The Importance of Betting Early*

Alessandro Innocenti
University of Siena, LabSi & Befinlab

Tommaso Nannicini
Bocconi University, IGIER & IZA

Roberto Ricciuti
University of Verona, LabSi & CESifo

This version, June 2014

Abstract

We evaluate the impact of decision timing on the outcome of the decision, when both the timing and the relevant decision are chosen under uncertainty. Betting markets provide the testing ground, as we exploit an original dataset containing more than one million online bets on games of the Italian Major Soccer League. We find that individuals perform systematically better when they place their bets farther away from the game day. The better performance of early bettors holds controlling for (time-invariant) unobservable ability, learning during the season, timing of the odds, unobservable game heterogeneity, and favorite team bias. We attribute this result to the increase of noisy information on game day, which hampers the capacity of late, non-professional bettors to use very simple prediction methods, such as team rankings or last game results. We also find that more successful bettors tend to bet in advance, focus on a smaller set of events, and prefer events associated with smaller betting odds.

JEL codes: D81, D83.
Keywords: decision timing, information overload, betting, sports forecasting.

*We thank seminar participants at IGIER-Bocconi, Rutgers University, University of East Anglia, and University of Siena for their insightful comments. Paolo Donatelli, Giacomo Magistretti, Nicola Pierri, and Vincenzo Scrutinio provided outstanding research assistance. We also thank *Microgame.it* and in particular Fabrizio D'Aloia for providing us with the data. Errors are ours and follow a random walk. Corresponding author: Tommaso Nannicini, Bocconi University, Department of Economics, Via Rontgen 1, 20136 Milan (Italy); phone: +39-02-58365241; e-mail: tommaso.nannicini@unibocconi.it.

1 Introduction

Decision timing is a key ingredient of decision making in many settings. Whenever the effect of a choice depends on the future state of the world—e.g., betting, financial markets, firm's strategy—agents face the additional choice of whether taking their decision close to or far from the future event. On the one hand, waiting for a last-minute decision may allow them to improve their information set. On the other hand, if they cannot efficiently process all inputs accruing in proximity to the event, information overload may be detrimental.

We study this tradeoff in the context of sports betting for two reasons. First, as we exploit a large dataset on repeated online bets, we can estimate the effect of the distance from the event on the probability of success by accommodating for the endogeneity of distance choice and without losing statistical accuracy. In particular, we control for unobservable heterogeneity at both the individual and event level, as well as for a number of time-varying confounding channels. Second, as we focus on a population of non-professional bettors, we isolate behavioral regularities that may extend beyond our context. The first element validates the internal validity and the second the external validity of our econometric strategy.

To test our hypothesis that decision timing matters, we analyze the winning probability of bets placed in two different seasons of the Italian Major Soccer League. The dataset contains more than one million online bets. The 7,093 individuals in our dataset are non-experts, who bet small amounts of money on multiple events to increase their potential payouts and only win if all the events happen. Betting on soccer relies on the availability of objective information, such as team rankings and win-loss records, which represent reasonably good predictors of game outcomes. For these reasons, we believe that the distance from game day is a significant factor among those determining how these non-professional bettors process and make use of the available information. The tradeoff highlighted above is clearly at work. Betting too early might force individuals to dismiss relevant information, such as players' injuries that happen close to the game. On the other hand, betting late faces individuals with a large amount of information, which increases with the public relevance of the event, comes from multiple sources, and may not be easy to handle.

In our empirical strategy, we control for individual fixed effects, therefore accommodating for (time-invariant) unobservable ability. Indeed, when we refer to "early" vs "late" bettor, we refer to the same individual placing different bets at a different time distance from the relevant event. As in some specifications we control for individual-times-team fixed effects, we also accommodate for the fact that individuals might systematically bet earlier or later on specific teams (e.g., their favorite one). To control for learning as individuals place more and more bets, we use flexible control functions. Moreover, to control for potential time-varying sources of omitted variable bias, we include the betting odds—which capture the strategic interaction between bettors and the other side of the market—and other attributes of the bet (e.g., financial amount, event's characteristics). Finally, in the most demanding specification, we include event fixed effects to accommodate for unobservable game heterogeneity.

According to our empirical results, for the same bettor, the probability of making a correct forecast is higher when the bet is made on the days that precede the event; in such cases, the chance of winning increases by 1.3 percentage points (that is, by about 3% with respect to the average) as opposed to bets on game day. The effect is larger when big teams or multiple bets are involved (about 5% in both cases). The relationship between betting early and winning is monotonic, as the probability of a correct forecast is larger the higher the number of days from the event, up to the maximum of 5 days. This evidence supports the hypothesis that information overload may occur; as the event becomes closer, individuals receive more information than they are able to handle, therefore increasing the probability of mistakes.

The estimated individual fixed effects show that successful (non-professional) bettors also tend to place their bets in advance. Furthermore, they are more selective, as they place a smaller number of bets in the same week, and tend to focus on events associated with lower betting odds, which are arguably easier to forecast.

The paper is organized as follows. Section 2 reviews the related literature. Section 3 describes our econometric strategy and the dataset. The empirical results are discussed in Section 4. Section 5 concludes.

2 Related literature

Since the 1970s, sports forecasting has been the object of extensive research motivated by two main reasons: (i) to ascertain if betting markets are informationally efficient and enable learning processes, and (ii) to check if experts make more accurate predictions than non-experts. Both strands of the literature aim at analyzing the conditions under which the availability of comprehensive information and professional advice is fully discounted by market prices (that is, betting odds) and rules out observable biases that could allow speculators to make higher-than-average returns.

A large body of empirical evidence supports the view that bettors' behavior does not conform to the rational choice model and is affected by a number of cognitive biases (Diecidue et al. 2004). First, bettors show a tendency to under bet favorites and over bet long-shots (Golec and Tamarkin 1995). Second, they exhibit decision biases such as confirmation, gambler's fallacy, and overconfidence related to inaccurate information processing (Blavatskyy 2009). Third, bettors adopt a series of heuristics whose suitability is context-dependent (Conlisk 1993). Finally, they are not effective enough in discounting the effect of noisy and redundant information and in reducing the impact of information overload (Bleichrodt and Schmidt 2002; Johnson et al. 2006).

A major strand of research concerns horse-race betting, which is a naturally occurring asset market in which the transmission of information from informed to uninformed traders is typically not smooth. Betting markets are efficient if they aggregate less-than-perfect information owned by all the participants and disseminates it to all the bettors. Figlewski (1979) investigates odds and forecasts of a number of bookmakers and experts concluding that racetrack betting markets discount quite well the available information, although bettors exhibit different degrees of accuracy depending on whether they are on-track or off-track bettors. Snyder (1978), Hausch et al. (1981), Asch et al. (1984), and Ziegelmeyer et al. (2004) provide evidence on the tendency to under bet favorites and to over bet long-shots.

Baseball, basketball, football, and soccer are sports in which the sources of insider information are less relevant than in racetrack. Pope and Peel (1989) analyze the fixed odds offered by bookmakers and the forecasts made by professional tipsters on UK soccer league games. They argue that betting markets are efficient in preventing bettors to gain abnormal returns on the basis of public information, but odds do not fully reflect all the available information. This finding is confirmed by Forrest and Simmons (2000), who consider newspaper tipsters offering professional advice on English and Scottish soccer games. They conclude that tipsters show a clear inadequacy in discounting the information publicly available on the newspapers. Moreover, their performance in predicting games is less successful than following the very simple strategy of betting on home wins.

The fact that the condition of being experts is not necessarily associated with a high degree of forecasting accuracy is extensively discussed by Camerer and Johnson (1991) for various domains (medical, financial, academic). Their conclusion is that experts' superiority in processing information is not strictly related to performance superiority, which is crucially affected by the matching of experts' cognitive abilities with "environmental demands" (Camerer and Johnson 1991, p. 213). An interpretation of this finding can be traced back to Oskamp (1965), who argues that the extent of collected information cannot be directly related to predictive accuracy. While predictive ability reaches a ceiling once a limited amount of information has been collected, confidence in the ability to make accurate decisions continues to grow proportionally (Davis et al. 1994). This induces overconfidence in decision-makers, who become more confident independently of the quality of collected information (Angner 2006). Bayesian-rational individuals tend to be relatively overconfident, overweight information of which they overestimate the precision, and underweight it in the opposite case (Van den Steen 2011). Further exposure to sources of information is consequently distorted by confirmation bias: once decision-makers devise a strong hypothesis, they will tend to misinterpret or even misread new information unfavorable to this hypothesis (Kahneman and Tversky 1973).

Goldstein and Gigerenzer (2009), Benartzi and Thaler (2001), Martignon and Hoffrage (2002), Rieskamp and Otto (2006), and Gigerenzer and Goldstein (2011) argue that decision making can be better explained by models of heuristics rather than by the standard rational decision model. Anderson et al. (2005) use recognition heuristics to account for non-experts' performance in soccer betting. According to Newell and Shanks (2004), recognition heuristics are assumed to demand little time, information, and cognitive effort, and exploit the relationship between a criterion value (e.g., success in home win) and its predictors (e.g., team rank position). Heuristics perform quite well in environments affected by noisy and redundant information such as sports forecasting. Noisy information is defined as an information structure in which not only can one signal indicates several states, but also several signals can occur in the same state (Crawford and Sobel 1982). In Dieckmann and Rieskamp (2007), redundant information is defined as information composed by pieces highly correlated with each other and supporting the same prediction (positive redundancy), or that contradict each other and suggest incompatible predictions (negative redundancy).

Glazer et al. (1992) claim that decision-making in noisy environments is only locally rational. If bettors are provided with a very rich source of information without activating a costly search process, confidence increases in relation to prior beliefs (Oskamp 1965). For example, Bettman et al. (1993) provide support for the notion that people also select strategies adaptively in response to information redundancy. They show that participants choosing between gambles search only for a subset of the available information when they encounter a redundant environment with positively correlated attributes. Negatively correlated attributes, in contrast, give rise to search patterns consistent with compensatory strategies that integrate more information. This cognitive bias is known as the illusion of knowledge, according to which beyond a certain threshold more information on the event increases self-confidence more than accuracy (Barber and Odean 2002).

This condition of "information overload" characterizes media information on Italian soccer, which provides the ground for our empirical analysis. The amount of information to be processed is greatly increased by the variety of communication systems on TV, the internet, and newspapers. Furthermore, much of the information is not original and watchers continuously process information received from other sources but differently presented.

Our dataset, which is described in the next section, includes small bets, generally evenly distributed across individuals. Therefore, it can be safely assumed that the individuals contained in our dataset are non-expert bettors.

3 Empirical strategy and data

Based on the literature surveyed in the previous section and on the available data, we test the following behavioral hypothesis.

H1 (information overload): As soon as the event approaches, the amount of noisy information available to bettors increases, therefore reducing their winning ability.

At the same time, we control for the following confounding hypothesis.

H2 (learning): Bettors improve their performance over time, as they get more acquainted with the environment and the relative strength of the teams.

We use a unique (large) dataset of online bets from a provider specialized in this field. The company is located in Southern Italy, but bets are made from all over the country.[1] Users have to register and then bet online through credit card payments. We were provided with bets on all games of 20 game weeks of the Italian Soccer Major League (*Serie A*), namely, the last 10 weeks of the 2004/05 season and the first 10 weeks of the 2005/06 season. Our dataset includes 1,205,597 single bets made by 7,093 registered users. Single bets may also be part of multiple bets including more than one event and may concern several events (e.g., which team wins, draw, goals scored, goals scored in the first half, and so forth). Multiple bets increase potential profits and are won only if all predicted events happen. In our analysis, we focus on the simplest events 1, X, 2, 12, 1X, and X2 (where 1 stands for home win, X for draw, and 2 for away win). These types of event account for 85% of all bets.[2]

The occurrence that bettor j correctly forecasts event i at game week t (W_{ijt}) can be modeled as follows:

$$W_{ijt} = \gamma_j + g(D_{ijt}) + f(t) + X'_{ijt}\beta + Z'_{it}\alpha + \epsilon_{ijt} \tag{1}$$

where γ_j are individual fixed effects (capturing all time-invariant characteristics of bettor j, including his/her intrinsic level of sophistication and ability); X_{ijt} is a vector of time-varying attributes linked to bettor j (such as the amount of money bet at game week t, or the number of other events linked to event i in a multiple bet); Z_{it} is a vector of time-varying attributes of event i (such as whether the home team or the favorite team won the game); $g(.)$ is a function of the distance from the day individual j places the bet to the day event i occurs (D_{ijt}); $f(.)$ is a function of game week t; and ϵ_{ijt} is an idiosyncratic error clustered at the event level.[3]

To test H1 (information overload), we consider three specifications of $g(.)$: linear function of D_{ijt} (which we call "betting distance"); dummy equal to one if the bet is placed before the game day and zero otherwise ("betting early"); non-parametric specification including a set of dummies for each value of D_{ijt} (which varies from zero for bets on game day to a maximum of 5 days). To control for H2 (learning), we introduce three specifications of $f(.)$: linear trend; quadratic trend; a full set of game week dummies. The inclusion of individual

[1] See the company website `www.microgame.it`.

[2] Using all bets does not change the results (available upon request).

[3] Estimation is by linear probability model, but the statistical significance of the results and the size of the marginal effects are almost unchanged with Probit or Logit models (available upon request).

fixed effects accommodates for all time-invariant bettors' characteristics potentially correlated with both the outcome and the treatment. The inclusion of betting odds in the vector Z_{it} controls for the decision of the other side of the market, that is, the betting company, which might strategically adjust the timing of the odds as the event approaches. The inclusion of other event's characteristics identified as relevant by the previous literature—such as the victory of the home team—controls for the fact that bettors might bet earlier on events easier to forecast.

Specifically, among the covariates related to event i, we consider the dummy "main teams," equal to one if the bet concerns at least one of the four leading teams during our sample period; the dummy "strong team wins," equal to one if the stronger team (measured by the relative ranking position in the league) wins; and the dummy "home team wins," equal to one if the home team wins. Among the attributes of each bettor j's decision, we consider the amount spent by the user in each game week ("amount by user"); the number of the other single bets associated with i ("other events"); and the official evaluation that the betting company gives to each event when the bet i is placed by individual j (betting "odds").[4]

To exploit the richness of our (large) dataset, we also augment the baseline specification in equation (1) in two ways:

$$W_{ijt} = \gamma_j + \gamma_j \times HT_i + g(D_{ijt}) + f(t) + X'_{ijt}\beta + Z'_{it}\alpha + \epsilon_{ijt} \quad (2)$$

$$W_{ijt} = \gamma_j + \rho_i + g(D_{ijt}) + f(t) + X'_{ijt}\beta + Z'_{it}\alpha + \epsilon_{ijt} \quad (3)$$

where HT_i is a set of dummies capturing the home team, and ρ_i are event fixed effects. In equation (2), the individual-times-team fixed effects are meant to accommodate for the fact that individuals might adopt different timing when betting on different teams (e.g., their favorite one).[5] In equation (3), the inclusion of event fixed effects controls for the fact that bettors might bet earlier on events easier to forecast according to unobservable features of the event itself. This specification, however, comes at the price of a reduced identifying variation, as it only exploits users who place multiple bets on the same event.

[4]To capture any systematic difference between the two seasons in our dataset, we also include a dummy for the 2005/06 season. This variable also controls for the so-called *Calciopoli* affair, the scandal that emerged in 2006 accusing some major teams of rigging games by selecting favorable referees in the season 2005/06, even though it is hard to imagine how this fraud (discovered only years later) might affect the ability of non-professional bettors to forecast the game outcomes (see Boeri and Severgnini 2011).

[5]We also estimated specifications with interactions between the individual fixed effects and a set of dummies capturing the away team, obtaining almost identical results (available upon request).

[Tables 1 and 2 here]

Table 1 reports the descriptive statistics of our variables. In our data, 45% of single bets are successful. This does not mean, however, that bettors have such a high winning rate, because single bets may be part of multiple bets (on average slightly more than 5 bets are made in each play, with considerable variability), and some of them may be wrong. Indeed, the winning rate in multiple bets is quite low: 5% on average. Most bettors place their bet on the same day of the event, while early bettors (i.e., those who play in the previous days) are about 32%. The average amount spent per bettor in a game week is 211 Euros, again with a large standard deviation. Almost 40% of bets are made on the main four teams.

Table 2 provides information on the above variables and on bettors' socio-economic characteristics by betting distance. We also test whether means are different between bets placed on game day and bets placed before. Thanks to the large sample size, a lot of differences are statistically significant, although most of the time economically small. Early bets tend to be placed on stronger teams, and to be associated with a larger number of multiple bets.

4 Empirical results and discussion

Tables 3, 4, and 5 report our baseline specifications as in equation (1). In the first three columns, we do not control for individual fixed effects, whereas this is done in the last three columns. The latter represent our preferred specifications, but it is instructive to compare results with and without fixed effects. As discussed above, to control for possible learning we use three specifications: linear trend in game week (columns 1 and 4); quadratic trend (columns 2 and 5); and a full set of game week dummies (columns 3 and 6). The difference between the three tables concerns how we model betting distance: linearly in Table 3; with the dummy "betting early" in Table 4; and with a full set of dummies for each value of the betting distance, which is measured in days, in Table 5.

Table 3 shows very similar results across all specifications. The coefficient of betting distance is significantly positive and very stable: the farther away from the event date the bet is, the higher the probability of winning. On average and for the same bettor, betting one day earlier increases the chance of winning by about 0.8 percentage points, that is, by about 1.8% with respect to the average probability of a correct forecast. Betting at the beginning of the week that leads to the event, as opposed to betting on the game day, increases the chance of winning by about 9%. This provides evidence of potential information overload.

Consistently with the previous literature, we find very strong effects for both dummies "home wins" and "strong wins" (equal to 40.8% and 60.9%, respectively, with respect to the average outcome). The ability of winning is positively and significantly affected by the monetary amount that each player bets, meaning that there is higher effort as long as more money is involved, with a large effect with respect to the average outcome (37.4% for an increase of the amount equal to its standard deviation). Betting for the main teams gives a higher probability of winning too. Columns 2 and 5 include the variable game week squared. We do not report its value since it is extremely small; therefore the linear specification is fairly good. As long as the season goes on, however, bettors worsen their performance, as highlighted by the significantly negative coefficients for the game week trend in both the linear and quadratic specifications. Finally, as we would also expect, higher odds are associated with a lower probability of winning (-46.0% for an increase of the odds equal to the standard deviation).

[Tables 3, 4, and 5 here]

In Table 4 the regressor of interest is the dummy "betting early," equal to one if the bet is placed on one of the 5 days preceding game day. This variable is significantly positive, meaning that the probability of making a correct forecast is higher when the bet is made in advance. On average and for the same bettor, the chance of winning increases by 1.3 percentage points (that is, by 2.9% with respect to the average). All the other variables confirm their behavior from a qualitative and quantitative point of view. Table 5 includes a full set of dummies for each value of betting distance. The effect of the distance from the event on the probability of winning is monotonic, as it increases to its maximum when individuals bet 5 days in advance. Wald tests on the equality of coefficients confirm this increasing effect as we move away from game day. Again, all of the other variables confirm their behavior.

Tables 6, 7, and 8 report the specifications augmented with additional fixed effects as in equations (2)—first three columns—and (3)—last three columns. The inclusion of individual-times-team fixed effects is meant to account for the fact that bettors might adopt different timing strategies with respect to different teams, such as their favorite one. The inclusion of event fixed effects is meant to capture unobservable game heterogeneity. We repeat these specifications for all measures of decision timing discussed above, that is, betting distance (Table 6), the dummy "betting early" (Table 7), and the full set of dummies for the days of distance (Table 8). The empirical results are almost unchanged. All point estimates remain statistically different from zero, even though their size is attenuated by the inclusion of event fixed effects. Here, however, we are skating on quite thin ice in terms of identifying variation, as we only rely on bettors who placed multiple bets on the same game.

[Tables 6, 7, and 8 here]

We also address heterogeneity issues, that is, we evaluate whether the effect of betting distance is stronger in specific subsamples. This is meant to further evaluate our information-overload interpretation of the (positive) effect of betting early. Specifically, in Table 9, we distinguish between bets on one of the main four teams of the *Serie A* and on all other teams. In Table 10, we discriminate between bets done on many events (that is, above the median of events associated between each other in multiple bets) or a few events (below the median). In the last row of each table, we report the p-value of the Wald-test on the equality of the estimated coefficients of betting distance for each pair of subsamples.

[Tables 9 and 10 here]

In Table 9, betting distance is always significantly positive, but the size of its coefficient is about three times larger when only the main teams are involved in the bet. This is consistent with our interpretation of the positive impact of betting early, because information overload on the event date is expected to be even more relevant for major teams. Compared with the previous estimations, another relevant variable changes its behavior: game week is usually positive in the linear specification when the main teams are included, and negative otherwise. Therefore, we observe some positive learning when the main teams—which are usually under the spotlights of newspapers—are involved. In Table 10, interestingly, the effect of betting early is quantitatively larger for bets linked to other bets in a multiple play. Again, in these circumstances, information overload is likely to exacerbate fallacies in decision making and to reduce the probability of winning.

[Figures 1, 2, and 3 here]

Finally, the estimated individual fixed effects allow us to shed some light on additional behavioral patterns in our data. Figure 1 shows that more successful bettors (that is, those with a larger fixed effect) also tend to bet in advance, from 3 to 5 days before the event takes place. This regularity, of course, does not affect the estimates discussed above, as they accommodate for unobservable heterogeneity, but it is an interesting finding per se. More skilled bettors seem to anticipate information overload and place their bets in advance. They are also more selective, as they place a smaller number of bets (Figure 2) and focus on bets associated with smaller betting odds (Figure 3), which are arguably easier to forecast.[6]

5 Conclusion

We find that decision timing matters. From the analysis of more than 1,250,000 online bets, we obtain an economically small but statistically very significant and stable difference in the winning probability of early versus late bettors. The estimated effect controls for bettors' unobservable heterogeneity, learning, betting odds, and unobservable characteristics of the event. Therefore, when we refer to "late" versus "early" bettors we are comparing the same individual making bets at different distances from each event. The poorer forecasting performance of late bettors is attributed to an inefficient processing of information, which is also consistent with the heterogeneity results that we are able to disclose.

The late bettors' decision process may be affected by various cues that, unknown to earlier bettors, have scarce relevance for predicting the outcomes. The excess of noisy information (especially harsh if the same individual decides to bet on the main teams or on multiple events) reduces the possibility of using very simple prediction methods, such as team rankings or home team winning. The use of these criteria and cues greatly improves the possibility of

[6]There is no clear pattern of association between ability and age or other observable bettors' characteristics (available upon request).

placing a winning bet. Some skilled bettors partly anticipate the issue, as individuals with larger fixed effects tend to bet from 3 to 5 days in advance.

We acknowledge two main limitations of our results. First, they are based on small stakes and we cannot rule out that when stakes are higher information processing could become more efficient, therefore bringing about positive learning and lower confusion from several sources of information. Second, we cannot rule out the fulfillment of other emotional objectives rather than standard profit maximization.

References

[1] Anderson, P., Edman, J., and Ekman, M. (2005). Predicting the World Cup 2002 in soccer: Performance and confidence of experts and non-experts. *International Journal of Forecasting*, 21, 565-576.

[2] Angner, E. (2006). Economists as Experts: Overconfidence in theory and practice. *Journal of Economic Methodology*, 13(1), 1-24.

[3] Asch, P., Malkiel B. G., and Quandt, R. E. (1984). Market efficiency in racetrack betting. *Journal of Business*, 57, 165-175.

[4] Barber, B. M., and Odean, T. (2002). Online investors: Do the slow die first? *Review of Financial Studies*, 15, 455-487.

[5] Benartzi, S., and Thaler, R. (2001). Naive diversification strategies in retirement saving plans. *American Economic Review*, 91, 79–98.

[6] Bettman, J. R., Johnson, E. J., Luce, M. F., and Payne, J. W. (1993). Correlation, conflict, and choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 931-951.

[7] Blavatskyy, P. R. (2009). Betting on own knowledge: Experimental test of overconfidence. *Journal of Risk and Uncertainty*, 38(1), 39-49.

[8] Bleichrodt, H., and Schmidt, U. (2002). A Context-Dependent Model of the Gambling Effect. *Management Science*, 48(6), 802-812.

[9] Boeri, T., and Severgnini, B. (2011), Match rigging and the career concerns of referees. *Labour Economics*, 18, 349–359.

[10] Camerer, C. F., and Johnson, E. J. (1991). The process–performance paradox in expert judgment: How can experts know so much and predict so badly? In: K. A. Ericsson, and J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits*. New York: Cambridge Press, 195–217.

[11] Conlisk, J. (1993). The utility of gambling. *Journal of Risk and Uncertainty*, 6(3), 255-275.

[12] Crawford, V. P., and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50, 1431–1451.

[13] Davis, F. D., Lohse, G. L., and Kottemann, J. E. (1994). Harmful effects of seemingly helpful information on forecasts of stock earnings. *Journal of Economic Psychology*, 15, 253–267.

[14] Diecidue, E., Schmidt, U., and Wakker, P.P. (2004). The Utility of Gambling Reconsidered. *Journal of Risk and Uncertainty*, 29(3), 241-259.

[15] Dieckmann, A. and Rieskamp, J. (2007). The influence of information redundancy on probabilistic inferences. *Memory & Cognition*, 35, 1801–1813.

[16] Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of Political Economy*, 87, 75–88.

[17] Forrest, D., and Simmons, R. (2000). Forecasting sport: the behaviour and performance of football tipsters. *International Journal of Forecasting*, 16, 317–331.

[18] Gigerenzer, G., and Goldstein, D.G. (2011). The recognition heuristic: A decade of research. *Judgment and Decision Making*, 6(1), 100-121.

[19] Glazer, R., Steckel, J. H., and Winer, R. S. (1992). Locally Rational Decision Making: The Distracting Effect of Information on Managerial Performance, *Management Science*, 38(2), 212-226.

[20] Goldstein, D. G. and Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, 25, 760-772.

[21] Golec, J., and Tamarkin, M. (1995). Do bettors prefer long shots because they are risk-lovers, or are they just overconfident? *Journal of Risk and Uncertainty*, 11(1), 51-64.

[22] Hausch, D. B., Ziemba, W. T., and Rubinstein, M. (1981). Efficiency of the market for racetrack betting. *Management Science*, 27, 1435–1452.

[23] Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80, 237–251.

[24] Johnson, J. E. V., Jones, O., and Tang, L. (2006). Exploring Decision Makers' Use of Price Information in a Speculative Market. *Management Science*, 52(6), 897-908.

[25] Martignon, L., and Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52, 29-71.

[26] Newell, B. R., and Shanks, D. R. (2004). On the role of recognition in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30, 923-935.

[27] Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology*, 2, 261–265.

[28] Pope, P. F., and Peel, D. A. (1989). Information, prices and efficiency in a fixed-odds betting market. *Economica*, 56, 323–341.

[29] Rieskamp, J., and Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135, 207–236.

[30] Snyder, W. W. (1978). Testing the efficient markets model. *Journal of Finance*, 33, 1109-1118.

[31] Ziegelmeyer, A., Broihanne, M.H., and Koessler F. (2004). Sequential Parimutuel Betting in the Laboratory. *Journal of Risk and Uncertainty*, 28(2), 165-186.

[32] Van den Steen, E. (2011). Overconfidence by Bayesian-Rational Agents. *Management Science*, 57(5), 884-896.

Tables and figures

Table 1 – Descriptive statistics

	Mean	Median	S.d.	Min	Max
Correct forecast	0.451	0.000	0.498	0.000	1.000
Betting distance	0.443	0.000	0.778	0.000	5.000
Betting early	0.317	0.000	0.465	0.000	1.000
Other events	5.151	5.000	2.061	0.000	13.000
Amount by user	0.211	0.153	0.242	0.003	6.018
Main teams	0.399	0.000	0.490	0.000	1.000
Home team wins	0.394	0.000	0.489	0.000	1.000
Strong team wins	0.366	0.000	0.482	0.000	1.000
Odds	2.216	1.900	1.033	1.050	18.000
2004/05 season	0.507	1.000	0.500	0.000	1.000

Notes. The number of observations is 1,205,575 for all variables. *Betting distance* is measured in days. *Betting early* is a dummy equal to one if the bet is placed before the game day, and zero otherwise. *Other events* captures the number of events associated with the single bet in a multiple bet. *Amount by user* is the amount bet by the user in the game week and is measured in thousand of Euros. All the other variables except *Odds* are dummies.

Table 2 – Conditional means by betting distance

	Conditional means by betting distance						*Conditional means and t-test*		
	0	1	2	3	4	5	Distance>0	Distance=0	P-value
Home team wins	0.392	0.396	0.408	0.410	0.368	0.355	0.398	0.392	0.000
Strong team wins	0.362	0.370	0.390	0.405	0.412	0.349	0.376	0.362	0.000
Other events	5.01	5.29	5.70	6.20	6.75	6.77	5.46	5.01	0.000
Amount by user	0.208	0.212	0.240	0.234	0.209	0.179	0.218	0.208	0.000
Main teams	0.384	0.421	0.445	0.472	0.461	0.379	0.429	0.384	0.000
Odds	2.24	2.19	2.11	2.03	1.92	1.92	2.16	2.24	0.000
2004/05 season	0.525	0.475	0.459	0.412	0.430	0.528	0.468	0.525	0.000
Female	0.135	0.132	0.143	0.134	0.118	0.084	0.133	0.135	0.003
Age	34.7	34.8	35.1	36	34.8	34.1	34.9	34.7	0.000
Lawyer	0.044	0.041	0.038	0.039	0.036	0.026	0.040	0.044	0.000
Bank employee	0.120	0.123	0.124	0.124	0.124	0.0921	0.123	0.120	0.000
Engineer & programmer	0.047	0.046	0.049	0.044	0.045	0.048	0.046	0.047	0.365
Architect	0.035	0.033	0.032	0.024	0.036	0.034	0.032	0.035	0.000
Clerk	0.032	0.033	0.043	0.033	0.029	0.036	0.035	0.032	0.000
Unemployed	0.006	0.006	0.004	0.005	0.009	0.007	0.005	0.006	0.000
Other profession	0.062	0.060	0.064	0.076	0.080	0.070	0.062	0.062	0.579
Missing profession	0.653	0.660	0.648	0.656	0.642	0.688	0.657	0.653	0.000
Observations	822,966	277,989	72,331	20,614	8,623	3,052	382,609	822,966	
Share	0.683	0.231	0.060	0.017	0.007	0.003	0.317	0.683	

Notes. The number of observations is 1,205,575 for all variables. *P-value* refers to a t-test of the equality of means in the subsamples of bets placed in the game day (*Distance=0*) vs. before the game day (*Distance>0*).

Table 3 – The impact of betting distance: baseline specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Betting distance	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Home team wins	0.184***	0.184***	0.182***	0.184***	0.184***	0.182***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Strong team wins	0.283***	0.283***	0.298***	0.284***	0.284***	0.297***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Game week	-0.003***	-0.007***		-0.004***	-0.007***	
	(0.000)	(0.000)		(0.000)	(0.001)	
Other events	0.008***	0.008***	0.008***	0.007***	0.007***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount by user	0.004	0.006	0.002	0.008	0.010*	0.006
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.004)
Main teams	0.048***	0.047***	0.046***	0.049***	0.048***	0.047***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Odds	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***	-0.095***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2004/05 season	-0.011***	-0.029***		-0.014***	-0.031***	
	(0.004)	(0.004)		(0.005)	(0.005)	
Game week squared		0.000***			0.000***	
		(0.000)			(0.000)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575
No. of individuals	7,093	7,093	7,093	7,093	7,093	7,093

Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (1). Standard errors clustered at the individual level are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 4 – The impact of betting early: baseline specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Betting early	0.013***	0.013***	0.011***	0.013***	0.013***	0.011***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Home team wins	0.184***	0.184***	0.182***	0.184***	0.184***	0.182***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Strong team wins	0.283***	0.283***	0.298***	0.284***	0.284***	0.297***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Game week	-0.003***	-0.007***		-0.004***	-0.007***	
	(0.000)	(0.000)		(0.000)	(0.001)	
Other events	0.008***	0.008***	0.008***	0.007***	0.008***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount by user	0.004	0.006	0.002	0.008	0.010*	0.006
	(0.004)	(0.003)	(0.003)	(0.006)	(0.006)	(0.004)
Main teams	0.048***	0.047***	0.046***	0.049***	0.048***	0.047***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Odds	-0.096***	-0.097***	-0.096***	-0.096***	-0.096***	-0.095***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2004/05 season	-0.011***	-0.029***		-0.014***	-0.031***	
	(0.004)	(0.004)		(0.005)	(0.005)	
Game week squared		0.000***			0.000***	
		(0.000)			(0.000)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575
No. of individuals	7,093	7,093	7,093	7,093	7,093	7,093

Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (1). Standard errors clustered at the individual level are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 5 – The impact of betting distance: non-parametric specifications

	(1)	(2)	(3)	(4)	(5)	(6)
1 day before	0.013***	0.013***	0.009***	0.013***	0.013***	0.010***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2 days before	0.010***	0.010***	0.011***	0.010***	0.010***	0.012***
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.003)
3 days before	0.021***	0.021***	0.021***	0.018***	0.018***	0.018***
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
4 days before	0.029***	0.030***	0.029***	0.024***	0.024***	0.024***
	(0.009)	(0.009)	(0.008)	(0.009)	(0.008)	(0.008)
5 days before	0.067***	0.067***	0.072***	0.062***	0.062***	0.069***
	(0.015)	(0.015)	(0.013)	(0.015)	(0.014)	(0.013)
Home team wins	0.184***	0.184***	0.182***	0.184***	0.184***	0.182***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Strong team wins	0.283***	0.283***	0.298***	0.284***	0.284***	0.297***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Game week	-0.003***	-0.007***		-0.004***	-0.007***	
	(0.000)	(0.000)		(0.000)	(0.001)	
Other events	0.008***	0.008***	0.008***	0.007***	0.008***	0.007***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount by user	0.004	0.006*	0.002	0.008	0.011*	0.006
	(0.004)	(0.003)	(0.003)	(0.006)	(0.006)	(0.004)
Main teams	0.048***	0.047***	0.046***	0.049***	0.048***	0.047***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Odds	-0.096***	-0.096***	-0.096***	-0.096***	-0.096***	-0.095***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2004/05 season	-0.010***	-0.029***		-0.014***	-0.031***	
	(0.004)	(0.004)		(0.005)	(0.005)	
Game week squared		0.000***			0.000***	
		(0.000)			(0.000)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575
No. of individuals	7,093	7,093	7,093	7,093	7,093	7,093
1 day = 2 days	*0.466*	*0.400*	*0.515*	*0.414*	*0.353*	*0.582*
2 days = 3 days	*0.074*	*0.071*	*0.075*	*0.214*	*0.204*	*0.219*
3 days = 4 days	*0.435*	*0.390*	*0.396*	*0.578*	*0.532*	*0.536*
4 days = 5 days	*0.017*	*0.016*	*0.003*	*0.015*	*0.015*	*0.003*

Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (1). Standard errors clustered at the individual level are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***. In italics, p-values for Wald tests on the equality of the coefficients of the betting-distance dummies.

Table 6 – The impact of betting distance: augmented specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Betting distance	0.008***	0.008***	0.007***	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Home wins	0.214***	0.215***	0.211***	-0.143***	-0.093***	0.279***
	(0.003)	(0.003)	(0.003)	(0.028)	(0.025)	(0.026)
Strong wins	0.273***	0.273***	0.292***	0.520***	0.570***	0.412***
	(0.003)	(0.003)	(0.002)	(0.027)	(0.024)	(0.042)
Game week	-0.003***	-0.001		0.002	-0.033***	
	(0.000)	(0.001)		(0.002)	(0.003)	
Game week squared		-0.000***			0.000***	
		(0.000)			(0.000)	
Other events	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount user	0.017***	0.016**	0.009	0.004	0.004	0.004
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Main teams	0.112***	0.113***	0.122***	-0.236***	-0.227***	-0.137***
	(0.003)	(0.003)	(0.003)	(0.018)	(0.016)	(0.042)
Odds	-0.097***	-0.097***	-0.096***	-0.097***	-0.097***	-0.097***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2004/2005 season	-0.019***	-0.008		-0.147**	-0.630***	
	(0.006)	(0.007)		(0.058)	(0.035)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	-	-	-	Yes	Yes	Yes
Indiv.×team fixed effects	Yes	Yes	Yes	No	No	No
Event fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575

Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (2)—columns 1-3—or (3)—columns 4-6. Standard errors clustered at the individual×team level in columns 1-3 and at the individual level in columns 4-6 are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 7 – The impact of betting early: augmented specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Betting early	0.013***	0.013***	0.011***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Home wins	0.214***	0.215***	0.211***	-0.145***	-0.093***	0.279***
	(0.003)	(0.003)	(0.003)	(0.028)	(0.025)	(0.026)
Strong wins	0.273***	0.273***	0.292***	0.518***	0.570***	0.406***
	(0.003)	(0.003)	(0.002)	(0.027)	(0.024)	(0.042)
Game week	-0.003***	-0.001		0.002	-0.031***	
	(0.000)	(0.001)		(0.002)	(0.003)	
Game week squared		-0.000***			0.000***	
		(0.000)			(0.000)	
Other events	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount user	0.017***	0.016**	0.009	0.004	0.004	0.004
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Main teams	0.112***	0.113***	0.122***	-0.231***	-0.225***	-0.143***
	(0.003)	(0.003)	(0.003)	(0.018)	(0.016)	(0.042)
Odds	-0.097***	-0.097***	-0.096***	-0.097***	-0.097***	-0.097***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2004/2005 season	-0.019***	-0.008		-0.135**	-0.610***	
	(0.006)	(0.007)		(0.058)	(0.035)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	-	-	-	Yes	Yes	Yes
Indiv.×team fixed effects	Yes	Yes	Yes	No	No	No
Event fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575

Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (2)—columns 1-3—or (3)—columns 4-6. Standard errors clustered at the individual×team level in columns 1-3 and at the individual level in columns 4-6 are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 8 – The impact of betting distance: augmented non-parametric specification

	(1)	(2)	(3)	(4)	(5)	(6)
1 day before	0.013***	0.013***	0.010***	0.005***	0.005***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2 days before	0.011***	0.011***	0.012***	0.007***	0.007***	0.007***
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
3 days before	0.015***	0.015***	0.015***	0.012***	0.012***	0.012***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
4 days before	0.025***	0.025***	0.029***	0.018***	0.018***	0.018***
	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)	(0.006)
5 days before	0.053***	0.053***	0.066***	0.039***	0.039***	0.039***
	(0.011)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)
Home wins	0.214***	0.215***	0.211***	-0.140***	-0.091***	0.279***
	(0.003)	(0.003)	(0.003)	(0.028)	(0.025)	(0.026)
Strong wins	0.273***	0.273***	0.292***	0.522***	0.570***	0.412***
	(0.003)	(0.003)	(0.002)	(0.027)	(0.024)	(0.042)
Game week	-0.003***	-0.001		0.002	-0.035***	
	(0.000)	(0.001)		(0.002)	(0.003)	
Game week squared		-0.000***			0.000***	
		(0.000)			(0.000)	
Other events	0.008***	0.008***	0.008***	0.008***	0.008***	0.008***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount user	0.018***	0.016**	0.009	0.004	0.004	0.004
	(0.006)	(0.006)	(0.006)	(0.004)	(0.004)	(0.004)
Main teams	0.112***	0.113***	0.122***	-0.238***	-0.227***	-0.136***
	(0.003)	(0.003)	(0.003)	(0.018)	(0.017)	(0.042)
Odds	-0.097***	-0.096***	-0.096***	-0.097***	-0.097***	-0.097***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2004/2005 season	-0.019***	-0.008		-0.157***	-0.642***	
	(0.006)	(0.007)		(0.057)	(0.036)	
Game week dummies	No	No	Yes	No	No	Yes
Individual fixed effects	-	-	-	Yes	Yes	Yes
Indiv.×home team fixed effects	Yes	Yes	Yes	No	No	No
Event fixed effects	No	No	No	Yes	Yes	Yes
No. of observations	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575	1,205,575

Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (2)—columns 1-3—or (3)—columns 4-6. Standard errors clustered at the individual×team level in columns 1-3 and at the individual level in columns 4-6 are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***.

Table 9 – Heterogeneity: main teams vs. others

	Main teams		Main teams		Main teams		Main teams		Main teams		Main teams	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Betting distance	0.012***	0.005***	0.012***	0.005***	0.010***	0.006***	0.013***	0.003***	0.013***	0.003**	0.012***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Home team wins	0.116***	0.244***	0.116***	0.244***	0.125***	0.240***	0.115***	0.244***	0.115***	0.244***	0.124***	0.240***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)
Strong team wins	0.478***	0.100***	0.480***	0.098***	0.485***	0.091***	0.479***	0.099***	0.481***	0.098***	0.486***	0.091***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Game week	0.004***	-0.007***	-0.002***	-0.012***			0.004***	-0.007***	-0.003***	-0.012***		
	(0.000)	(0.000)	(0.001)	(0.001)			(0.000)	(0.000)	(0.001)	(0.001)		
Other events	0.002***	0.013***	0.002***	0.013***	0.002***	0.012***	0.002***	0.012***	0.002***	0.012***	0.002***	0.011***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Amount by user	-0.003	0.006	-0.000	0.008*	0.000	0.003	-0.000	0.009	0.005	0.013**	0.003	0.006
	(0.005)	(0.005)	(0.005)	(0.005)	(0.004)	(0.003)	(0.007)	(0.007)	(0.008)	(0.006)	(0.005)	(0.005)
Odds	-0.103***	-0.088***	-0.103***	-0.088***	-0.103***	-0.088***	-0.102***	-0.086***	-0.102***	-0.086***	-0.102***	-0.086***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2004/05 season	0.196***	-0.150***	0.160***	-0.173***			0.204***	-0.151***	0.162***	-0.171***		
	(0.004)	(0.005)	(0.006)	(0.005)			(0.006)	(0.006)	(0.007)	(0.007)		
Game week squared			0.000***	0.000***					0.000***	0.000***		
			(0.000)	(0.000)					(0.000)	(0.000)		
Game week dummies	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Individual fixed effects	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	480,534	725,041	480,534	725,041	480,534	725,041	480,534	725,041	480,534	725,041	480,534	725,041
No. of individuals	6,814	6,905	6,814	6,905	6,814	6,905	6,814	6,905	6,814	6,905	6,814	6,905
Wald test p-value	*0.000*		*0.000*		*0.005*		*0.000*		*0.000*		*0.000*	

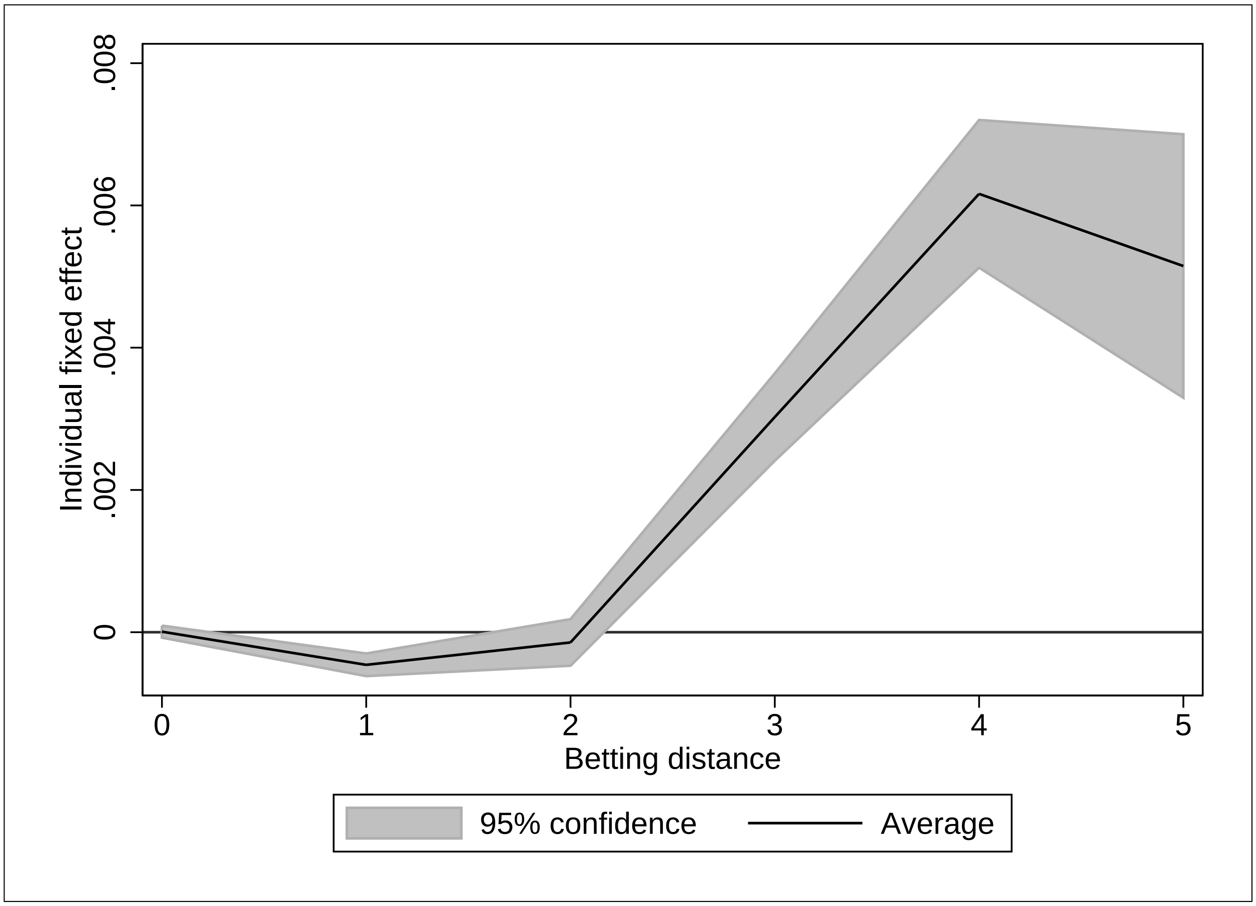
Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (1) in separate subsamples (events involving main teams vs. events involving other teams). "Main teams" refer to the four leading teams during our sample period (namely, *F.C. Internazionale*, *Juventus F.C.*, *A.C. Milan*, and *A.S. Roma*). Standard errors clustered at the individual level are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***. The *Wald test p-value* captures the significance of the difference of the coefficients of betting distance in the two subsamples.

Table 10 – Heterogeneity: bets linked to many events vs. others

	Many events		Many events		Many events		Many events		Many events		Many events	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Betting distance	0.009***	0.006***	0.008***	0.006***	0.009***	0.005***	0.008***	0.006***	0.008***	0.006***	0.009***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Home team wins	0.274***	0.134***	0.274***	0.134***	0.270***	0.133***	0.274***	0.134***	0.274***	0.134***	0.271***	0.133***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Strong team wins	0.371***	0.233***	0.372***	0.233***	0.385***	0.246***	0.373***	0.233***	0.374***	0.233***	0.385***	0.246***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)	(0.003)
Game week	-0.001***	-0.004***	-0.008***	-0.008***			-0.001***	-0.004***	-0.008***	-0.008***		
	(0.000)	(0.000)	(0.001)	(0.000)			(0.000)	(0.000)	(0.001)	(0.001)		
Main teams	0.012***	0.064***	0.010***	0.063***	0.011***	0.063***	0.014***	0.065***	0.013***	0.064***	0.013***	0.063***
	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)
Amount by user	0.006	0.002	0.009*	0.004	0.000	0.002	0.014*	0.003	0.018**	0.006	0.006	0.004
	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.003)	(0.008)	(0.007)	(0.008)	(0.007)	(0.006)	(0.005)
Other events	0.011***	0.003***	0.011***	0.003***	0.009***	0.004***	0.010***	0.003***	0.010***	0.003***	0.008***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Odds	-0.120***	-0.092***	-0.120***	-0.092***	-0.119***	-0.091***	-0.117***	-0.091***	-0.117***	-0.091***	-0.116***	-0.091***
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)
2004/05 season	0.047***	-0.032***	0.017***	-0.051***			0.044***	-0.033***	0.016*	-0.051***		
	(0.006)	(0.004)	(0.006)	(0.005)			(0.008)	(0.006)	(0.008)	(0.006)		
Game week squared			0.000***	0.000***					0.000***	0.000***		
			(0.000)	(0.000)					(0.000)	(0.000)		
Game week dummies	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Individual fixed effects	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
No. of observations	691,104	514,471	691,104	514,471	691,104	514,471	691,104	514,471	691,104	514,471	691,104	514,471
No. of individuals	6,119	6,243	6,119	6,243	6,119	6,243	6,119	6,243	6,119	6,243	6,119	6,243
Wald test p-value	*0.160*		*0.196*		*0.025*		*0.202*		*0.213*		*0.013*	

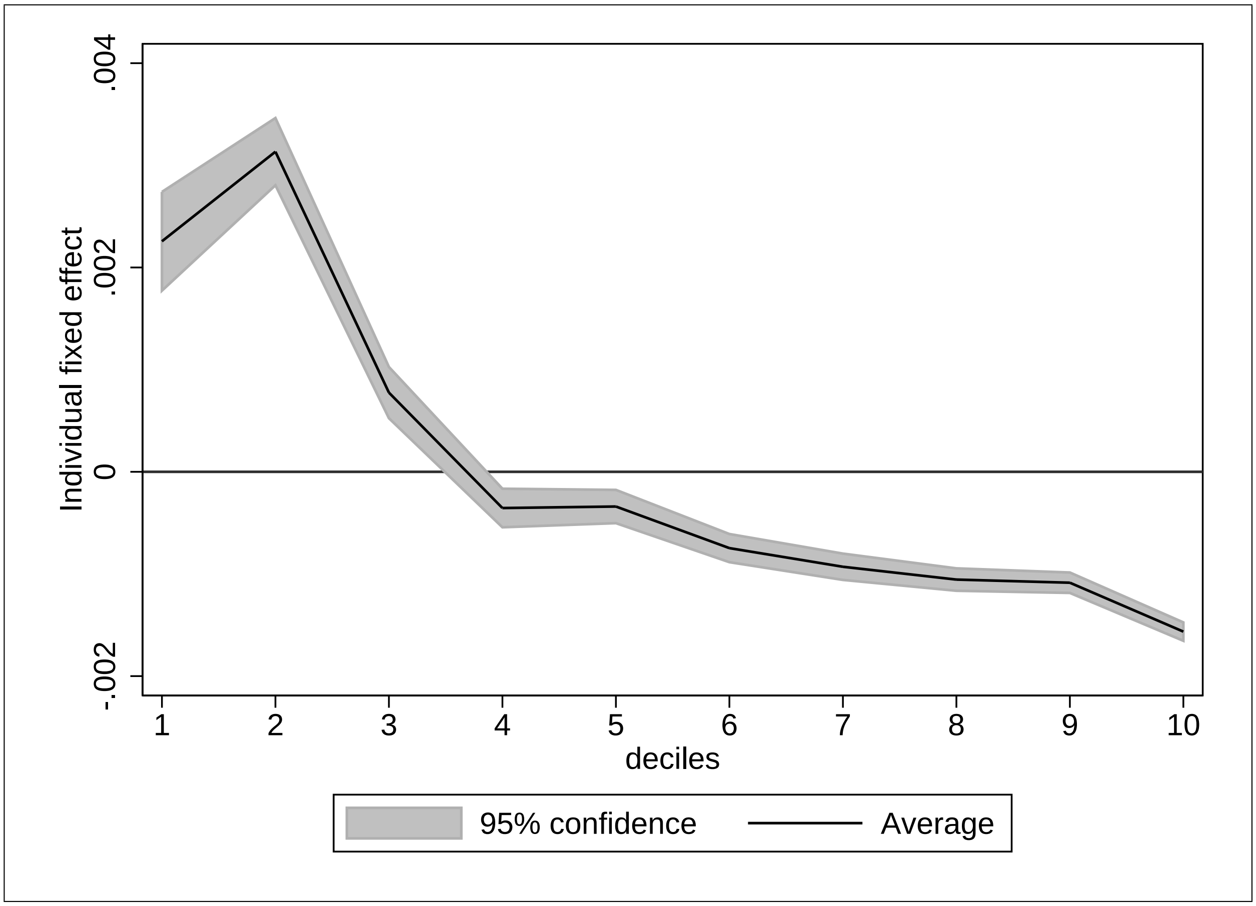
Notes. Dependent variable: probability of correctly forecasting the single event (included in either a single or multiple bet). Estimation method: linear probability model as in equation (1) in separate subsamples (bets linked to a higher-than-median number of multiple bets vs. others). Standard errors clustered at the individual level are reported in parentheses. Significance at the 10% level is represented by *, at the 5% level by **, and at the 1% level by ***. The *Wald test p-value* captures the significance of the difference of the coefficients of betting distance in the two subsamples.

Figure 1 – Individual fixed effects by betting distance



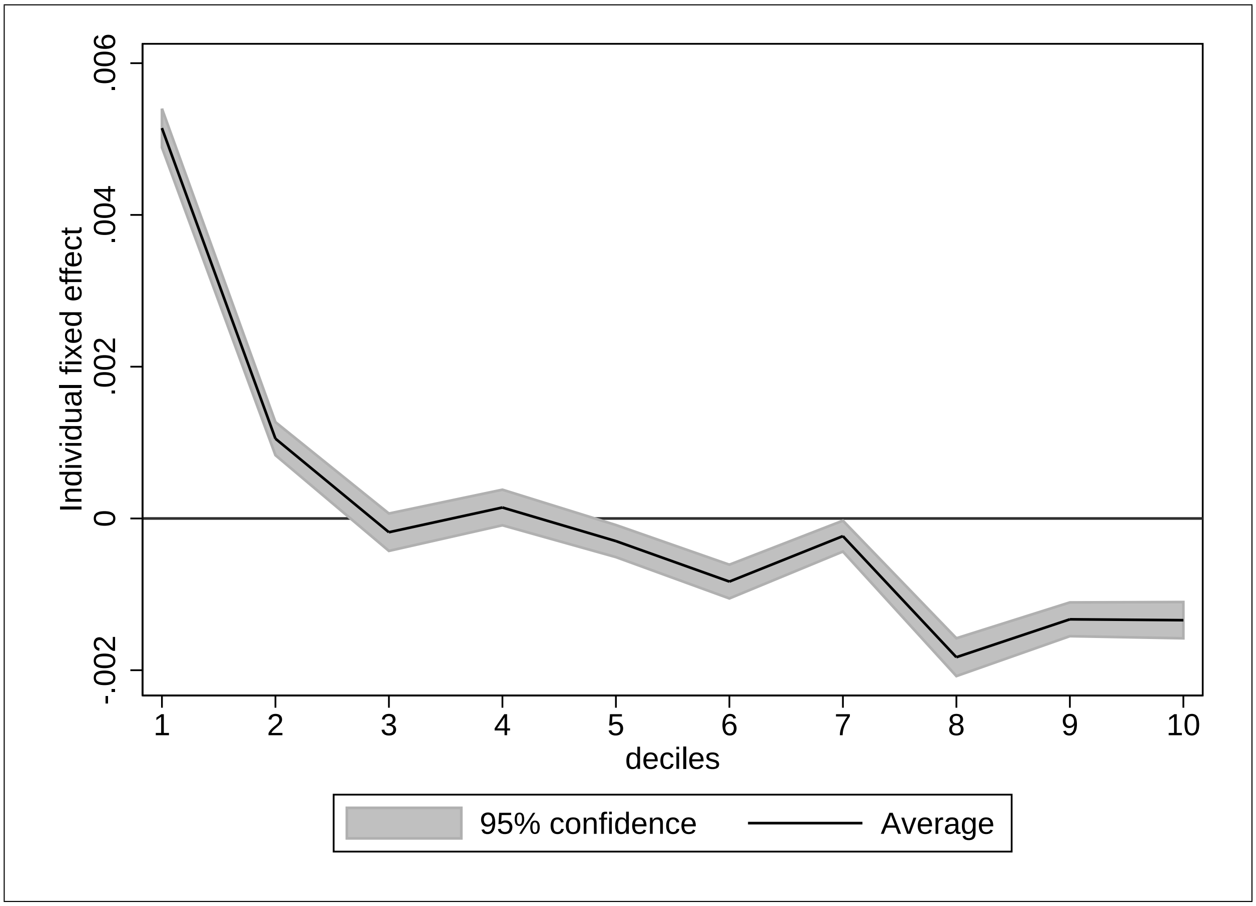
Notes. Distribution of individual fixed effects from the estimation of equation (1). Share of observations by betting distance: 0 [0.683]; 1 [0.231]; 2 [0.060]; 3 [0.017]; 4 [0.007]; 5 [0.003].

Figure 2 – Individual fixed effects by deciles of number of bets per game week



Notes. Distribution of individual fixed effects from the estimation of equation (1). The intervals for the deciles are: 1 [1-14]; 2 [14-29]; 3 [29-53]; 4 [53-81]; 5 [81-110]; 6 [110-139]; 7 [139-169]; 8 [169-208]; 9 [208-258]; 10 [258-1,137].

Figure 3 – Individual fixed effects by deciles of odds



Notes. Distribution of individual fixed effects from the estimation of equation (1). The intervals for the deciles are: 1 [1.05-1.25]; 2 [1.25-1.43]; 3 [1.43-1.6]; 4 [1.6-1.75]; 5 [1.75-1.9]; 6 [1.9-2.2]; 7 [2.2-2.65]; 8 [2.65-2.85]; 9 [2.85-3.25]; 10 [3.25-18].