Does Discretized Information Induce Misperception? Evidence from a Sports Betting Market

July 28, 2026

Abstract

Does the format of information influence how markets process it? In theory, transformations of existing information, such as discretization into ratings, should not alter beliefs. Yet under cognitive constraints, discretized information can distort perceptions even when it conveys no new information. Using a betting market, we develop a novel test that cleanly distinguishes these competing views by isolating the cognitive channel from coordination, regulatory, and outcome-driven explanations, a challenge unaddressed by prior research. We find that discretization induces misperception, particularly under low economic stakes or high reliance on heuristics. Our findings have implications for how regulators and intermediaries disclose information.

"Financial markets and sports gambling are merging, and some big financial firms now go around acting like there is no difference between them."
— Matt Levine, "Sports Bets at the Stock Exchange", *Bloomberg Opinion*, October 2025

1 Introduction

Information disclosure constitutes a central component of regulation and intermediation in financial markets (Goldstein and Yang, 2017; Goldstein and Leitner, 2020). While existing rational models have largely emphasized what information to disclose (e.g., Verrecchia, 1983; Dye, 1985; Diamond and Verrecchia, 1991; Morris and Shin, 2002; Bond and Goldstein, 2015), the question of how to disclose it can be equally important if investors face cognitive constraints (Tversky and Kahneman, 1974; Sims, 2003). A prominent example is the communication of information in discrete formats, such as letter-based credit ratings or mutual fund star ratings, to reduce the cognitive burden of processing complex information. However, such discretization may distort perceptions if individuals tend to amplify differences between members of different categories—a cognitive shortcut known as categorical thinking (Mullainathan, 2002; Mullainathan *et al.*, 2008; Huang, 2019; Shue *et al.*, 2024).[1] Prior research shows that economic decisions often exhibit discontinuities in response to changes in discretized information, even when these changes carry no intrinsic informational content.[2] Although such discontinuities are typically interpreted as evidence of investor misperception, this inference requires careful scrutiny.

The empirical challenge is that in conventional financial settings examined in prior research, discontinuities in economic behavior may reflect rational expectations of the real effects that discretized information can generate (hereafter, the *real effects channel*). Consider the case of corporate credit ratings. Two firms with nearly identical fundamentals may be perceived as having very different credit risk if they fall on opposite sides of a rating threshold. Yet such a discontinuity in beliefs is not necessarily evidence of misperception, since even fully rational lenders may respond to a mechanical change in credit rating for at least two reasons. First, credit ratings serve as coordination devices for investor beliefs (Boot *et al.*, 2006), so a lender may curtail funding in anticipation that other lenders will

[1] The cognitive shortcut of categorical thinking often leads individuals to perceive members of a category as more homogeneous than they actually are, while exaggerating differences across categories. For a review of categorical thinking in the psychology literature, see Macrae and Bodenhausen (2000).

[2] On the effects of credit ratings, see, among others, Kliger and Sarig (2000), Adelino *et al.* (2017), and Cornaggia *et al.* (2018). On the effects of mutual fund performance ratings and ESG ratings, see, among others, Guercio and Tkac (2008), Hartzmark and Sussman (2019), Evans and Sun (2020), Huang *et al.* (2020), Ben-David *et al.* (2021), and Reuter and Zitzewitz (2021).

respond similarly. Second, rating changes often carry regulatory consequences that constrain some lenders' ability to extend credit (Ellul *et al.*, 2011; Chen *et al.*, 2014). Thus, even when rating changes are orthogonal to firm fundamentals, lenders may rationally adjust their behavior in expectation of tighter credit conditions and higher borrowing costs for the firm (Manso, 2013; Purnanandam and Wirth, 2023). Discontinuities in fund flows around mutual fund performance ratings or ESG ratings may likewise reflect the real effects channel rather than misperception, because fund flows can affect subsequent fund performance (Berk and Green, 2004) and managers' risk-taking incentives (Chevalier and Ellison, 1997), while the ratings themselves can induce funds to adjust their portfolios (Gantchev *et al.*, 2024).

This paper introduces a novel non-laboratory test that addresses this empirical challenge and cleanly isolates misperceptions arising from information discretization. Our key innovation is to exploit a setting in which the real effects channel is absent, meaning that changes in discretized information do not affect the actual probability of future outcomes. The South Korean horse racing betting market provides such a setting, owing to its distinctive institutional features. Crucially, in this market the absence of real effects can be validated ex post. Using this setting, we show that even when changes in discretized information carry neither informational value nor real effects, bettors nonetheless exhibit systematic discontinuities in their economic decisions. This result underscores the need for regulators and intermediaries to consider carefully how information is presented to investors.

Horse race betting markets provide a particularly suitable empirical setting, as they allow researchers to infer individuals' perceived probabilities of future events from observed betting behavior—an advantage rarely available in other contexts.[3] First, the traded assets in these markets are state-contingent claims with a well-defined termination point (Thaler and Ziemba, 1988), enabling the direct elicitation of perceived probabilities from real-stakes decisions and facilitating comparison between perceived and actual outcome probabilities. By contrast, other empirical designs typically require strong assumptions to infer beliefs or rely on surveys that lack financial consequences. Second, betting decisions based on perceived probabilities do not influence the realization of outcomes—a feature uncommon in most economic settings (Moskowitz, 2021). For example, in credit markets, a creditor's belief about a firm's default risk can affect its access to capital and thus its actual likelihood of default, complicating the identification of misperception. In horse racing, by contrast, this feedback channel is absent, allowing for cleaner identification of belief distortions. Moreover, recent research highlights the prevalence of gambling preferences among investors and documents significant overlap between participants in equity and betting markets (Dorn

[3]For a literature review highlighting the merits of sports as empirical settings for economics research, see Palacios-Huerta (2023).

et al., 2015; Gao and Lin, 2015; Baker *et al.*, 2024). This suggests that findings in betting markets may generalize to investor behavior in broader financial contexts.

Our empirical approach investigates whether bettors rely on ranking information about horses even when precise race times are publicly available and easily accessible. Rankings are simply discretized representations of continuous race times and therefore contain no incremental information beyond what is already conveyed by the race time data. In such settings, standard economic theory predicts that ranking information should not influence the perceptions of rational, unbiased bettors. To implement this approach, we focus on pairs of horses that are adjacently ranked within the same race and exhibit nearly identical race times. We then analyze betting behavior in the subsequent race, testing for discontinuities at ranking thresholds by comparing wagers on the horse ranked just above the threshold with those on its immediately lower-ranked counterpart.

However, discontinuities in wagers around rankings may reflect real effects of rankings on future performance rather than bettor misperception. For example, rankings can influence the level of competition a horse faces in subsequent races. This is indeed the case in many horse racing markets around the world, where a *rating system* is used to ensure competitive balance within each race.[4] Within these systems, horses are typically assigned performance-based ratings that are updated after each race. These ratings then determine both the handicap weights they carry and the caliber of competition they face in subsequent races. As a result, rankings—despite being merely discretized representations of race times—can shape a horse's future competitive environment and thereby influence subsequent performance. These institutional features complicate the interpretation of discontinuities in betting behavior as evidence of misperception, since observed shifts may instead reflect rational expectations of real effects.

A key advantage of the South Korean horse racing market for this analysis is its relatively late adoption of a rating framework. Prior to 2015, the market operated under a *tiering system* in which horses were classified based on the cumulative number of top-five finishes they had achieved, a metric that determined both handicap weights and race assignments. Under this system, two horses that failed to place in the top five in a given race would face identical competitive conditions in their next race, and their owners would receive nearly identical monetary compensation. Exploiting this institutional feature, we focus on pairs of adjacently ranked horses with nearly identical race times, both of which placed sixth or lower, and examine bettors' beliefs about their future performance. By restricting attention to these horses, we ensure that treatment and control groups are comparable in both compensation

[4]The list of countries that employ rating systems is extensive, including the United States, the United Kingdom, Ireland, Hong Kong, and Australia.

and the level of competition faced in their subsequent race. Our identification strategy tests for discontinuities in perceived winning probabilities around ranking thresholds in a setting where (i) institutional features imply no expected differences in actual winning probabilities *ex ante*, and (ii) this absence of differences can be verified *ex post* using realized outcomes. The absence of real effects allows us to interpret bettors' tendency to assign higher winning or placing probabilities to the higher-ranked horse—despite equivalent prior performance—as evidence of misperception driven by reliance on redundant categorical information.

Importantly, our inference relies only on the assumption that bettors place more win bets on horses they believe are more likely to win, without requiring assumptions about their risk preferences. This represents a key advantage, as many existing tests of investor rationality depend on specific assumptions regarding risk preferences. For instance, a prominent empirical regularity in betting markets is the favorite–longshot bias, whereby equilibrium prices indicate that bettors systematically overvalue longshots and undervalue favorites relative to their true winning probabilities. Competing explanations for this bias attribute it either to locally risk-loving preferences or to cognitive errors. Our approach provides a sharper test of rationality, as risk-loving preferences cannot account for discontinuities in beliefs around classification thresholds. Moreover, by focusing on local discontinuities, we avoid imposing parametric assumptions on the mapping from bets to beliefs. This model-free approach mitigates concerns about joint hypothesis testing and reduces the risk of erroneously rejecting the null due to model misspecification.

Using detailed race-level data from South Korea spanning 1993 to 2024, we identify pairs of adjacently ranked horses in the same race with nearly identical race times—specifically, those separated by a margin of no more than a nose (approximately three inches)—and compare their outcomes in their respective subsequent races. We first document that paired horses are balanced on various characteristics that can affect race outcomes, such as jockey ability and handicap weight. We then show that they exhibit nearly identical ex post probabilities of winning or placing. These patterns indicate that, conditional on race time, ranking provides no incremental information about future performance in this setting. By contrast, betting behavior diverges sharply from realized outcomes: despite nearly identical ex post performance, higher-ranked horses attract significantly more wagers than their lower-ranked counterparts. This pattern suggests that bettors rely on ranking information and perceive higher-ranked horses as more likely to win or place, even though such rankings are redundant conditional on race times. The difference in win-bet shares between higher- and lower-ranked horses can reach 0.46 percentage points, equivalent to over one-tenth of the average win-bet share of 4.5 percent. This misperception of winning probabilities has economically meaningful consequences. For instance, consider a horse with a true winning

probability of 4 percent. If the proportion of win bets placed on the horse accurately reflects this probability, the expected gross return on a win bet is $(1-0.2) \times 0.04 \times (\frac{1}{0.04}) = 0.8$ after the 20 percent house take—that is, an expected loss of 20 percent. However, if bettors overestimate the horse's chances and the betting share rises to 4.4 percent, the expected return falls to $(1-0.2) \times 0.04 \times (\frac{1}{0.044}) \approx 0.727$, an expected loss of 27.3 percent. Misperception thus raises the expected loss by 7.3 percentage points, or roughly 36 percent of the baseline 20 percent loss.

We next explore potential mechanisms underlying the observed misperception. One possibility is that bettors rely on ranking information because acquiring precise race time data is costly. While theoretically plausible, this explanation appears unlikely, as both race times and rankings from previous races are publicly available and easily accessible in our setting. Nonetheless, we empirically test this hypothesis by investigating whether the ranking effect has diminished over time, particularly following technological advancements that reduced information acquisition costs. If such costs were the primary driver of the effect, we would expect it to be stronger in earlier periods. Contrary to this expectation, we find that the ranking effect is more pronounced in recent years and either weak or absent in earlier periods, suggesting that information acquisition costs are unlikely to explain the observed misperception.

A second potential mechanism is that bettors may rationally disregard race time information due to the cognitive costs of processing it—a phenomenon known as rational inattention (Sims, 2003).[5] To test this hypothesis, we examine whether the ranking effect diminishes as economic stakes increase. The rationale is that larger wagers provide stronger incentives for bettors to allocate cognitive resources toward processing information. Consistent with this prediction, we find that the ranking effect weakens as the average bet size increases.

Third, we examine whether reliance on heuristics contributes to the observed misperception.[6] While the cognitive costs of processing race time information should, in principle, be uniform across ranking thresholds, heuristic reliance may vary with a horse's relative standing. We find that the ranking effect is substantially more pronounced among higher-ranked horses, such as those in 6th place, while it is weak or altogether absent among lower-ranked horses, such as those ranked 12th, which are generally longshots with minimal chances of winning. This asymmetry is consistent with evidence from the psychology literature showing that individuals struggle to distinguish between small probabilities

[5] Maćkowiak *et al.* (2023) provide a review of the literature on rational inattention.

[6] For a review of heuristic decision-making in the psychology literature, see Gigerenzer and Gaissmaier (2011) and the references therein.

(e.g., those associated with a 12th-ranked horse) and extremely small probabilities (e.g., those associated with a 13th-ranked horse) (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996; Snowberg and Wolfers, 2010). Overall, our findings support the interpretation that bettor misperception is primarily driven by rational inattention and heuristic decision-making.

So far, we have isolated the effects of discretization-driven misperception in a setting where discretized information has no real effects. As a final empirical exercise, we investigate the extent to which our findings generalize to settings where the discretization of information has real effects, as it often does in financial markets. We examine this question through a calibration exercise using horse pairs in a setting where prior rank affects the competitive environment in the subsequent race. We find that this differential competitive environment leads higher-ranked horses to have lower probabilities of winning or placing in their subsequent races. Higher-ranked horses also receive smaller win- or place-bet shares, indicating that bettors account for these real effects. However, perceived probabilities of winning or placing inferred from betting behavior indicate that bettors still overestimate the chances of higher-ranked horses relative to their realized performance. This overestimation could reflect bettors' miscalibration of the real effects, discretization-driven misperception, or other channels. Our calibration implies that the discretization-driven misperception we isolate can explain 69 to 77 percent of this overestimation. These results indicate that this channel remains economically important even in settings where discretization has real consequences.

Finally, we discuss the implications of our findings for other settings. Several features of sports betting markets make them informative about financial markets. Both settings feature high trading volumes, broad information availability, professional participants, active market making, arbitrage, and substantial return volatility (Moskowitz, 2021). They also attract participants with similar behavioral tendencies: investors with a greater propensity to gamble disproportionately hold lottery-like stocks (Barberis and Shleifer, 2003; Kumar, 2009; Dorn and Sengmueller, 2009; Grinblatt and Keloharju, 2009), while professional sports bettors increasingly rely on quantitative trading strategies. Consistent with this connection, Baker *et al.* (2024) show that households substitute sports betting for brokerage investment following legalization. These similarities support the relevance of our findings for financial markets. Nonetheless, important institutional differences caution against direct extrapolation.

This paper contributes to the literature on investor rationality and market efficiency, in which betting markets have increasingly been employed as laboratory settings (Rosett, 1965; Ali, 1977; Snyder, 1978; Moskowitz and Vasudevan, 2022; Moskowitz, 2021). We

extend this literature by providing, to our knowledge, the first clean empirical evidence that information discretization can mislead market participants and distort their decision-making. Although discontinuities in economic behavior at classification thresholds have been well documented—particularly in the contexts of credit ratings and mutual fund ratings[7]—identifying the underlying mechanism has proven challenging. Instead of reflecting misperceptions, such discontinuities may represent rational responses to the real effects that classifications exert on future outcomes. We address this gap by exploiting an empirical setting in which classifications have no consequences for subsequent outcomes, thereby providing clean evidence that the observed discontinuities reflect perceptual distortions induced by information discretization.

Furthermore, we contribute to the literature on betting markets by introducing a novel test of models incorporating cognitive constraints. Prior research has primarily focused on the longshot bias, an empirical regularity that has proven difficult to interpret because it is consistent with both rational models and those invoking cognitive limitations. Preference-based models often reconcile these patterns with investor rationality by extending standard utility theory to accommodate locally risk-loving behavior, whereas behavioral models attribute them to cognitive constraints (e.g., Thaler and Ziemba, 1988). Our approach circumvents this identification challenge by developing a test that does not rely on assumptions regarding risk preferences, thereby providing cleaner evidence of the cognitive limitations faced by market participants.

Finally, the evidence that information discretization induces misperception has important policy implications for how regulators and intermediaries disclose information to investors. Disclosure of information in financial markets lies at the core of regulatory efforts to promote market quality and stability (Greenstone *et al.*, 2006; Goldstein and Yang, 2017). While prior research has largely focused on what information should be disclosed, our findings highlight that the format in which information is disclosed is equally important.[8]

[7]For studies on credit ratings, see Kisgen (2006); Sufi (2007); Goel and Thakor (2015); Adelino *et al.* (2017); Almeida *et al.* (2017); Cornaggia *et al.* (2018); Goldstein and Huang (2020); Gredil *et al.* (2022); Purnanandam and Wirth (2023); for related work on mutual fund ratings, see Guercio and Tkac (2008); Evans and Sun (2020); Huang *et al.* (2020); Ben-David *et al.* (2021); Reuter and Zitzewitz (2021).

[8]For studies on the optimal content of information disclosure, see Verrecchia (1983); Dye (1985); Diamond and Verrecchia (1991); Morris and Shin (2002); Bond and Goldstein (2015); Bouvard *et al.* (2015); Shapiro and Skeie (2015); Faria-e Castro *et al.* (2017); Monnet and Quintin (2017); Goldstein and Leitner (2018); Leitner and Williams (2023); Quigley and Walther (2024).

2 Framework

The primary objective of our empirical analysis is to assess whether horses' past rankings—despite being redundant given the availability of precise race time information—systematically influence bettors' expectations about winning probabilities in subsequent races. To address this question, we begin by developing a framework that infers bettors' perceived winning probabilities for each horse from their betting behavior. We then present an empirical strategy that compares perceived and actual winning probabilities for pairs of horses that differ in ranking but exhibit nearly identical race times. Finally, we highlight the distinct advantages of betting markets over alternative settings for implementing this identification approach.

2.1 Eliciting Bettors' Perceived Probabilities

Following Snowberg and Wolfers (2010), we model bettors as expected utility maximizers with subjective beliefs and utility function $u(\cdot)$. In equilibrium, bettors must be indifferent across all horses in a race when deciding where to place their bets:

$$\hat{p}_A \times u(1 + o_A) = \hat{p}_B \times u(1 + o_B), \tag{1}$$

where $\hat{p}_i$ denotes bettors' perceived probability that horse i wins, and o_i represents the betting odds for horse i (i.e., the market price of the corresponding state-contingent claim). $u(0)$ is normalized to 0.

We require only that the utility function be strictly increasing—that is, $u' > 0$—so that more is always preferred to less. Under this assumption, $o_A > o_B$ implies $\hat{p}_A < \hat{p}_B$: the order of bettors' perceived winning probabilities can be directly inferred from observed betting odds.[9]

Importantly, our inference of perceived probabilities does not require any assumption about bettors' risk aversion, u''. This is a key advantage, as existing empirical tests often yield conflicting conclusions depending on the assumptions imposed on risk preferences. A well-documented regularity in betting markets is the favorite–longshot bias, whereby bettors systematically overvalue longshots and undervalue favorites relative to their objective winning probabilities. Interpreting this bias, however, has proven challenging: it is consistent with both rational models and behavioral explanations, depending on how one interprets bettors' attitudes toward risk.[10]

[9] Recovering the levels of $\hat{p}_i$ requires a stronger assumption, which we impose only for the calibration in Section 6.1.

[10] For reviews of the favorite–longshot bias literature, see Thaler and Ziemba (1988) and Snowberg and

2.2 Identifying the Effects of Ranking Information on Perceived Probabilities

We denote horses' past race time records as X and their race rankings as N. Since X spans all information that can be generated by its discretized counterpart N (i.e., $\sigma(N) \subseteq \sigma(X)$), standard economic models suggest that, conditional on observing X, bettors should not rely on N when forming expectations or making betting decisions. The goal of our empirical analysis is to test whether bettors' perceived winning probability $\hat{p}$, and thus their betting behavior, is nonetheless influenced by N despite its informational redundancy. To this end, we test the following null hypothesis:

$$\mathbf{H_0}: \quad \hat{p}(X|N=n) = \hat{p}(X) \quad \text{for all } n. \tag{2}$$

If this null hypothesis is rejected, it suggests that bettors' perceived probability $\hat{p}$ is influenced by N. However, without knowing the functional form of $\hat{p}$, this test faces challenges due to joint hypothesis testing: a rejection may result from misspecification of the functional form of $\hat{p}$. To address this concern, we adopt a model-free approach. Specifically, we compare pairs of horses with nearly identical race times but different rankings and examine whether bettors assign significantly different winning probabilities to these horses. In other words, we test for discontinuities in bettors' perceived probabilities around ranking thresholds:

$$\mathbf{H_0'}: \quad \hat{p}(X|N=n) = \hat{p}(X+\varepsilon|N=n+1) \quad \text{for all } n \text{ and all sufficiently small } \varepsilon > 0. \tag{3}$$

A key challenge with this approach is that rejecting the null hypothesis in Equation (3) implies rejection of the original hypothesis in Equation (2) *only if* the perceived probability function is continuous in X so that $\hat{p}(X) \approx \hat{p}(X+\varepsilon)$ for all sufficiently small $\varepsilon > 0$. Otherwise, discontinuities in perceived probabilities around ranking thresholds could arise even if bettors' perceptions do not depend on N. In particular, if horses' actual winning probabilities jump at ranking thresholds, bettors who correctly anticipate this real effect may exhibit discontinuities in their perceived probabilities as well.

To address this issue, we restrict our analysis to cases where the actual winning probabilities of horses with nearly identical race times are approximately equal—that is, $p(X) \approx p(X+\varepsilon)$ for small $\varepsilon > 0$, where $p(\cdot)$ denotes a horse's actual winning probability.[11]

Wolfers (2008).

[11] Section 3 describes how we identify such a setting.

Under this condition, any discontinuities in bettors' perceived winning probabilities around ranking thresholds are unlikely to occur unless bettors rely on ranking information. Alternative explanations must then justify why such discontinuities appear in perceived probabilities despite the absence of corresponding differences in actual probabilities.

2.3 Advantages of Betting Markets

One might be skeptical about using betting markets and wonder why testing for discontinuities around ranking thresholds could not be applied in other financial market settings. We highlight two empirical challenges that make such applications difficult elsewhere. First, betting markets trade state-contingent claims with pre-specified payoffs, enabling researchers to elicit bettors' perceived probabilities of certain states from real-stakes decisions with minimal assumptions. Moreover, actual probabilities of these states can be observed and compared to perceived probabilities. In contrast, other financial settings typically require strong assumptions to infer perceived probabilities or rely on survey-based measures, which often lack real economic stakes.

Second, even if measurement challenges are addressed, discontinuities around ranking thresholds can generate actual discontinuities in outcomes. For example, consider an investor with full access to the underlying, latent information that determines credit ratings. This investor seeks to estimate default probabilities for two firms with nearly identical fundamentals but different credit ratings. Importantly, the actual default probability may differ between these firms because lower credit ratings can influence other investors' behavior and constrain the firm's borrowing capacity (Manso, 2013; Cornaggia *et al.*, 2018; Purnanandam and Wirth, 2023). These effects are particularly pronounced since credit ratings serve as coordination devices for investor expectations (Boot *et al.*, 2006) and carry important regulatory implications (Ellul *et al.*, 2011; Chen *et al.*, 2014). As a result, even if the provision of discretized information does not lead to investor misperception, we may still observe discontinuities in perceived default probabilities around rating thresholds. Huang *et al.* (2020) document analogous challenges in the context of mutual fund star ratings. These real effects create identification challenges by introducing an alternative explanation based on rational expectations of these effects, rather than misguided beliefs.

3 Institutional Setting and Data

The previous discussion highlights the importance of identifying cases where horses with nearly identical race times do not differ in their actual winning probabilities. A key advantage

of the South Korean horse racing market as an empirical setting is its late adoption of a rating system, which provides a large sample of cases suitable for this analysis. In this section, we present an overview of the institutional features of South Korean horse racing. Building on this context, we carefully exclude instances where differences in rankings between similarly performing horses could influence their future outcomes by altering their competitive environment—such as through race assignments or handicap weights. Although the institutional setting supports our ex-ante expectation of equivalent performance, we formally validate this assumption empirically in the subsequent analysis.

3.1 Institutional Details

Background The Korea Racing Authority is the sole institution authorized to conduct horse racing in South Korea. This state-owned entity operates three racecourses located in Seoul, Jeju, and Busan, with races held twice weekly. On each race day, typically 16 to 17 races take place at each venue, featuring between 8 and 14 horses per race. Private owners employ trainers and enter their horses into these competitions. The market is substantial: in 2019, total sales (betting turnover) reached 4.95 billion euros, ranking South Korea eighth globally, with 12.7 million betting entrants as reported by the Korea Racing Authority.[12]

Tiering System Many countries with large, developed horse racing markets adopt rating systems to organize races.[13] Under this system, each horse is assigned a rating that reflects its ability, with the goal of making races competitive and equalizing each horse's chances of winning. Horses compete against others with similar ratings, and handicap weights are imposed based on each horse's rating. Handicappers, officials responsible for assigning these weights, update ratings after each race based on performance. Therefore, a horse's performance and ranking today significantly affect its competitive environment in the next race, a mechanism we refer to as the real effects channel. As previously discussed, this link between performance and future competition introduces a significant identification challenge, rendering betting markets in such settings unsuitable for our empirical analysis.

Unlike most countries, the Korea Racing Authority did not adopt a rating system until February 2015. Prior to that, it relied on a tiering system that grouped horses into six tiers based on their *cumulative* number of top-five finishes in past races. Horses competed against others within the same tier, broadly ensuring a comparable level of ability among

[12] See International Federation of Horseracing Authorities.

[13] This list includes the United Kingdom, Ireland, Hong Kong, France, Germany, Italy, Australia, New Zealand, South Africa, the UAE, Brazil, Argentina, Singapore, Macao, Turkey, India, Malaysia, Uruguay, and Zimbabwe (Korea Racing Authority).

competitors. Notably, the primary basis for handicapping was the cumulative number of top-five finishes, meaning that the competitive environment in the subsequent race for two horses that both failed to finish in the top five would be comparable. Equally important, monetary compensation under this system consisted of two components: a nearly fixed participation payment awarded to all entrants and a performance-based prize given only to those finishing in the top five. As a result, horses that finished outside the top five received virtually identical compensation, as their owners were paid solely for participation. These institutional features of the South Korean horse racing market are central to our empirical strategy, as they mitigate concerns about endogenous variation in future competition.

In February 2015, the Korea Racing Authority introduced a rating system. Under this system, horse ratings are updated after each race based on performance, regardless of whether the horse finishes in the top five. Horses continue to be classified into six tiers based on their rating and compete against others within the same tier. However, the primary basis for handicapping is now each horse's rating, which is updated after every race—even for horses that fail to finish in the top five. As a result, past rankings can have real effects on a horse's future rankings by influencing its rating and future competitive environment. Interestingly, however, horses in the sixth tier are exempt from the rating system and continue to be governed by the original tiering rules. This institutional feature implies that, even after February 2015, horses in the sixth tier that do not place in the top five remain suitable for our empirical analysis.

Race Organization At the end of each month, the Korea Racing Authority announces the race plan for the upcoming month. This plan specifies the conditions for participation in each race (e.g., tier and age), the award amounts for top-placed horses, race distance, and the handicapping method. Horse owners can submit applications to enter a race one week before it takes place, provided their horses meet the participation criteria. The Korea Racing Authority finalizes the list of participating horses based on a predetermined order of selection, with lanes randomly assigned. The horse owner is responsible for selecting the jockey.

Betting Odds Bettors place bets on horses before each race begins, with two main types of wagers available. The first is straight wagers, where bets are placed on a single horse to either win or place (i.e., finish 1st to 3rd) in a race. The second type is horizontal exotic wagers, which involve bets on multiple horses in a single race. In our analysis, we focus on straight wagers to elicit bettors' beliefs about each horse.

After betting closes, the pari-mutuel system finalizes the betting odds. Let v_i represent

the amount of bets placed on horse i to win in a race. Under this system, the betting odds o_i for horse i to win is given by:

$$1 + o_i = (1 - \tau) \frac{\sum_{i \in \mathcal{I}} v_i}{v_i} \tag{4}$$

where τ represents the tax and house take rates, and $\mathcal{I}$ is the set of all horses in the race. In South Korea, τ for win and place bets is set at 20 percent. For each dollar bet, a bettor will receive $(1 + o_i)$ dollars if the horse wins the race. The betting odds for a horse to place are determined similarly.

The share of bets on horse i, relative to the total amount of bets, can be derived from the betting odds as follows:

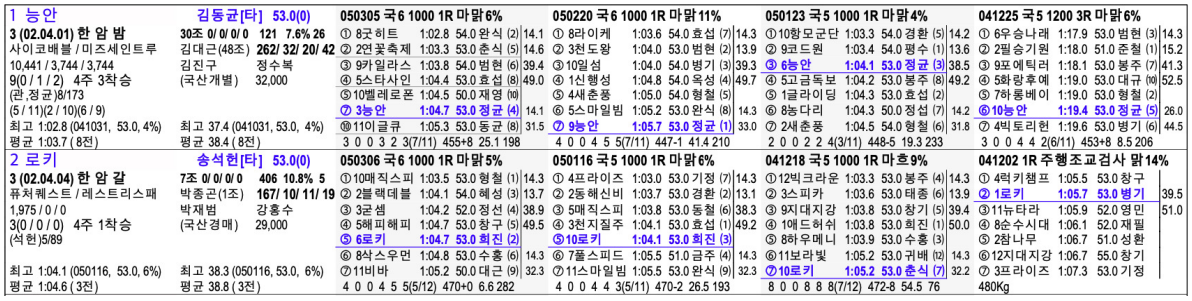
$$\frac{v_i}{\sum_{i \in \mathcal{I}} v_i} = \left(\frac{1}{1 + o_i} \right) / \left(\sum_{i \in \mathcal{I}} \frac{1}{1 + o_i} \right). \tag{5}$$

In our analysis, we focus on these shares of bets, as they are bounded between 0 and 1, whereas betting odds often exhibit extreme values.

The Informational Environment On each race day, the Korea Racing Authority distributes a physical pamphlet titled *Today's Race*, which contains information on the past performance records of horses competing in that day's events. Historically, this pamphlet has served as a low-cost information source for bettors. Until November 2012, it was distributed free of charge to attendees, after which it was sold for 600 KRW (approximately 0.50 USD). Nonetheless, it has remained freely accessible on the Korea Racing Authority's website.

Figure 1 presents an excerpt from the Today's Race pamphlet, distributed for the second race held at Seoul Racecourse on April 2, 2005. The first and second rows provide information on the horses starting from gates 1 and 2, respectively. The first column summarizes each horse's basic details, while the second column contains information about the jockey. Beginning with the third column, the pamphlet displays each horse's past rankings and race times, along with records of other horses they competed against. For example, the blue-highlighted entry in the third column of the first row indicates that the horse starting from gate 3 finished 7th (marked by a circled numeral) with a time of 1:04.7 in its previous race on March 5, 2005. Notably, however, the pamphlet does not report race times for *all* horses in prior races, which limits bettors' ability to accurately assess performance gaps between competitors.

Alternatively, bettors can access complete historical race information through the Korea



1 능안	김동균[타] 53.0(0)	050305 국6 1000 1R 마맑 6%					050220 국6 1000 1R 마맑 11%					050123 국5 1000 1R 마맑 4%					041225 국5 1200 3R 마맑 6%				
3 (02.04.01) 한 암 밤	30조 0/ 0/ 0/ 0 121 7.6% 26	① 8곳히트	1:02.8	54.0 완식 (2)	14.1		① 8라이켄	1:03.6	54.0 효섭 (7)	14.3		①10항모군단	1:03.3	54.0 경환 (5)	14.2		① 6우승나래	1:17.9	53.0 범현 (3)	14.3	
사이코배블 / 미즈세인트루	김대근(48조) 262/ 32/ 20/ 42	② 2연꽃축제	1:03.3	53.0 춘식 (5)	14.6		② 3천도왕	1:04.0	53.0 범현 (2)	13.9		② 9코드원	1:03.4	54.0 평수 (1)	13.6		② 2밀승기원	1:18.0	51.0 준철 (1)	15.2	
10,441 / 3,744 / 3,744	김진구 정수복	③ 9카일라스	1:03.8	54.0 범현 (6)	39.4		③10일섬	1:04.0	54.0 병기 (3)	39.3		③ 6능안	1:04.1	53.0 정균 (3)	38.5		③ 9포에틱러	1:18.1	53.0 봉주 (7)	41.3	
9(0 / 1 / 2) 4주 3착승	(국산개별) 32,000	④ 5스타사인	1:04.4	53.0 효섭 (8)	49.0		④ 1신행성	1:04.8	54.0 옥성 (4)	49.7		④ 5고금독보	1:04.2	53.0 봉주 (8)	49.2		④ 5화랑후예	1:19.0	53.0 대규 (10)	52.5	
(관,정균)8/173		⑤10벨레로폰	1:04.5	50.0 재영 (10)			⑤ 4새춘풍	1:05.0	54.0 형철 (5)			⑤ 1글라이딩	1:04.3	53.0 효섭 (2)			⑤ 7하롱베이	1:19.0	53.0 형철 (2)		
(5 / 11)(2 / 10)(6 / 9)		⑦ 3능안	1:04.7	53.0 정균 (4)	14.1		⑥ 5스마일빔	1:05.2	53.0 완식 (8)	14.3		⑥ 8농다리	1:04.3	50.0 정섭 (7)	14.2		⑥10능안	1:19.4	53.0 정균 (5)	26.0	
최고 1:02.8 (041031, 53.0, 4%)	최고 37.4 (041031, 53.0, 4%)	⑩11이글큐	1:05.3	53.0 동균 (8)	31.5		⑦ 9능안	1:05.7	53.0 정균 (1)	33.0		⑦ 2새춘풍	1:04.5	54.0 형철 (6)	31.8		⑦ 4씩토리헌	1:19.6	53.0 병기 (6)	44.5	
평균 1:03.7 (8전)	평균 38.4 (8전)	3 0 0 3 2 3(7/11) 455+8 25.1 198					4 0 0 4 5 5(7/11) 447-1 41.4 210					2 0 0 2 2 4(3/11) 448-5 19.3 233					3 0 0 4 4 2(6/11) 453+8 8.5 206				
2 로키	송석헌[타] 53.0(0)	050306 국6 1000 1R 마맑 5%					050116 국5 1000 1R 마맑 6%					041218 국5 1000 1R 마흐 9%					041202 1R 주행조교검사 맑 14%				
3 (02.04.04) 한 암 갈	7조 0/ 0/ 0/ 0 406 10.8% 5	①10매직스피	1:03.5	53.0 형철 (1)	14.3		① 4프라이즈	1:03.0	53.0 기정 (7)	14.3		①12빅크라운	1:03.3	53.0 봉주 (4)	14.3		① 4럭키챔프	1:05.5	53.0 창구		
퓨처퀘스트 / 레스트리스패	박종곤(1조) 167/ 10/ 11/ 19	② 2블랙데블	1:04.1	54.0 혜성 (3)	13.7		② 2동해신비	1:03.7	53.0 경환 (2)	13.1		② 3스피카	1:03.6	53.0 태종 (6)	13.9		② 1로키	1:05.7	53.0 병기	39.5	
1,975 / 0 / 0	박재범 강홍수	③ 3군셈	1:04.2	52.0 정선 (4)	38.9		③ 5매직스피	1:03.8	53.0 동철 (6)	38.3		③ 9지대지강	1:03.8	53.0 창기 (5)	39.4		③11뉴타라	1:05.9	52.0 영민	51.0	
3(0 / 0 / 0) 4주 1착승	(국산경매) 29,000	④ 5헤피해피	1:04.7	53.0 창구 (5)	49.5		④ 3천지질주	1:04.1	53.0 효섭 (1)	49.2		④ 1애드허쉬	1:03.8	53.0 희진 (1)	50.0		④ 8순수시대	1:06.1	52.0 재필		
(석현)5/89		⑤ 6로키	1:04.7	53.0 희진 (2)			⑤10로키	1:04.1	53.0 희진 (3)			⑤ 8하우메니	1:03.9	53.0 수홍 (3)			⑤ 2참나무	1:06.7	51.0 성환		
		⑥ 8삭스우먼	1:04.8	53.0 수홍 (6)	14.3		⑥ 7폴스피드	1:05.5	51.0 금주 (4)	14.3		⑥11보라빛	1:05.2	53.0 귀배 (12)	14.3		⑥12지대지강	1:06.7	55.0 창기		
최고 1:04.1 (050116, 53.0, 6%)	최고 38.3 (050116, 53.0, 6%)	⑦11비바	1:05.2	50.0 대근 (9)	32.3		⑦11스마일빔	1:05.5	53.0 완식 (9)	32.3		⑦10로키	1:05.2	53.0 춘식 (7)	32.2		⑦ 3프라이즈	1:07.3	53.0 기정		
평균 1:04.6 (3전)	평균 38.8 (3전)	4 0 0 4 5 5(5/12) 470+0 6.6 282					4 0 0 4 4 3(5/11) 470-2 26.5 193					8 0 0 8 8 8(7/12) 472-8 54.5 76					480Kg				

Figure 1: Example of a Physical Pamphlet Distributed by the Korea Racing Authority

This figure shows an excerpt from *Today's Race*, a physical pamphlet distributed by the Korea Racing Authority for the second race at Seoul Racecourse on April 2, 2005. The first and second rows present information on the horses starting from gates 1 and 2, respectively. The first column summarizes each horse's basic characteristics, while the second column provides details about the jockey. Starting from the third column, the pamphlet lists each horse's past performance, including rankings and race times, as well as records of the other horses they competed against. For example, the entry highlighted in blue in the third column of the first row indicates that the horse starting from gate 3 finished 7th (marked by a circled numeral) with a time of 1:04.7 in its most recent race on March 5, 2005.

Racing Authority's website (http://race.kra.co.kr). Since November 1999, the KRA has publicly disclosed detailed race data on this platform, including each horse's race time measured to the 0.1 second, with updates provided in real time. In addition, private data vendors compile and publish race information in booklet form, often accompanied by their own predictions.

3.2 Data and Sample Selection

Source We collect all horse race records from January 1993—the earliest available date—through October 2024 from the Korea Racing Authority's official website. The dataset comprises 766,947 race entries for 46,425 horses across 74,012 races. It contains detailed information for each race, including the horse's weight on the race date, finishing time (measured to the 0.1 second), distance behind the preceding horse, and finalized betting odds.

Figure 2 summarizes the distribution of betting odds in the Korean horse racing market. Panel A presents the density plots of win and place betting odds as determined by the pari-mutuel system, showing that win odds frequently exceed 100. Panel B displays a box plot of win betting odds by horses' rankings in their most recent race. This pattern suggests that horses ranked lower in their previous race are perceived as longshots with low winning

probabilities. Panel C shows the average realized gross return per dollar wagered, stratified by betting odds. The results reveal a clear pattern consistent with the long-shot bias: bets placed on horses with low winning probabilities systematically yield lower average returns, aligning with prior studies (e.g., Thaler and Ziemba, 1988; Snowberg and Wolfers, 2008).

We also collect data on the size distribution of bets from the Public Data Portal (`http://data.go.kr`), maintained by the South Korean government. This dataset spans the period from January 1995 to May 2024 and reports the total amount wagered and the number of bets placed at each racecourse on each race date, categorized by bet size (in KRW) into the following intervals: (0–100], (100–200], ..., (900–1,000], (1,000–2,000], (2,000–3,000], ..., (9,000–10,000], (10,000–20,000], (20,000–30,000], ..., (90,000–99,900], and 100,000, the maximum allowable bet. Using this dataset, we compute the average bet size for each racecourse–race date:

$$\text{Bet Size}_i = \frac{\sum_{b\in\mathcal{B}} V_{i,b}}{\sum_{b\in\mathcal{B}} N_{i,b}}, \tag{6}$$

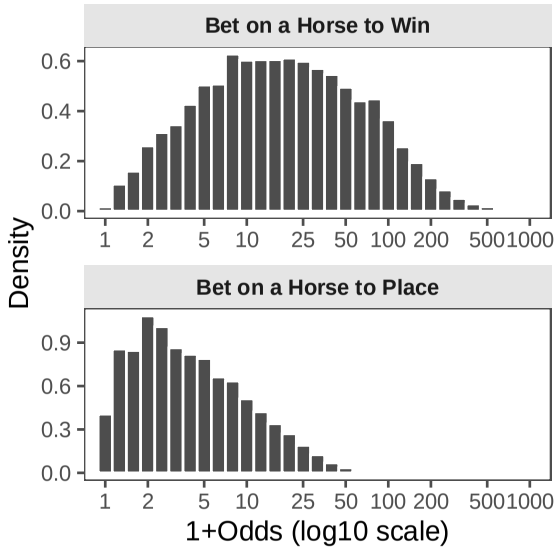
where Bet Size_i is the average bet size for racecourse–race date i, $V_{i,b}$ and $N_{i,b}$ are, respectively, the amount wagered and the number of bets placed on racecourse–race date i with bet size in interval b, and $\mathcal{B}$ is the set of bet-size intervals. This measure serves as a proxy for bettors' economic stakes in predicting winning and placing outcomes. As an alternative, we also compute an amount-weighted bet size for each racecourse–race date:

$$\text{Bet Size (Amount-Weighted)}_i = \frac{\sum_{b\in\mathcal{B}} V_{i,b} \times S_b}{\sum_{b\in\mathcal{B}} V_{i,b}}, \tag{7}$$

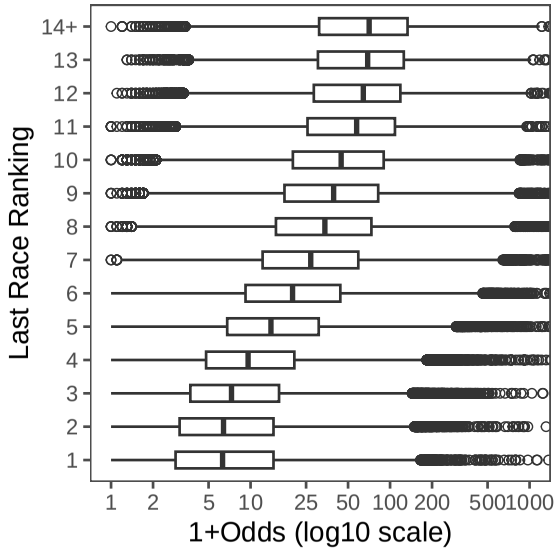
where S_b is the midpoint of bet-size interval b. Table O2 of the Online Appendix shows that our main results remain robust when we use this alternative measure of bet size.

Sample Construction Using the raw datasets described above, we construct a dataset at the race×ranking×sign level, where race×ranking identifies a ranking threshold. Each ranking threshold within a race defines a *pair* of adjacently ranked horses, and each horse in the pair contributes one observation, with *sign* indicating whether it is the higher- or lower-ranked member. Because adjacent pairs overlap, all horses—except those ranked first or last in a race—appear twice in the dataset. Specifically, a horse ranked n in race i appears once as the higher-ranked horse in the pair around the nth ranking threshold (with the horse ranked $n+1$), and once as the lower-ranked horse in the pair around the $(n-1)$th ranking threshold (with the horse ranked $n-1$).

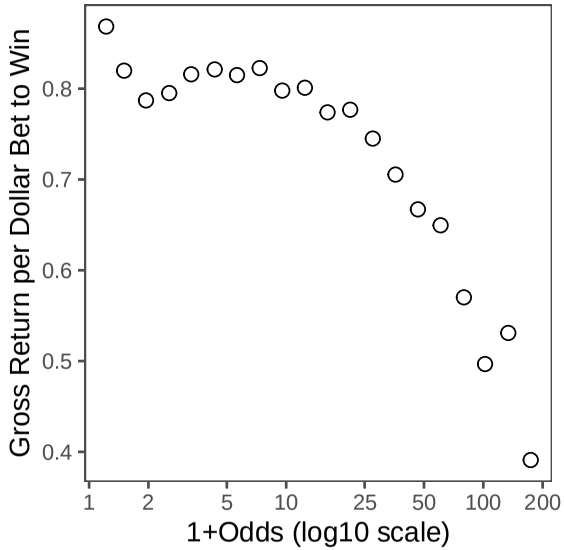
For each pair, we analyze differences in the two horses' subsequent race performance. We exclude observations with missing win betting odds and cases in which a horse was scratched



(A) Density Plot of Win and Place Betting Odds



(B) Box Plot of Win Betting Odds by Last Ranking



(C) Long-Shot Bias

Figure 2: Betting Odds

This figure presents the distribution of betting odds and realized returns in the Korean horse racing market. Panel A displays the density plots of win and place betting odds. Panel B shows a box plot of win betting odds by horses' rankings in their most recent race. Panel C presents the average realized gross return per dollar wagered on a win, grouped by betting odds.

or did not finish its subsequent race. These exclusions reduce the number of observations from 1,289,924 to 1,220,324. We refer to the resulting dataset as the *full sample* and present its summary statistics in Table 1.

To identify and exclude from the full sample pairs of horses where the competitive environment in the next race may differ between the higher- and lower-ranked horse, we apply the following filtering steps:

1. **Rating System**: We exclude races governed by the rating system to avoid the mechanical influence of past rankings on future competitive environments and, consequently, on race outcomes. After this step, the sample includes all tiers before February 2015 and only sixth-tier horses after February 2015, yielding 923,252 observations. Additionally, before February 2015, some races featured handicap weights assigned at the handicappers' discretion. Excluding these races further reduces the sample to 766,459 observations.

2. **Unawarded Horses**: We then restrict the sample to horses that were unawarded—i.e., did not finish in the top five—in their previous race. This helps ensure that both horses in a pair are likely to face similarly ranked opponents and comparable handicap weights in their next race. This step reduces the sample to 357,755 observations.

3. **Special Cases and Outliers**: Finally, we exclude four types of cases: (i) pairs whose race time gap in their most recent race is greater than 5 seconds;[14] (ii) horses that competed in a "special" race between their previous and subsequent races;[15] (iii) horses whose racing tier changed without a documented reason between the two races; and (iv) horses with inter-race date gaps below the 1st percentile or above the 99th percentile of the observed distribution. These exclusions remove outliers and special cases, yielding a final sample of 320,344 observations.

We refer to this refined dataset as the *regression sample* and use it for our empirical analysis. Figure 3 reports the number of observations in both the full sample and the regression sample, stratified by ranking in the prior race. In addition, the summary statistics for the regression sample are presented in Table 2.

[14] We impose this restriction to mitigate the influence of outliers on the MSE-optimal bandwidth of Calonico *et al.* (2014). Pairs with a race time gap exceeding 5 seconds account for only 0.34 percent of all pairs.

[15] Unlike standard races, which are held routinely and run roughly 100 times per week, special races take place only about 40 times per year and are designed exclusively to showcase top-tier horses, jockeys, and trainers.

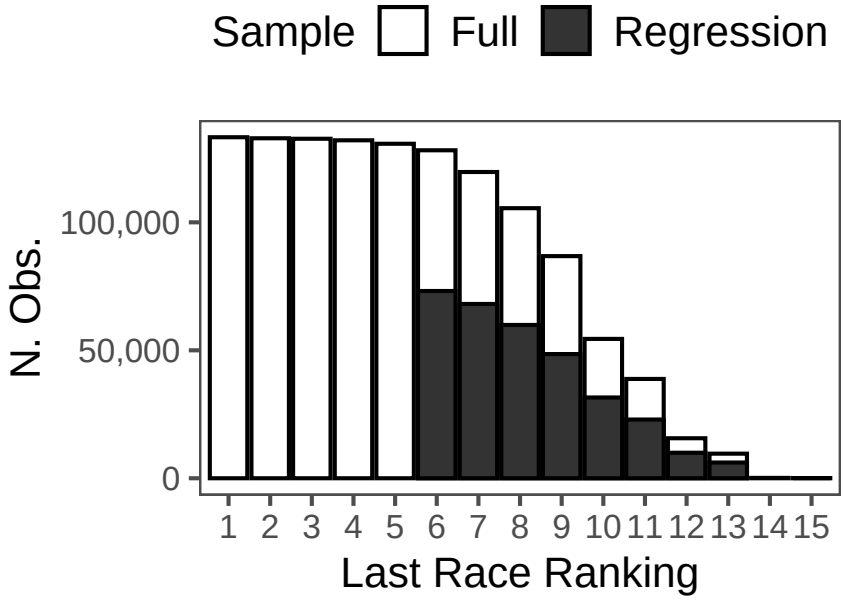


Figure 3: Sample Size by Prior-Race Ranking

This figure displays the number of observations in both the full sample and the regression sample, grouped by ranking in the previous race.

4 Empirical Model

Using the regression sample, we estimate the following specification within a regression discontinuity design (RDD) framework:

$$Y_{h,i,r,n} = \tau D_{h,r,n} + f(\tilde{X}_{h,r,n}) + \mu_{r,n} + \varepsilon_{h,i,r,n}, \tag{8}$$

where $Y_{h,i,r,n}$ denotes the outcome for horse h in its subsequent race i; race r is the horse's previous race, in which it was ranked either n or $n+1$ depending on its position in the pair. Let $X_{r,n}$ denote the race time of the nth-ranked horse in race r. The running variable $\tilde{X}_{h,r,n}$ for the pair around ranking threshold n in race r is the gap between horse h's race time and its pair partner's: if horse h was the higher-ranked member of the pair around threshold n (i.e., ranked n), then $\tilde{X}_{h,r,n} = X_{r,n} - X_{r,n+1} \leq 0$; if horse h was the lower-ranked member (i.e., ranked $n+1$), then $\tilde{X}_{h,r,n} = X_{r,n+1} - X_{r,n} \geq 0$.[16]

The variable $D_{h,r,n}$ is an indicator equal to one if horse h was the higher-ranked horse around ranking threshold n in race r. The function $f(\tilde{X}_{h,r,n})$ controls flexibly for the effect of

[16]Since race time is recorded in 0.1-second increments, the two horses in a pair may have the same recorded race time even when they are not tied, leading to a race time gap of zero. The official ranking is determined by the actual finish order, so $D_{h,r,n}$ (defined below) is well-defined even when the recorded gap is zero.

the running variable $\tilde{X}_{h,r,n}$. The parameter of interest, τ, captures the difference in outcomes between higher- and lower-ranked horses around the ranking threshold. Finally, we include race-by-ranking fixed effects, $\mu_{r,n}$, which serve as pair fixed effects and restrict identification to comparisons between the two horses in the same pair.

We employ both non-parametric and parametric approaches to specify $f(\tilde{X}_{h,r,n})$. In the non-parametric approach, we restrict the sample to horse pairs separated by no more than a nose (approximately 3 inches) in their previous race. For these pairs, the race time gap is less than 0.1 seconds and is therefore recorded as zero, so $\tilde{X}_{h,r,n} = 0$ for both horses and $f(\tilde{X}_{h,r,n})$ drops out of the specification. In the parametric approach, we model $f(\tilde{X}_{h,r,n})$ as a cubic polynomial in $\tilde{X}_{h,r,n}$ fully interacted with the treatment indicator $D_{h,r,n}$:

$$f(\tilde{X}_{h,r,n}) = \gamma_1 \tilde{X}_{h,r,n} + \gamma_2 \tilde{X}^2_{h,r,n} + \gamma_3 \tilde{X}^3_{h,r,n} + \gamma_4 \left(\tilde{X}_{h,r,n} \times D_{h,r,n} \right) + \gamma_5 \left(\tilde{X}^2_{h,r,n} \times D_{h,r,n} \right) + \gamma_6 \left(\tilde{X}^3_{h,r,n} \times D_{h,r,n} \right). \quad (9)$$

This parametric approach accommodates potential nonlinearities and allows the slope of the outcome function to differ on either side of the threshold. We estimate the model within the MSE-optimal bandwidth of Calonico *et al.* (2014) for the race-time gap.[17] In addition, we examine the robustness of the results to alternative bandwidth choices and polynomial orders, with the findings reported in the Online Appendix.

One remaining concern is that selection into the subsequent race may invite alternative interpretations of our empirical findings. In particular, higher- and lower-ranked horses could be systematically paired with jockeys of differing ability. To address this concern, we follow Abowd *et al.* (1999) and estimate each jockey's ability using the raw race record dataset:

$$Y_{h,j,r} = \alpha_h + \lambda_j + \delta_r + \varepsilon_{h,j,r}, \quad (10)$$

where $Y_{h,j,r}$ denotes the ranking of horse h, ridden by jockey j in race r. The term α_h captures horse fixed effects, λ_j represents jockey fixed effects, and δ_r accounts for race fixed effects. The coefficient λ_j captures jockey j's ability, controlling for both horse ability and the competitive environment of race r. Using these estimated jockey abilities, we test whether higher- and lower-ranked horses are systematically matched with jockeys of differing ability in their subsequent race. We also test for balance in other observable characteristics of the

[17] Specifically, because race times are recorded in 0.1-second increments, pairs with a race-time gap below 0.1 seconds have a running variable equal to zero, i.e., exactly at the cutoff. For bandwidth selection, we therefore apply a sign-assigning perturbation ($\pm 10^{-6}$ seconds) to the running variable based on the observed ranking. We employ a uniform kernel so that all observations within the bandwidth receive equal weight. Given the measurement granularity of race times, we round the optimal bandwidth down to the nearest 0.1 second.

subsequent race.

A separate concern arises when the two horses in a pair compete in the same subsequent race. In that case, their realized outcomes and bet shares are mechanically linked within the race: exactly one horse wins, a fixed number of horses place, and the shares of bets in each betting pool sum to one. This dependence induces a negative correlation between paired horses' outcomes, as an improvement in one horse's outcome comes at the expense of the other's. We address this concern by re-estimating our baseline specifications after excluding pairs whose two members compete in the same subsequent race. This restriction removes the direct within-pair mechanical dependence in both realized outcomes and bet shares.

5 Results

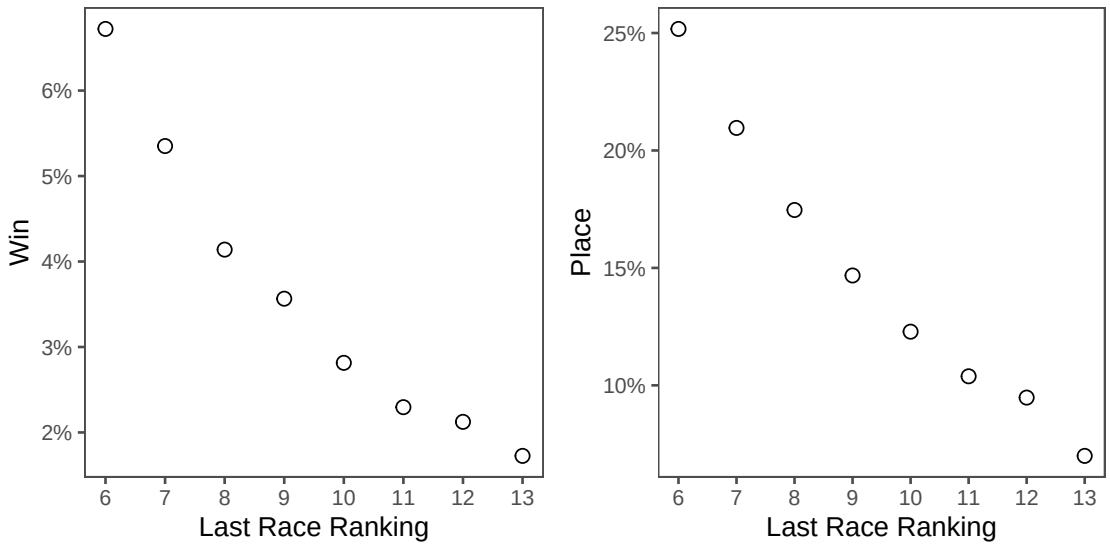
5.1 Redundancy of Ranking Information

We begin by evaluating the predictive value of a horse's past ranking for its subsequent performance, using the regression sample. As noted in Section 3.2, this sample is restricted to unawarded horses—that is, those that did not finish in the top five in their previous race (horses ranked 6th through 13th). Figure 4 displays the probability of a horse winning or placing in its next race as a function of its prior ranking. The figure shows that, when not conditioning on race time, past rankings are meaningfully predictive of future outcomes. For instance, horses ranked sixth in their previous race win approximately 6.7 out of 100 times in their next start, while those ranked 13th win only about 1.7 out of 100 times.

We conduct three analyses to test whether horses' past rankings, conditional on race time information, remain useful for forecasting future performance. First, we perform a univariate analysis comparing horses designated as higher- or lower-ranked in their previous race. We restrict the sample to horse pairs that finished within a nose (roughly three inches) of each other. As previously defined, for each race×ranking, we refer to the two adjacent horses around the ranking as a "pair."

Table 3 presents the results. We find no significant differences in horse characteristics—such as gender, gelding status, age, or weight—between higher- and lower-ranked horses. Likewise, race characteristics and gate assignments do not differ significantly across the two groups. Definitions of all variables are provided in Appendix A.

Second, we test whether higher- and lower-ranked horses differ in handicap weights or the abilities of their jockeys in the subsequent race. To do so, we estimate Equation (8), using either handicap weights or estimated jockey ability as the dependent variable. Table 4 reports the results in two panels. In both panels, columns (1) and (2) focus on the subsample of pairs



(A) Probability of Winning **(B)** Probability of Placing

Figure 4: Predictive Value of Past Rankings

This figure examines the extent to which a horse's prior ranking predicts its performance in a subsequent race. The x-axis in both panels denotes the horse's ranking in its previous race. Panel A plots the probability of winning the next race as a function of past ranking, while Panel B presents the probability of placing in the subsequent race.

with a distance gap of less than or equal to a nose. Columns (3) and (4) extend the analysis to pairs with a race-time gap within the MSE-optimal bandwidth. These specifications control for the gap between a horse's race time and the threshold around its ranking, using a cubic polynomial of the race time gap interacted with $D_{r,n}$, an indicator variable for the higher-ranked horse, as specified in Equation (9). Columns (2) and (4) also include pair fixed effects.

Panel A of Table 4 reports the results for handicap weights. Across all four specifications, the coefficient estimates are not statistically distinguishable from zero, indicating that higher- and lower-ranked horses are not assigned different weights in their next race. This finding supports the idea that being marginally above or below a ranking threshold has no effect on handicap assignments.

Panel B of Table 4 examines whether the ability of jockeys riding higher- and lower-ranked horses in their next race systematically differs. This is an important consideration, as jockeys are selected by horse owners, and systematic differences in jockey skill could rationalize the observed betting behavior. In other words, we may observe differences in perceived winning probabilities for higher- and lower-ranked horses in their next race, even in the absence of bettor misperception. To address this, we first estimate jockey ability using Equation (10)

and then re-estimate Equation (8) with estimated jockey ability as the outcome variable. The results show that across all four specifications, the coefficient of interest remains statistically insignificant. Therefore, we find no evidence of systematic differences in jockey ability riding higher- and lower-ranked horses in their subsequent races, making it unlikely that jockey ability explains the observed betting patterns.

Finally, we test whether actual performance differs between higher- and lower-ranked horses in their next race. We estimate Equation (8) using two outcome variables: an indicator equal to one if the horse wins, multiplied by 100, and an indicator equal to one if the horse places, multiplied by 100. Table 5 presents the findings. As in prior analyses, columns (1) and (2) focus on nose-gap pairs, while columns (3) and (4) include pairs with a race-time gap within the MSE-optimal bandwidth, incorporating a polynomial function of the race-time gap interacted with the higher-ranked indicator, as outlined in Equation (9).

The results are reported in Table 5. Panel A examines winning probabilities, and Panel B analyzes placing probabilities. In both cases, none of the coefficient estimates is statistically significant across the four specifications. These results confirm that, conditional on race time, there is no difference in future performance between higher- and lower-ranked horses.

Taken together, the evidence in Tables 3, 4, and 5 shows that, once race time information is accounted for, ranking information no longer predicts future performance. This supports the validity of our sample construction and reinforces the conclusion that rankings, in and of themselves, do not generate real effects on subsequent performance outcomes.

5.2 Ranking Information and Misperception

In the previous section, we show that once race time information is accounted for, a horse's prior ranking no longer predicts its actual probability of winning or placing in its next race. In this section, we examine whether bettors' perceived probabilities for a horse to win or place in a race vary based on the horse's ranking in its previous race.

Specifically, we test whether perceived winning and placing probabilities differ between higher- and lower-ranked horses. To do so, we first infer these perceived probabilities using the share of bets placed on each horse, as outlined in Section 2. We then estimate Equation (8) using two outcome variables: Bet to Win (%) and Bet to Place (%), defined as the share of bets placed on a horse to win or to place, respectively, multiplied by 100.

Mirroring the structure in Table 5, Table 6 presents results in two panels. In both panels, columns (1) and (2) focus on the subsample of pairs with a distance gap of less than or equal to a nose. Columns (3) and (4) extend the analysis to pairs with a race-time gap within the MSE-optimal bandwidth. These specifications control for the gap between a

horse's race time and the threshold around its ranking, using a polynomial of the race-time gap interacted with $D_{r,n}$, an indicator variable for the higher-ranked horse, as specified in Equation (9). Columns (2) and (4) also include pair fixed effects.

Panel A examines perceived winning probabilities. Across all four specifications, we find consistent evidence that bettors perceive the higher-ranked horse as more likely to win its next race. Across all four regression specifications, the coefficient estimates are statistically significant and indicate that bettors place between 0.26 and 0.46 percentage points more win bets on higher-ranked horses compared to their lower-ranked counterparts. Relative to the average win bet share, this represents an increase of approximately 6 to 10 percent, despite no corresponding difference in actual performance.

The economic implications of bettor misperception are substantial. Consider, for example, a horse with a true winning probability of 4 percent. If the share of bets placed on this horse matches this probability, the expected gross return is $(1-0.2)\times 0.04\times(\frac{1}{0.04})=0.8$ after the 20 percent house take—that is, an expected loss of 20 percent. However, if bettors overestimate the horse's chances and the share of bets increases by 0.4 percentage points (to 4.4 percent), the expected return falls to $(1-0.2)\times 0.04\times(\frac{1}{0.044})\approx 0.727$, an expected loss of 27.3 percent. Misperception thus raises the expected loss by 7.3 percentage points, or roughly 36 percent of the baseline 20 percent loss.

Panel B of Table 6 reveals similar patterns for place bets. Once again, the coefficient estimates are positive and statistically significant across all four regression specifications. Consistent with the findings for win bets shown in Panel A, bettors allocate between 0.29 and 0.40 percentage points more place bets to higher-ranked horses relative to their lower-ranked counterparts. Given the average share of place bets, this corresponds to an increase of approximately 4.7 to 6.2 percent.

Figure 5 summarizes the main empirical findings. In each panel, the x-axis plots the difference in race times between horses—that is, between higher- and lower-ranked horses with nearly identical race times around a given ranking. Panels A and B compare the actual probabilities of winning and placing (finishing in the top three), respectively, for the two horses in their subsequent race. The results show no meaningful difference in either outcome. These patterns indicate that, conditional on race time, ranking provides no incremental information about future performance.

By contrast, the betting behavior shown in Panels C and D of Figure 5 reveals a pronounced divergence from realized outcomes. These panels compare the proportions of win and place bets placed on higher- versus lower-ranked horses. Despite exhibiting nearly identical ex post performance, higher-ranked horses attract significantly more wagers than their lower-ranked counterparts. This pattern suggests that bettors rely on ranking

information and perceive higher-ranked horses as more likely to win or place, even though rankings are redundant conditional on race times. We show that the results for both actual outcomes and perceived probabilities are robust to alternative bandwidths and polynomial orders; see Figures O1–O6 and Table O1 in the Online Appendix.

Overall, the results provide robust empirical evidence that bettors perceive higher-ranked horses as more likely to win or place in their next race, despite these horses exhibiting identical performances ex post. The absence of real effects of rankings on future performance, as demonstrated in Section 5.1, makes alternative explanations based on rational expectations unlikely drivers of this misperception. For such explanations to hold, they would need to account for both (i) the observed divergence in perceived probabilities and (ii) the lack of corresponding differences in actual outcomes.

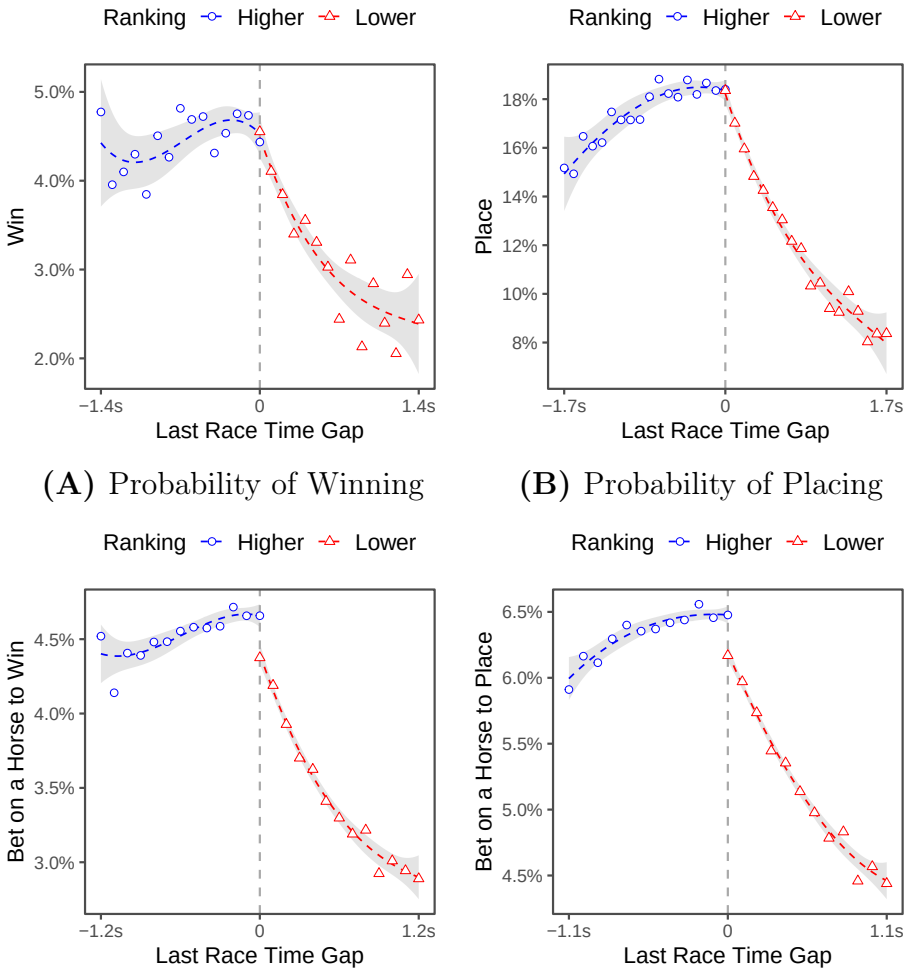
5.3 Addressing Mechanical Interdependence Within Pairs

One concern about our main results is that the two horses that competed in the previous race may compete again in the same subsequent race, where outcomes are mechanically dependent: exactly one horse wins each race, the number of placing positions is fixed, and the shares of win and place bets each sum to one within their respective betting pools. For example, if one member of a pair wins the race, the other cannot; likewise, a larger bet share for one horse mechanically reduces the shares of the remaining horses, including its pair partner. The observed differences between paired horses in their subsequent race might therefore partly reflect this mechanical negative dependence rather than discretization-driven misperception.

To address this concern, we re-estimate the analyses in Tables 5 and 6 after restricting the sample to pairs in which the two horses compete in *different* subsequent races. Because the within-race constraints on realized outcomes and bet shares do not operate across races, this restriction removes the direct mechanical dependence between the two horses in each pair. It therefore allows us to assess whether the divergence between betting behavior and realized outcomes persists absent this dependence.

The results are reported in Table 7. Panel A presents the results for actual winning and placing probabilities, while Panel B presents the results for the shares of win and place bets. Within each panel, the odd-numbered columns implement a non-parametric approach by restricting the sample to pairs with a distance gap of less than or equal to a nose, while the even-numbered columns use the MSE-optimal bandwidth and the polynomial specification in Equation (9).

The estimates in Panel A remain statistically indistinguishable from zero, indicating



(A) Probability of Winning **(B)** Probability of Placing

(C) Shares of Win Bets **(D)** Shares of Place Bets

Figure 5: Ranking, Performance, and Betting Behavior

This figure examines the extent to which past ranking, conditional on race time, influences subsequent race performance and bettor behavior. The x-axis shows the difference in race times between adjacently ranked horses, calculated as the race time of the higher-ranked horse (*Higher*) relative to that of the horse ranked immediately below (*Lower*). Panels A and B compare the actual probabilities of winning and placing, respectively, between higher- and lower-ranked horses in their subsequent race. Panels C and D compare the shares of win and place bets placed on each horse. The sample in each panel is restricted to the MSE-optimal bandwidth selected following the procedure in Calonico *et al.* (2014). Circles and triangles represent unconditional means by previous-race time gap, while the dashed lines show fitted values from a cubic polynomial. The shaded gray areas denote 95% confidence intervals for the fitted lines, with standard errors clustered at the horse-pair level.

no difference in actual winning or placing probabilities between higher- and lower-ranked horses in the different-subsequent-race sample. By contrast, Panel B shows that higher-ranked horses continue to attract significantly more bets. For win bets, for example, higher-ranked horses receive 0.46 percentage points more win bets according to the non-parametric approach and 0.26 percentage points more win bets according to the parametric approach. Relative to the mean of the dependent variable, these estimates correspond to increases of approximately 10.3 percent and 6.3 percent, respectively. Importantly, these magnitudes are close to those reported in Table 6, suggesting that the observed divergence in betting behavior is not driven by mechanical interdependence within pairs.

Overall, the findings in Table 7 are both qualitatively and quantitatively consistent with those reported in Tables 5 and 6. This robustness analysis reinforces the conclusion that the observed differences between higher- and lower-ranked horses in their next race capture the effects of misperception arising from information discretization.

5.4 Mechanisms

The previous section demonstrates that ranking information can systematically distort bettor perceptions. Specifically, we show that bettors rely on ranking information when making wagering decisions, even though such rankings are redundant in the presence of readily available race time data. Horses whose race times place them just above a ranking threshold are perceived as more likely to win their next race than those just below, despite having virtually identical ex ante performance and no observable difference in ex post outcomes. In this section, we explore the underlying mechanisms that drive this misperception.

5.4.1 Information Acquisition Costs

Why do bettors rely on ranking information despite its redundancy? One potential explanation is that acquiring race time information may involve costs (Grossman and Stiglitz, 1980). This explanation seems unlikely, however, as both race times and rankings from past performances are publicly available and readily accessible for all horses. Nonetheless, we empirically assess this possibility by examining whether the magnitude of the ranking effects has declined over time, particularly in response to technological advancements that have substantially lowered information acquisition costs in recent decades. If such costs were the primary driver of the ranking effects, we would expect the effects to be more pronounced in earlier periods—especially prior to the diffusion of information technologies around the year 2000.

To formally test this hypothesis, we estimate the ranking effects separately by decade—

the 1990s, 2000s, 2010s, and 2020s—using the following regression specification:

$$Y_{h,i,r,n} = \sum_{k \in \mathcal{K}} \tau_k \left(D_{h,r,n} \times 1\left\{ i \in k \right\} \right) + g\left(\tilde{X}_{h,r,n} \right) + \mu_{r,n,k(i)} + \varepsilon_{h,i,r,n}, \tag{11}$$

where $\mathcal{K} = \{1990\text{s}, 2000\text{s}, 2010\text{s}, 2020\text{s}\}$, $k\left(i\right)$ denotes the decade in which race i takes place, and $1\left\{ i \in k \right\}$ is an indicator variable equal to one if the year of race i falls within decade k. For the parametric approach, the function $g(\tilde{X}_{h,r,n})$ is a cubic polynomial in $\tilde{X}_{h,r,n}$ fully interacted with $D_{h,r,n}$ and the decade indicators $1\{i \in k\}$:

$$\begin{aligned} g(\tilde{X}_{h,r,n}) = \sum_{k \in \mathcal{K}} \Big[& \gamma_{1k} \left(\tilde{X}_{h,r,n} \times 1\left\{ i \in k \right\} \right) + \gamma_{2k} \left(\tilde{X}^2_{h,r,n} \times 1\left\{ i \in k \right\} \right) \\ & + \gamma_{3k} \left(\tilde{X}^3_{h,r,n} \times 1\left\{ i \in k \right\} \right) + \gamma_{4k} \left(\tilde{X}_{h,r,n} \times D_{h,r,n} \times 1\left\{ i \in k \right\} \right) \\ & + \gamma_{5k} \left(\tilde{X}^2_{h,r,n} \times D_{h,r,n} \times 1\left\{ i \in k \right\} \right) + \gamma_{6k} \left(\tilde{X}^3_{h,r,n} \times D_{h,r,n} \times 1\left\{ i \in k \right\} \right) \Big]. \end{aligned} \tag{12}$$

All other covariates are defined as in Equation (8). Since every regressor, including the fixed effects, is interacted with the decade indicators, the coefficients τ_k from this model are identical to those from estimating Equation (8) separately by decade.

The results are presented in Table 8. Columns (1) and (2) report estimates for perceived winning probabilities, while Columns (3) and (4) focus on perceived placing probabilities. To account for the running variable, we employ both non-parametric and parametric approaches. Columns (1) and (3) restrict the sample to pairs with a distance gap of less than or equal to a nose, whereas Columns (2) and (4) extend the analysis to pairs with a race-time gap within the MSE-optimal bandwidth. These specifications control for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race-time gap interacted with an indicator for the higher-ranked horse and the decade indicators, as specified in Equation (12).

The findings reveal that the ranking effects are concentrated in the more recent decades and are either absent or significantly weaker prior to 2000. This pattern runs counter to what one would expect if information acquisition costs were the primary driver of the results. Taken together, the results in Table 8 suggest that information acquisition costs are unlikely to explain bettors' reliance on ranking information.

5.4.2 Rational Inattention

Another potential mechanism underlying the documented bettor misperception is rational inattention, whereby bettors may optimally choose to ignore race time records due to the

cognitive costs associated with processing such information (Sims, 2003). To test this hypothesis, we examine whether the magnitude of ranking effects declines with the size of bettors' wagers. The rationale behind this test is that higher economic stakes provide stronger incentives for bettors to allocate cognitive resources toward processing more information. If rational inattention is driving the reliance on ranking information, we would expect the ranking effects to diminish as wager size increases.

To investigate this, we estimate the ranking effects separately for low and high bet sizes using the following regression specification:

$$Y_{h,i,r,n} = \tau_L \left(D_{h,r,n} \times \text{Low Bet Size}_i\right) + \tau_H \left(D_{h,r,n} \times \text{High Bet Size}_i\right) + h(\tilde{X}_{h,r,n}) + \mu_{r,n,k(i)} + \varepsilon_{h,i,r,n}, \quad (13)$$

where Low Bet Size$_i$ is an indicator for whether the average bet size on the race date of race i falls within the bottom 50 percent of the distribution for that racecourse in the given year, and High Bet Size$_i$ is an indicator for the top 50 percent. We also include fixed effects for pairs of horses that competed in race r around ranking threshold n, interacted with the bet-size group of each horse's subsequent race i: $\mu_{r,n,k(i)}$, where $k\left(i\right) \in \{\text{Low}, \text{High}\}$. Since the bet-size group $k\left(i\right)$ is assigned based on the date of each horse's subsequent race i, the two horses in a pair may fall into different bet-size groups when their subsequent races occur on different dates.[18] One caveat of this structure is that such pairs are split across fixed-effect cells and therefore do not contribute to identification, which reduces the effective sample size.

The function $h(\tilde{X}_{h,r,n})$ is a cubic polynomial in $\tilde{X}_{h,r,n}$ fully interacted with $D_{h,r,n}$ and the bet-size group indicators:

$$\begin{aligned} h(\tilde{X}_{h,r,n}) = \sum_{k \in \{\text{Low,High}\}} \Big[& \gamma_{1k} \left(\tilde{X}_{h,r,n} \times k \text{ Bet Size}_i\right) + \gamma_{2k} \left(\tilde{X}^2_{h,r,n} \times k \text{ Bet Size}_i\right) \\ & + \gamma_{3k} \left(\tilde{X}^3_{h,r,n} \times k \text{ Bet Size}_i\right) + \gamma_{4k} \left(\tilde{X}_{h,r,n} \times D_{h,r,n} \times k \text{ Bet Size}_i\right) \\ & + \gamma_{5k} \left(\tilde{X}^2_{h,r,n} \times D_{h,r,n} \times k \text{ Bet Size}_i\right) + \gamma_{6k} \left(\tilde{X}^3_{h,r,n} \times D_{h,r,n} \times k \text{ Bet Size}_i\right) \Big]. \end{aligned} \quad (14)$$

All other covariates are defined as in Equation (8). As in the previous specification, the coefficients τ_L and τ_H are identical to split-sample estimates.

The results are presented in Table 9. Mirroring the structure of Table 8, the dependent

[18] As discussed in Section 3, there are three racecourses in Korea: Seoul, Busan, and Jeju. Horses run almost exclusively at their home racecourse, so both horses in a pair that competed in Seoul, for example, also run their subsequent races in Seoul. Within-pair variation in average bet size therefore comes from differences in race dates, not racecourses.

variable in Columns (1) and (2) is perceived winning probability, while in Columns (3) and (4), it is perceived placing probability. To account for the effects of the running variable, we employ both non-parametric and parametric approaches. Columns (1) and (3) restrict the sample to pairs with a distance gap of less than or equal to a nose. Columns (2) and (4) extend the analysis to pairs with a race-time gap within the MSE-optimal bandwidth. These specifications control for the running variable by fitting a polynomial function, as specified in Equation (14).

The findings indicate that ranking effects are concentrated in races with low bet sizes and are either absent or significantly attenuated in races with high bet sizes. The point estimates for the low bet-size group are 35 to 219 percent larger than those for the high bet-size group across outcomes and specifications, although the differences are not statistically significant at the 5 percent level, potentially reflecting limited power. This pattern supports the notion that the degree to which bettors attend to race time information depends on the economic stakes involved in their wagering decisions. The results are consistent with what one would expect if rational inattention were a key mechanism underlying bettors' reliance on ranking information.

As a robustness check, Table O2 in the Online Appendix repeats the analysis using the alternative measure of average bet size defined in Equation (7). The resulting estimates are similar to those in Table 9, indicating that the findings are not sensitive to how average bet size is measured. Taken together, the evidence suggests that rational inattention can, at least in part, account for the observed bettor misperception.

5.4.3 Heuristic Reliance on Ranking Information

A final mechanism that may help explain bettors' reliance on ranking information, despite its redundancy, is heuristic decision-making. The literature on decision-making under risk suggests that individuals often struggle to distinguish between small and very small probabilities (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992; Wu and Gonzalez, 1996; Snowberg and Wolfers, 2010). In the context of horse racing, where lower-ranked horses are typically longshots, bettors who rely on simple heuristics may perceive little difference between, for example, the 12th- and 13th-ranked horses. In contrast, they may perceive a more substantial difference between the 6th- and 7th-ranked horses, even when the underlying difference in race times is comparable. Therefore, although the cognitive cost of processing race time information should be uniform across ranking thresholds, the impact of reliance on heuristics, such as ranking information, may vary depending on the rank position.

To investigate this hypothesis, we revisit the analysis presented in Figure 5, this time

allowing the ranking effects to vary by ranking position. Specifically, we estimate the effect of being higher-ranked separately for each ranking threshold. The results are illustrated in Figure 6. Consistent with our earlier findings, we again observe no discontinuity in actual performance: higher- and lower-ranked horses are equally likely to win or place across all ranking thresholds. However, bettor misperception—the effect of ranking information on perceived winning and placing probabilities—is disproportionately concentrated around higher rankings (e.g., 6th and 7th places), while such discontinuities are either absent or negligible at lower rankings. These results are consistent with what one would expect if bettors rely on simple heuristics—such as ranking information—when making decisions under uncertainty. Taken together, the evidence suggests that heuristic decision-making can, at least in part, account for the observed misperception in bettor behavior.

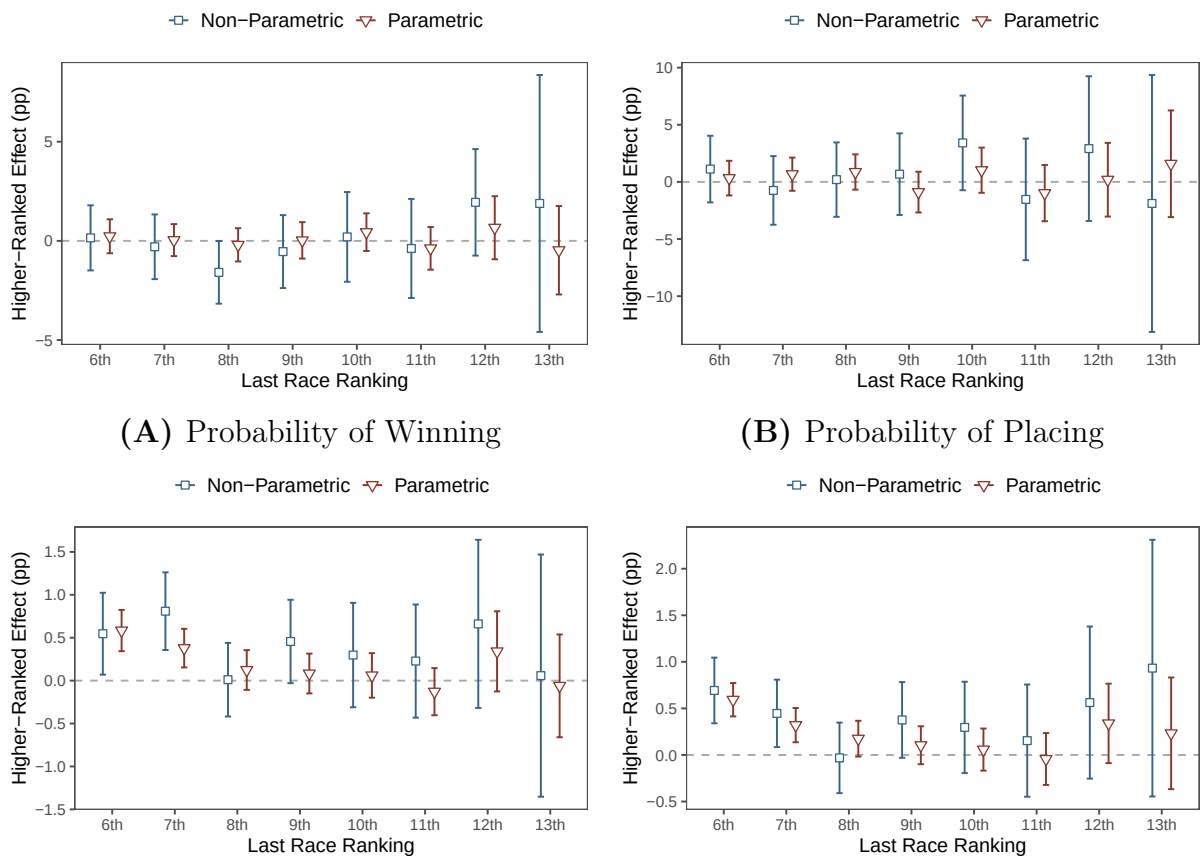
6 Generalizability to Other Settings

6.1 Misperception Effects in the Presence of Real Effects

So far, we have isolated the effects of discretization-induced misperception using pairs of horses in which the higher-ranked horse finishes sixth or lower. For these horses, ranking does not affect tier assignments, compensation, or subsequent competitive conditions. As discussed above, the absence of a real-effect channel allows us to isolate the effects of misperception from any real consequences of information discretization around ranking thresholds. However, in many economically important settings, including financial markets, discrete classifications have real effects on future outcomes. This raises a natural question: whether, and to what extent, the effects of discretization-induced misperception identified in our baseline sample remain quantitatively important when such real effects are also present.

We address this question through a calibration exercise using horse pairs in which the higher-ranked horse finished among the top five in the previous race. In this setting, prior rank affects the competitive environment that a horse faces in its subsequent race. We first verify this institutional feature by testing whether being the higher-ranked horse within a pair is associated with moving to a higher competition tier and with lower probabilities of winning and placing in the subsequent race. Panel A of Table 10 reports the results. Consistent with the institutional feature, being the higher-ranked horse within a pair is significantly associated with a change in tier and with lower probabilities of winning and placing in the subsequent race. These findings are consistent across the parametric and non-parametric approaches.

Do bettors anticipate these real effects and incorporate them into their betting decisions?



(A) Probability of Winning

(B) Probability of Placing

(C) Shares of Win Bets

(D) Shares of Place Bets

Figure 6: Ranking Effects by Ranking Position

This figure revisits the ranking effects analyzed in Figure 5, disaggregated by ranking position. Panels A and B display the actual probabilities of winning and placing, respectively, in the horse's next race. Panels C and D report the shares of win and place bets, which proxy bettors' perceived probabilities of winning and placing. The x-axis in all panels represents the horse's finishing position in its previous race, ranging from 6th to 13th place. Each panel plots the estimated coefficient on $D_{r,n}$—an indicator for the higher-ranked horse—estimated separately for each ranking position. Two specifications are shown for each rank: The non-parametric approach uses the nose-length sample with pair fixed effects; The parametric approach uses observations within MSE-optimal bandwidth selected following Calonico *et al.* (2014) with a cubic polynomial in the race-time gap interacted with $D_{r,n}$ (see Equation (9)) and pair fixed effects. The vertical lines denote the 95 percent confidence intervals for the fitted lines. Standard errors are clustered at horse-pair level.

To answer this question, we next examine how being the higher-ranked horse within a pair affects the shares of win and place bets a horse attracts in the subsequent race. Panel B of Table 10 reports the results. Across all four regression specifications, the coefficient estimates are negative and statistically significant, indicating that bettors allocate significantly smaller shares of win and place bets to higher-ranked horses than to their lower-ranked counterparts. In the non-parametric specification, for example, the win-bet share of higher-ranked horses is 0.72 percentage points lower, while their place-bet share is 0.35 percentage points lower.

While the above results indicate that bettors incorporate the real effects of ranking into their bets, the question remains whether they nonetheless overestimate the winning or placing probabilities of higher-ranked horses. To answer this question, we first infer bettors' perceived winning and placing probabilities from their betting behavior. Following Ali (1977) and Snowberg and Wolfers (2010), we consider an equilibrium condition in which risk-neutral bettors bet according to their subjective probabilities,[19] and show that the perceived winning probability of a horse coincides with its share of win bets. Specifically, for risk-neutral bettors who bet on horses to win, the equilibrium condition in Equation (1) implies that

$$\hat{q}_i = \lambda \times \left(\frac{1}{1 + o_i} \right) \tag{15}$$

for some $\lambda > 0$, where $\hat{q}_i$ is the perceived winning probability of horse i in a race, and o_i is the betting odds on horse i. Since the perceived winning probabilities of horses in a race sum to one, we have

$$\hat{q}_i = \left(\frac{1}{1 + o_i} \right) / \left(\sum_{j \in \mathcal{I}} \frac{1}{1 + o_j} \right), \tag{16}$$

where $\mathcal{I}$ is the set of horses competing in the race. As shown in Section 3, this expression equals the share of win bets placed on horse i. Similarly, the perceived placing probability of a horse is given by its share of place bets multiplied by the number of place slots in the race, since the perceived placing probabilities of all horses in a race sum to the number of place slots.

Based on these perceived probabilities inferred from betting behavior, for each horse, we construct the gap between perceived and actual probabilities of winning or placing. A higher value of this gap indicates that bettors overestimate the horse's winning or placing probability. We then examine whether the gap is larger for higher-ranked horses than for lower-ranked counterparts. While the gap may contain measurement error in perceived probabilities (e.g., due to the risk-neutrality approximation), there is no strong reason to

[19] While our main discontinuity test does not require risk neutrality, this additional assumption is introduced to quantitatively map betting behavior to perceived probabilities.

expect such error to be discontinuous around the ranking threshold.

Panel C of Table 10 presents the results. We find that the gap between the perceived and actual winning probabilities is 0.80 percentage points larger for the higher-ranked horse than for the lower-ranked horse in the nose-gap sample and 0.77 percentage points larger in the parametric specification. The former estimate is statistically significant at the 10 percent level, whereas the latter is significant at the 1 percent level. The corresponding discrepancies for placing probabilities are 2.65 and 2.34 percentage points, respectively, both of which are statistically significant at the 1 percent level. Overall, these results suggest that bettors overestimate the winning and placing probabilities of higher-ranked horses even in a setting where ranking information has real effects, which could reflect bettors' miscalibration of the real effects, discretization-driven misperception, or other channels.

How much of this overestimation can be explained by discretization-driven misperception? While we cannot directly estimate the misperception component in the top-five sample, we use our estimate of the misperception effect between the sixth- and seventh-ranked horses, where ranking has no real effects, as a benchmark for this calibration. This benchmark is likely to be a lower bound on the misperception effect in the top-five sample: Figure 6 shows that the misperception effect tends to be larger at higher rankings, so the corresponding effect for top-five horses likely exceeds that for the sixth- and seventh-ranked pair. For winning probabilities, the benchmark is 0.55 percentage points in the nose-gap sample and 0.58 percentage points in the parametric specification, corresponding to 68.6 and 75.8 percent, respectively, of the win-probability discrepancy observed in the top-five sample. For placing probabilities, the benchmark estimates of 2.05 and 1.77 percentage points correspond to 77.1 and 75.6 percent of the observed discrepancy. Overall, the calibration results reported in Table 10 suggest that the documented misperception effect remains economically important even when discrete classifications also generate real effects.

6.2 Implications for Financial Markets

The calibration exercise in Table 10 provides a bridge to financial markets, where discrete classifications can shape both investor beliefs and subsequent outcomes. As discussed in Section 2, classifications such as credit ratings and mutual fund star ratings summarize underlying information and reduce processing costs, but they can also generate real effects. Investors may therefore respond rationally to a category change even when the underlying continuous signal changes only marginally. Our calibration shows, however, that such responses need not eliminate the misperception induced by the classification itself. Bettors adjust their wagers in the direction implied by the real effects of rank, yet they continue

to overestimate the winning and placing probabilities of higher-ranked horses. Thus, discontinuities in investor beliefs around financial classification thresholds may reflect both a rational response to the consequences of crossing a threshold and an excessive response to the categorical label itself.

Several features of betting markets make our findings plausibly relevant to financial markets. Both markets aggregate dispersed beliefs into prices under real economic stakes and feature broad information availability, repeated participation, sophisticated traders, arbitrage opportunities, and substantial payoff variation (Moskowitz, 2021). The two settings also attract participants with related behavioral tendencies. Investors with stronger gambling preferences trade more actively and disproportionately hold lottery-like securities (Kumar, 2009; Dorn and Sengmueller, 2009; Grinblatt and Keloharju, 2009; Dorn *et al.*, 2015; Gao and Lin, 2015), while access to legal sports betting appears to reduce household investment through brokerage accounts (Baker *et al.*, 2024). More directly, evidence from financial markets indicates that categorization and limited attention affect prices. When a stock is reclassified, its returns comove more strongly with its new category and less strongly with its former category, consistent with category-based demand (Barberis *et al.*, 2005). Related theory shows that limited attention can lead investors to process information at the category level rather than at the security level (Peng and Xiong, 2006). Moreover, investors react less promptly to earnings announcements when attention is limited or diverted (DellaVigna and Pollet, 2009; Hirshleifer *et al.*, 2009), and systematic behavioral biases persist even among professional futures traders (Coval and Shumway, 2005). Together, this evidence suggests that public information, repeated participation, and market sophistication alone need not eliminate discretization-induced misperceptions.

These parallels do not imply that the magnitudes estimated in our setting carry over directly to financial markets. Financial securities have longer horizons, noisier and more endogenous payoffs, and richer information sets, features that may strengthen investors' reliance on coarse classifications. Conversely, short selling, hedging, and arbitrage may attenuate the price effects of such misperception, although they need not eliminate them because arbitrageurs face constraints (Shleifer and Vishny, 1997). Our evidence therefore supports the qualitative implication that discretized information can affect investors' beliefs even when it contains no incremental information.

7 Conclusion

This paper examines how discretized information influences market prices by shaping market participants' beliefs, even when it contains no additional informational content. Using the

South Korean horse racing market as our empirical setting, we compare horses positioned on opposite sides of ranking thresholds despite having virtually identical race times. Because ranking is a monotonic discretization of race times, it contains no information beyond the underlying times themselves. We then test whether betting prices on these horses' future performance differ significantly. The unique institutional features of this setting allow us to identify a baseline sample in which changes in ranking per se have no effect on the probability of future outcomes. We can therefore isolate the effects of perceptual bias arising from discretized information on betting prices from other mechanisms—an empirical challenge that is rarely surmountable in conventional financial settings.

Our findings indicate that discretized information can systematically mislead market participants. Bettors consistently overestimate the winning probabilities of horses just above a ranking threshold relative to those just below it, despite no significant difference in actual performance. The disparity in win-bet shares reaches 0.46 percentage points, equivalent to over one-tenth of the average win-bet share of 4.5 percent. This perceptual bias therefore has economically meaningful consequences: relative to the baseline expected loss of 20 percent implied by the house take, it increases the expected loss on bets to 27.2 percent, an increase of as much as 36 percent. A complementary calibration exercise shows that the effect of discretization-driven misperception remains important even in settings where rankings also have real effects.

Turning to the mechanisms underlying these empirical patterns, we find that the bias persists even when information-acquisition costs are minimal, suggesting that such costs are unlikely to be the primary driver. The ranking effect is concentrated in races with smaller wagers and is absent or substantially attenuated in races with larger wagers. Moreover, the effect is amplified in settings in which bettors are more likely to rely on heuristics such as coarse rankings when making decisions. Taken together, these results are most consistent with explanations based on rational inattention and heuristic decision-making.

Our paper has important implications for understanding investor rationality. The findings demonstrate that investors operate under cognitive constraints—a feature largely absent from standard rational models—and that these constraints make the format in which information is disclosed consequential. Accordingly, our results inform the design of disclosure policies by regulators and financial intermediaries. Information disclosure is central to regulatory efforts aimed at improving market quality and stability (Greenstone *et al.*, 2006; Goldstein and Yang, 2017). Although simplifying information can reduce cognitive burdens and facilitate decision-making, it may also distort investor perceptions and generate inefficient outcomes. Moreover, when simplification induces judgment errors, other market participants may respond strategically to exploit these biases (Almeida

et al., 2016; Cohen *et al.*, 2024), with adverse welfare consequences: strategic responses to misguided investors can distort corporate policies and exacerbate capital misallocation.

References

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999) High Wage Workers and High Wage Firms, *Econometrica*, **67**, 251–333.

Adelino, M., Cunha, I. and Ferreira, M. A. (2017) The Economic Effects of Public Financing: Evidence from Municipal Bond Ratings Recalibration, *The Review of Financial Studies*, **30**, 3223–3268.

Ali, M. M. (1977) Probability and Utility Estimates for Racetrack Bettors, *Journal of Political Economy*, **85**, 803–815.

Almeida, H., Cunha, I., Ferreira, M. A. and Restrepo, F. (2017) The Real Effects of Credit Ratings: The Sovereign Ceiling Channel, *The Journal of Finance*, **72**, 249–290.

Almeida, H., Fos, V. and Kronlund, M. (2016) The real effects of share repurchases, *Journal of Financial Economics*, **119**, 168–185.

Baker, S. R., Balthrop, J., Johnson, M. J., Kotter, J. and Pisciotta, K. (2024) Gambling Away Stability: Sports Betting's Impact on Vulnerable Households, *NBER Working Paper*.

Barberis, N. and Shleifer, A. (2003) Style investing, *Journal of financial Economics*, **68**, 161–199.

Barberis, N., Shleifer, A. and Wurgler, J. (2005) Comovement, *Journal of financial economics*, **75**, 283–317.

Ben-David, I., Li, J., Rossi, A. and Song, Y. (2021) What Do Mutual Fund Investors Really Care About?, *The Review of Financial Studies*, **35**, 1723–1774.

Berk, J. and Green, R. (2004) Mutual fund flows and performance in rational markets, *Journal of Political Economy*, **112**, 1269–1295.

Bond, P. and Goldstein, I. (2015) Government intervention and information aggregation by prices, *The Journal of Finance*, **70**, 2777–2812.

Boot, A. W., Milbourn, T. T. and Schmeits, A. (2006) Credit ratings as coordination mechanisms, *The Review of Financial Studies*, **19**, 81–118.

Bouvard, M., Chaigneau, P. and Motta, A. D. (2015) Transparency in the financial system: Rollover risk and crises, *The Journal of Finance*, **70**, 1805–1837.

Calonico, S., Cattaneo, M. D. and Titiunik, R. (2014) Robust nonparametric confidence intervals for regression-discontinuity designs, *Econometrica*, **82**, 2295–2326.

Chen, Z., Lookman, A. A., Schürhoff, N. and Seppi, D. J. (2014) Rating-based investment practices and bond market segmentation, *The Review of Asset Pricing Studies*, **4**, 162–205.

Chevalier, J. and Ellison, G. (1997) Risk taking by mutual funds as a response to incentives, *Journal of political economy*, **105**, 1167–1200.

Cohen, L., Kim, D. S. and So, E. C. (2024) Box jumping: Portfolio recompositions to achieve higher morningstar ratings, *Available at SSRN*.

Cornaggia, J., Cornaggia, K. J. and Israelsen, R. D. (2018) Credit Ratings and the Cost of Municipal Financing, *The Review of Financial Studies*, **31**, 2038–2079.

Coval, J. D. and Shumway, T. (2005) Do behavioral biases affect prices?, *The Journal of Finance*, **60**, 1–34.

DellaVigna, S. and Pollet, J. M. (2009) Investor inattention and friday earnings announcements, *The journal of finance*, **64**, 709–749.

Diamond, D. W. and Verrecchia, R. E. (1991) Disclosure, liquidity, and the cost of capital, *The journal of Finance*, **46**, 1325–1359.

Dorn, A. J., Dorn, D. and Sengmueller, P. (2015) Trading as Gambling, *Management Science*, **61**, 2376–2393.

Dorn, D. and Sengmueller, P. (2009) Trading as entertainment?, *Management science*, **55**, 591–603.

Dye, R. A. (1985) Disclosure of nonproprietary information, *Journal of accounting research*, pp. 123–145.

Ellul, A., Jotikasthira, C. and Lundblad, C. T. (2011) Regulatory pressure and fire sales in the corporate bond market, *Journal of Financial Economics*, **101**, 596–620.

Evans, R. B. and Sun, Y. (2020) Models or Stars: The Role of Asset Pricing Models and Heuristics in Investor Risk Adjustment, *The Review of Financial Studies*, **34**, 67–107.

Faria-e Castro, M., Martinez, J. and Philippon, T. (2017) Runs versus lemons: Information disclosure and fiscal capacity, *The Review of Economic Studies*, **84**, 1683–1707.

Gantchev, N., Giannetti, M. and Li, R. (2024) Sustainability or performance? ratings and fund managers' incentives, *Journal of Financial Economics*, **155**, 103831.

Gao, X. and Lin, T.-C. (2015) Do Individual Investors Treat Trading as a Fun and Exciting Gambling Activity? Evidence from Repeated Natural Experiments, *The Review of Financial Studies*, **28**, 2128–2166.

Gigerenzer, G. and Gaissmaier, W. (2011) Heuristic decision making, *Annual review of psychology*, **62**, 451–482.

Goel, A. M. and Thakor, A. V. (2015) Information reliability and welfare: A theory of coarse credit ratings, *Journal of Financial Economics*, **115**, 541–557.

Goldstein, I. and Huang, C. (2020) Credit Rating Inflation and Firms' Investments, *The Journal of Finance*.

Goldstein, I. and Leitner, Y. (2018) Stress tests and information disclosure, *Journal of Economic Theory*, **177**, 34–69.

Goldstein, I. and Leitner, Y. (2020) Stress tests disclosure: Theory, practice, and new perspectives, *Handbook of financial stress testing*, pp. 208–223.

Goldstein, I. and Yang, L. (2017) Information disclosure in financial markets, *Annual Review of Financial Economics*, **9**, 101–125.

Gredil, O. R., Kapadia, N. and Lee, J. H. (2022) On the information content of credit ratings and market-based measures of default risk, *Journal of Financial Economics*, **146**, 172–204.

Greenstone, M., Oyer, P. and Vissing-Jorgensen, A. (2006) Mandated disclosure, stock returns, and the 1964 securities acts amendments, *The Quarterly Journal of Economics*, **121**, 399–460.

Grinblatt, M. and Keloharju, M. (2009) Sensation seeking, overconfidence, and trading activity, *The Journal of finance*, **64**, 549–578.

Grossman, S. J. and Stiglitz, J. E. (1980) On the Impossibility of Informationally Efficient Markets, *The American Economic Review*, **70**, 393 408.

Guercio, D. D. and Tkac, P. A. (2008) Star Power: The Effect of Morningstar Ratings on Mutual Fund, *Journal of Financial and Quantitative Analysis*, **43**, 907–936.

Hartzmark, S. M. and Sussman, A. B. (2019) Do investors value sustainability? a natural experiment examining ranking and fund flows, *The Journal of Finance*, **74**, 2789–2837.

Hirshleifer, D., Lim, S. S. and Teoh, S. H. (2009) Driven to distraction: Extraneous events and underreaction to earnings news, *The journal of finance*, **64**, 2289–2325.

Huang, C., L, F. and Wen, X. (2020) Star Ratings and the Incentives of Mutual Funds, *The Journal of Finance*, **75**, 1715–1765.

Huang, X. (2019) Mark twain's cat: Investment experience, categorical thinking, and stock selection, *Journal of Financial Economics*, **131**, 404–432.

Kahneman, D. and Tversky, A. (1979) Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, **47**, 263–292.

Kisgen, D. J. (2006) Credit Ratings and Capital Structure, *The Journal of Finance*, **61**, 1035–1072.

Kliger, D. and Sarig, O. (2000) The information value of bond ratings, *The journal of finance*, **55**, 2879–2902.

Kumar, A. (2009) Who gambles in the stock market?, *The journal of finance*, **64**, 1889–1933.

Leitner, Y. and Williams, B. (2023) Model secrecy and stress tests, *The Journal of Finance*, **78**, 1055–1095.

Maćkowiak, B., Matějka, F. and Wiederholt, M. (2023) Rational inattention: A review, *Journal of Economic Literature*, **61**, 226–273.

Macrae, C. N. and Bodenhausen, G. V. (2000) Social cognition: Thinking categorically about others, *Annual review of psychology*, **51**, 93–120.

Manso, G. (2013) Feedback effects of credit ratings, *Journal of Financial Economics*, **109**, 535–548.

Monnet, C. and Quintin, E. (2017) Rational opacity, *The review of financial studies*, **30**, 4317–4348.

Morris, S. and Shin, H. S. (2002) Social value of public information, *american economic review*, **92**, 1521–1534.

Moskowitz, T. J. (2021) Asset pricing and sports betting, *The Journal of Finance*, **76**, 3153–3209.

Moskowitz, T. J. and Vasudevan, K. (2022) Betting Without Beta, *Available at SSRN 3845505*.

Mullainathan, S. (2002) Thinking through categories, Tech. rep., Working Paper, Harvard University.

Mullainathan, S., Schwartzstein, J. and Shleifer, A. (2008) Coarse thinking and persuasion, *The Quarterly journal of economics*, **123**, 577–619.

Palacios-Huerta, I. (2023) The beautiful dataset, *Available at SSRN 4665889*.

Peng, L. and Xiong, W. (2006) Investor attention, overconfidence and category learning, *Journal of Financial Economics*, **80**, 563–602.

Purnanandam, A. and Wirth, A. (2023) Can Credit Rating Affect Credit Risk? Causal Evidence from an Online Lending Marketplace, *SSRN Electronic Journal*.

Quigley, D. and Walther, A. (2024) Inside and outside information, *The Journal of Finance*, **79**, 2667–2714.

Reuter, J. and Zitzewitz, E. (2021) How Much Does Size Erode Mutual Fund Performance? A Regression Discontinuity Approach*, *Review of Finance*, **25**, 1395–1432.

Rosett, R. N. (1965) Gambling and Rationality, *Journal of Political Economy*, **73**, 595–607.

Shapiro, J. and Skeie, D. (2015) Information management in banking crises, *The Review of Financial Studies*, **28**, 2322–2363.

Shleifer, A. and Vishny, R. W. (1997) The limits of arbitrage, *The Journal of Finance*, **52**, 35–55.

Shue, K., Townsend, R. and Wang, C. (2024) Categorical thinking about interest rates, Tech. rep., CESifo Working Paper.

Sims, C. A. (2003) Implications of rational inattention, *Journal of Monetary Economics*, **50**, 665–690.

Snowberg, E. and Wolfers, J. (2008) Handbook of Sports and Lottery Markets, *Part III: Favorite-Longshot Bias in the Win Market*, pp. 103–136.

Snowberg, E. and Wolfers, J. (2010) Explaining the Favorite-Long Shot Bias: Is it Risk-Love or Misperceptions?, *Journal of Political Economy*, **118**, 723–746.

Snyder, W. W. (1978) Horse Racing: Testing the Efficient Market Model, *The Journal of Finance*, **33**, 1109–1118.

Sufi, A. (2007) The Real Effects of Debt Certification: Evidence from the Introduction of Bank Loan Ratings, *Review of Financial Studies*, **22**, 1659–1691.

Thaler, R. H. and Ziemba, W. T. (1988) Anomalies: Parimutuel Betting Markets: Racetracks and Lotteries, *Journal of Economic Perspectives*, **2**, 161–174.

Tversky, A. and Kahneman, D. (1974) Judgment under Uncertainty: Heuristics and Biases, *Science*, **185**, 1124–1131.

Tversky, A. and Kahneman, D. (1992) Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and uncertainty*, **5**, 297–323.

Verrecchia, R. E. (1983) Discretionary disclosure, *Journal of accounting and economics*, **5**, 179–194.

Wu, G. and Gonzalez, R. (1996) Curvature of the probability weighting function, *Management science*, **42**, 1676–1690.

Table 1: Summary Statistics: Full Sample

This table reports summary statistics for the full sample. Observations with missing win betting odds, as well as those involving horses that were scratched or did not finish their subsequent race, are excluded. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." In this dataset, a ranking of n in race i generates two observations based on sign. That is, a horse ranked n, unless it is at the top or bottom of the ranking, appears once as the higher-ranked horse in the pair $(n, n+1)$ and once as the lower-ranked horse in the pair $(n-1, n)$. The sample period spans January 1993 to October 2024. The exchange rate is typically 1,200 KRW per 1 USD. For definitions of all variables, see Appendix A.

	Obs.	Mean	Median	Std.	P25	P75
Horse Characteristics						
Gelding	1,220,324	0.23	0.00	0.42	0.00	0.00
Male	1,220,324	0.32	0.00	0.47	0.00	1.00
Female	1,220,324	0.45	0.00	0.50	0.00	1.00
Age	1,220,324	3.86	4.00	1.38	3.00	4.00
Days from Last Race	1,220,324	34.12	28.00	32.65	21.00	35.00
Change in Tier	1,220,324	0.08	0.00	0.27	0.00	0.00
Higher-Ranked in Last Race	1,220,324	0.50	1.00	0.50	0.00	1.00
Last Race Time Gap	1,220,324	−0.01	0.00	0.80	−0.30	0.20
Last Race Distance Gap ≤ Nose	1,220,324	0.05	0.00	0.21	0.00	0.00
1 + Win Betting Odds	1,220,324	69.85	16.90	554.65	6.70	44.70
1 + Place Betting Odds	1,220,090	27.89	3.40	465.69	1.90	7.10
Win	1,220,324	0.09	0.00	0.29	0.00	0.00
Place	1,220,324	0.28	0.00	0.45	0.00	1.00
Race Characteristics						
Age-Sex Based Handicapping	1,220,324	0.30	0.00	0.46	0.00	1.00
Award Based Handicapping	1,220,324	0.36	0.00	0.48	0.00	1.00
Discretionary Handicapping	1,220,324	0.34	0.00	0.47	0.00	1.00
Race Distance (meter)	1,220,324	1301.87	1200.00	325.27	1000.00	1400.00
N. of Horses	1,220,324	11.01	11.00	1.85	10.00	12.00
Seoul Stadium	1,220,324	0.52	1.00	0.50	0.00	1.00
Jeju Stadium	1,220,324	0.27	0.00	0.45	0.00	1.00
Busan Stadium	1,220,324	0.21	0.00	0.40	0.00	0.00
Avg. Bet Size (1K KRW)	1,019,897	17.80	15.13	7.27	12.68	21.05
Sales (1B KRW)	1,170,704	3.01	2.93	2.12	1.46	4.34

Table 2: Summary Statistics: Regression Sample

This table reports summary statistics for the regression sample. This sample is constructed by removing cases from the full sample in which the competitive environment faced by the higher- and lower-ranked horses within a pair is likely different (see Subsection 3.2). For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." In this dataset, a ranking of n in race i generates two observations based on sign. That is, a horse ranked n, unless it is at the top or bottom of the ranking, appears once as the higher-ranked horse in the pair $(n, n+1)$ and once as the lower-ranked horse in the pair $(n-1, n)$. The sample period spans January 1993 to October 2024. The exchange rate is typically 1,200 KRW per 1 USD. For definitions of all variables, see Appendix A.

	Obs.	Mean	Median	Std.	P25	P75
Horse Characteristics						
Gelding	320,344	0.19	0.00	0.39	0.00	0.00
Male	320,344	0.29	0.00	0.45	0.00	1.00
Female	320,344	0.52	1.00	0.50	0.00	1.00
Age	320,344	3.64	3.00	1.22	3.00	4.00
Days from Last Race	320,344	32.03	28.00	22.89	21.00	35.00
Change in Tier	320,344	0.00	0.00	0.00	0.00	0.00
Higher-Ranked in Last Race	320,344	0.51	1.00	0.50	0.00	1.00
Last Race Time Gap	320,344	−0.02	0.00	0.79	−0.30	0.30
Last Race Distance Gap ≤ Nose	320,344	0.04	0.00	0.20	0.00	0.00
1 + Win Betting Odds	320,344	104.71	40.10	586.45	16.70	88.30
1 + Place Betting Odds	320,216	29.42	6.40	441.09	3.30	12.60
Win	320,344	0.04	0.00	0.20	0.00	0.00
Place	320,344	0.16	0.00	0.37	0.00	0.00
Race Characteristics						
Age-Sex Based Handicapping	320,344	0.50	1.00	0.50	0.00	1.00
Award Based Handicapping	320,344	0.50	0.00	0.50	0.00	1.00
Race Distance (meter)	320,344	1233.23	1200.00	278.64	1000.00	1400.00
N. of Horses	320,344	11.18	11.00	1.90	10.00	12.00
Seoul Stadium	320,344	0.62	1.00	0.49	0.00	1.00
Jeju Stadium	320,344	0.20	0.00	0.40	0.00	0.00
Busan Stadium	320,344	0.18	0.00	0.39	0.00	0.00
Avg. Bet Size (1K KRW)	260,110	19.30	16.46	7.31	14.33	23.23
Sales (1B KRW)	310,996	3.08	2.95	2.14	1.51	4.49

Table 3: Univariate Analysis

This table compares the characteristics of horses, race conditions, and gate assignments for horses designated as higher- and lower-ranked in their previous race, and examines whether there are any significant differences across these dimensions between the two groups. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." In this dataset, a ranking of n in race i generates two observations based on sign. That is, a horse ranked n, unless it is at the top or bottom of the ranking, appears once as the higher-ranked horse in the pair $(n, n+1)$ and once as the lower-ranked horse in the pair $(n-1, n)$. We restrict our attention to pairs with a distance gap of less than or equal to a nose (approximately three inches) in their last races. The sample period spans January 1993 to October 2024. For definitions of all variables, see Appendix A.

	Lower-Ranked			Higher-Ranked			Lower − Higher	
	Obs	Mean	Std.	Obs	Mean	Std.	Diff.	t-stat
Horse Characteristics								
Gelding	6,541	0.19	0.39	6,555	0.19	0.39	−0.00	(−0.12)
Male	6,541	0.29	0.45	6,555	0.30	0.46	−0.01	(−1.26)
Female	6,541	0.52	0.50	6,555	0.51	0.50	0.01	(1.25)
Age	6,541	3.68	1.18	6,555	3.67	1.16	0.00	(0.19)
Horse Weight	6,541	426.23	71.90	6,555	426.71	71.68	−0.48	(−0.38)
ΔHorse Weight	6,541	−0.35	6.92	6,555	−0.19	6.73	−0.16	(−1.34)
Days from Last Race	6,541	31.45	20.77	6,555	31.36	21.43	0.10	(0.26)
Race Characteristics								
Race Distance (meter)	6,541	1253.35	270.69	6,555	1250.59	268.13	2.77	(0.59)
N. of Horses	6,541	11.29	1.92	6,555	11.30	1.92	−0.01	(−0.25)
Sunny Day	6,541	0.73	0.45	6,555	0.73	0.44	−0.01	(−0.86)
Gate Assignment								
Gate 1	6,541	0.09	0.28	6,555	0.09	0.29	−0.01	(−1.43)
Gate 2	6,541	0.08	0.26	6,555	0.07	0.26	0.00	(1.01)
Gate 3	6,541	0.05	0.21	6,555	0.05	0.22	−0.00	(−0.86)
Gate 4	6,541	0.04	0.20	6,555	0.04	0.20	−0.00	(−0.06)
Gate 5	6,541	0.02	0.13	6,555	0.02	0.14	−0.00	(−0.69)
Gate 6	6,541	0.01	0.12	6,555	0.01	0.11	0.00	(0.62)
Gate 7	6,541	0.10	0.30	6,555	0.09	0.29	0.01	(1.57)
Gate 8	6,541	0.09	0.29	6,555	0.09	0.28	0.01	(1.05)
Gate 9	6,541	0.09	0.29	6,555	0.09	0.29	0.00	(0.94)
Gate 10	6,541	0.09	0.29	6,555	0.10	0.29	−0.01	(−1.29)
Gate 11	6,541	0.10	0.29	6,555	0.10	0.30	−0.00	(−0.32)
Gate 12	6,541	0.09	0.28	6,555	0.09	0.28	−0.00	(−0.15)
Gate 13	6,541	0.08	0.28	6,555	0.08	0.28	0.00	(0.20)
Gate 14	6,541	0.07	0.26	6,555	0.08	0.27	−0.00	(−0.79)

Table 4: Do Higher-Ranked Horses Receive Lower Handicap Weights or Better Jockeys in Their Next Race?

This table compares the handicap weights and jockey abilities of higher- and lower-ranked horses in their next race. Panel A analyzes the handicap weights assigned to each horse, while Panel B examines the ability of the jockey riding each horse, estimated using Equation (10). For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." In each panel, columns (1) and (2) focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Columns (3) and (4) broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race-time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Columns (2) and (4) also include pair fixed effects. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using standard errors clustered at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

(A) Handicap Weights

Dep. Var.:	*Handicap Weight*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.0139	0.0363	0.0110	0.0169
	(0.51)	(1.25)	(0.78)	(1.15)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	1.1s	1.1s
Obs	13,096	11,095	285,714	239,741
Adj. R^2	-0.00	0.20	0.00	0.21
Mean of Dep. Var.	53.94	53.95	53.91	53.91

(B) Jockey Ability

Dep. Var.:	*Jockey Ability*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.0002	0.0038	0.0024	0.0015
	(0.02)	(0.42)	(0.54)	(0.32)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	1.1s	1.1s
Obs	13,089	11,081	285,494	239,387
Adj. R^2	-0.00	0.40	0.00	0.43
Mean of Dep. Var.	6.13	6.15	6.12	6.13

Table 5: Are Higher-Ranked Horses More Likely to Win or Place in Their Next Race?

This table compares the winning and placing probabilities of higher- and lower-ranked horses in their next race. Panel A analyzes differences in the likelihood of winning, while Panel B focuses on differences in the likelihood of placing. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." In each panel, columns (1) and (2) focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Columns (3) and (4) broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race-time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Columns (2) and (4) also include pair fixed effects. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using standard errors clustered at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

(A) Win

Dep. Var.:	*Win (%)*			
	(1)	(2)	(3)	(4)
Higher-Ranked	−0.3456	−0.3365	0.0421	0.0809
	(−0.97)	(−0.88)	(0.25)	(0.44)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	1.4s	1.4s
Obs	13,096	11,095	297,825	249,288
Adj. R^2	-0.00	0.03	0.00	0.01
Mean of Dep. Var.	4.49	4.41	4.10	4.10

(B) Place

Dep. Var.:	*Place (%)*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.5405	0.5288	0.1342	0.2715
	(0.82)	(0.74)	(0.45)	(0.84)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	1.7s	1.7s
Obs	13,096	11,095	304,442	254,317
Adj. R^2	-0.00	0.06	0.00	0.06
Mean of Dep. Var.	18.62	18.88	16.55	16.71

Table 6: Do Bettors Perceive Higher-Ranked Horses as More Likely to Win or Place in Their Next Race?

This table compares the perceived winning and placing probabilities of higher- and lower-ranked horses in their next race. Panel A analyzes differences in the shares of bets placed on higher- and lower-ranked horses to win, while Panel B focuses on differences in the shares of bets placed on them to place in their next race. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." In each panel, columns (1) and (2) focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Columns (3) and (4) broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race-time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Columns (2) and (4) also include pair fixed effects. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using standard errors clustered at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

(A) Winning Bet

Dep. Var.:	*Bet to Win (%)*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.3908***	0.4609***	0.2556***	0.3023***
	(3.86)	(4.29)	(5.06)	(5.62)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	1.2s	1.2s
Obs	13,096	11,095	291,124	244,058
Adj. R^2	0.00	0.16	0.01	0.15
Mean of Dep. Var.	4.56	4.53	4.20	4.18

(B) Placing Bet

Dep. Var.:	*Bet to Place (%)*			
	(1)	(2)	(3)	(4)
Higher-Ranked	0.3466***	0.3984***	0.2872***	0.3183***
	(4.30)	(4.66)	(6.98)	(7.26)
Pair FEs		✓		✓
Higher×Cubic			✓	✓
Sample: $\Delta \leq$	Nose	Nose	1.1s	1.1s
Obs	13,090	11,085	285,613	239,599
Adj. R^2	0.00	0.23	0.01	0.21
Mean of Dep. Var.	6.38	6.39	5.99	5.99

Table 7: Robustness to Horse Pairs Competing in Different Subsequent Races

This table revisits the analyses of Tables 5 and 6 using the subsample of horse pairs in which the two horses compete in different subsequent races, to avoid the mechanical interdependence between outcomes of horses competing in the same subsequent race. Panel (A) presents results for actual winning and placing probabilities; Panel (B) presents results for win and place bet shares. In each panel, Columns (1) and (3) focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Columns (2) and (4) broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race-time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). All specifications include pair fixed effects and cluster standard errors at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

(A) Actual Probabilities

Dep. Var.:	*Win (%)*		*Place (%)*	
	(1)	(2)	(3)	(4)
Higher-Ranked	−0.4485	0.0007	0.2446	0.1854
	(−1.11)	(0.00)	(0.32)	(0.52)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	1.3s	Nose	1.4s
Obs	9,810	218,496	9,810	220,986
Adj. R^2	0.04	0.02	0.07	0.06
Mean of Dep. Var.	4.32	4.08	18.63	16.68

(B) Perceived Probabilities

Dep. Var.:	*Bet to Win (%)*		*Bet to Place (%)*	
	(1)	(2)	(3)	(4)
Higher-Ranked	0.4600***	0.2571***	0.4009***	0.2784***
	(4.00)	(4.52)	(4.35)	(5.93)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	1.3s	Nose	1.1s
Obs	9,810	218,496	9,800	212,386
Adj. R^2	0.14	0.14	0.22	0.21
Mean of Dep. Var.	4.48	4.15	6.37	5.96

Table 8: Are the Effects Weaker in Recent Periods?

This table estimates the effect of ranking information on bettors' perceived probabilities of horses winning or placing in their subsequent race, as analyzed in Table 6, separately by decade. The analysis is conducted for each of the four decades in the sample period: the 1990s, 2000s, 2010s, and 2020s. The variables 1990s, 2000s, 2010s, and 2020s are indicator variables equal to one if the race occurred in a year within the corresponding decade. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." Columns (1) and (3) focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Columns (2) and (4) broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race-time gap interacted with an indicator for the higher-ranked horse and the decade indicators (see Equation (12)). All regression specifications include pair fixed interacted with decade fixed effects. The row H_0: 1990s − 2020s reports the difference between the Higher-Ranked × 1990s and Higher-Ranked × 2020s coefficients, together with the associated t-statistics. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using standard errors clustered at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

Dep. Var.:	*Bet to Win (%)*		*Bet to Place (%)*	
	(1)	(2)	(3)	(4)
Higher-Ranked × 1990s	0.0883	0.0786	0.0648	0.0764
	(0.39)	(0.68)	(0.34)	(0.77)
Higher-Ranked × 2000s	0.5242***	0.4193***	0.4376***	0.3987***
	(2.88)	(4.68)	(3.09)	(5.58)
Higher-Ranked × 2010s	0.6026***	0.3068***	0.5384***	0.3396***
	(3.33)	(3.44)	(3.78)	(4.68)
Higher-Ranked × 2020s	0.3923	0.1216	0.4017	0.3408**
	(1.08)	(0.58)	(1.30)	(1.96)
H_0: 1990s −2020s	-0.3040	-0.0430	-0.3368	-0.2644
	(-0.71)	(-0.18)	(-0.93)	(-1.32)
Pair FE ×Decade	✓	✓	✓	✓
Higher×Cubic ×Decade		✓		✓
Sample: $\Delta \leq$	Nose	1.2s	Nose	1.1s
Obs	11,055	242,985	11,045	238,552
Adj. R^2	0.16	0.15	0.23	0.21
Mean of Dep. Var.	4.53	4.18	6.39	5.99

Table 9: Are the Effects Weaker Among Bets with Higher Stakes?

This table estimates the effect of ranking information on bettors' perceived probabilities of horses winning or placing in their subsequent race, as analyzed in Table 6. Within each racecourse-year, race dates are assigned to two groups (Low and High) based on the average bet size per bet. The average bet size per bet is computed as total purchase value divided by the total number of bets. The variable Low Bet Size is an indicator equal to one if the average bet size on a given race date falls within the bottom 50 percent of the distribution for that racecourse in that year, and High Bet Size equals one if it falls in the top 50 percent. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." Columns (1) and (3) focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Columns (2) and (4) broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race-time gap interacted with $D_{r,n}$, an indicator for the higher-ranked horse and the bet-size group indicators (see Equation (14)). All regression specifications include pair fixed effects interacted with bet-size-group fixed effects. The row H_0: Low − High reports the difference between the Higher-Ranked × Low Bet Size and Higher-Ranked × High Bet Size coefficients, together with the associated t-statistics. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using standard errors clustered at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

Dep. Var.:	*Bet to Win (%)*		*Bet to Place (%)*	
	(1)	(2)	(3)	(4)
Higher-Ranked ×Low Bet Size	0.7283***	0.4451***	0.4942***	0.4130***
	(3.86)	(4.62)	(3.30)	(5.22)
Higher-Ranked ×High Bet Size	0.2282	0.2479**	0.3000*	0.3054***
	(1.07)	(2.34)	(1.84)	(3.69)
H_0: Low −High	0.5001*	0.1972	0.1942	0.1077
	(1.75)	(1.38)	(0.88)	(0.94)
Pair FE ×Bet Size	✓	✓	✓	✓
Higher×Cubic×Bet Size		✓		✓
Sample: $\Delta \leq$	Nose	1.2s	Nose	1.1s
Obs	6,590	142,079	6,588	139,598
Adj. R^2	0.16	0.15	0.24	0.22
Mean of Dep. Var.	4.64	4.29	6.49	6.13

Table 10: Real Effects, Bettors' Perceptions, and Discrepancy: Top-5 Finishers

This table examines real effects, bettors' perceptions, and the discrepancy between them for horse pairs in which the higher-ranked horse finished 5th or better in their previous race. Panel A tests whether the horse changes competition tier, actual winning probability, or actual placing probability in its next race. Panel B tests differences in the shares of bets placed on the horse to win and to place. Panel C reports the discrepancy between the share of bets placed on a horse and its actual winning or placing probability, where the share of place bets is scaled by the number of place slots in each race. *Discretization Effect* in Panel C reports coefficients estimated on the subsample of pairs in which the higher-ranked horse ranked exactly 6th in their previous race, using the scaled bet-share outcomes as dependent variables. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." In each panel, odd-numbered columns focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Even-numbered columns broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the race-time gap by including a cubic polynomial interacted with D, an indicator variable denoting higher-ranked horses (see Equation (9)). The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using standard errors clustered at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

(A) Real Effects

Dep. Var.:	*Tier Change (%)*		*Win (%)*		*Place (%)*	
	(1)	(2)	(3)	(4)	(5)	(6)
Higher-Ranked	7.3706***	7.4260***	−1.5158***	−1.5113***	−3.6158***	−3.2886***
	(21.46)	(41.03)	(−3.30)	(−6.34)	(−5.40)	(−9.59)
Pair FEs	✓	✓	✓	✓	✓	✓
Higher×Cubic		✓		✓		✓
Sample: $\Delta \leq$	Nose	1.0s	Nose	1.1s	Nose	1.1s
Obs	19,185	384,534	19,185	388,994	19,185	388,994
Adj. R^2	0.37	0.35	0.04	0.04	0.08	0.08
Mean of Dep. Var.	10.04	11.32	11.94	13.33	36.73	38.20

(B) Bettors' Perceptions

Dep. Var.:	*Bet to Win (%)*		*Bet to Place (%)*	
	(1)	(2)	(3)	(4)
Higher-Ranked	−0.7186***	−0.6770***	−0.3526***	−0.3182***
	(−4.38)	(−7.80)	(−4.02)	(−7.11)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	0.9s	Nose	1.0s
Obs	19,185	375,771	19,181	384,442
Adj. R^2	0.22	0.25	0.22	0.26
Mean of Dep. Var.	12.64	13.32	12.27	12.58

(C) Discrepancy

Dep. Var.:	Perceived −Actual			
	Win (%)		Place (%)	
	(1)	(2)	(3)	(4)
Higher-Ranked	0.7972*	0.7697***	2.6534***	2.3445***
	(1.83)	(3.48)	(4.23)	(7.32)
Pair FEs	✓	✓	✓	✓
Higher×Cubic		✓		✓
Sample: $\Delta \leq$	Nose	1.2s	Nose	1.1s
Discretization Effect	0.5468**	0.5837***	2.0464***	1.7720***
Obs	19,185	394,660	19,181	388,902
Adj. R^2	-0.01	-0.02	0.00	-0.00
Mean of Dep. Var.	0.70	0.06	-0.56	-1.09

Appendix

A Variable Definitions

A.1 Horse Characteristics

- **Gelding**: Indicator variable equal to one for gelded horses.
- **Male**: Indicator variable equal to one for male horses.
- **Female**: Indicator variable equal to one for female horses.
- **Age**: Age of the horse.
- **Horse Weight**: Horse's weight on the race date.
- **ΔHorse Weight**: Change in a horse's weight on the race date relative to its weight on the previous race date.
- **Days from Last Race**: Number of days between the race date and the previous race date.
- **Change in Tier**: Indicator variable equal to one if there is a change in racing tier between the race date and the previous race date.
- **Higher-Ranked in Last Race**: Indicator variable equal to one for the horse with the higher rank in the pair in the previous race.
- **Last Race Time Gap**: Difference in race time between the horse and its pair in the previous race.
- **Last Race Distance Gap $\leq$ Nose**: Indicator variable equal to one for pairs with a distance gap of less than or equal to a nose (approximately three inches) in the previous race.
- **Win Betting Odds**: Odds for the horse to win the race.
- **Place Betting Odds**: Odds for the horse to place in the race.

A.2 Race Characteristics

- **Age-Sex Based Handicapping**: Indicator variable equal to one for races where handicap weights are determined by age and sex.
- **Award Based Handicapping**: Indicator variable equal to one for races where handicap weights are determined by each horse's accumulated awards.
- **Discretionary Handicapping**: Indicator variable equal to one for races where handicap weights are determined at the handicapper's discretion.
- **Race Distance**: Distance of the track in a given race.
- **N. of Horses**: Total number of horses participating in a race.
- **Seoul Racecourse**: Indicator variable equal to one for races held at the Seoul Racecourse.
- **Jeju Racecourse**: Indicator variable equal to one for races held at the Jeju Racecourse.
- **Busan Racecourse**: Indicator variable equal to one for races held at the Busan Racecourse.
- **Avg. Bet Size**: Average bet size for races held at a given racecourse on a specific race date.
- **Sales**: Total betting turnover (i.e., total sales) for a race.
- **Sunny Day**: Indicator variable equal to one if weather conditions are sunny on the race date.
- **Gate 1–Gate 14**: Indicator variables for horses starting from gates 1 through 14, respectively.

A.3 Outcome Variables

- **Handicap Weight**: Weight assigned to a horse as part of the handicapping system.
- **Jockey Ability**: Ability level of the jockey riding a horse, estimated using Equation (10).

- **Win (%)**: Indicator variable equal to one for horses winning the race, multiplied by 100.
- **Place (%)**: Indicator variable equal to one for horses placing in the race, multiplied by 100.
- **Bet to Win (%)**: Share of bets on a horse to win, multiplied by 100.
- **Bet to Place (%)**: Share of bets on a horse to place, multiplied by 100.

Online Appendix for

"Does Discretized Information Induce Misperception? Evidence from a Sports Betting Market"

A Robustness Check

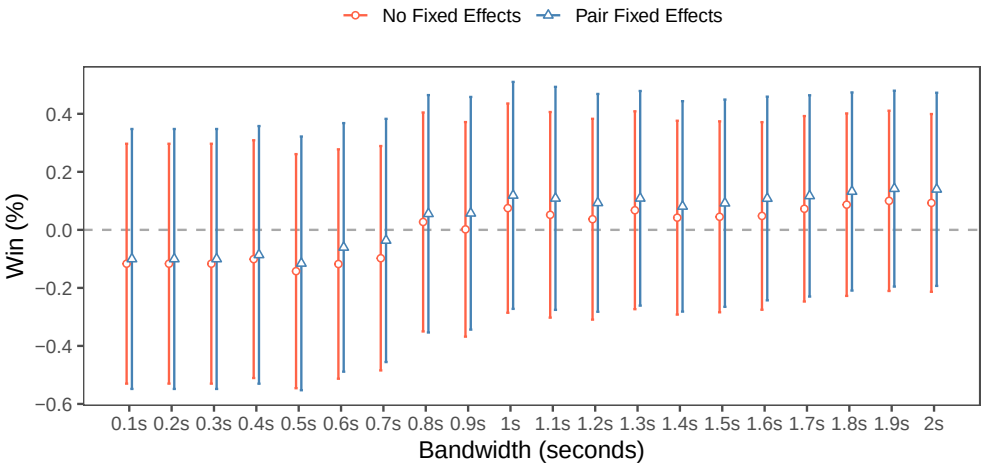


Figure O1: Bandwidth Sensitivity: Win

This figure presents a robustness test for Panel (A) of Table 5 by examining the sensitivity of the estimated effect of being higher-ranked on the probability of winning the next race to alternative bandwidth choices. The x-axis shows the bandwidth used to restrict the regression sample around the ranking threshold. The y-axis reports the estimated coefficient controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Circles and triangles correspond to the same cubic polynomial specification estimated without and with pair fixed effects, respectively. The vertical bars denote 95 percent confidence intervals. Standard errors are clustered at the horse pair level.

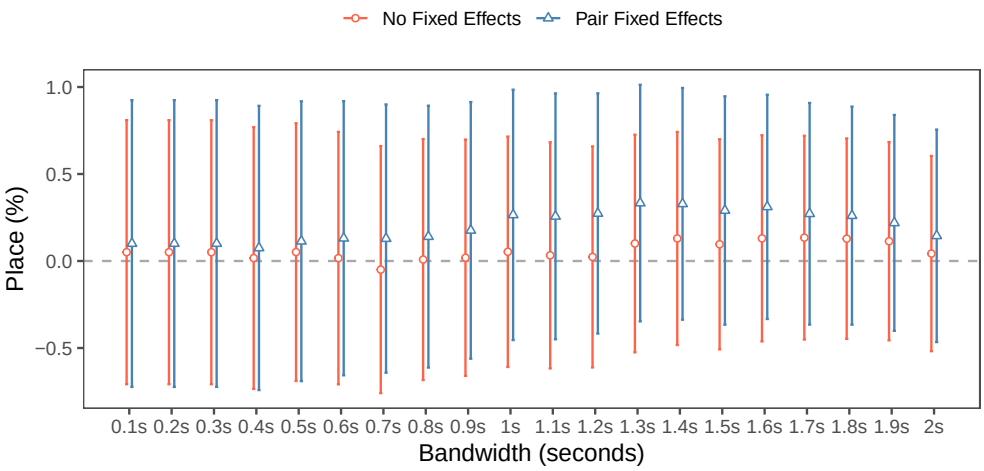


Figure O2: Bandwidth Sensitivity: Place

This figure presents a robustness test for Panel (B) of Table 5 by examining the sensitivity of the estimated effect of being higher-ranked on the probability of placing the next race to alternative bandwidth choices. The x-axis shows the bandwidth used to restrict the regression sample around the ranking threshold. The y-axis reports the estimated coefficient controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Circles and triangles correspond to the same cubic polynomial specification estimated without and with pair fixed effects, respectively. The vertical bars denote 95 percent confidence intervals. Standard errors are clustered at the horse pair level.

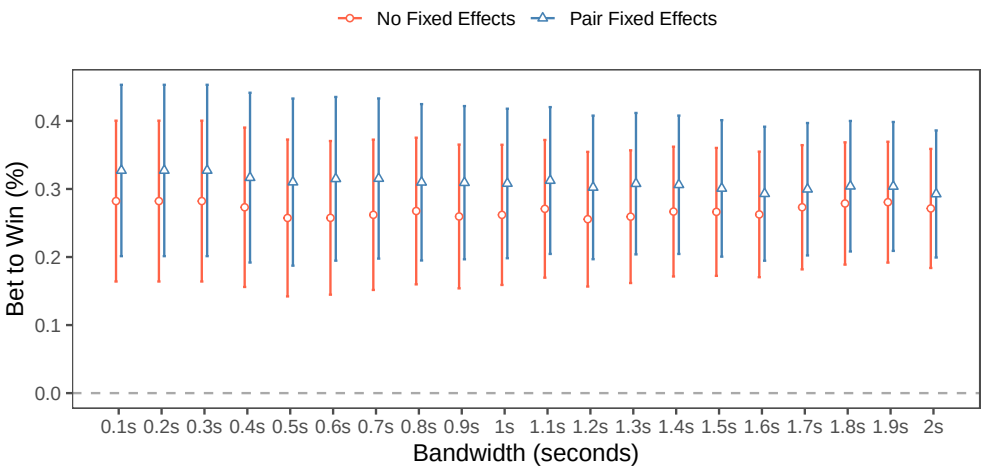


Figure O3: Bandwidth Sensitivity: Win Bet

This figure presents a robustness test for Panel (A) of Table 6 by examining the sensitivity of the estimated effect of being higher-ranked on the share of win bets, a proxy for bettors' perceived probability of winning, to alternative bandwidth choices. The x-axis shows the bandwidth used to restrict the regression sample around the ranking threshold. The y-axis reports the estimated coefficient controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Circles and triangles correspond to the same cubic polynomial specification estimated without and with pair fixed effects, respectively. The vertical bars denote 95 percent confidence intervals. Standard errors are clustered at the horse pair level.

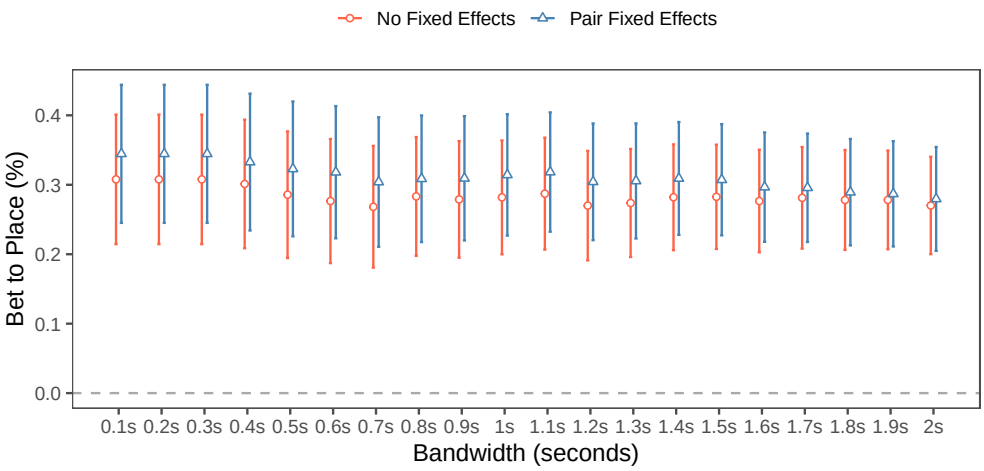


Figure O4: Bandwidth Sensitivity: Place Bet

This figure presents a robustness test for Panel (B) of Table 6 by examining the sensitivity of the estimated effect of being higher-ranked on the share of place bets, a proxy for bettors' perceived probability of placing, to alternative bandwidth choices. The x-axis shows the bandwidth used to restrict the regression sample around the ranking threshold. The y-axis reports the estimated coefficient controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Circles and triangles correspond to the same cubic polynomial specification estimated without and with pair fixed effects, respectively. The vertical bars denote 95 percent confidence intervals. Standard errors are clustered at the horse pair level.

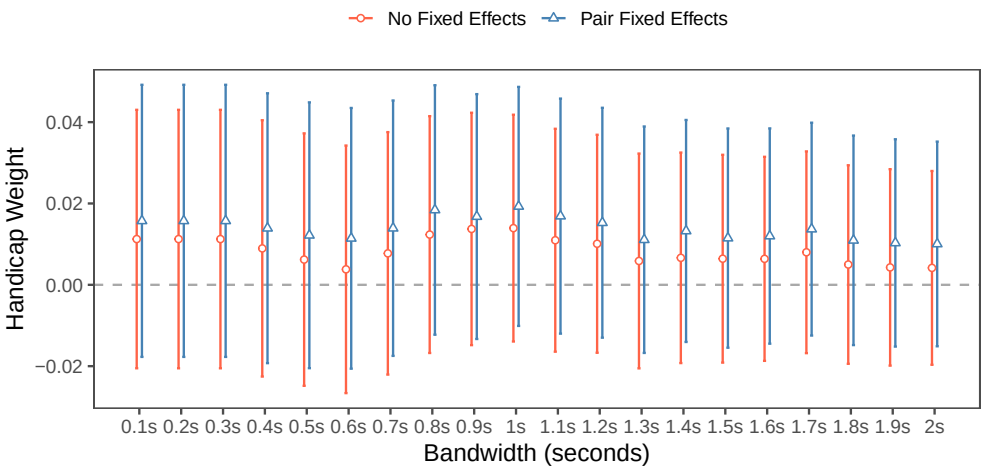


Figure O5: Bandwidth Sensitivity: Handicap

This figure presents a robustness test for Panel (A) of Table 4 by examining the sensitivity of the estimated effect of being higher-ranked on handicap weight assigned to the horse's next race to alternative bandwidth choices. The x-axis shows the bandwidth used to restrict the regression sample around the ranking threshold. The y-axis reports the estimated coefficient controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Circles and triangles correspond to the same cubic polynomial specification estimated without and with pair fixed effects, respectively. The vertical bars denote 95 percent confidence intervals. Standard errors are clustered at the horse pair level.

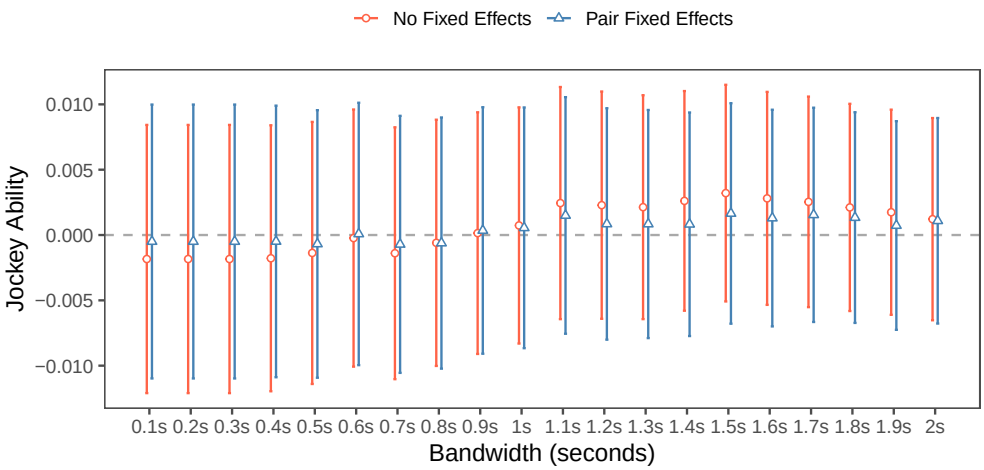


Figure O6: Bandwidth Sensitivity: Jockey Ability

This figure presents a robustness test for Panel (B) of Table 4 by examining the sensitivity of the estimated effect of being higher-ranked on jockey ability to alternative bandwidth choices. The x-axis shows the bandwidth used to restrict the regression sample around the ranking threshold. The y-axis reports the estimated coefficient controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race time gap interacted with $D_{r,n}$, an indicator variable denoting higher-ranked horses (see Equation (9)). Circles and triangles correspond to the same cubic polynomial specification estimated without and with pair fixed effects, respectively. The vertical bars denote 95 percent confidence intervals. Standard errors are clustered at the horse pair level.

Table O1: Sensitivity of RD-Design to Functional Form

This tables presents bias-corrected estimates using the optimal bandwidth selector of Calonico *et al.* (2014). Panel (A) revisits the results of Table 5, while Panel (B) revisits the results of Table 6. Columns vary the local polynomial degree from 1 to 3. z-statistics in parentheses are based on standard errors clustered at the pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

(A) Actual Probabilities

Dep. Var.:	*Win (%)*			*Place (%)*		
	(1)	(2)	(3)	(4)	(5)	(6)
Higher-Ranked	0.1224	0.0421	0.0134	0.1312	0.1132	0.0313
	(0.82)	(0.24)	(0.07)	(0.44)	(0.37)	(0.09)
Bias-corrected	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial Degree	1	2	3	1	2	3
Sample: $\Delta \leq$	0.7s	1.1s	1.7s	0.6s	1.4s	1.6s
Obs	255,713	285,714	304,442	245,893	297,825	302,864
Mean of Dep. Var.	4.20	4.14	4.08	17.26	16.65	16.62

(B) Perceived Probabilities

Dep. Var.:	*Bet to Win (%)*			*Bet to Place (%)*		
	(1)	(2)	(3)	(4)	(5)	(6)
Higher-Ranked	0.2718***	0.2694***	0.2585***	0.2816***	0.2791***	0.2801***
	(5.42)	(5.42)	(4.67)	(6.77)	(6.62)	(5.98)
Bias-corrected	Yes	Yes	Yes	Yes	Yes	Yes
Polynomial Degree	1	2	3	1	2	3
Sample: $\Delta \leq$	0.5s	1.3s	1.7s	0.5s	1.3s	1.6s
Obs	228,758	294,068	304,442	228,688	293,960	302,750
Mean of Dep. Var.	4.31	4.19	4.17	6.10	5.97	5.95

Table O2: Robustness: Are the Effects Weaker Among Bets with Higher Stakes?

This table revisits Table 9 using an alternative estimate of average bet size. Within each racecourse-year, race dates are assigned to two groups (Low and High) based on an amount-weighted measure of average bet size (see Equation (7)). Specifically, average bet size is constructed as the weighted average of the midpoints of purchase-value bands, where the weights are each band's share of total purchase value on a given race date and racecourse. The variable Low Bet Size is an indicator equal to one if the average bet size on a given race date falls within the bottom 50 percent of the distribution for that racecourse in that year, and High Bet Size equals one if it falls in the top 50 percent. For each race×ranking, we refer to the two adjacent horses around the ranking as a "pair." Columns (1) and (3) focus on the subsample of pairs with a distance gap of less than or equal to a nose (approximately three inches). Columns (2) and (4) broaden the analysis to pairs whose time gap falls within the MSE-optimal bandwidth, selected using the procedure of Calonico *et al.* (2014), controlling for the gap between a horse's race time and the race time threshold around its ranking by including a polynomial of the race time gap interacted with $D_{r,n}$, an indicator for the higher-ranked horse and the bet-size group indicators (see Equation (14)). All regression specifications include pair fixed effects interacted with bet-size-group fixed effects. The row H_0: Low − High reports the difference between the Higher-Ranked × Low Bet Size and Higher-Ranked × High Bet Size coefficients, together with the associated t-statistics. The sample period spans January 1993 to October 2024. t-statistics, reported in parentheses, are calculated using standard errors clustered at the horse-pair level. *, **, and *** indicate statistical significance at the 10 percent, 5 percent, and 1 percent levels, respectively. For definitions of all variables, see Appendix A.

Dep. Var.:	*Bet to Win (%)*		*Bet to Place (%)*	
	(1)	(2)	(3)	(4)
Higher-Ranked ×Low Bet Size	0.7302***	0.4126***	0.5314***	0.4209***
	(3.88)	(4.24)	(3.43)	(5.20)
Higher-Ranked ×High Bet Size	0.3771*	0.3142***	0.3825**	0.3385***
	(1.69)	(2.91)	(2.26)	(4.01)
H_0: Low −High	0.3531	0.0984	0.1490	0.0823
	(1.21)	(0.68)	(0.65)	(0.70)
Pair FE ×Bet Size	✓	✓	✓	✓
Higher×Cubic×Bet Size		✓		✓
Sample: $\Delta \leq$	Nose	1.2s	Nose	1.1s
Obs	6,220	136,358	6,218	133,949
Adj. R^2	0.16	0.15	0.23	0.22
Mean of Dep. Var.	4.64	4.27	6.51	6.11