Journal of Economic Behavior & Organization 107 (2014) 614–629

Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

The role of surprise: Understanding overreaction and underreaction to unanticipated events using in-play soccer betting market

Darwin Choi[a], Sam K. Hui[b,*]

[a] Hong Kong University of Science and Technology, Hong Kong
[b] Stern School of Business, New York University, USA

ARTICLE INFO

Article history:
Received 31 December 2012
Received in revised form 9 February 2014
Accepted 13 February 2014
Available online 22 February 2014

JEL classification:
D84
G13
G14

Keywords:
Over/underreaction
Unanticipated events
Surprise
Conservatism
Prediction market
Bayesian dynamic model

ABSTRACT

Previous research in finance has found evidences of both overreaction and underreaction to unanticipated events, but has yet to explain why investors overreact to certain events while underreacting to others. In this paper, we hypothesize that while market participants generally underreact to new events due to conservatism, the extent of underreaction is moderated by "surprise," thus causing market participants to overreact to events that are highly surprising. We test our hypothesis using data from an in-play soccer betting market, where new events (goals) are clearly and exogenously defined, and the degree of "surprise" can be directly quantified (goals scored by underdogs are more surprising). We provide both statistical and economic evidences in support of our hypothesis.

1. Introduction

Financial researchers have long been interested in whether market participants react to unanticipated events in an unbiased manner,[1] or whether they exhibit behavioral biases such as overreaction (Brooks et al., 2003) or underreaction (Chan, 2003). While previous research has found evidences for both overreaction (Brooks et al., 2003; Coleman, 2011) and underreaction (Klibanoff et al., 1998; Chan, 2003), some of these evidences are somewhat conflicting or even contradictory. For instance, Brooks et al. (2003) show that markets tend to overreact to industrial disasters and CEO deaths, while Chan (2003) claims that investors underreact to headline news. In general, "clean" empirical evidences are difficult to find in financial markets, as many phenomena can be viewed as evidences of opposite claims. For example, momentum in stock price (Jegadeesh and Titman, 1993) can be viewed as both the result of underreaction (Barberis et al., 1998; Hong and Stein,

* Corresponding author at: Tisch Hall 910, 40 West 4th Street, New York, NY 10012, USA. Tel.: +1 201 2169763.
E-mail addresses: dchoi@ust.hk (D. Choi), khui@stern.nyu.edu (S.K. Hui).

[1] Similar to Brooks et al. (2003), we focus on events that are "unanticipated" both in terms of the timing and the nature of the event. Thus, we exclude scheduled events (e.g., earning announcements, layoffs) from our consideration, as some market participants may have gained (partial) access to such announcements ahead of time and hence create information asymmetry (Gil and Levitt, 2007).

http://dx.doi.org/10.1016/j.jebo.2014.02.009
0167-2681/

1999) and overreaction (Daniel et al., 1998). More importantly, previous research has not specified the type of events (and the conditions under which they happen) where overreaction or underreaction is more likely to occur. This has been viewed as a serious limitation to behavioral finance theories by Fama (1998), who states that behavioral theories "must specify biases in information processing that cause the same investors to underreact to some types of events and overreact to others."

We develop a behaviorally motivated hypothesis of how investors underreact or overreact to unanticipated events. Specifically, we hypothesize that overreaction and underreaction are driven by conservatism (Barberis et al., 1998) and "surprise" (Reisenzein et al., 2012). When reacting to an event that is expected or only moderately surprising, market participants insufficiently update their prior beliefs due to conservatism and hence underreact (Barberis et al., 1998). In contrast, a more "surprising" event, i.e., an event that strongly violates prior expectations, attracts more cognitive processing (Meyer et al., 1997) as participants attempt to "make sense" of the incongruence between the observed event and their current schema (Pezzo, 2003). This in turn amplifies the surprising event and attenuates other sources of information (i.e., their prior beliefs), thereby resulting in a higher "weight" being put on the new event when forming a judgment. Thus, underreaction is moderated by surprise; for extremely surprising events, market participants overweight the new information, leading to overreaction.

Testing our hypothesis on financial market data is challenging because it is difficult to a priori quantify how "surprising" an events is (Barberis et al., 1998) and to unambiguously measure its impact on equity prices (Fama, 1998). Thus, we turn to data from an "in-play" soccer betting market to test our hypothesis (Avery and Chevalier, 1999). In an in-play sports betting market, participants place bets while a match is still under way. It offers the ideal setting to test our hypothesis for several reasons: first, the arrival of a goal is apparent and its impact on odds can be objectively assessed with actual match outcomes; this circumvents the problem in financial markets where market efficiency has to be jointly tested with a model of expected "normal" return (Fama, 1998). Second, unlike in financial markets where there can be a long delay between media reporting and the occurrence of an unanticipated event (Coleman, 2011) and hence involve information asymmetry issues, goals are reported as soon as they are scored and immediately become public knowledge among all market participants. Third, we can clearly define how "surprising" a goal is by comparing the strengths of the two teams: a goal scored by the "underdog" is more surprising than a goal scored by the "favorite." Fourth, goals are exogenous shocks, which may not be the case in financial markets.[2] Finally, real money is at stake in a betting market and transactional volume is condensed within a short time horizon, which provides a large sample of events in a real-world setting to test our hypothesis.

Our dataset is comprised of second-by-second transaction records in 2017 soccer matches obtained from Betfair, an online betting exchange. The total betting volume in our sample amounts to around £3 billion. While a match is under way, participants may bet on the outcomes "team1 win", "draw", or "team2 win"; we focus on the in-play odds of the scoring team. We operationalize how "surprising" a goal is by the difference between the implied winning probability of the non-scoring team and that of the scoring team, measured right before the goal. We study overreaction and underreaction to the first goal of the match using a sequence of logistic regressions and a Bayesian structural model. In general, we find that market participants underreact to goals that are expected or only moderately surprising. This underreaction is moderated by the degree of surprise, resulting in overreaction to very surprising goals. We find that these biased reactions attenuate over time and disappear at around 5 min after the goal, but are unrelated to transactional volume. In addition, we explore the economic size of the mispricing bias by developing a strategy that bets on the scoring team when underreaction is predicted, and against the scoring team if overreaction is predicted. Through a split-sample analysis, we find that such strategy earns a profit of 2.46% ($p = 0.03$) after commissions if the bets are placed at 2 min after the goal, suggesting that underreaction and overreaction are economically significant.

The remainder of this paper is organized as follows. Section 2 briefly reviews previous research on overreaction and underreaction to unanticipated events, the role of surprise in judgment and decision making, and the sport betting markets. Section 3 describes our hypothesis of how conservatism and surprise drive underreaction and overreaction. Section 4 describes our dataset along with key summary statistics. In Section 5, we present statistical and economic evidences on overreaction and underreaction. Finally, Section 6 concludes with implications for financial markets.

2. Background and conceptual development

2.1. Overreaction and underreaction in financial markets

As discussed earlier, previous research has found evidences for both overreaction and underreaction. For instance, Brooks et al. (2003) study stock market reactions to 21 unanticipated events, e.g., the sudden death of CEOs. They find that the initial price reaction to an unanticipated event tends to be partially reversed in 90 min after the event, suggesting initial overreaction. Similarly, Coleman (2011) analyzes initial stock market response to 60 "shock" corporate events such as fatal industrial disasters. He finds that the price of the affected firms fall by an average of 3% within two or three trading hours, but

[2] Coleman (2011) shows that the timing of initial announcement following an unanticipated event depends on firm-specific factors such as capital intensity and age of the CEO. More generally, company information release may depend on managers' perceptions of the firm being over or undervalued, and thus can be endogenous.

half of the price impact is reversed by the market close on the day of the event. Coleman (2011) interprets this observation as market overreaction, as risk-averse investors overreact to the uncertainty created by the shock.

On the other hand, other researchers have found that investors underreact to various new events, e.g., share repurchase (Ikenberry et al., 1995), dividend initiation and omissions (Michaely et al., 1995), and stock splits (Ikenberry et al., 1996; Ikenberry and Ramnath, 2002). For instance, Klibanoff et al. (1998) demonstrate that investors tend to underreact to news that is relevant to economic fundamentals. They also show that the "prominence" or "salience" of a news event, as measured by the column width on the front page of the *New York Times* covering the news story, reduces the extent of underreaction. Similarly, Chan (2003) collects a database of firm-specific headlines to examine how the market reacts to public news. By comparing two sets of stocks with or without headlines, he finds price "drift" patterns only in the set of stocks that appear on the headline, which suggests that investors underreact to news. He also finds that underreaction seems stronger in lower priced and more illiquid stocks.

While previous research has found evidences for both underreaction and overreaction, researchers have yet to specify the conditions under which underreaction or overreaction is more likely to occur. We aim to shed some light on this issue by exploring how underreaction and overreaction is driven by "surprise", which we review next.

2.2. The role of surprise in judgment and decision making

Surprise is a universal human emotion (Maguire et al., 2011; Plutchik, 1980). By definition, surprise is a peculiar state of mind, often of brief duration, elicited by unexpected events of all kinds, i.e., events that disconfirm, contradict, or violate an expectation or belief (Reisenzein et al., 2012; Teigen and Keren, 2003). Formally, from the perspective of schema theory in psychology, surprising events are those that are incongruent with currently activated schema, thereby activating a "schema-discrepancy detector" which results in key behavioral changes (Meyer et al., 1997; Reisenzein et al., 2012). The level of surprise experienced for an event is associated with the subjective difficulty of integrating it with an existing cognitive representation (Maguire et al., 2011). A higher level of surprise results from a higher level of contrast between the observed and expected outcome (Teigen and Keren, 2003), and leads to a higher extent of adaptive behavior.

Surprise plays an important role in how people update their judgments following a new event. If an event conforms to prior expectation, there is "congruence" between a person's schema and the event, and thus no urgent need to revise the schema (Reisenzein et al., 2012). The update of belief following an expected event thus occurs automatically and effortlessly. In contrast, a "surprising" event violates the person's current belief and forces the person to change his or her beliefs (Itti and Baldi, 2009) in order to "make sense" of the unexpected event, and to adapt to the new observed evidence (Pezzo, 2003). The feeling of surprise serves as a cue to make aware of the fact that outcome information is different from and inconsistent with their previous belief (Ofir and Mazursky, 1997). This inconsistency results in an increased need for "sense making," i.e., the motivation to resolve representative discrepancies (Maguire et al., 2011), by redistributing cognitive processing resources to the unexpected event. In turn, the focusing of attention on the unexpected event (Baldi and Itti, 2010) amplifies its importance and attenuates interfering information from other sources (Reisenzein et al., 2012), e.g., prior belief, during the updating process. As a result, the person puts more "weight" on surprising events, compared to events that are expected, when forming an overall judgment.

Note that the aforementioned role of surprise on judgment and decision making is also consistent with the behavioral theory put forth by Griffin and Tversky (1992), which motivates the behavioral model of investors' behavior in Barberis et al. (1998), showing that people in general overweight the "strength" (or "extremity") and underweight the "reliability" of an evidence when forming a judgment. Viewed under the theoretical framework of Griffin and Tversky (1992), "surprising" events are more "extreme" as they are incongruent with current expectation and belief, and can thus be thought of events of higher "strength." This again leads to an overweighting of surprising events when updating one's subjective belief, resulting in the attenuation of underreaction to surprising events and overreaction to extremely surprising events.

As discussed earlier, testing the role of surprise in driving overreaction and underreaction in financial markets is extremely challenging. Thus, we now turn to the in-play soccer betting market to test our hypothesis.

2.3. The in-play sports betting market as a test setting

In an in-play soccer betting market, market participants trade contracts which have a certain payoff that is contingent on the match outcome. Clearly, betting markets share many structural similarities with financial markets (Oliven and Rietz, 2004; Wolfers and Zitzewitz, 2004). As discussed earlier, testing our hypothesis using an in-play soccer betting market circumvents several confounding factors in financial markets. In addition, compared to a purely artificial laboratory-based market setting (List, 2009), a sports betting market is more realistic as it involves higher monetary stakes, thus alleviating some of the undesirable "artificiality" of laboratory experiments (Levitt and List, 2007).

Most previous research on sports betting market focuses on "pre-match" betting markets (Gandar et al., 1988; Golec and Tamarkin, 1991; Durham et al., 2005; Durham and Santhanakrishnan, 2008), where bettors are allowed to place bets only before the match starts. Researchers have found that the probabilities inferred from the "consensus forecast" in the pre-match betting markets are very close to the actual outcome probabilities. For instance, by studying a betting market on horse racing in the U.K., Smith et al. (2009) conclude that betting exchange odds accurately estimate actual outcome probabilities and have higher predictive value than the corresponding bookmaker odds.

Obviously, research on pre-match markets cannot study real-time reactions to information shocks, e.g., goals that happen *during* a soccer match. Only a handful of papers have studied in-play sport betting markets. For instance, Gil and Levitt (2007) analyze minute-to-minute transactional data (on intrade.com) for 50 soccer matches in the 2002 FIFA World Cup. They find that while markets respond strongly to goals being scored, prices continue to trend higher 10–15 min after a goal, suggesting that bettors underreact to goals. This phenomenon is interpreted differently by Croxson and Reade (2013), who argue that efficient in-play odds should continue drifting after a goal is scored. To see this, suppose that Team 1 scores the first goal of the match. If the odds are efficient, the odds that Team 1 wins should decrease till the end of the match if no additional goal is scored, because Team 1 will win by 1-0. To test for market efficiency, Croxson and Reade (2013) assemble a dataset of around 1200 soccer matches from Betfair (the same market where we obtain our data). They conclude that semi-strong market efficiency cannot be rejected and that prices update "swiftly and fully." An important insight drawn from their research is that in order to accurately assess reactions to goals, one must incorporate a model of how efficient prices drift as time passes in the absence of goals. We build upon this insight when developing our Bayesian structural model in Section 5.4 to study overreaction and underreaction.

3. Hypotheses: how conservatism and surprise drives overreaction and underreaction

We now develop our key hypotheses of how conservatism and "surprise" jointly drive underreaction and overreaction to new events. These cognitive biases are particularly relevant because when reacting to new information, participants in an in-play betting market have to make a split-second judgment, and thus are more likely to rely on behavioral heuristics since there is not enough time to make a fully informed judgment.

We use the following stylized model for illustration. Upon receiving new information, a market participant has to incorporate the new information by putting "weights" on her prior belief (w_1) and on the new information (w_2) to form her "new belief." That is, New Belief = w_1 × Prior Belief + w_2 × New Information.

If market participants react to new information in an unbiased manner, w_1 and w_2 should correspond to the appropriate weights under Bayesian updating. Following Barberis et al. (1998), we hypothesize that market participants tend to underreact to news due to conservatism (Edwards, 1968), which states that individuals rely too much on their prior beliefs and make insufficient adjustments. Specifically, participants typically update their posteriors in the right direction, but the magnitude of the update is too small when compared to a Bayesian framework (Edwards, 1968). Thus, for news that is not surprising or only moderately surprising, we expect w_1 to be too large and w_2 to be too small compared to their normative values, resulting in underreaction. Note that this is the same behavioral explanation given in Barberis et al. (1998) for underreaction in general.

Next, underreaction is moderated by how "surprising" the new information is. As discussed in Section 2.2, a "surprising" event forces market participants to focus their cognitive processing on the conflicting evidence to "make sense" of the incongruence between the observed event and their current schema. This amplifies the conflicting signal (i.e., the surprising event) and attenuates other sources of information (their current beliefs), forcing them to change their current beliefs to "adapt" to the new evidence (Reisenzein et al., 2012). Thus, we hypothesize that for new information that is more surprising, w_1 should get closer to its normative value. Finally, taking the above argument further, if the new information is extremely surprising, the overweighting of the surprising event (i.e., higher w_2) will more than offset the effect of conservatism, resulting in overreaction.

We now turn to the in-play soccer betting market. In soccer betting, the odds of the bets specify the payoffs bettors would receive if a certain outcome happens. For example, a bet on a certain team winning the match with decimal odds of 5.0 will receive $5 for every $1 wagered if the team wins (including the original $1). The implied probability of this event is 0.2 in a fair game with zero expected return. Therefore, betting odds reflect the aggregate market's belief (or "consensus forecast" in Barberis et al., 1998) of the winning probabilities of the teams (Wolfers and Zitzewitz, 2004), and can be used to gauge whether bettors overreact or underreact to goals. More specifically, if bettors overreact to a goal, they will bid down the odds of the scoring team, inflating the implied probability that the scoring team wins. In contrast, if they underreact to a goal, the odds of the scoring team will be too high and the implied probability will be too low compared to the actual outcome probability.[3]

As discussed, we operationalize how "surprising" a goal is using the relative strengths of the two teams: the favorite (underdog) is defined as the team having a higher (lower) implied winning probability shortly before the first goal of the match is scored. Note that we restrict our attention to only the first goal of each match to maintain statistical independence across observations (matches), which we depend upon in all our statistical analyses later. If we analyze multiple goals from each match, goals from the same match may not be independent of each other. Further, the timing of the first goal can be identified more accurately from our data than other goals (as will be discussed in Section 4), which reduces the extent of measurement errors in the data.

[3] While we focus on the scoring team, the hypothesis is unchanged if we look at the draw odds and the non-scoring team's odds. When a goal is scored by Team 1, Team 1's odds should decrease, while draw and Team 2's win odds should increase. Our hypothesis defines underreaction as insufficient adjustments in the scoring team's odds. Since the sum of 1/odds over all outcomes is very close to 1 in the data at all times, this is equivalent to insufficient adjustments in the draw and the non-scoring team's odds. The same argument applies to overreaction.

Table 1
Summary statistics of our dataset.

	N	Mean	Median	S.D.	Min	Max
Matched volume (£)	2017	1,481,627	917,054	1,627,070	9654	13,400,407
Matched volume ([+2, +6 min] after goal)	2017	152,523	93,779	175,621	103	1,708,552
First goal (min)	2017	35.84	25.97	30.47	0.18	113.07
Scoring team is favorite	2017	0.670	1.000	0.470	0.000	1.000
Scoring team wins	2017	0.712	1.000	0.453	0.000	1.000
Implied probability (+2 min after goal)	1938	0.686	0.720	0.187	0.017	0.994
Implied probability (+3 min after goal)	1981	0.691	0.726	0.190	0.002	0.995
Implied probability (+6 min after goal)	1975	0.694	0.730	0.192	0.034	0.997

We state our hypotheses in the context of soccer betting as follows:

H1a *(General underreaction)*. Due to conservatism, market participants rely too much on their prior beliefs, resulting in underreaction to the first goal if the goal is not surprising or only moderately surprising.

H1b *(Surprise reduces underreaction)*. A surprising goal increases the "weight" that is put on it when forming a judgment of winning probability. Thus, we expect lower underreaction for more surprising goals.

H1c *(Extreme surprise leads to overreaction)*. Taking H1b one step further, a goal that is extremely surprising triggers overreaction, as bettors tend to overweight the extremely surprising goal relative to their prior belief.

We also explore how the biases in H1a–c vary with time lag and betting volume. Given more time to evaluate the new information, bettors should rely less on heuristics; thus we expect that the bias in H1a–c should attenuate over time. In addition, higher transactional volume may attract more "professional" bettors who are presumably less susceptible to biases. Hence, we expect the biases in H1a–c to be moderated by higher transactional volume. Formally, we specify H2 and H3 as follows and empirically test them in Section 5.

H2 *(Time moderation)*. The biases in H1a–c attenuate over time after the first goal is scored.

H3 *(Volume moderation)*. The biases in H1a–c decrease with higher betting volume.

4. Data

We test our hypotheses using data from Betfair, an online betting exchange. We purchase historical data from Betfair through an authorized third-party vendor, Fracsoft. We obtain all available matches of major international and European soccer tournaments, both at the club and national team levels, from August 2006 to March 2011. Specifically, our dataset is comprised of a total of 2687 soccer matches in the following competitions: FIFA World Cup 2006 and 2010, five seasons of UEFA Champions League in Europe and Barclays Premier League in England (2006/2007–2010/2011), and one season of La Liga in Spain, Serie A in Italy, and Bundesliga in Germany (2010/2011). Note that around 30% of all matches are unavailable; Fracsoft attributes the missing data to technical issues when recording the bets from Betfair. Thus, we believe that the selection is random and does not materially affect our findings.

Betfair allows participants to submit "back" and "lay" orders to bet on or against a certain outcome, respectively. As discussed earlier, a "back" bet on a certain team winning the match with decimal odds of 5.0 will receive $5 for every $1 wagered (including the original $1) if the team wins, while a corresponding "lay" bet will lose $5 if the team wins and earn $1 otherwise. As an exchange, Betfair operates like a limit order market and matches the "back" and "lay" orders, hence the matched bets are zero-sum games between the "back" and "lay" bettors. Given that there is no bookmaker, any biases we observe from Betfair odds should come from market participants. Betfair charges a commission of up to 5% on net winnings, with a "loyalty bonus" that reduces the commission to as low as 2% for frequent bettors. The commissions are not reflected in the decimal odds.

Our data record the volume, decimal odds, timestamp (to the second), and outcomes betting "in favor of" or "against" for all the bets in the "Match Odds" market. Unlike the data collected by Gil and Levitt (2007), our data do not record the identities of individual bettors, but only transactions (matched bets) at the market level. In our data, all transactions within a second are aggregated by Fracsoft. As discussed earlier, participants can place bets before the game starts as well as when the match is underway. During the match, Betfair suspends trading briefly (for up to 2 min) and clears all unfilled back and lay orders when the game starts and when a "material event" has happened. Material events include goals, penalty kicks awarded, and red cards (which result in a player's dismissal). However, while we observe the time when trading is suspended, the Betfair data do not record the actual material event leading up to the suspension. We therefore collect corresponding event information from ESPN FC (soccernet.espn.go.com).

Out of all matches, 2160 of them have at least one goal scored and the first goal's scoring time could be identified from the Betfair data and matched to ESPN FC. We exclude first goals that are scored from penalty kicks, because Betfair suspends trading when penalties are awarded, not when they are scored. This results in a final sample of 2017 matches. Table 1 presents some summary statistics. The matched in-play volume per match is around £1.5 million on average. The first goal

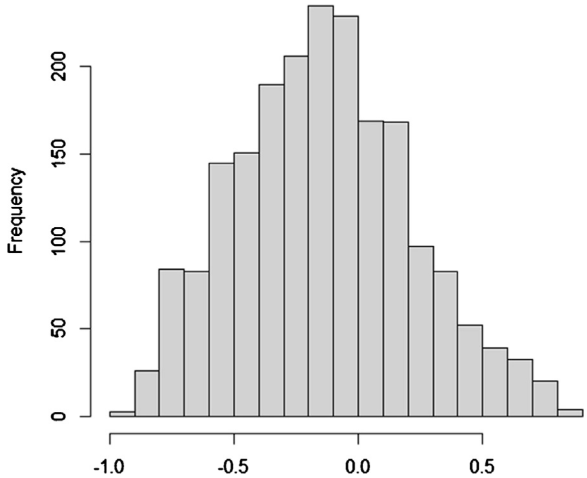

Fig. 1. Histogram of the surprise metric across 2017 matches.

of the match is, on average, scored around 36 min after the match has started. We also calculate the betting volume in the window [+2, +6 min] after the first goal. The average is £0.15 million, or around 10% of the total in-play volume. 67% of the first goals are scored by the favorite; 71% of the scoring teams eventually win the match. We also show the summary statistics of the implied probabilities at 2, 3, and 6 min after the first goal, which will be discussed in Section 5.2.

5. Empirical analysis of overreaction and underreaction

5.1. Operationalization of the surprise metric

We quantify the degree of "surprise" of the first goal as follows. First, we define the implied winning probability of each team by taking the reciprocal of the decimal odds, adjusted by a normalizing constant that is very close to one (Asch et al., 1984; Wolfers and Zitzewitz, 2004). [4] Then, we define s_i, the "surprise" metric, as the implied winning probability of the "non-scoring team" minus that of the "scoring team", where both implied probabilities are measured at 1 min before the first goal of the match. Clearly, a positive (negative) s_i means that the scoring team is weaker (stronger) than the non-scoring team, and that the goal is surprising (expected).

The above definition of surprise captures the market's "consensus perception" of the relative strengths of the two teams, right before a goal is scored. The advantage of using the prevailing implied probabilities is that they have already incorporated other relevant information during the match, e.g., the percentage of ball possession, injuries, form of the players. We note that, however, if the first goal does not occur until very late in the match, the differences in the implied probabilities of the two teams will tend to be small. Thus, as a robustness check, we also develop an alternative definition of surprise that utilizes the implied probabilities based on pre-match odds. The results using this alternative surprise metric are very similar to those presented in the later sections and are available upon request.[5] In our sample, s_i ranges from −0.92 to 0.88, with a mean of −0.14 and a median of −0.14. Note that a negative mean and median are expected because the favorite is more likely to score first. A histogram of s_i across all 2017 matches in our dataset is shown in Fig. 1.

For illustrative purposes, we show how the implied probabilities of the scoring team reacts to the first goal using two matches in our sample. The top panel in Fig. 2 shows a "surprising" goal ($s_i = 0.75$), scored by New Zealand (the underdog) against Italy in the FIFA World Cup 2010; while the bottom panel in Fig. 2 shows an "expected" goal ($s_i = -0.72$), scored by Arsenal (the favorite) against Watford in the Barclays Premier League 2006/2007. The implied probability goes up after the goal in both cases, reflecting the increased likelihood that the scoring team will win given that it is now leading by one goal. However, by looking at the graphs alone, we cannot infer whether there is overreaction or underreaction, as the true (unobserved) conditional winning probability of the scoring team is unknown. Thus, in the following subsections, we present statistical evidences of overreaction and underreaction by comparing the odds across a set of matches to the corresponding set of match outcomes.

[4] Using odds at the [+2, +6 min] window after the first goal, we calculate the following statistics for the normalizing constant: mean = 1.0026, median = 1.0015, standard deviation = 0.0192.

[5] As an additional robustness check, we have also utilized another alternative definition of the surprise metric, where s_i^{ALT} = Pr(Non-scoring team wins)/(Pr(Scoring team wins) + Pr(Non-scoring team wins)). We find that s_i^{ALT} is almost perfectly correlated with s_i ($r = .99$; $p < .001$) and thus our results are substantially unchanged. We thank an anonymous reviewer for this suggestion.

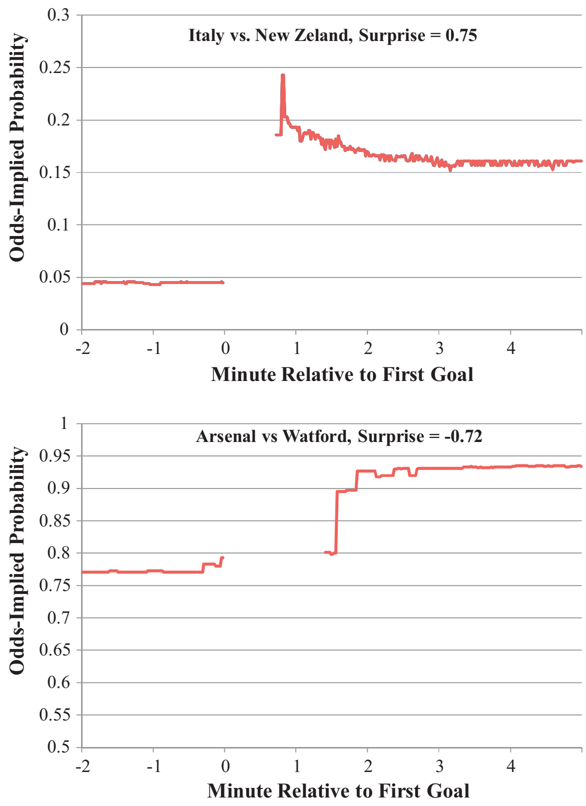

Fig. 2. Two examples of reaction in implied probabilities of the scoring team after the first goal.

5.2. A non-parametric analysis of over- and underreaction

We conduct a non-parametric analysis that explores the relationship between actual match outcomes and implied probabilities that the scoring team wins at different levels of "surprise" (s_i). Recall that after the first goal, the betting market is suspended for up to 2 min before it reopens. Thus, to assess how bettors update their beliefs after the first goal, we compare the in-play odds 2 min after the first goal is scored, i.e., the last transacted odds before the 2-min mark, versus the actual match outcomes.

First, we calculate the implied probability p_i that the scoring team wins using odds 2 min after the first goal. We let y_i be an indicator variable that takes the value of 1 if the scoring team in match i wins, and 0 otherwise. If bettors' reactions to the first goal are efficient, p_i should be equal to the conditional probability that $[y_i = 1]$ given all information available at that time. Thus, we compare the expected count that the scoring team wins from these implied probabilities ($\sum_i p_i$) versus the actual count ($\sum_i y_i$) that the scoring team wins, and use the difference as a statistical measure of overreaction and underreaction. If there is overreaction (underreaction), the expected count should be larger (smaller) than the actual count.

To explore how overreaction and underreaction are driven by surprise, we divide the data into two non-overlapping sets: "expected" goals, which correspond to the condition ($s_i < -c$), and "surprising" goals, which correspond to the condition ($s_i > c$). We vary the threshold value c from 0.0 to 0.6, and in each case compare the total number of times that the scoring team wins with its expected value under the assumption that reactions to first goals are efficient; the results are shown in Table 2. A sequence of p-values, which indicate whether the observed data are significantly different from their expected values, are generated using the following Monte Carlo procedure (Wasserman, 2010):

1. The null hypothesis H_0 is that the in-play odds at the 2-min mark (p_i) accurately reflect the conditional probability of the match outcome (y_i), i.e., H_0: $\Pr(y_i = 1) = p_i$.
2. The test statistic employed here is the difference between the expected number of times that the scoring team wins and the actual number of wins, i.e., $d = \sum_i p_i - \sum_i y_i$.
3. The null distribution of the test statistics d is generated by a Monte Carlo simulation under H_0, i.e., y_i's are independent and each follows a Bernoulli distribution with probability p_i.

As can be seen in the left panel of Table 2, when the first goal is "expected" ($s_i < -c$), p_i underestimates the conditional probability $Pr(y_i = 1)$, indicating underreaction to an "expected" goal (H1a). This effect is statistically significant ($p < .05$) for

Table 2
Nonparametric analysis of overreaction and underreaction.

c	"Expected" goals ($s_i < -c$)				"Surprising" goals ($s_i > c$)			
	N	Expected # of wins (%)	Actual # of wins (%)	p-Value	N	Expected # of wins (%)	Actual # of wins (%)	p-Value
0.0	1299	1014.3 (78.1%)	1067 (79.7%)	0.000*	638	313.8 (49.2%)	320 (50.2%)	0.312
0.1	1090	868.5 (79.7%)	918 (84.2%)	0.000*	477	214.0 (44.9%)	212 (44.4%)	0.451
0.2	860	704.5 (81.9%)	735 (85.5%)	0.003*	318	123.4 (38.8%)	124 (39.0%)	0.505
0.3	664	558.3 (84.1%)	578 (87.0%)	0.016*	224	75.9 (33.9%)	78 (34.8%)	0.392
0.4	484	416.6 (86.1%)	428 (88.4%)	0.077^	145	41.7 (28.8%)	38 (26.2%)	0.288
0.5	336	295.8 (88.0%)	306 (91.1%)	0.040*	95	23.8 (25.1%)	21 (22.1%)	0.302
0.6	191	173.4 (90.8%)	183 (95.8%)	0.006*	56	11.9 (21.3%)	9 (16.1%)	0.219

* $p<.05$.
^ $p<.10$.

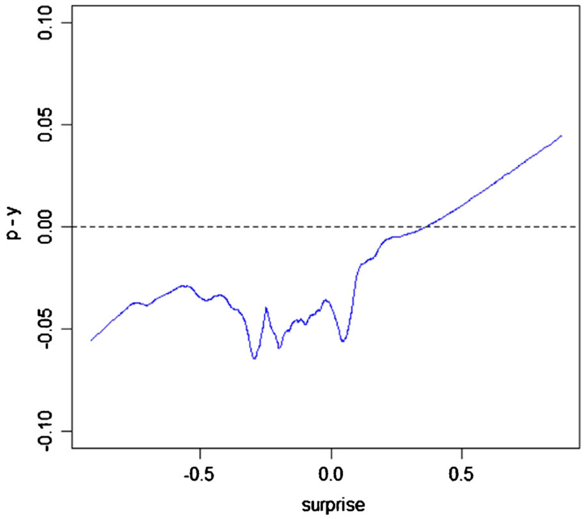

Fig. 3. A non-parametric smoothing spline between $(p_i - y_i)$ and s_i.

most values of the threshold c. By contrasting this with the right panel of Table 2, we find that this significant underreaction disappears when the first goal is a "surprise" ($s_i > c$) (H1b). Interestingly, the deviations of actual vs. expected number of wins by the scoring teams go in the opposite direction when the goal is extremely surprising ($s_i > 0.4$): the expected number of wins by the scoring team is larger than the actual observed number, providing initial evidence of overreaction (H1c). Note, however, that none of the p-values under the "surprise" condition is statistically significant, presumably because of the small sample sizes which result in the non-parametric tests having low power (Wasserman, 2010).

Next, Fig. 3 shows a smoothing spline (Venables and Ripley, 2002) that non-parametrically captures the relationship between $(p_i - y_i)$ and s_i. As can be seen, for small values of s_i, the nonparametric estimate of $(p_i - y_i)$ is negative, again indicating underreaction. However, as s_i becomes larger (around $s_i > 0.4$), the non-parametric estimate of $(p_i - y_i)$ turns positive, suggesting the presence of overreaction. Note also that the smoothing spline is fairly linear and (approximately) monotonically increasing. To formally assess H1a–c and to study the moderating factors of time lag and transaction volume proposed in H2 and H3 (respectively), we follow up with more elaborate statistical models with additional parametric assumptions.

5.3. A sequence of logistic regressions

We now estimate the magnitude of overreaction and underreaction as a function of s_i, and study how the biases are moderated by time lag, as hypothesized in H2. Toward that end, we compute the implied probabilities for each match at $t = 2, 3, \ldots, 15$ min after the first goal (again we use the last transacted odds before each minute-mark).[6] We then estimate the following sequence of logistic regressions:

$$y_i \sim Bernoulli(\pi_{it}) \tag{1}$$

[6] Note that other goals and material events may occur during this timeframe. We "filter" our sample based on whether other events occur or not; that is, if another goal has occurred at $t = 4$, such an observation is not included in the logistic regression at $t = 4, 5, \ldots, 15$.

Table 3
Results of a sequence of logistic regressions.

Time after first goal (min)	α	p-Value ($\alpha=0$)	β	p-Value ($\beta=0$)	p-Value ($\alpha=\beta=0$)
2	0.165*	0.004	−0.385*	0.032	0.000*
3	0.111^	0.051	−0.297^	0.097	0.020*
4	0.091	0.115	−0.253	0.159	0.070^
5	0.069	0.236	−0.230	0.204	0.167
6	0.060	0.302	−0.167	0.356	0.321
7	0.066	0.267	−0.126	0.492	0.372
8	0.057	0.342	−0.083	0.652	0.530
9	0.041	0.500	−0.086	0.644	0.676
10	0.027	0.657	−0.075	0.691	0.809
11	0.018	0.777	−0.112	0.560	0.783
12	0.021	0.741	−0.142	0.468	0.693
13	0.025	0.699	−0.145	0.465	0.670
14	0.021	0.746	−0.186	0.360	0.584
15	0.015	0.824	−0.173	0.400	0.654

* $p<.05$.
^ $p<.10$.

$$\text{logit}(\pi_{it}) = \text{logit}(p_{it}) + \alpha_t + s_i\beta_t \qquad (2)$$

where as before, y_i denotes the outcome of match i, which takes the value of 1 if the scoring team wins, and 0 otherwise. p_{it} denotes the implied probability (that the scoring team will win) at the tth minute after the first goal, and π_{it} denotes the "true" conditional probability that $y_i = 1$ given all the information available at time t. The term $(\alpha_t + s_i\beta_t)$ represents the bias as a function of s_i. Clearly, $\alpha_t + s_i\beta_t > 0$ means that the true probability π_{it} is larger than the implied probability p_{it}, indicating underreaction to the first goal. Conversely, $\alpha_t + s_i\beta_t < 0$ indicates overreaction. The models in Eq. (2) are estimated using maximum likelihood and the results are summarized in Table 3.

For each value of t, we perform a likelihood ratio test (Nelder and Wedderburn, 1972) to compare the model described by Eq. (2) against the "efficient" model, which refers to the joint null hypothesis that $\{\alpha = \beta = 0\}$. Under the "efficient" model, $\pi_{it} \equiv p_{it}$ and hence the implied probability accurately reflects the conditional probability of the match outcome. Thus, a small p-value provides statistical evidence for overreaction or underreaction, depending on the sign of $\alpha_t + s_i\beta_t$. Further, if the systematic bias is attenuated by time lag, we should find that as time goes on, the p-values should increase and at some point become statistically insignificant.

This is exactly borne out by the results shown in Table 3. First, we find that the best fitting model at $t=2$ (i.e., 2 min after the first goal, when the betting market re-opens for almost all matches) is $\text{logit}(\pi_{it}) = \text{logit}(p_{it}) + 0.165 - 0.385s_i$, with a p-value of 0.000, suggesting that this model fits significantly better than the "efficient" model. This rejects the null hypothesis of efficient in-play odds (i.e., $\alpha_t = \beta_t = 0$), in favor of the alternative hypothesis that in-play odds are subject to systematic bias, describe by the term $\alpha_t + s_i\beta_t$. As shown in Fig. 1 and in Section 5.1, the surprise metric s_i ranges from −0.92 to 0.88, so $\alpha_t + s_i\beta_t$ (for $t=2$) ranges from 0.51 to −0.18.[7] This indicates that we observe underreaction generally ($\alpha_t > 0$; H1a); underreaction is moderated by surprise ($\beta_t < 0$; H1b), and when $s_i > 0.43$, we observe overreaction (H1c). Thus, our findings here corroborate with the initial findings in the non-parametric analysis in Section 5.2: when the first goal is highly surprising, bettors tend to overreact and overestimate the likelihood that the underdog will win.

In addition, we find that, consistent with H2, the biases are attenuated over time. We see that the p-values shown in the last column of Table 3 increases over time. The p-value is still below 0.05 by the 3rd minute, and becomes marginally significant ($p=0.07$) at the 4th minute and insignificant ($p=0.17$) after 5 min, when the "efficient" hypothesis can no longer be rejected. This pattern holds also for individual p-values that test $\alpha_t = 0$ and $\beta_t = 0$; both p-values increase over time and are significant only for the first 3 min after the first goal. Together, this suggests that the systematic bias in H1 is corrected in around 5 min.[8]

5.4. A Bayesian structural model of in-play odds

Up to this point, we analyze the data at the aggregate level and obtain some evidence for H1 and H2. We did not, however, provide an estimate of the *rate* at which the biases are reduced over time. Further, the logistic regression approach does not explicitly take into account the timing of the first goal, nor control for the occurrence of other material events (e.g.,

[7] The overreaction and underreaction are quoted in logit scale; thus, the magnitude of the bias terms depends on where the true probability is on the logit scale. For instance, if we assume that the true probability is 0.9, an underreaction of 0.51 in logit scale would translate to an underestimation of 0.056 in probability (from 0.9 to 0.844), while if the true probability is 0.3, an overreaction of −0.18 would translate to an overestimation of 0.039 in probability (from 0.3 to 0.339).

[8] To explore the robustness of our results, we have conducted several additional analyses that control for (i) goal timing (by restricting the sample to only first goals that occur in the first half), (ii) the presence of outliers, and (iii) transaction volume. We find that our results are fairly consistent across all robustness checks. Details are available upon request.

red cards, another goal) after the first goal. In this section, we analyze the transaction data at the second-by-second level using a Bayesian structural model, which allows us to estimate the extent to which the biases decrease over time and the relationship between the biases and transaction volume (H3). In addition, the Bayesian model incorporates the "efficient drift" of odds, leading to more accurate estimation of model parameters. By estimating the (latent) true probability, the analysis can also examine overreaction and underreaction for each individual match.

As before, let i index matches ($i = 1, \ldots, I$) and y_i denote the match outcome ($y_i = 1$ if the scoring team wins, and 0 otherwise). Because we now analyze data at the second-by-second level, t indexes time in *seconds* rather than minutes. For the ith match, we let t_i^0 denote the match time that the first goal is scored, and t_i^r denotes the match time when the market reopens following the first goal. Further, let V_{it} denote the mean-centered log transaction volume at time t.

Let $q_{it} = Pr(y_i = 1|F_{it})$ be the true (unobserved) conditional probability that the scoring team in match i will win, given all the information (F_{it}) available up to time t. Next, we specify a prior distribution for q_{it} by incorporating a model of soccer scoring based on an independent bivariate Poisson process (Maher, 1982). Formally, we specify the following model:

$$\text{logit}(q_{it}) = \text{logit}(q_{i,t-1}) + g_{it} + \varepsilon_{it} \quad (t > t_i^r) \tag{3}$$

$$\varepsilon_{it} \sim N(0, \sigma_{it}^2), \quad \text{where } \sigma_{it}^2 = \begin{cases} \sigma^2 & \text{if } E_{it} = 0 \\ \tau^2 & \text{if } E_{it} = 1 \end{cases} \quad (\sigma^2 < \tau^2), \tag{4}$$

where g_{it} denotes the "efficient drift" in logit(q_{it}), given the current match time t and the events that occur between time $t-1$ and t.[9] Importantly, g_{it} takes into account the remaining amount of time left in the match, as goals scored toward the end of the match are more impactful. The computation of g_{it} is described in Appendix A. E_{it} is an indicator variable that takes the value of 1 if another "material event" (e.g., red card, another goal, penalty awarded) occurs at time t. Thus, the specification in Eq. (4) reflects that, in the absence of another "material event" at time t, the absolute magnitude of the modeling error ε_{it} should be smaller than if another material event has occurred ($\sigma^2 < \tau^2$).

Next, we model the observed in-play implied probabilities as the sum of the true conditional probability, a systematic bias, and random error:

$$\underbrace{\text{logit}(p_{it})}_{\text{implied probability}} = \underbrace{\text{logit}(q_{it})}_{\text{true probability}} - \underbrace{b_{it}}_{\text{systematic bias}} + \underbrace{\zeta_{it}}_{\text{error}}, \quad \zeta_{it} \sim N(0, w_{it}^2), \tag{5}$$

Note that we specify $-b_{it}$ instead of $+b_{it}$ so that the sign of the bias terms is consistent with the previous sections. We then parameterize g_{it} as follows:

$$b_{it} = (\alpha + s_i\beta)e^{-\delta(t-t_i^0)-\gamma V_{it}}. \tag{6}$$

The parameterization of the bias is comprised of two terms: an additive term ($\alpha + s_i\beta$), which, as before, specifies the direction and magnitude of the bias immediately after the first goal as a function of surprise, and a multiplicative term $e^{-\delta(t-t_i^0)-\gamma V_{it}}$ that estimates the extent to which the biases are moderated by time lag ($e^{-\delta(t-t_i^0)}$) and by higher transaction volume ($e^{-\gamma V_{it}}$). A positive value for δ and γ indicates that the bias is attenuated by longer time lag (H2) and higher transaction volume (H3), respectively.

To complete our model specification, we specify a set of standard, weakly informative prior distributions for our model parameters (Gelman et al., 2003), and sample from the posterior distributions of model parameters using a MCMC procedure (Johannes and Polson, 2009) discussed in Appendix B. The results from the Bayesian model are consistent with most of our earlier findings. First, the posterior mean of α and β are 0.231 (95% interval = [0.225, 0.235]) and −0.413 (95% interval = [−0.423, −0.403]), respectively, which implies underreaction in general and overreaction when the first goal is extremely surprising ($s_i > 0.56$). Fig. 4 plots the term $\alpha + s_i\beta$ versus s_i, along with pointwise 95% posterior intervals. As can be seen, because the Bayesian structural model takes into account the efficient drift and the occurrence of other events, the model parameters are estimated with much higher accuracy, and hence the posterior intervals are quite narrow. We find that for large values of s_i, the 95% posterior intervals do not cover zero, indicating that the overreaction is statistically significant.

Our Bayesian model also allows us to assess overreaction and underreaction for each individual match, by looking at the posterior mean estimate of the quantity ($q_{it_i^r} - p_{it_i^r}$), i.e., the estimated true conditional probability minus observed implied probability for each match. A histogram of the posterior mean estimate of $q_{it_i^r} - p_{it_i^r}$ across all matches is shown in Fig. 5. As can be seen, we find that there are both cases of overreaction and underreaction, with an average absolute percentage bias ($|q_{it_i^r} - p_{it_i^r}|/q_{it_i^r}$) of 1.44% and 2.09% when overreaction and underreaction occurs, respectively.

[9] If a goal is scored after 40 min in the first half, the window we analyze may cover the halftime break, where g_{it} should be very close to zero as there is little amount of information. However, from the Betfair and ESPN FC data we are unable to pinpoint the time when the halftime break starts. We do not analyze these goals differently and believe that the impact should be minimal: given that our estimate of g_{it} is typically very small and that only 137 (6.8%) of the 2017 goals in our sample are scored after 40 min in the first half.

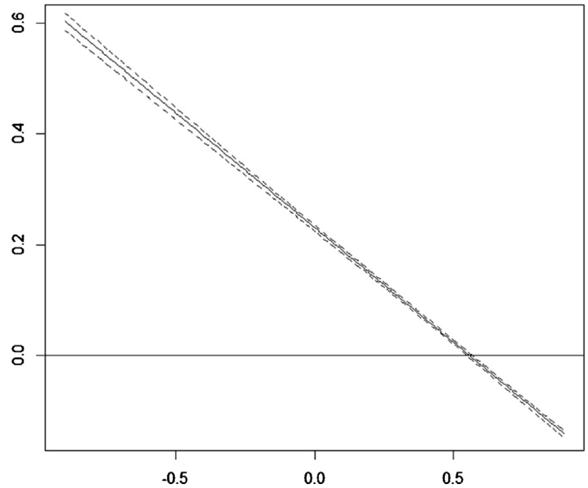

Fig. 4. The bias ($\alpha + s_i\beta$), shown on the y-axis, as a function of surprise (x-axis).

Further, the time-decay parameter δ is estimated to be around 0.009 (with a tight 95% interval of [0.009, 0.009]), indicating that, consistent with H2, the systematic bias decreases over time ($\delta > 0$). Specifically, for each minute after the first goal, the bias is reduced by about 40%. By the end of the 5th minute, the bias decays to around 10% of its original magnitude; this is consistent with the findings in Section 5.3 that the bias more or less vanishes 5 min after the first goal. Finally, we find that the posterior mean of γ is very close to 0, with a 95% interval that covers zero. Thus, H3 is not supported, suggesting that the observed biases are largely independent of betting volume.

5.5. *Economic size of the mispricing bias*

Having established statistical evidences in support of H1a–c and H2, we now explore the economic size of the mispricing bias by developing a strategy that exploits the identified bias. That is, we "back" the outcome "scoring team wins" when the first goal is expected or only moderately surprising ($s_i < c$), and "lay" the outcome "scoring team wins" when the first goal is extremely surprising ($s_i > c$), executed at 2, 3, and 6 min after the first goal. The only remaining issue is how "surprise threshold" c should be chosen. Toward that end, we use a "split-sample" approach where the data is randomly split into two halves. Half of the data are used as a "training sample" to select threshold c (by estimating the logistic regression in Eq. (2) at 2 min after the first goal and then selecting $c = (-\alpha/\beta)$). The profitability of the strategy, which corresponds to the economic size of the mispricing bias, is then evaluated using the remaining half of the data. The split sample analysis is repeated 1000 times to assess the distribution of returns. That is, in each time, a different random split is used, which allows us to generate confidence intervals and hence p-values for the estimated trading profits.

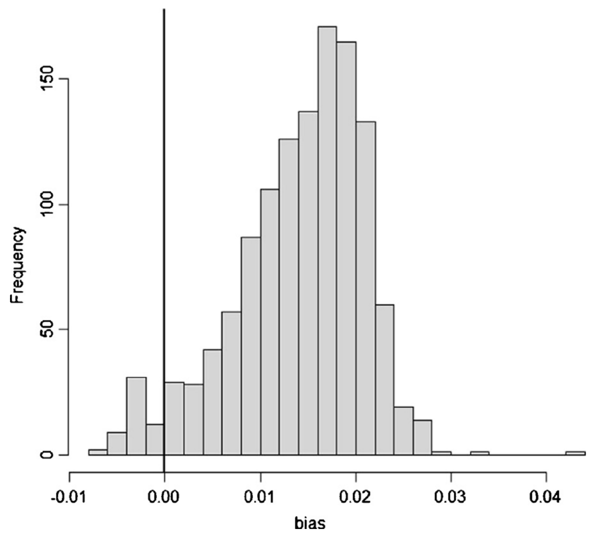

Fig. 5. Histogram of the posterior mean estimates of the bias (estimated true conditional probability minus implied probability) across matches.

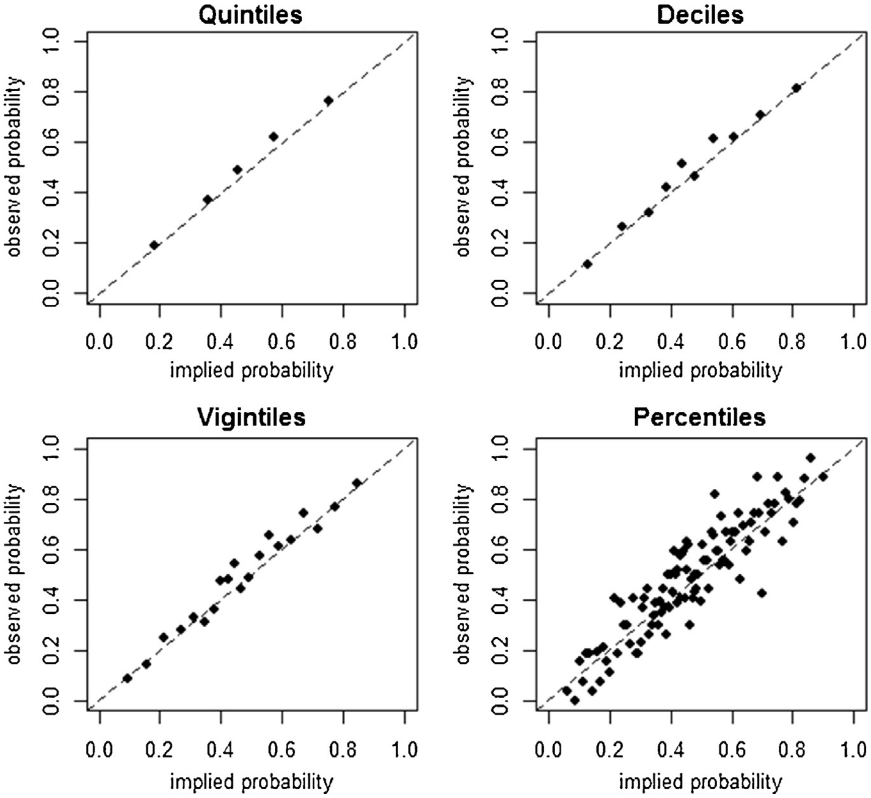

Fig. 6. Implied probability (computed using pre-match odds) versus observed probability.

In each bet, we wager an amount of £1/odds so that we either win or lose £1 in each bet; essentially, we create a portfolio of bets which utilize value-weighting (Fama, 1998). After taking into account the 5% commission charged by Betfair, the estimated profit from our betting strategy is around 2.46% (p = .028) if executed at 2 min after the goal. It decreases to 1.59% (p = .097) if executed at 3 min after the goal, and becomes statistically insignificant (0.73%, p = .28) at 6 min after the goal. As a comparison, the profits we identified here are in line with those in Gray and Gray (1997), who report a strategy that earns around 4% after commission in the National Football League (NFL) betting market, indicating that the size of the mispricing biases due to underreaction and overreaction is economically significant.[10]

5.6. Comparison to the favorite-longshot bias

In this section, we briefly explore the favorite-longshot bias, which may offer an alternative explanation for some of our findings. Specifically, the favorite-longshot bias states that bettors tend to undervalue favorites (Thaler and Ziemba, 1988), presumably due to their risk-loving preference or misperception of probability (Snowberg and Wolfers, 2010). The favorite-longshot bias may offer an alternative explanation for the observed underreaction because after the favorite scores, its winning probability becomes even higher; if bettors indeed undervalue favorites, the odds for the favorite after it scores the first goal would tend to underestimate the probability that the scoring team wins, which is consistent with our findings of underreaction.

To explore the magnitude of the favorite-longshot bias in the soccer betting exchange market, we first analyze the accuracy of the pre-match implied probabilities with respect to actual match outcomes. Specifically, we divide the pre-match implied probabilities in quintiles, deciles, vigintiles, and percentiles, and plot the implied probabilities versus the observed outcome probabilities. The results are shown in Fig. 6.

As can be seen, in all four panels of Fig. 6, the implied probabilities from the betting exchange closely track the actual observed outcome probabilities, as all points on the scatterplots are fairly close to 45° line. This is consistent with the finding in Smith et al. (2009) that pre-match odds accurately reflect the outcome probabilities. In particular, the favorite-longshot bias does not appear to the significant for implied probabilities that are within the range of (0.1, 0.9), as can be seen in

[10] Note that we do not take into account the back-lay spread when computing the above profits. Thus, the estimated profits here should be viewed as an upper bound for the actual profit of the betting strategy.

Table 4
Summary of results across different statistical tests.

Statistical test	H1a (General underreaction)	H1b (Surprise reduces underreaction)	H1c (Extreme surprise leads to overreaction)	H2 (Time moderation)	H3 (Volume moderation)
Nonparametric test (Section 5.2)	Supported	Supported	Directionally consistent	Not tested	Not tested
Logistic regression (Section 5.3)	Supported	Supported	Directionally consistent	Supported	Not tested
Bayesian structural model (Section 5.4)	Supported	Supported	Supported	Supported	Not supported

the "percentile" plot in the lower right panel of Fig. 6, where the two points on the scatterplot that correspond to the 1st percentile (the leftmost point) and the 99th percentile (the rightmost point) are both very close to the 45° line, indicating that the implied probabilities are consistent with the observed outcome probabilities.

In order to further rule out the favorite-longshot bias as an alternative explanation of observed underreaction, we re-estimate the logistic regression model in Section 5.3 (for 2 min after the first goal) by including only observations that have implied probabilities that are within the range of (0.1, 0.9) (where we know from Fig. 6 that the favorite-longshot bias is not significant). The parameter estimates from the logistic regression are $\alpha = 0.135$ ($p = .02$) and $\beta = -0.422$ ($p = .03$), which is fairly similar to the parameter estimates obtained using all observations. Thus, this result suggests that the observed underreaction in our data is not primarily driven by the favorite-longshot bias.

Further, recall that in the logistic regression and the Bayesian structural model, we find the observed biases attenuate in around 5 min (which is consistent with H2). These "time reversal" patterns are inconsistent with both the risk-loving preference or probability misperception mechanisms in the favorite-longshot bias, as risk preference and probability perception tend to be time invariant and thus are unlikely to reverse over time. Thus, taken together, we believe that our results cannot be fully explained by alternative explanations based on the favorite-longshot bias.

6. Discussion and conclusion

In this paper, we develop a behaviorally motivated hypothesis of how conservatism and surprise drive underreaction and overreaction to unanticipated events. Specifically, we hypothesize that, if the new event is expected or only moderately surprising, conservatism will dominate and market participants tend to adjust too little from their prior beliefs, leading to underreaction to the event. A surprising event, on the other hand, induces market participants to overweight its importance due to discrepancies with their current schema. As a result, we predict that while market participants underreact to new events in general, underreaction decreases with surprise, and overreaction occurs when the event is extremely surprising. We test our hypothesis using data from Betfair, an in-play soccer betting market, using several different statistical models. Our results are summarized in Table 4. As can be seen, we find general support for H1a ("General underreaction") and H1b ("Surprise reduces underreaction") across all three statistical tests. For H1c ("Extreme surprise leads to overreaction"), the results of the Bayesian structural model support our hypothesis, while the other two statistical tests produce directionally consistent yet statistically insignificant results. In addition, consistent with H2, we find that the bias found in H1a–c attenuates over time, with the biases disappearing around 5 min after the first goal. Finally, the biases are unrelated to transaction volume, and thus H3 is not supported. In addition to the statistical evidence in Table 4, we also demonstrate the economic significance of the mispricing bias by constructing a strategy that exploits the identified bias.

Our results contrast with that of Croxson and Reade (2013), who conclude that Betfair market is semi-strong form efficient. They analyze the drift in odds after a goal is scored, focusing only on goals that arrive shortly before the half-time break, when there is little new information about the match outcome. The absence of drift at the half-time break suggests that odds reflect goal-related information swiftly and fully. In contrast, our paper develops formal tests that make use of actual match outcomes and (latent) true probabilities. Further, Croxson and Reade (2013) do not differentiate between goals that are "expected" or "surprising", a key driver of our results in Table 4.

Our findings are related to Klibanoff et al. (1998), who empirically show that investors tend to underreact to country-specific economic news, but such underreaction disappears if the news appears on the front page of the *New York Times*. Klibanoff et al. (1998) thus propose that underreaction is caused by the limited attention investors paid to news, as some investors may not be exposed to relevant economic news if such news does not appear on a major newspaper. Our research suggests that even after controlling for exposure to new information (presumably, all bettors in the in-play markets are closely following the soccer match while they are betting, and hence all goals receive full exposure), underreaction and overreaction are still present. We show that underreaction and overreaction are driven by the level of surprise, which can be viewed as a behavioral antecedent to attention paid to the event (Baldi and Itti, 2010), the key driver of underreaction in Klibanoff et al. (1998).

A key feature of our current study is that we use the in-play soccer betting market to understand financial markets in general, which also points to an obvious limitation. Our inferences are only valid and generalizable to financial markets to the extent that the behavior of bettors is comparable to investors in financial markets. Clearly, the set of participants and monetary stakes in betting markets are quite different from those in the stock market, which raise questions about the applicability of experimental or quasi-experimental settings to financial markets (Levitt and List, 2007). This concern is somewhat alleviated because rather large monetary stakes are involved in Betfair, which brings our data collection setting a bit closer to real world financial markets than other purely laboratory-based experiments. Still, in betting markets expected returns are usually negative, and participants' risk preferences may differ substantively from participants in stock market in general. Thus, we believe that the current paper only provides some suggestive evidence and that future research is certainly needed to study the role of surprise in driving underreaction and overreaction in financial markets.

Future research may also look into how certain market microstructure and transactional details of the in-play betting market affect the extent of biased reactions. For instance, future research may study how trading activities differ across favorites and underdogs, and how such differences may be related to price changes. In addition, it may be interesting to study whether a transaction is back- or lay-initiated and whether that affects the extent of the biased reactions. Finally, one could study the role of arbitrageurs in in-play betting markets in more details, similar to the analysis by Forsythe et al. (1992) who studied the Iowa Presidential Stock Market. Such analysis, however, may require a much richer dataset that also records the identities of traders and their trading behavior during the course of a match, which are unavailable in our current data.

Acknowledgements

We acknowledge the General Research Fund of the Research Grants Council of Hong Kong (Project number: 641011) for financial support.

Appendix A. Computation of the bivariate Poisson process and the "efficient drift" g_{it}

First, we assume that goal scoring by the two teams follows an independent bivariate Poisson process (Maher, 1982) with rates λ_{i1} and λ_{i2}, respectively. Then, given the Poisson rate parameters (λ_{i1}, λ_{i2}), the current goal difference (z_{it}), and the current match time (t), the conditional probability that the scoring team will win can be computed numerically. We denote this conditional probability as $f(\lambda_{i1}, \lambda_{i2}, z_{it}, t)$ (note that a closed-form expression for f is not available). Next, we compute the efficient drift by:

$$g_{it} = \text{logit}(f(\lambda_{i1}, \lambda_{i2}, z_{it}, t)) - \text{logit}(f(\lambda_{i1}, \lambda_{i2}, z_{it}, t-1)) \tag{A1}$$

For concreteness, we provide the following numerical example. Let $\lambda_{i1} = 1.0$, $\lambda_{i2} = 0.7$, and that team 1 (the scoring team) is leading by 1-0 at $t = 1800$ (30 min into the match); we assume that a match, including injury time, lasts 95 min. For the rest of the match, the number of goals scored by team 1 and team 2 follow Poisson distributions with rates $((5700 - 1800)/5700)\lambda_{i1} = 0.68$ and $((5700 - 1800)/5700)\lambda_{i2} = 0.48$, respectively. Thus, the conditional probability that team 1 will win the match is $f(1.0, 0.7, 1, 1800) = 0.777496$. Similarly, if no additional goal occurs in the next second, we can compute the conditional probability that team 1 wins as $f(1.0, 0.7, 1, 1801) = 0.777517$. Clearly, because team 1 is in the lead, the passage of time without a goal increases the conditional probability that team 1 will win, resulting in a positive drift. Thus, we have $g_{it} = \text{logit}(0.777517) - \text{logit}(0.777496) = +0.00012$.

The only computation left is in the estimation of (λ_{i1}, λ_{i2}). We calibrate the (λ_{i1}, λ_{i2}) parameters using the pre-match implied probabilities. That is, the values of (λ_{i1}, λ_{i2}) are chosen such that the (team 1 win, draw, team 2 win) probabilities are in maximal agreement with the pre-match implied probabilities. The details are available from the authors upon request.

One may argue that the independent bivariate Poisson scoring process is an over-simplification of the scoring process in an actual soccer match. Other refinements that build upon the bivariate Poisson model have been proposed in the literature (e.g., Dixon and Coles, 1997). While these refinements may allow the function f to capture the "true" conditional probabilities more accurately, we prefer our current specification for three reasons. First, since we only use the first difference of f, the difference between models is likely to be minimal. Second, in Eqs. (3) and (4) of our model we have already included an idiosyncratic error term ε_{it} to capture any modeling error; our estimation shows that the size of the error term tends to be small (posterior mean of $\sigma^2 = 0.00088$), indicating that the bivariate Poisson distribution is an adequate description of the evolution of the true conditional probability. Finally, the independent bivariate Poisson assumption provides vast computational advantage over more elaborate models; specifically, it allows us to calibrate (λ_{i1}, λ_{i2}) easily, which is not the case for other models.

Appendix B. Prior specification and MCMC sampling

We specify a uniform distribution on $q_{it_i^r}$, diffuse $N(0, 100^2)$ distributions for the parameters (α, β, δ, γ), and weakly informative $Inv - \chi^2(0.001, 1)$ distribution for the variance parameters σ^2, τ^2, ω^2 (Gelman et al., 2003). The MCMC procedure utilizes the forward filtering, backward sampling (FFBS) algorithm (Carter and Kohn, 1994), which allows us to explore the posterior distribution of model parameters very efficiently and hence minimizes the number of iterations needed for convergence (Carter and Kohn, 1994). The details of our computation procedure, including the source code, are available upon request. In each MCMC iteration, we draw from the full conditional distribution of each model parameter in the following order:

1. *Drawing from the full conditional distribution of $q_{it_i^r}$*: A Gaussian random-walk Metropolis–Hastings algorithm is used to draw logit($q_{it_i^r}$). The scale of the Gaussian random walk proposal distribution is tuned to achieve an acceptance rate of around 40% (Gelman et al., 2003).
2. *Drawing from the full conditional distribution of q_{it} $(t > t_i^r)$*: For logit(q_{it}) other than the first period, the Gaussian form of Eqs. (3) and (5) allows us to sample from its full conditional distribution using FFBS algorithm (Carter and Kohn, 1994). Basically, we compute the conditional mean and variance of logit(q_{it}) for each t using a modified Kalman filter (Kalman, 1960), then sample all logit(q_{it}) jointly. This allows us to sample from the full conditional distribution of logit(q_{it}) very efficiently (Carter and Kohn, 1994).
3. *Drawing from the full conditional distribution of $(\alpha, \beta, \delta, \gamma)$*: We again use a random walk Metropolis–Hastings algorithm. We sample each parameter one-by-one, and adapt the scale of the random-walk proposal distribution to achieve an acceptance rate of around 40%.
4. *Drawing from the full conditional distribution of $(\sigma^2, \tau^2, \omega^2)$*: Because the variance parameters are given conjugate, weakly informative priors, we sample from the full conditional distribution of $(\sigma^2, \tau^2, \omega^2)$ using standard conjugate computations (Gelman et al., 2003).

The above procedure is repeated for 2000 iterations, with the first 1000 draws discarded as burn-in. Standard diagnostics confirm that the Markov chain has converged with respect to key parameters. Thus, the last 1000 draws are used to summarize the posterior distribution of model parameters.

References

Asch, P., Malkiel, B.G., Quandt, R.E., 1984. Market efficiency in racetrack betting. Journal of Business 57, 165–175.
Avery, C., Chevalier, J., 1999. Identifying investor sentiment from price paths: the case of football betting. Journal of Business 72, 493–521.
Baldi, P., Itti, L., 2010. Of bits and wows: a Bayesian theory of surprise with applications to attention. Neural Networks 23, 649–666.
Barberis, N., Shleifer, A., Vishny, R., 1998. A model of investor sentiment. Journal of Financial Economics 49, 307–343.
Brooks, R.M., Patel, A., Su, T., 2003. How the equity market responds to unanticipated events. Journal of Business 76, 109–133.
Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.
Chan, W.S., 2003. Stock price reaction to news and no-news: drift and reversal after headlines. Journal of Financial Economics 70, 223–260.
Coleman, L., 2011. An exploratory analysis of factors influencing initial market response and media reports following shock corporate events. The Financial Review 46, 313–336.
Croxson, K., Reade, J.J., 2013. Information and efficiency: goal arrival in soccer betting. The Economic Journal (in press).
Daniel, K., Hirshleifer, D., Subrahmanyam, A., 1998. Investor psychology and security market under- and overreactions. Journal of Finance 53, 1839–1885.
Dixon, M.J., Coles, S.G., 1997. Modelling association football scores and inefficiencies in football betting market. Journal of Royal Statistical Association (Series C) 46, 265–280.
Durham, G.R., Hertzel, M.G., Martin, J.S., 2005. The market impact of trends and sequences in performance: new evidence. Journal of Finance 60, 2551–2569.
Durham, G.R., Santhanakrishnan, M., 2008. The relative importance of strength and weight in processing new information in the college football betting market. Journal of Prediction Markets 2, 13–28.
Edwards, W., 1968. Conservatism in human information processing. In: Kleinmutz, B. (Ed.), Formal Representation in Human Judgment. John Wiley and Sons, New York, pp. 17–52.
Fama, E.F., 1998. Market efficiency, long-term return, and behavioral finance. Journal of Financial Economics 49, 283–306.
Forsythe, R., Nelson, F., Neumann, G.R., Wright, J., 1992. Anatomy of an experimental political stock market. American Economic Review 82, 1142–1161.
Gandar, J., Zuber, R., O'Brien, T., Russo, B., 1988. Testing rationality in the point spread betting market. Journal of Finance 43, 995–1008.
Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. Bayesian Data Analysis, 2nd ed. Chapman & Hall.
Gil, R., Levitt, S.D., 2007. Testing the efficiency of markets in the 2002 World Cup. Journal of Prediction Markets 1, 255–270.
Golec, J., Tamarkin, M., 1991. The degree of inefficiency in the football betting market. Journal of Financial Economics 30, 311–323.
Gray, P.K., Gray, S.F., 1997. Testing market efficiency: evidence from the NFL sports betting market. Journal of Finance 52, 1725–1737.
Griffin, D., Tversky, A., 1992. The weighting of evidence and the determinants of confidence. Cognitive Psychology 24, 411–435.
Hong, H., Stein, J.C., 1999. A unified theory of underreaction, momentum trading, and overreaction in asset markets. Journal of Finance 54, 2143–2184.
Ikenberry, D., Lakonishok, J., Vermaelen, T., 1995. Market underreaction to open market share repurchases. Journal of Financial Economics 39, 181–208.
Ikenberry, D., Ramnath, S., 2002. Underreaction to self-selected news events: the case of stock splits. Review of Financial Studies 15, 489–526.
Ikenberry, D., Rankine, G., Stice, E., 1996. What do stock splits really signal? Journal of Finance and Quantitative Analysis 31, 357–375.
Itti, L., Baldi, P., 2009. Bayesian surprise attracts human attention. Vision Research 49, 1295–1306.
Jegadeesh, N., Titman, S., 1993. Returns to buying winners and selling losers: implications for stock market efficiency. Journal of Finance 48, 65–91.
Johannes, M., Polson, N., 2009. MCMC methods for continuous-time financial econometrics. In: Ait-Shahlia, Y., Hansen, L.P. (Eds.), Handbook of Financial Econometrics. Applications, vol. 2. North Holland, Amsterdam, pp. 1–72.
Kalman, R.E., 1960. A new approach to linear filtering and prediction problems. Transactions of ASME: Journal of Basic Engineering 82 (D), 35–45.
Klibanoff, P., Lamont, O., Wizman, T.A., 1998. Investor reaction to salient news in closed-end country funds. Journal of Finance 53, 673–699.
Levitt, S.D., List, J.A., 2007. What do laboratory experiments measuring social preferences reveal about the real world? Journal of Economic Perspectives 21, 153–174.
List, J.A., 2009. An introduction to field experiments in economics. Journal of Economic Behavior and Organization 70, 439–442.
Maguire, R., Maguire, P., Keane, M.T., 2011. Making sense of surprise: an investigation of the factors influencing surprising judgments. Journal of Experimental Psychology: Learning, Memory, and Cognition 37, 176–186.
Maher, M.J., 1982. Modeling association football scores. Statistica Neerlandica 36, 109–118.
Meyer, W., Reisenzein, R., Schutzwohl, A., 1997. Toward a process analysis of emotions: the case of surprise. Motivation and Emotion 21, 251–274.
Michaely, R., Thaler, R.H., Womack, K., 1995. Price reactions to dividend initiations and omissions: overreaction or drift? Journal of Finance 50, 573–608.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. Journal of Royal Statistical Society (Series A) 135, 370–384.
Oliven, K., Rietz, T., 2004. Suckers are born but markets are made: individual rationality, arbitrage, and market efficiency on an electronic futures market. Management Science 50, 336–351.
Ofir, C., Mazursky, D., 1997. Does a surprising outcome reinforce or reverse hindsight bias. Organizational Behavior and Human Decision Processes 69, 51–57.
Pezzo, M., 2003. Surprise, defence, or making sense: what removes hindsight bias? Memory 11, 421–441.
Plutchik, R., 1980. Emotion: A Psychoevolutionary Synthesis. Harper and Row, New York.

Reisenzein, R., Meyer, W., Niepel, M., 2012. Surprise. In: Ramachandran, V.S. (Ed.), Encyclopedia of Human Behavior. , 2nd ed. Academic Press, Boston.
Smith, M.A., Paton, D., Williams, L.V., 2009. Do bookmakers possess superior skills to bettors in prediction outcomes? Journal of Economic Behavior and Organization 71, 539–549.
Snowberg, E., Wolfers, J., 2010. Explaining the favorite-long shot bias: is it risk-love or misperceptions? Journal of Political Economy 118, 723–746.
Teigen, K.H., Keren, G., 2003. Surprise: low probabilities or high contrasts? Cognition 87, 55–71.
Thaler, R.H., Ziemba, W.T., 1988. Anomalies: Parimutuel betting markets: racetracks and lotteries. Journal of Economic Perspectives 2, 167–174.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics with S. Springer, New York.
Wasserman, L., 2010. All of Statistics: A Concise Course in Statistical Inference. Springer, New York.
Wolfers, J., Zitzewitz, E., 2004. Prediction markets. Journal of Economic Perspectives 18, 107–126.