suggest that there is marginal over-reaction to moves on longshots, this result is plausible because plunges on longshots could be in part the result of herding behaviour by bettors suspecting an informed plunge. However, the larger the move on these longshots, the closer the coefficient is to unity. The marginal reaction to positive moves on horses with *OPR* of around 20% and 50% is correct. There is marginal under-reaction to positive moves on horses with OPR of 30% and horses with 40% starting probability. The picture will become clearer when the overall move is considered.

Plotting this function onto (*DPR*,*NPR*) space as discussed in Section 3.7, Figure 3.A.1 is obtained. The two panels show different views of the same figure. The curved plane measures the normalised change in the expected win-probability relative to the implied change for all *OPR* in the relevant range. If the market reacts correctly, then in the absence of a FL-Bias the curved plane should lie directly on the 45 degree plane for all *OPR*, Figure 3.A.2 takes cross sections from Figure 3.A.1 to allow this to be done more easily.

It is clear from Figure 3.A.1 that the function exhibits significant variation across the range of *OPR*. The function is relatively flat for low and extremely high opening implied win-probability (*OPR*) horses and steep for horses with 40-50% *OPR* (favourites). Note from the discussion in Section 3.7 that the flat regions signify marginal over-reaction and the steep regions under-reaction, there seems to be over-reaction for extreme favourites and longshots. The figure also clearly demonstrates the functional form that is being imposed; this shape arises because it is that of the logistic function.

To investigate the results implied by the logit model more closely, cross-sections of the curved plane are investigated for various levels of *OPR*, these curves are shown in Figure 3.A.2. Starting with the longshots with opening probability of 10% (panel a), which is close to the overall mean probability of 9.2%, the curve is relatively flat and always between the 45 degree line and the *x*-axis. There is (both marginal and total effect) over-reaction to positive and negative moves on these longshots. For example, a horse with 5% point *DPR*'s chances of winning are only 2.5% points better than a horse with zero *DPR*'s chances of winning, indicating a (total effect) over-reaction for positive moves. This is consistent with the favourite-longshot effect outlined in Section 3.4; these horses' win-probabilities are already overestimated by about half a percentage point (see the previous chapter). The curve is censored to rule out negative probabilities. The curve becomes steeper towards the right hand side indicating that the marginal effect of big plunges on these longshots increases.

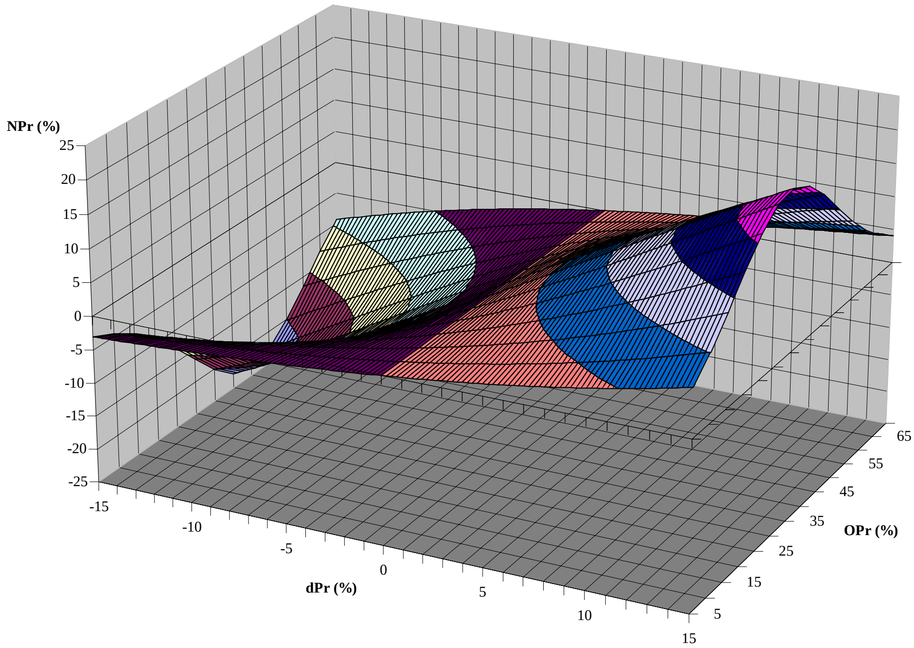
For horses with *OPR*=20%, the curve lies on the 45 degree line, so the market reacts correctly. It is worth noting that traditionally this is the region where estimates of the true winning probability are unbiased (i.e. there is no FL-Bias), plots of implied win probability against the true win probability cross the 45 degree line in this region (see Cain, Law and Peel (2003)), so there is no favourite-longshot effect acting here. However, for negative moves, there is still over-reaction.

For *OPR*=30% (odds of around 2/1), negative moves are quite accurate, however the slope is continually increasing from around 0.5 for large negative moves to unity when *DPR* is equal to zero. For plungers, a 10% increase in *DPR* leads to the horse actually having a 15% higher chance of winning compared to if *DPR* was zero, the market has not reacted enough. The case is similar for plunges on horses with *OPR*=40%. For horses with *OPR* in this vicinity with positive moves, the situation is consistent with the *favourite-longshot effect*, the under-reaction (the actual increase in the chances of the horse winning is greater than that implied by the market) helps compensate for the initial underestimation of the probability of about 5% points. Under-reaction is more evident now for negative moves.

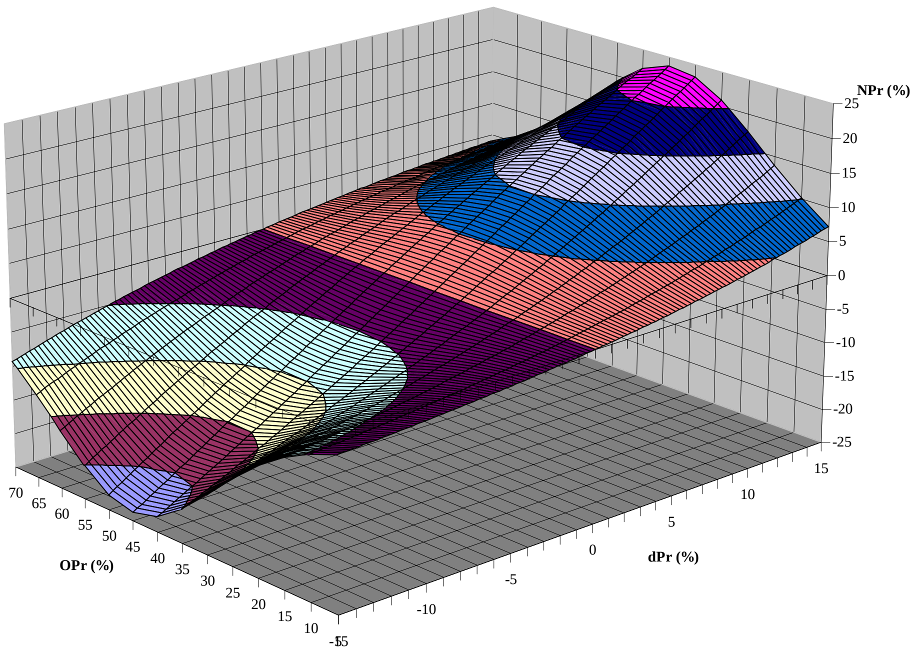
For *OPR*=50% positive moves are very accurate and negative moves exhibit similar under-reaction to 40% *OPR* horses. For positive moves the pattern is similar to that hypothesised in Figure 3.2, where small moves are under-reactions and larger moves become overreactions, although plunges of over +15% *DPR* are rare in the dataset. Finally for odds-on favourites with *OPR*=60%, negative moves are accurate and any positive moves are over-reactions.

FIGURE 3.A.1
MARKET REACTION WITH LOGIT MODEL

a)

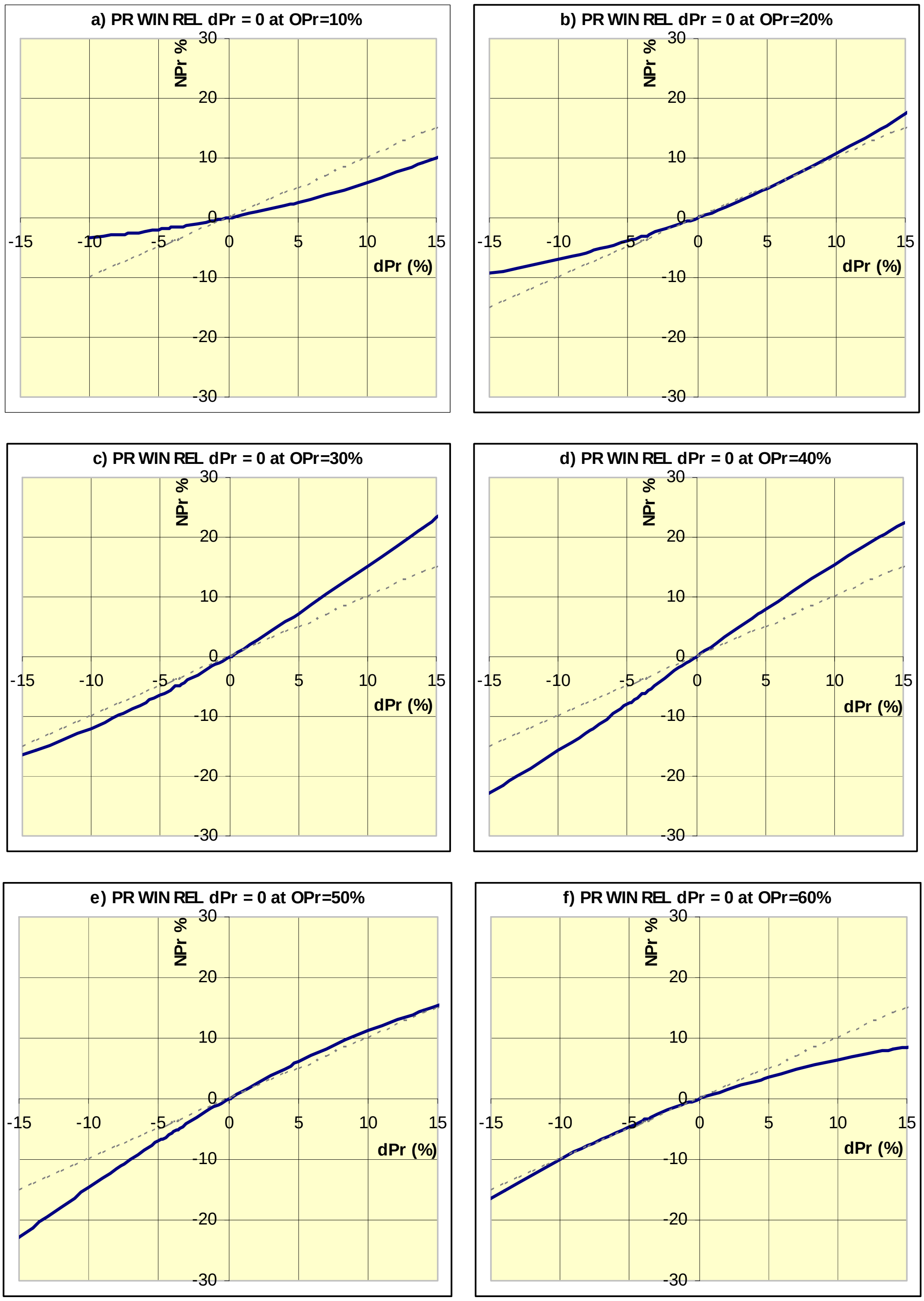


b)



Results from a logit regression of *WIN* against *OPR* and *DPR*. *NPR* measures the expected change in the chances of the horse winning compared to a horse with *DPR* equal to zero. 39,137 observations from 3,590 races.

FIGURE 3.A.2
MARKET REACTION WITH LOGIT MODEL (CROSS SECTIONED)



Rubric: As for Figure 3.2.

3.A.2 – LPM CORRECTING FOR THE INITIAL BIAS

The probability movements considered earlier could partially correspond to a correction of an initial bias. The effect on the parameter estimates is referred to as the *favourite-longshot effect*. Positive moves on favourites are likely to reflect a correction of the bias, and are thus not as informative as the same move on a longshot (whose probability should fall if the bias is being corrected hence larger β_{DPR}). To compensate for this effect, another probability measure is proposed: the adjusted *DPR*, *ADPR*. To obtain this measure, unbiased opening and starting probability estimates are calculated, these are referred to *AOPR* and *ASPR* respectively, the difference between these will be *ADPR*. *AOPR* and *ASPR* are defined as the estimated/observed win-probabilities from Chapter 2,

$$AOPR = \alpha_0 + \alpha_1 OPR + \alpha_2 * N * OPR \tag{3.11}$$

$$ASPR = a_0 + a_1 SPR + a_2 * N * SPR \tag{3.12}$$

$$ADPR = ASPR - AOPR\ . \tag{3.13}$$

AOPR and *ASPR* are unbiased estimates of the true win probability so the difference between them will be free from any favourite-longshot effect. In other words *ADPR* should be more reflective of DPR_I in equation (3.7). *ADPR* is calculated using different parameters for Class A&B, C&D, E and F&G observations. The results generated by regressing the binary *WIN* variable against *AOPR* and *ADPR* are presented in Table 3.A.2, whose interpretation is the same as that for Table 3.1. The following assumes that the correction parameters are known to everybody. As a result *the analysis only provides the descriptive outcome of the situation.*

It is clear that the results from Section 3.5 are carried through to this setting with the *favourite-longshot effect* free probability measure[19]. The hypothesis that the market reaction is correct, i.e. a one percentage point increase in the implied win-probability implies a one percentage point higher chance of the horse winning, cannot be rejected for all classes of races. For Class A&B races, due to the large standard errors, the hypothesis that moves are insignificant cannot be rejected at the 5% level

[19] A regression of specification *i)* using the bootstrap employed earlier (not shown) delivers very similar point estimates (β_{AOPR} = 1.009 and β_{ADPR} = 0.767) to the parameters in Table 3.A.2.

of significance, neither can a hypothesis of under-reaction (e.g. a coefficient of *ADPR* = 1.2) be rejected.

At the same time the hypotheses of market over-reaction, e.g. a coefficient of 0.8 on *ADPR* cannot be rejected for Class C&D, E and F&G races at the 5% level of significance. It is interesting to note that the point estimate on coefficient of *ADPR* for Class E races is greater than unity, this supports the result found earlier, when investigating for a switchover, suggesting that there is under-reaction in Class E races. A hypothesis of market under-reaction, e.g. a coefficient of 1.2 on *ADPR* cannot be rejected for all classes except for Class C&D races (and the pooled regression, where the upper boundary of the 95% confidence interval of the estimate is 1.07).

TABLE 3.A.2

REACTION TO MARKET MOVERS: LPM WITH ADJUSTED DPRS

Dependent Variable: Win (Binary)

Independent Variable	**i) ALL**	ii) A & B	iii) C & D	iv) E	v) F & G
AOPR	1.004 (0.017)***	1.006 (0.047)***	1.006 (0.023)***	0.991 (0.035)***	1.009 (0.046)***
ADPR	0.862 (0.110)***	0.725 (0.372)*	0.756 (0.170)***	1.16 (0.206)***	0.783 (0.234)***
constant	0.000 -0.002	-0.001 -0.002	-0.001 -0.004	0.001 -0.003	0.001 -0.004
Races	3590	451	1719	863	557
Runners	39137	4798	17245	10073	7021
R^2	0.141	0.131	0.155	0.126	0.128

Rubric: Same as Table 3.1 except that *AOPR* is the estimated (unbiased) opening win-probability (hence its estimated coefficient is equal to unity) and *ADPR* is the alternative measure of the move which is free from the favourite-longshot effect described in Section 3.4. The coefficient on *AOPR* is always not significantly different to unity and the coefficient on *ADPR* is never significantly different to unity at the 5% level for these regressions.

Alternative specifications with the squared and/or cubic exponents of *ADPR* have also been tested (the results are not shown) and once again their estimated coefficients are not significant.

In Section 3.6, the effect of different size fields, attendance levels and the number of bookmakers present on market movers were investigated. In this section, the methodology used in Section 3.6 to investigate these issues will be repeated using the adjusted probability measure. The results are presented in Table 3.A.3 and, as with the original probability measure, suggest that degree of reaction to market moves is independent of the three factors.

Finally, the effect of omitting non-movers using the alternative probability measure is considered. As in Section 3.6, regressions are run omitting runners with -0.01 < *ADPR* < 0.01; (24,534 runners out of the 39,137 horses are omitted), and two further specifications are run keeping the positive or negative movers only. The results are

TABLE 3.A.3
THE EXTENT OF MARKET REACTION FOR DIFFERENT SIZED FIELDS: ADJUSTED PROBABILITY MEASURE

Dependent Variable: Win (Binary)

Independent Variable	N	HIGHN-Dummy	ATTENDANCE	BOOKMAKERS
AOPR	1.008 (0.023)***	1.008 (0.023)***	1.008 (0.023)***	1.008 (0.023)***
N*AOPR	0.001 (0.002)	0.001 (0.002)		
ADPR	0.521 (0.297)*	0.812 (0.135)***	0.859 (0.140)***	0.887 (0.181)***
N*ADPR	0.032 (0.025)			
HIGHN*ADPR		0.151 (0.223)		
ATTENDANCE*ADPR			0.001 (0.015)	
BOOKMAKERS*ADPR				0.001 (0.003)
constant	-0.003 (0.002)	-0.003 (0.002)	-0.003 (0.002)*	-0.003 (0.002)*
Races	3590	3590	3498	3498
Runners	39137	39173	38163	38163
R^2	0.141	0.141	0.142	0.142

Notes: As Table 3.4 and 3.6, but these adjusted specifications use *AOPR* and *ADPR* (and adjusted interactions) as the dependent variables as opposed to *OPR* and *DPR*. Regressions run with clustered (by race) standard errors, displayed in parentheses.

TABLE 3.A.4
THE EXTENT OF MARKET REACTION OMITTING NON-MOVERS: ADJUSTED PROBABILITY MEASURE

Pooled	ADPR < -0.01 & ADPR > 0.01	ADPR < -0.01	ADPR > 0.01
βADPR			
1.004	0.843	1.046	0.927
(0.017)***	(0.112)***	(0.293)***	(0.254)***
Runners	14603	7373	7230

Notes: Same as Table 3.9, but results are from regressions using the alternative measure. The hypothesis that the coefficient of *ADPR* is equal to unity cannot be rejected at the 5% level of significance.

presented in Table 3.A.4, and once again, the estimates of the coefficient on *ADPR* are not significantly different to unity. The hypothesis that the estimates of *ADPR* are close to unity, because of the overwhelming presence of non-movers, can be rejected because the regressions omitting the non-movers yield the same results.