The Role of Surprise:

Understanding Over- and Underreactions in Prediction Markets

Darwin Choi and Sam K. Hui *

September 9, 2012

* Darwin Choi (dchoi@ust.hk) is an Assistant Professor of Finance in Hong Kong University of Science and Technology, and Sam K. Hui (khui@stern.nyu.edu) is an Assistant Professor of Marketing in the Stern School of Business at New York University. We acknowledge the General Research Fund of the Research Grants Council of Hong Kong (Project Number: 641011) for financial support. Corresponding author: Sam K. Hui.

The Role of Surprise:

Understanding Over- and Underreactions in Prediction Markets

Abstract

Researchers typically use prediction markets to conduct “event studies” by comparing the price of a contract before and after an event occurs. However, the underlying assumption that market participants react to new events in an unbiased manner is rarely tested. In this paper, we study over- and underreactions in prediction markets using in-play soccer betting data. Our results suggest that over- and underreactions are driven by how surprising the event is: while market participants in general underreact to new events, they tend to overreact to events that are highly surprising. We propose a behavioral explanation based on conservatism and information salience, and discuss how researchers can deal with these biases when conducting prediction market event studies.

Keywords: prediction market, event study, over- and underreactions, surprise, conservatism, salient information, Bayesian model.

1. Introduction

Recently, prediction markets have found widespread use in many business settings. Successful applications of prediction markets include, for example, business forecasting (Spann and Skiera 2003), new product development (Dahan et al. 2010; Ho and Chen 2007), product concept evaluation (Chen et al. 2009; Dahan et al. 2011), prediction of movie box office revenue (Spann and Skiera 2003), and financial forecasting (Berg et al. 2009).

Prediction markets are often used to conduct event studies (Kothari and Warner 2006), defined as empirical research that quantifies the impact of an event (e.g., outcome of the U.S. presidential election) on a certain metric of interest (e.g., stock market return). Using prediction markets for event studies is especially appropriate, and sometimes the only feasible method, when relevant financial data are not publicly available (Slamka et al. 2008). For instance, Elberse (2007) studies the impact of casting announcement on box office performance using data from Hollywood Stock Exchange (HSX), an online prediction market, and concludes that stars are worth approximately $3million in theatrical box office revenue. Similarly, Henderson et al. (2010) proposes a prediction market event study to assess the impact of death penalty legislation on homicide rates. In the above applications, researchers utilize the change in the price of a certain contract (e.g., expected box office performance of the movie *Cold Mountain*) in the prediction market *before* and *after* an event (Tom Cruise dropped out) occurs to assess the impact of that event.

Using prediction markets for event studies, however, involves the untested assumption that market participants, on aggregate, react in an "unbiased" manner to the focal event (Elberse 2007; Srinivasan and Bharadwaj 2005). Any systematic over- and underreactions to new events would clearly cause biases in the assessment of an event's impact. For instance, if market participants on HSX become overly pessimistic about the prospect of the movie *Cold Mountain* after Tom Cruise dropped out, the corresponding event study will overestimate the impact of Tom Cruise's (non-) involvement and overstate his "worth." While some research studies over- and underreactions in financial markets (discussed in Section 2.1), not much research has been conducted to understand biased reactions in prediction markets. This is partly because the study of prediction markets is still a rather nascent area, and partly because "clean" settings that allow one to conduct empirical testing are generally unavailable.

In this paper, we hypothesize that over- and underreactions to new events are driven by how "surprising" the event is.[1] Our argument is built upon two widely-accepted cognitive biases: conservatism (Edwards 1988) and overweighting of salient information (Griffin and Tversky 1992). We conjecture that, when reacting to an event that is in line with expectations or only moderately surprising, market

[1] Throughout this paper, we use the terms "new event," "new information," "information shock", and "news" interchangeably.

participants tend to be conservative, i.e., they rely too much on their prior beliefs and hence underreact. In contrast, news that is more surprising (i.e., that strongly violates prior beliefs) attract more attention and become more salient in the decision making process (Ofir and Mazursky 1997; Dunlosky et al. 2000). Thus, we expect that underreaction decreases with surprise, as market participants assign more "weight" to the more surprising new information when updating their beliefs. If the new information is extremely surprising, market participants may even overweight the new information, leading to overreaction.

We turn to data from an in-play soccer betting market, where participants place bets while a match is still under way, to empirically test our hypotheses. The in-play soccer betting market offers several advantages for studying biased reactions in prediction markets. First, the arrival of information (goals) is apparent and its impact on odds can be compared with actual match outcomes. Second, it is easy to define how "surprising" a goal is by comparing the relative strength of the two teams: goals scored by the "underdog" are more surprising than goals scored by the "favorite." Third, goals are clearly exogenous; this may not be the case in other settings. Further, in a soccer betting market, real money is at stake and transactional volume is concentrated within a short time horizon (about 95 minutes), which provides a strong test for the existence of biases in prediction markets. That is, if we find biases in the soccer betting market, we can reasonably expect such biases to be even stronger (or take longer to correct) in other settings that use "play money" (Slamka et al. 2008) or last for a longer duration (e.g., in new production development applications). We return to this issue in Section 6.

Our dataset is comprised of second-by-second transaction records in 2,017 international soccer matches obtained from Betfair, an online sports betting exchange. The total betting volume in our sample over the period 2006-2011 amounts around £3 million. We operationalize how "surprising" a goal is by the implied probability that the non-scoring team wins minus that of the scoring team, measured right before the goal. We document over- and underreactions to the first goal of the match using a sequence of logistic regressions of actual match outcomes on the implied probability that the scoring team wins. Taking one step further, we develop a Bayesian model that describes how these biases change over time and vary with betting volume. We find that market participants underreact to goals that are not surprising or only moderately surprising. Further, underreaction is moderated by the degree of surprise, and the results point to overreaction if the goal is very surprising. These biases are economically significant, as we demonstrate using a profitable betting strategy.

In summary, we make three main contributions. First, our paper is one of the first to use the in-play betting market to study over- and underreactions in prediction markets. Second, we demonstrate that market participants underreact to events that are expected or only moderately surprising, but overreact to very surprising events. Third, we provide several suggestions on how researchers can alleviate these biases when conducting prediction market event studies.

2. Background and literature review

2.1. Over- and underreactions in financial markets

Financial markets share several structural similarities with prediction markets (Wolfers and Zitzewitz 2004). In Table 1, we summarize the previous literature on over- and underreactions in financial markets in terms of the setting or the type of information, whether over- and underreaction is found, and the proposed theoretical explanation.

[Insert Table 1 about here]

Table 1 reveals several limitations of the previous literature. First, even in financial markets, "clean" empirical evidence is generally difficult to find, as some phenomena can be viewed as evidence of opposite claims. For instance, momentum in stock price (Jegadeesh and Titman 1993) can be viewed as both the result of underreaction (Barberis et al. 1998; Hong and Stein 1999) and overreaction (Daniel et al. 1998). Second, most of the previous literature focuses on the medium to long time horizon (typically in months or quarters), which is much longer than the typical "event windows" considered in event studies. Further, previous research has not considered the role of surprise in driving over- and underreactions.

2.2. Sports betting market

Most research on sports betting market focuses on "pre-match" betting markets, where bettors are allowed to place bets only before the match starts. Examples include Golec and Tamarkin (1991), Gandar et al. (1998), Durham et al. (2005), and Durham and Santhanakrishnan (2008). Obviously, research on pre-match markets cannot study real-time reactions to information shocks.

In an in-play sports betting market, bettors can place bets while the match is still under way. Gil and Levitt (2007) collect data on fifty Soccer World Cup matches in 2002 from intrade.com. They find that the odds that a certain team wins continue to drift ten to fifteen minutes after a goal, which they interpret as violations of market efficiency. This claim is subsequently disputed by Croxson and Reade (2011), who argue that even if the in-play odds are efficient, they should continue drifting after a goal is scored.[2] They assemble a dataset of over 1,200 soccer matches from Betfair, along with the corresponding scoring data, to examine the efficiency of the in-play betting market. An important insight drawn from their research is that in order to accurately assess market efficiency, one must incorporate a model of how efficient prices drift as time passes. We build upon this insight when developing our Bayesian structural model in Section 5.4 to study over- and underreactions.

[2] Suppose that Team 1 scores the first goal of the match. Under the assumption of efficient odds, 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.

3. Hypotheses

In this section, we develop our key hypotheses about how "surprise" drives biased reacctions to new information. Our hypotheses are based on two cognitive biases, conservatism and information salience. These cognitive biases are particularly relevant because when reacting to new information, participants in a prediction market have to make a split-second judgment, and thus are more likely to rely on behavioral heuristics.

We use the following stylized model of belief update to illustrate our hypotheses. Upon receiving new information, a market participant has to incorporate 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,

$$\text{New Belief} = w_1 \times \text{Prior Belief} + w_2 \times \text{New Information}.$$

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. We now develop our hypotheses with reference to w_1 and w_2 . First, 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.

Second, we hypothesize that this general pattern of underreaction is moderated by how "surprising" the new information is. Previous research in psychology finds that people tend to assign more "weight" to salient information; i.e., events that are distinctive, obvious, and prominent affect judgment disproportionately. For instance, Griffin and Tversky (1992) show that people overreact to the strength or extremeness of the evidence (e.g., the warmth of a recommendation letter) and underreact to weight (e.g., the credibility of the person who writes the letter). Further, by tracking participants' visual focuses, Itti and Baldi (2009) and Baldi and Itti (2010) show that surprising information draws more attention. Ranganath and Rainer (2003) review research in neuroscience and suggest how novel events cause neural responses that result in greater attention to those events. Thus, we hypothesize that underreaction is moderated by surprise, as under-adjustment due to conservatism is offset by the overweighting due to the salience of the new information. That is, for new information that is more surprising, w_1 should get closer to its normative value, resulting in a reduction of underreaction.

Third, taking the above argument one step further, if the new information is extremely surprising, the overweighting of salient information (i.e., higher w_2) will more than offset the effect of conservatism, resulting in overreaction.

We now turn to the context of the in-play soccer betting market. In soccer betting, the implied probability computed from the reciprocal of betting odds (Wolfers and Zitzewitz 2004) can be used to gauge biased reactions to goals: if bettors overreact (underreact) to a goal, the implied winning probability of the scoring team will be too high (low).

Based on our hypotheses, bettors' reactions to the first goal of a match depends on how "surprising" that goal is, operationalized using the relative strength of the two teams: the favorite is defined as the team having a higher (lower) winning probability shortly before the first goal of the match is scored. Note that we study only the first goal of each match because the timing of the first goal can be identified more accurately from our data than other goals (as will be discussed in Section 4).

Thus, we re-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 if the goal is not surprising or only moderately surprising.

H1b (Surprise reduces underreaction): A surprising goal attracts attention and is more salient. Thus, we expect lower underreaction for more surprising goals.

H1c (Very high surprise leads to overreaction): A goal that is extremely surprising triggers overreaction, as bettors tend to overweight the salient shock.

Next, we propose two other hypotheses (H2 and H3) on how the biases in H1abc vary with time lag and betting volume. We expect that the biases will be corrected over time as arbitrageurs may form strategies to profit from over- and underreactions. Our objective is to formally test the correction and document the amount of time it takes.

H2 (Time moderation): The bias found in H1abc attenuates over time after the first goal is scored.

Finally, we also examine if higher betting volume corresponds to lower bias. Although our data do not record bettors' identities, one possible way to distinguish between professional bettors and small bettors is to examine volume. The intuition is that higher volume involves higher financial stakes and is associated with more rational or informed bettors (Easley and O'Hara 1987). Our last hypothesis tests whether there is any systematic relationship between the biased reactions and betting volume:

H3 (Volume moderation): The bias found in H1abc decreases with higher betting volume.

4. Data

We test our hypotheses using data from Betfair, an online sports betting exchange. We obtain all available matches of major international and European soccer tournaments, 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 (2006/07 to 2010/11), as well as one season of La Liga in Spain, Serie A in Italy, and Bundesliga in Germany (2010/11).

Betfair allows participants to submit "back" and "lay" orders to bet on or against an outcome (respectively). A "back" bet on a certain team winning the match with decimal odds of 5 will receive $5 for every $1 wagered (including the original $1), 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. Thus, any biases we observe from Betfair odds come from the market participants, as the odds are not set by a bookmaker. Betfair charges a commission of up to 5% on net winnings.

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. All data within a second are aggregated. Participants can place bets before the game starts as well as when the match is underway. During the match, Betfair suspends trading briefly (for around two minutes) 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 augment our dataset by collecting corresponding event information from ESPN Soccernet (soccernet.espn.go.com).

Out of all matches, 2,160 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 Soccernet. 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 2,017 matches. Table 2 presents some summary statistics of our dataset. The matched in-play volume per match is around £1.5 million on average. The first goal of the match is, on average, scored around 36 minutes after the match has started. We also calculate the betting volume in the window [+2, +6 minutes] 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, and 71% of the scoring teams eventually win the match. We also show the statistics of the implied probabilities at 2, 3, and 6 minutes after the first goal (these probabilities are described in Section 5.1).

[Insert Table 2 about here]

5. Empirical analysis of over- and underreactions

5.1. Operationalization of the surprise metric

We begin by quantifying the degree of "surprise" of the first goal. We define s_i, the "surprise" metric, as the implied probability that the "non-scoring team" wins the match minus the implied probability that the "scoring team" wins, where both implied probabilities are measured at one minute before the first goal of the match. A positive (negative) s_i means that the scoring team is relatively weaker (stronger) than the non-scoring team, and that the goal is surprising (expected).

Following Wolfers and Zitzewtiz (2004), the implied probabilities are computed by taking the reciprocal of the decimal odds (adjusted by a very small normalizing constant so that the probabilities sum up to 1). 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.

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

We conduct a non-parametric analysis that explores the relationship between actual match outcomes and implied probabilities for different values of s_i. Specifically, we first calculate the implied probability p_i that scoring team wins using odds two minutes after the first goal. We let y_i be an indicator variable that takes a value of 1 if the scoring team in match *i* wins, and 0 otherwise. If bettors' reaction to the first goal is unbiased, 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 ($\sum_i y_i$), and use the difference as a statistical measure of biased reactions. If there is over-(under)reaction, the expected count should be larger (smaller) than the actual count.

We divide the data into two sets: "expected" goal, which corresponds to the condition ($s_i < -c$), and "surprise" goals, which corresponds to the condition ($s_i > c$). The results are shown in Table 3, along with a sequence of p-values, generated using Monte Carlo simulations (Wasserman 2010) that assess whether the observed data are significantly different from their expected values.

[Insert Table 3 about here]

As can be seen in the left panel of Table 3, when the first goal is "expected," p_i underestimates the conditional probability $P(y_i = 1)$, indicating underreaction to an "expected" shock (H1a). This effect is statistically significant for most values of the threshold *c*. However, as seen in the right panel of Table 3, this effect disappears when the first goal is a "surprise" (H1b). Interestingly, the deviations of actual vs.

expected number of wins by the scoring teams go in the opposite direction when the goal is very 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 for the "surprise" condition is statistically significant, presumably because of the small sample size which results in the non-parametric tests having low power (Wasserman 2010).

Next, Figure 1 shows a smoothing spline (Venables and Ripley 2002) that 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, indicating underreaction. As s_i becomes larger, the non-parametric estimate of $(p_i - y_i)$ turns positive, suggestion the presence of overreaction. Note also that the smoothing spline is fairly linear and roughly monotonically increasing, which provides some validation for the parametric assumptions that we will make in the subsequent analyses.

[Insert Figure 1 about here]

5.3. A sequence of logistic regressions

To estimate the magnitude of over- and underreactions as a function of surprise, we estimate a sequence of logistic regression:

$$y_i \sim \text{Bernoulli}(\pi_{it}) \qquad [1]$$

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

where, as before, y_i denotes the outcome of match *i*; p_{it} denotes the implied probability (that the scoring team will win) at the *t*-th minute after the first goal. The term ($\alpha_t + s_i \beta_t$) represents the bias; clearly, a positive (negative) value of $\alpha_t + s_i \beta_t$ indicates under-(over)reaction. The models in [2] are estimated using the package *glm*() in R and the results are summarized in Table 4.

[Insert Table 4 about here]

First, we find that the estimated model at *t* = 2 is $\text{logit}(\pi_{it}) = \text{logit}(p_{it}) + 1.57 - 0.385 s_i$, with a p-value of 0.001, showing that this model fits significantly better compared to the model that assumes "efficient odds" (i.e., $\alpha_t = \beta_t = 0$). Given that s_i ranges from -0.92 to 0.88, $\alpha_t + s_i \beta_t$ ranges from 0.51 to -0.18. Thus, we observe underreactions generally ($\alpha_t > 0$; H1a). Underreaction is moderated by surprise ($\beta_t < 0$; H1b), and when s_i > 0.41, we observe overreaction (H1c).

Second, we find that, consistent with H2, the biases are attenuated over time. We see that the p-values in the second-last column of Table 4 increases with time. The p-value is still below 0.05 by the 4th minute, but becomes marginally significant (*p* = 0.07) at the 5th minute and insignificant (p = 0.11) after six minutes, indicating that 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 three or four minutes after the first goal. Together, this suggests that the systematic bias in H1 is corrected in around five minutes.

Further, the last column of Table 4 shows the result of a Hosmer-Lemeshow (HL) test (Hosmer and Lemeshow 2000), which assesses the robustness of our parametric assumption and helps rule out some alternative explanations. The HL test compares a probability assessment versus actual outcome without making parametric assumptions. Basically, the HL-test divides the data into *k* "bins" (chosen to be *k* = 20 here), and check the accuracy of the probability assessment in each bin; the test statistics follows a χ^2_{k-2} distribution. Consistent with our earlier results with logistic regression, the HL-test rejects the "efficient" hypotheses at 2 to 4 minutes after the first goal, and becomes insignificant after the 5^{th} minute. Further, we find that the HL-test for pre-match implied probabilities is also insignificant, indicating that the pre-match odds are quite efficient. This helps rule out alternative explanations based on bettors' "preferences," which exist before the match and are unrelated to reaction to information shocks: e.g., the "favorite-longshot bias" (Thaler and Ziemba 1988), or preference for "positive skewness." (Kumar 2009).

5.4. A formal Bayesian model of in-play odds

In the previous section, 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 bias is reduced over time. We also did not control for the occurrence of other material events (e.g., red cards, another goal) after the first goal. In this section, we analyze the 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). As discussed earlier, the Bayesian model incorporates the "efficient drift" of odds, leading to more accurate estimation.

As before, let i index matches (i = 1, …, I) and y_i denotes 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* index time in *seconds* rather than minutes. For the *i*-th 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 of q_{it} by incorporating a model of soccer scoring based on an independent bivariate Poisson process (Maher 1982). Formally, we specify the following model:

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

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

where g_{it} denotes the "efficient drift" in $\text{logit}(q_{it})$, given the events that occur between time *t*-1 and *t*. The computation of g_{it} is described in Appendix I. E_{it} is an indicator variable that takes a value of 1 if another "material event" (e.g., red card) occurs at time *t*. Thus, the specification in Equation [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:

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

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

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

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 and by higher transaction volume. 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 specification, we specify a set of standard, weakly informative prior distributions for our model parameters, and sample from the posterior distributions of model parameters using MCMC (Johannes and Polson 2009), details of which are discussed in Appendix II. The results from the Bayesian model confirm 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 very surprising ($s_i > 0.56$). Figure 2 plots the term $\alpha + s_i\beta$ versus s_i, along with pointwise 95% posterior intervals. As can be seen, for large values of s_i, the 95% posterior intervals do not cover zero, indicating that the overreaction is statistically significant.

[Insert Figure 2 about here]

Our Bayesian model allows us to assess over- and underreaction for each individual match, by looking at the posterior 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. We find that there are both cases of over- and underreaction, with an average bias $|q_{it_i^r} - p_{it_i^r}|/q_{it_i^r}$ of 1.44% and 2.09% when overreaction and underreaction occur, respectively.

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. 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 roughly consistent with the findings in Section 5.3. In addition, 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; the biases are largely independent of betting volume.

To sum up, we find that results for H1abc are generally consistent with those shown in Sections 5.2 and 5.3, suggesting that our results are robust to the choice of statistical test and the level of data aggregation (match-level, minute-by-minute, and second-by-second).

5.5. Economic evidence

We demonstrate the economic significance of our findings by developing a betting strategy that exploits the identified biases. Specifically, we "back" the outcome "scoring team wins" when the first goal is expected or moderately surprising ($s_i < 0.4$) and "lay" the outcome "scoring team wins" when the first goal is very surprising ($s_i > 0.4$), executed at 2, 3, and 6 minutes after the first goal. In both cases, we bet an amount of £1/odds so that we either win or lose £1 in each bet. Table 5 shows the result of this betting strategy, along with p-values generated using non-parametric bootstrap (Efron and Tibshirani 1993). As can be seen, after commissions, the profit from our betting strategy is around 2.79% (p = .02) if executed at two minutes after the goal. It decreases to 1.85% (p = .07) if executed at three minutes after the goal, and becomes statistically insignificant (0.82%, p = .26) at six minutes after the goal. Interestingly, the profits we identified here is in line with those in Gray and Gray (1997), who report a strategy that earns around 4% after commission in the NFL betting market.

[Insert Table 5 about here]

Table 5 also reports the average actual transaction volume in one-minute windows after we execute the strategies. For example, at the [+2, +3 minutes] window, the average volume per match is

£42,789. The actual betting activity is generally higher when the goal is less surprising (not reported in the table), and when the time window is closer to the goal. These findings suggest that there is enough liquidity to execute our strategies, and that the biases we find are unlikely a result of delayed responses to the goals.

6. Discussion and conclusion

Using in-play soccer betting data, we show that participants in prediction markets do not always react to events in an unbiased manner. Instead, market participants over- and underreact to information shocks depending on the level of surprise. Our findings are consistent with two behavioral biases: conservatism and information salience. If the new event is expected, market participants tend to adjust too little from their previous beliefs. A surprising event, on the other hand, is likely to be salient and induces market participants' overweighting of its importance. As a result, underreaction decreases with surprise and overreaction occurs when the event is very surprising. We also demonstrate the economic significance of our findings.

Our findings lead to several implications for researchers who want to conduct event studies using prediction markets. Importantly, we demonstrate that market participants in a prediction market are subject to biases similar to those in financial markets. The statistical and economic significance of such biases suggest that the underlying assumption behind prediction market event studies, that participants react to new events in an "unbiased" manner, does not always hold. Thus, researcher should be cautious when interpreting estimates from prediction market event studies. Further, we outline several strategies that researchers can use to minimize the impact of over- and underreaction biases.

First, the choice of an appropriate "event window" is crucial for any prediction market event study. Ideally, because of the biases that market participants exhibit, researchers should pick an event window that is long enough for such biases to correct, so that prices in the prediction market have already settled back at the "efficient" level. Obviously, this involves a tradeoff: the event window cannot be too large because other events may happen in the interim and cause confounding issues (Snowberg et al. 2011). Even in the in-play soccer betting market studied in this paper where real money is at stake and the transaction volume is huge, we find that over- and underreactions last up to 5 minutes after a goal, which is around 5-6% of the total duration that the market is open (around 95 minutes). For other prediction market settings where the market is open for longer durations, with lower transaction volume (e.g., the Iowa political market), or where "play money" is used (e.g., Hollywood Stock Exchange), one may expect that the biases will take even longer to settle. Future research may look into this issue in more details to determine the most appropriate event window under different market settings.

Second, suppose we find that in some prediction markets, e.g., play-money markets with low transaction volume, over- and underreactions take a long time to settle. What can researchers do in that case? One interesting possibility is to formally model market-level biased reactions, using an approach similar to the Bayesian structural model proposed here. Towards that end, researchers can directly incorporate a "bias" term that is exponentially decreasing over time (see Equation [6]) into their event study model, and empirically estimate the "time decay" parameter δ. This would allow researchers to further disentangle the over- and underreaction bias from the actual impact of the event.

Finally, for event studies that are planned in advance, researchers may be able to "de-bias" the reactions of the prediction markets using information extracted from a separate prediction market that measures the "prior probability" of an event. This is closely related to the idea of "contingent market" described in Henderson et al. (2010). For instance, if a researcher plans in advance to conduct an event study on the impact of Tom Cruise (potentially) dropping out of *Cold Mountain*, she may design two separate markets that (respectively) predict (i) the box office revenue of *Cold Mountain*, and (ii) whether or not Tom Cruise will leave the cast. This will allow researchers to assess whether the second event is a surprise to market participants. The researcher can then determine whether over- or underreaction should be expected, and hence correct for the bias statistically. We believe that this is an important issue for future research and would further enhance the validity of prediction market event studies.

Reference

Baldi, Pierre, and Laurent Itti (2010), "Of Bits and Wows: A Bayesian Theory of Surprise with Applications to Attention," *Neural Networks*, 23, 649-666.

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny (1998), "A Model of Investor Sentiment," *Journal of Financial Economics*, 49, 307-343.

Berg, Joyce, George R. Neumann, and Thomas A. Rietz (2009), "Searching for Google's Value: Using Prediction Markets to Forecast Market Capitalization Prior to an Initial Public Offering," *Management Science*, 55(3), 348-361.

Brooks, Raymond M., Ajay Patel, and Tie Su (2003), "How the Equity Market Responds to Unanticipated Events," *Journal of Business*, 76, 109-133.

Carter, C.K., and R. Kohn (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541-553.

Chen, Li, Paul Goes, James Marsen, and Zhongju Zhang (2009), "Design and Use of Preference Markets for Evaluation of Early Stage Technologies," *Journal of Management Information Systems*, 26(3), 45-70.

Croxson, Karen, and J. James Reade (2011), "Information and Efficiency: Goal Arrival in Soccer Betting," *Working Paper*.

Dahan, Ely, Adlar J. Kim, Andrew W. Lo, Tomaso Paggio, and Nicholas Chan (2011), "Securities Trading of Concepts (STOC)," *Journal of Marketing Research*, 48(3), 497-517.

Dahan, Ely, Arina Soukhoroukova, and Martin Spann (2010), “New Product Development 2.0: Preference Markets – How Scalable Securities Markets Identify Winning Product Concepts and Attributes,” *Journal of Product Innovation Management*, 27(7), 937-954.

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam (1998), “Investor Psychology and Security Market Under- and Overreactions,” *Journal of Finance*, 53, 1839-1885.

Dixon, Mark J., and Stuart G. Coles (1997), “Modelling Association Football Scores and Inefficiencies in Football Betting Market,” *Journal of Royal Statistical Association (Series C)*, 46, 265-280.

Dunlosky, John, R. Reed Hunt, and Elizabeth Clark (2000), “Is Perceptual Salience Needed in Explanation of the Isolation Effect?” *Journal of Experimental Psychology*, 26, 649-657.

Durham, Gregory R., Michael G. Hertzel, and J. Spencer Martin (2005), “The Market Impact of Trends and Sequences in Performance: New Evidence,” *Journal of Finance*, 60, 2551-2569.

Durham, Gregory R., and Mukunthan Santhanakrishnan (2008), “The Relative Importance of Strength and Weight in Processing New Information in the College Football Betting Marketing,” *Journal of Prediction Markets*, 2, 13-28.

Easley, David, and Maureen O’Hara (1987), “Price, Trade Size, and Information in Securities Markets,” *Journal of Financial Economics*, 19, 69-90.

Edwards, Ward (1968), “Conservatism in Human Information Processing,” in Kleinmutz B. (Eds), *Formal Representation in Human Judgment*, John Wiley and Sons, New York, 17-52.

Efron, Bradley, and R. J. Tibshirani (1993), *An Introduction to Bootstrap*, Chapman and Hall/CRC.

Elberse, Anita (2007), “The Power of Stars: Do Star Actors Drive the Success of Movies?” *Journal of Marketing*, 71, 102-120.

Frazzini, Andrea (2006), “The Disposition Effect and Underreaction to News,” *Journal of Finance*, 61, 2017-2046.

Gandar, John, Richard Zuber, Thomas O’Brien, and Ben Russo (1988), “Testing Rationality in the Point Spread Betting Market,” *Journal of Finance*, 43, 995-1008.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003), *Bayesian Data Analysis, 2nd Edition*, Chapman & Hall.

Gil, Ricard, and Steven D. Levitt (2007), “Testing the Efficiency of Markets in the 2002 World Cup,” *Journal of Prediction Markets*, 1, 255-270.

Golec, Joseph, and Maurry Tamarkin (1991), “The Degree of Inefficiency in the Football Betting Market,” *Journal of Financial Economics*, 30, 311-323.

Gray, Philip K., and Stephen F. Gray (1997), “Testing Market Efficiency: Evidence from the NFL Sports Betting Market,” *Journal of Finance*, 52, 1725-1737.

Griffin, Dale, and Amos Tversky (1992), “The Weighting of Evidence and the Determinants of Confidnece,” *Cognitive Psychology*, 24, 411-435.

Grinblatt, Mark, and Bing Han (2005), “Prospect Theory, Mental Accounting, and Momentum,” *Journal of Financial Economics*, 78, 311-339.

Henderson, M. Todd, Justin Wolfers, and Eric Zitzewitz (2010), “Predicting Crime,” *Arizona Law Review*, 51(15), 19-31.

Ho, Tech-Hua, and Kay-Yut Chen (2007), “New Product Blockbusters: The Magic and Science of Prediction Markets,” *California Management Review*, 50(1), 144-158.

Hong, Harrison, and Jermey C. Stein (1999), “A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets,” *Journal of Finance*, 54, 2143-2184.

Hosmer, David W., and Stanley Lemeshow (2000), *Applied Logistic Regression*, New York: Wiley.

Itti, Laurent, and Pierre Baldi (2009), “Bayesian Surprise Attracts Human Attention,” *Vision Research*, 49, 1295-1306.

Jegadeesh, Narasimhan, and Sheridan Titman (1993), “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,” *Journal of Finance*, 48, 65-91.

Johannes, Michael, and Nick Polson (2009), “MCMC Methods for Financial Econometrics,” in *Handbook of Financial Econometrics*, Yacine Ait-Shahlia and Lars Peter Hansen (Eds), 2, 1-72, North Holland.

Klibanoff, Peter, Owen Lamont, and Thierry A. Wizman (1998), “Investor Reaction to Salient News in Closed-End Country Funds,” *Journal of Finance*, 53, 673-699.

Kothari, S. P., and J. B. Warner (2006), “Econometrics of Event Studies,” in B. Espen Eckbo (Ed.) *Handbook of Corporate Finance: Empirical Corporate Finance Vol. A*, Elsevier, North Holland.

Kumar, Alok (2009), “Who Gambles in the Stock Market,” *Journal of Finance*, 64, 1889-1933.

Maher, M. J. (1982), “Modeling Association Football Scores,” *Statistica Neerlandica*, 36, 109-118.

Ofir, Chezy, and David Mazursky (1997), “Does a Surprising Outcome Reinforce or Reverse Hindsight Bias?” *Organizational Behavior and Human Decision Processes*, 69, 51-57.

Poteshman, Allen M. (2001), “Underreaction, Overreaction, and Increasing Misreaction to Information in the Options Market,” *Journal of Finance*, 56, 851-876.

Ranganath, Charan, and Gregor Rainer (2003), “Neural Mechanisms for Detecting and Remembering Novel Events,” *Nature Reviews Neuroscience*, 4, 193-202.

Slamka, Christian, Arina Soukhoroukova, and Martin Spann (2008), “Event Studies in Real- and Play-money Prediction Markets,” *Journal of Prediction Markets*, 2(2), 53-70.

Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz (2011), “How Prediction Markets can Save Event Studies,” *NBER Working Paper*, 16949.

Spann, Martin, and Bernd Skiera (2003), “Internet-based Virtual Stock Markets for Business Forecasting,” *Management Science*, 49(10), 1310-1326.

Srinivasan, Raji, and Sundar Bharadwaj (2005), “Event Studies in Marketing Strategy Research,” in *Assessing Marketing Strategy Performance*, Christine Moorman and Donald R. Lehmann (Eds.), Marketing Science Institute, Cambridge, MA.

Tetlock, Paul C. (2011), “All the News That’s Fit to Reprint: Do Investors React to Stale Information?” *Review of Financial Studies*, 24, 1481-1512.

Thaler, Richard H., and William T. Ziemba (1988), “Anomalties: Parimutuel Betting Markets: Racetracks and Lotteries,” *Journal of Economic Perspectives*, 2, 167-174.

Venables, W.N., and B. D. Ripley (2002), *Modern Applied Statistics with S*, Springer.

Wasserman, Larry (2010), *All of Statistics: A Concise Course in Statistical Inference*, Springer.

Wolfers, Justin, and Eric Zitzewitz (2004), “Prediction Markets,” *Journal of Economic Perspectives*, 18, 107-126.

Appendix

I. 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 match time (t), the conditional probability that the scoring team will win can be computed numerically. We denote this conditional probability as $f(\lambda_{i1}, \lambda_{i2}, z_{it}, t)$. Hence, we compute the efficient drift by:

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

For concreteness, we provide a numerical example. Let λ_{i1}=1.0, λ_{i2}=0.7, and that team 1 (the scoring team) is leading by 1-0 at t = 1800 (30 minutes into the match); we assume that a match, including injury time, lasts 95 minutes. For the rest of the match, the number of goals scored by team 1 and team 2 follow Poisson distributions with rates $\frac{5700-1800}{5700}\lambda_{i1} = 0.68$ and $\frac{5700-1800}{5700}\lambda_{i2} = 0.48$, 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 ($\lambda_{i1}, \lambda_{i2}$). We calibrate the ($\lambda_{i1}, \lambda_{i2}$) parameters using the pre-match implied probabilities. That is, the values of ($\lambda_{i1}, \lambda_{i2}$) 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 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 Equation [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 σ^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 ($\lambda_{i1}, \lambda_{i2}$) easily, which is not the case for other models.

II. Prior specification and MCMC sampling

We specify a uniform distribution on $q_{it_i^r}$, diffuse N(0, 100^2) distributions for the parameters ($\alpha, \beta, \delta, \gamma$), and weakly informative $Inv-\chi^2(0.001,1)$ distribution for the variance parameters $\sigma^2, \tau^2, \omega^2$ (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 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 $\text{logit}(q_{it_i^r})$. The scale of the Guassian 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 $\text{logit}(q_{it})$ other than the first period, the Gaussian form of Equations [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 $\text{logit}(q_{it})$ for each t using a modified Kalman filter, then sample all $\text{logit}(q_{it})$ jointly. This allows us to sample from the full condition distribution of $\text{logit}(q_{it})$ 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% (Gelman et al. 2003).

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 2,000 iterations, with the first 1,000 draws discarded as burn-in. Standard diagnostics confirm that the Markov chain has converged with respect to key parameters. Thus, the last 1,000 draws are used to summarize the posterior distribution of model parameters.

Paper	Major findings	Explanations
Barberis et al. (1998)	Underreaction to earning news and overreaction to a series of earning news; price momentum occurs because of underreaction.	Conservatism and representativeness
Hong and Stein (1999)	Underreaction to information due to slow diffusion and overreaction due to trend chasing; price momentum occurs because of underreaction.	Bounded rationality
Daniel et al. (1998)	Underreaction to public information and overreaction to private information; price momentum occurs because of continuing overreaction, followed by long-run correction.	Overconfidence and biased self-attribution
Klibanoff et al. (1998)	Underreaction (of closed-end fund prices) to changes in Net Asset Value; underreaction is reduced if there is news appearing on the front page of *The New York Times.*	Salience
Brooks et al. (2003)	Overreaction to surprising news such as CEO's sudden death and plane crashes.	Empirically focused; no theoretical explanation given
Grinblatt and Han (2005); Frazzini (2006)	Underreaction to information and earnings news; equilibrium stock prices are lower than fundamental values.	Disposition effect
Poteshman (2001)	Option prices show daily underreaction and multiple-day overreaction to shifts in volatility.	Conservatism and representativeness
Tetlock (2011)	Overreaction to stale information.	Failure to distinguish between old and new information

Table 1. Related literature on over- and underreactions in financial markets.

	Mean	Median	S.D.	Min	Max
Matched volume (£)	1,481,627	917,054	1,627,070	9,654	13,400,407
Matched volume ([+2, +6 minutes] after goal)	152,523	93,779	175,621	103	1,708,552
First goal (minutes)	35.84	25.97	30.47	0.18	113.07
Scoring team is favorite	0.670	1.000	0.470	0.000	1.000
Scoring team wins	0.712	1.000	0.453	0.000	1.000
Implied probability (+2 minutes after goal)	0.686	0.720	0.187	0.017	0.994
Implied probability (+3 minutes after goal)	0.691	0.726	0.190	0.002	0.995
Implied probability (+6 minutes after goal)	0.694	0.730	0.192	0.034	0.997

Table 2. Summary statistics of our dataset.

	"Expected" goals				"Surprise" goals			
c	N	Expected # of wins	Actual # of wins	p-value	N	Expected # of wins	Actual # of wins	p-value
0.0	1299	1014.3	1067	0.000*	638	313.8	320	0.312
0.1	1090	868.5	918	0.000*	477	214.0	212	0.451
0.2	860	704.5	735	0.003*	318	123.4	124	0.505
0.3	664	558.3	578	0.016*	224	75.9	78	0.392
0.4	484	416.6	428	0.077^	145	41.7	38	0.288
0.5	336	295.8	306	0.040*	95	23.8	21	0.302
0.6	191	173.4	183	0.006*	56	11.9	9	0.219

Table 3. Nonparametric analysis of over- and underreactions (*: p < .05; ^: p < .10).

Time	α	p-value ($\alpha = 0$)	β	p-value ($\beta = 0$)	p-value ($\alpha = \beta = 0$)	HL-statistics
Pre-match	N/A	N/A	N/A	N/A	N/A	23.20
2	0.157*	0.005	-0.385*	0.029	0.001*	32.35*
3	0.109^	0.051	-0.291^	0.095	0.019*	27.46^
4	0.086	0.122	-0.294^	0.092	0.042*	26.45^
5	0.073	0.193	-0.285	0.104	0.073^	16.64
6	0.060	0.285	-0.279	0.112	0.112	23.30
7	0.054	0.337	-0.229	0.192	0.208	21.97
8	0.055	0.334	-0.236	0.180	0.196	22.81

Table 4. Results of logistic regressions along with Hosmer-Lemeshow test (*: p < .05; ^: p < .10).

Time	N	Volume (£)	Net earnings (£)	Net earnings (%)	p-value	Average volume at [t, t+1] (£)
2	1937	1339.45	37.43	2.79%*	0.018	42,789
3	1980	1373.32	25.47	1.85%^	0.073	33,795
6	1975	1377.13	11.28	0.82%	0.260	24,356

Table 5. Profit and loss based on our betting strategy (*: p < .05; ^: p < .10).

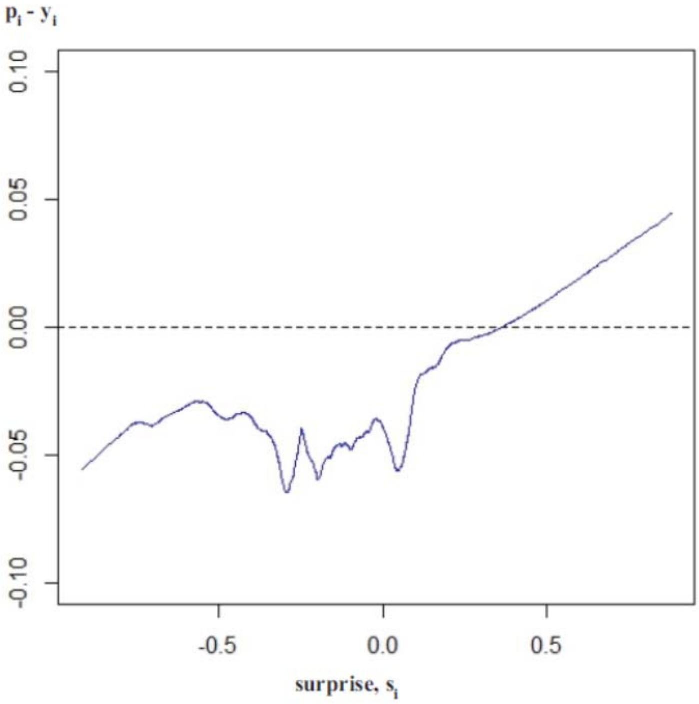

Figure 1. A smoothing spline between $(p_i - y_i)$ and s_i.

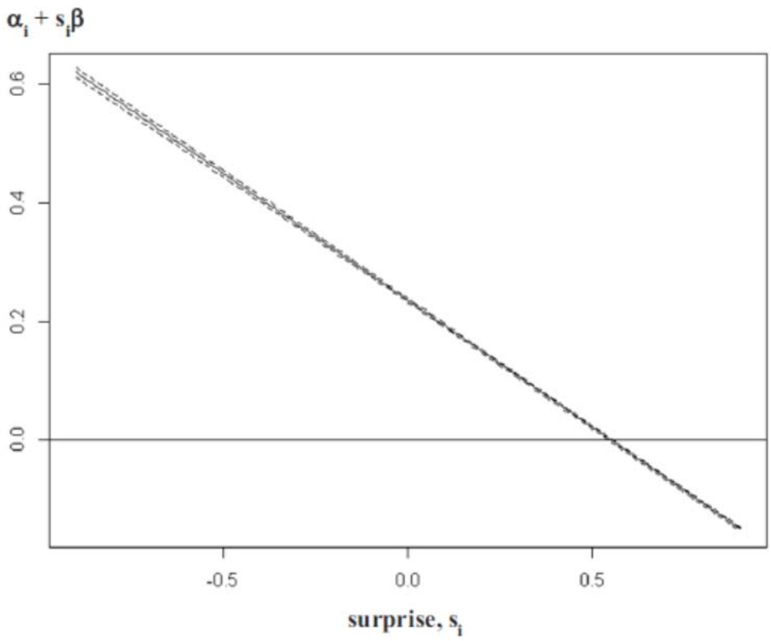

Figure 2. The bias ($\alpha + s_i\beta$) as a function of surprise.